

# Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées

Document de synthèse  
présenté à  
l'Université Lille 1

par **Radu S. Stoica**

pour l'obtention de

**l'Habilitation à Diriger des Recherches**

Spécialité : Mathématiques

Soutenue le 22 Octobre 2014 devant le jury composé de :

P. Calka	Professeur (Université de Rouen), <i>Rapporteur</i>
Yu. Davydov	Professeur (Université Lille 1), <i>Garant</i>
D. Dereudre	Professeur (Université Lille 1)
D. Hestroffer	Astronome (IMCCE - Observatoire de Paris)
A. Kretschmar	Directeur de Recherche (INRA Avignon)
C. Lantuéjoul	Directeur de Recherche (Ecole des Mines), <i>Rapporteur</i>
A. Philippe	Professeur (Université de Nantes)
S. Zuyev	Professeur (Université de Chalmers, Suède), <i>Rapporteur</i>

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Plan du mémoire . . . . .	7
1.2	Projet de recherche . . . . .	7
<b>I</b>	<b>Modélisation probabiliste</b>	<b>12</b>
<b>2</b>	<b>Processus ponctuels marqués</b>	<b>15</b>
2.1	Construction et définition . . . . .	15
2.2	Caractérisation des processus ponctuels . . . . .	16
2.3	Le processus ponctuel de Poisson . . . . .	18
2.4	Processus ponctuels avec interaction . . . . .	23
2.4.1	Construction de la densité de probabilité . . . . .	24
2.4.2	Processus ponctuels markoviens . . . . .	28
<b>3</b>	<b>Contributions</b>	<b>33</b>
3.1	Modélisation des filaments galactiques : les processus Candy et Bisous . . . . .	36
3.1.1	Modélisation de l'interaction entre les objets . . . . .	36
3.1.2	Modélisation de l'attache aux données . . . . .	42
3.2	Modélisation d'agrégats en épidémiologie animale . . . . .	47
3.2.1	Présentation des données et formulation du problème . . . . .	47
3.2.2	Modélisation de l'attache aux données . . . . .	49
3.2.3	Modélisation de l'interaction entre objets . . . . .	50
3.3	Modélisation du morcellement pour un environnement . . . . .	53
3.4	Combinaison des étiquettes dans des modèles de champ aléatoire . . . . .	58
3.5	Modélisation de paysages à l'aide de pavages en $T$ . . . . .	66
3.5.1	L'espace des $T$ -pavages . . . . .	66
3.5.2	Le $T$ -pavage complètement aléatoire . . . . .	69
3.5.3	Modèles de Gibbs pour les $T$ -pavages . . . . .	75

<b>II</b>	<b>Simulation de type MCMC</b>	<b>79</b>
<b>4</b>	<b>Chaînes de Markov : notions, propriétés et algorithmes</b>	<b>81</b>
4.1	Cadre théorique pour la simulation MCMC "classique" . . . . .	81
4.2	Algorithme de type Metropolis-Hastings pour les processus ponctuels marqués . . . . .	89
4.2.1	Construction de l'algorithme . . . . .	89
4.2.2	Propriétés de convergence . . . . .	93
4.3	Cadre théorique pour la simulation "parfaite" . . . . .	97
4.4	Algorithme CFTP pour les processus ponctuels marqués . . . . .	100
4.4.1	Processus spatiaux de naissance et de mort . . . . .	100
4.4.2	Algorithme CFTP pour les processus ponctuels . . . . .	103
<b>5</b>	<b>Contributions</b>	<b>107</b>
5.1	Construction des dynamiques adaptées . . . . .	107
5.1.1	Etude par simulation de la dynamique adaptée au mo- dèle Candy . . . . .	109
5.2	Dynamique Monte Carlo pour la simulation d'un modèle de pavage en $T$ . . . . .	114
5.3	Analyse des algorithmes de simulation exacte . . . . .	119
5.3.1	Clan des ancêtres . . . . .	119
5.3.2	Convergence de l'algorithme clan des ancêtres . . . . .	120
5.3.3	Metropolis-Hastings exact . . . . .	121
5.3.4	Convergence de l'algorithme Metropolis-Hastings exact . . . . .	124
5.3.5	L'échantillonneur de Gibbs exact . . . . .	129
5.3.6	Convergence de l'échantillonneur de Gibbs . . . . .	133
5.3.7	Etude par simulation . . . . .	134
<b>III</b>	<b>Inférence statistique</b>	<b>142</b>
<b>6</b>	<b>Outils mathématiques pour l'inférence statistique de pro- cessus ponctuels marqués</b>	<b>145</b>
6.1	Détection de la forme cachée dans les données . . . . .	145
6.2	Estimation des paramètres de la forme . . . . .	147
6.2.1	Maximum de la pseudo-vraisemblance . . . . .	147
6.2.2	Maximum de vraisemblance . . . . .	149
6.2.3	Algorithme EM : données incomplètes . . . . .	153
<b>7</b>	<b>Contributions</b>	<b>156</b>
7.1	Convergence du recuit simulé pour des processus ponctuels marqués . . . . .	156
7.1.1	Détection de filaments galactiques : premiers pas . . . . .	160
7.2	Maximum de vraisemblance pour le modèle Candy . . . . .	162

7.3	Approche bayésienne pour la détection des structures . . . . .	169
7.3.1	Détection d'agrégats en épidémiologie animale : obtention d'une solution . . . . .	170
7.3.2	Détection de filaments galactiques : obtention d'une solution . . . . .	173
7.4	Evaluation des résultats. Caractérisation morpho-statistique des structures. . . . .	177
7.4.1	Détection d'agrégats en épidémiologie animale : vérification et analyse des résultats . . . . .	177
7.4.2	Détection de filaments galactiques : vérification et analyse des résultats . . . . .	179
7.5	Construction d'une forme moyenne par un estimateur d'ensembles de niveaux . . . . .	183
7.5.1	Ensembles de niveaux : un outil pour moyennner des ensembles aléatoires . . . . .	184
7.5.2	Estimation des ensembles de niveaux . . . . .	186
7.5.3	Application de l'espérance de Vorob'ev pour les champs de lacets browniens . . . . .	189
7.5.4	Application de l'espérance de Vorob'ev aux filaments galactiques . . . . .	189
7.6	Ensembles de niveau, statistiques d'ordre et lois à queue régulière pour l'étude des perturbations planétaires sur les comètes . . . . .	192
7.6.1	Statistiques d'ordre . . . . .	193
7.6.2	Lois stables . . . . .	196
7.6.3	Théorie d'Öpik . . . . .	200

## **IV Conclusion et perspectives 202**

### **8 Bilan et nouvelles directions de recherche 204**

8.1	Analyse générale de l'approche . . . . .	204
8.1.1	Et les autres méthodes ? . . . . .	204
8.1.2	Pour quoi ça marche ? . . . . .	205
8.2	Questions concernant la modélisation de forme . . . . .	206
8.2.1	Extension à l'infini, percolation et transition de phase . . . . .	206
8.2.2	Processus de Cox, processus permanents et déterminants . . . . .	208
8.3	Questions concernant la simulation MCMC . . . . .	209
8.3.1	Méthodes classiques de simulation . . . . .	210
8.3.2	Simulation exacte . . . . .	211
8.4	Questions concernant l'inférence statistique . . . . .	212
8.4.1	Détection conjointe de la forme et des paramètres . . . . .	212
8.4.2	Simulation de la loi a posteriori . . . . .	213
8.4.3	Choix de modèle . . . . .	213



8.5	Projet : processus stochastiques et géométrie aléatoire . . . .	215
8.5.1	Modélisation spatio-temporelle . . . . .	216
8.5.2	Simulation de processus stochastiques des formes . . .	217
8.5.3	Statistiques pour les processus stochastiques des formes	218
8.5.4	Outils probabilistes et d'inférence statistique pour caractériser le chaos . . . . .	218

## Remerciements

Ce travail a été possible grâce à des gens remarquables que j'ai pu rencontrer lors de mon périple. Je remercie mes professeurs, mes amis et ma famille. Pour être plus dans le concret, j'ai essayé de faire une liste avec leur noms. De par certaines limitations, cette liste est incomplète. Cependant, dans mon coeur, tout le monde y est.

### Science

P. Calka, Yu. Davydov, D. Dereudre, D. Hestroffer, A. Kretzschmar, C. Lantuéjoul, A. Philippe, S. Zuyev. X. Descombes, J. Zerubia. M. N. M. van Lieshout. P. Gregori, J. Mateu. V. Martinez, E. Saar. D. Allard, F. Aries, J. Chadoeuf, R. Senoussi. F. Baccelli, D. Stoyan. M. Bodea. K. Adamczyk-Chauvat, G. Castellan-Guerin, F. Chatelain, M. Emily, M. Fouchard, J. Francos, E. Gay, P. Heinrich, K. Kieu, S. Liu, L. J. Liivamägi, M. El Machkouri, H. Monod, L. Roques, M. Sigelle, E. Tempel, V. C. Tran, G. Valsecchi, A. Vienne. M.-F. Barme, C. Biernacki, D. Coupier, S. Cuenca, J.-J. Derycke, M. Fradon, S. Huart, S. Hertsoen, J. Jacques, J. Khan, B. Lecocq, L. Marsalle, C. Suquet, M.-C. Viano. L'Equipe Probabilités et Statistiques, l'Amicale des personnels de l'UFR de Mathématiques – Lille 1, l'Université Lille 1, le Laboratoire Paul Painlevé. M. Braileanu, J.-F. Deant, G. Ducrocq, P. Ducrocq, Y. Hatzfeld, E. Loubier, J.-C. Olmo, D. Veret, S. Veret.

### Conscience

Paul, Luisa et Hugues, Nathalie, Cedric, Frederique et Guillaume, Teresa et Guy, Vincent, Serge, l'Idée Halles, les Diables Bleus, Thierry, Thierry, Caroline et Fabrice, la Croix-Rousse, Raphaelle et Christophe, Sabine, Anais et Arnaud, Andreea et Guillaume, Severine. Gaelle et Lilian, Paqui et Jesus, Lucia, Vicent, le Bario. Emmanuelle, Mehdi, Marie et Nicolas, Neila et Michel, Laetitia et Antoine, le Va Zen. Les Dojos. Anca et Ciprian, Melissa et Adrian, Alina et Mihai, Ozi et Ioana, Raluca et Bebe, Aura et Horia, Rodica et Virgil, Corina et Valentin, Oana et Bogdan, Silvia, Bogdan, Cristina, Monica, Simona et Sorin. Martine et Guy, Annie, Olivier, Jocelyne, Nina, Florica. D2, la Fraternité et Combe Martin.

### Âme

Irina, Barbu, Ana Alexandra, Simona et surtout mon frère, Andrei Mircea.

# Chapitre 1

## Introduction

### 1.1 Plan du mémoire

En juin 2006, j'ai été recruté à l'Université Lille 1 sur un poste de maître de conférences en mathématiques dans le domaine des probabilités et des statistiques. J'avais participé à ce concours avec un projet scientifique et pédagogique intitulé "Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées".

Ce mémoire décrit les étapes réalisées au cours de ce projet, se rapportant à la partie recherche. Nous commençons par la présentation du projet et par la structure du document.

### 1.2 Projet de recherche

Nous appelons données spatialisées un ensemble dont chaque élément comporte deux composantes : une position et une caractéristique. Les images numériques sont un exemple de données spatialisées : le pixel est caractérisé par un index dans une matrice et sa valeur numérique indique la couleur associée à cette position dans l'image. Les catalogues des galaxies en astronomie, les bases de données recensant des élevages laitiers sont également des exemples des données spatialisées.

Un catalogue de galaxies contient la position dans notre univers et les caractéristiques (masse, luminosité, déplacement vers le rouge, forme, etc.) de chaque galaxie. Une base de données recensant des élevages laitiers contient la position dans le territoire et les caractéristiques de chaque ferme (espèce et nombre d'animaux, filiation et lieu de naissance de chaque animal, etc.). Dans ces derniers exemples, les positions des éléments ne sont plus disposées sur une grille régulière, et les caractéristiques correspondantes se présentent

sous la forme de structures de données très riches et très complexes.

*Le caractère spatialisé des données implique que les réponses possibles aux questions que les données suggèrent, comportent une forte composante morphologique.*

Quelques exemples de questions : extraire les réseaux routiers présents dans une image satellitaire, caractériser la répartition spatiale à large échelle des galaxies de notre Univers, déterminer la tendance à l'agrégation des fermes ayant une grande prédisposition pour une certaine maladie. En effet, dans de nombreuses situations, pouvoir analyser des données spatialisées revient à être capable de répondre à la question suivante : quelle est la forme qui se trouve "cachée" dans les données ?

L'objectif de ce projet est d'étudier et proposer des outils probabilistes et statistiques permettant de répondre à cette question. Plus précisément, nous nous sommes focalisés sur le développement des outils suivants :

- modélisation stochastique : qu'est-ce que la forme que l'on cherche ?
- simulation de type MCMC : comment mettre en oeuvre la forme ?
- inférence statistique et évaluation : comment détecter la forme, quels sont les paramètres de la forme, la forme trouvée existe-t-elle vraiment ?

## **Structure du document**

Le mémoire est organisé en quatre parties. Les trois premières parties correspondent chacune à une de ces questions. Chaque partie contient deux chapitres. Le premier chapitre introduit les outils mathématiques utilisés. Le deuxième chapitre présente nos contributions comme réponses à la question du projet. La quatrième et dernière partie du mémoire est dédiée aux conclusions et aux perspectives.

# Liste des publications

- [T1] X. Descombes, R. S. Stoica, L. Garcin, and J. Zerubia. A RJMCMC algorithm for object processes in image processing. *Monte Carlo Methods and Applications*, 7(1-2) :149–156, 2001.
- [T2] X. Descombes, R. S. Stoica, and J. Zerubia. Two Markov point processes for simulating line networks. In *International Conference on Image Processing Proceedings*, Kobe, 1999. IEEE.
- [T3] X. Descombes, M. N. M. van Lieshout, R. S. Stoica, and J. Zerubia. Parameter estimation by a markov chain monte carlo technique for the candy model. In *Statistical Signal Processing Proceedings*, Singapore, 2001. IEEE.
- [T4] P. Gregori, J. Mateu, and R. S. Stoica. A marked point process for modelling three dimensional patterns. In A. Baddeley, P. Gregori, J. Mateu, R. S. Stoica, and D. Stoyan, editors, *Spatial Point Process Modelling and its Applications*, Castellon, Spain, 2004. Publicacions de la Universitat Jaume I.
- [T5] P. Heinrich, R. S. Stoica, and V. C. Tran. Level sets estimation and vorob’ev expectation of random compact sets. *Spatial Statistics*, 2 :47–61, 2012.
- [T6] K. Kiêu, K. Adamczyk-Chauvat, H. Monod, and R. S. Stoica. A completely random T-tessellation model and Gibbsian extensions. *Spatial Statistics*, 6 :118–138, 2013.
- [T7] M. N. M. van Lieshout and R. S. Stoica. The Candy model revisited : properties and inference. *Statistica Neerlandica*, 57 :1–30, 2003.
- [T8] M. N. M. van Lieshout and R. S. Stoica. Perfect simulation for marked point processes. *Computational Statistics and Data Analysis*, 51 :679–698, 2006.
- [T9] M. N. M. van Lieshout and R. S. Stoica. A note on pooling of labels in random fields. *Statistics and Probability Letters*, 80 :1431–1436, 2010.
- [T10] L. Roques and R. S. Stoica. Species persistence decreases with habitat fragmentation : an analysis in periodic stochastic en-

- vironments. *Journal of Mathematical Biology*, 55(2) :189–205, 2007.
- [T11] R. S. Stoica, F. Chatelain, and M. Sigelle. Inférence paramétriques pour les processus ponctuels marqués en analyse d’images. In X. Descombes, editor, *Applications de la géométrie stochastique à l’analyse d’images (Série Signal et Image IC2)*. Hermès Lavoisier, 2011.
- [T12] R. S. Stoica, F. Chatelain, and M. Sigelle. Parametric inference for marked point processes in image analysis. In X. Descombes, editor, *Stochastic Geometry for Image Analysis (Digital Signal and Image Processing series)*. John Wiley and Sons, 2012.
- [T13] R. S. Stoica, X. Descombes, M. N. M. van Lieshout, and J. Zerubia. An application of marked point processes to the extraction of linear networks from images. In J. Mateu and F. Montes, editors, *Spatial statistics through applications*. WIT Press, Southampton, UK, 2002.
- [T14] R. S. Stoica, X. Descombes, and J. Zerubia. Road extraction in remote sensed images using a stochastic geometry framework. In *Bayesian inference and maximum entropy methods in science and engineering : 20th International Workshop*, volume 560, pages 531–542. AIP Conference Proceeding, 2001.
- [T15] R. S. Stoica, X. Descombes, and J. Zerubia. A Gibbs point process for road extraction in remotely sensed images. *International Journal of Computer Vision*, 57 :121–136, 2004.
- [T16] R. S. Stoica, E. Gay, and A. Kretzschmar. Cluster detection in spatial data based on Monte Carlo inference. *Biometrical Journal*, 49(2) :1–15, 2007.
- [T17] R. S. Stoica, P. Gregori, and J. Mateu. Simulated annealing and object point processes : tools for analysis of spatial patterns. *Stochastic Processes and their Applications*, 115 :1860–1882, 2005.
- [T18] R. S. Stoica, P. Gregori, and J. Mateu. A versatile MCMC strategy for sampling posterior distributions of analytically intractable models. *Publications IRMA Lille*, 67(IV), 2007.
- [T19] R. S. Stoica, S. Liu, Yu. Davydov, M. Fouchard, A. Vienne, and G. B. Valsecchi. Order statistics and heavy-tailed distributions for planetary perturbations on Oort cloud comets. *Astronomy and Astrophysics*, 513(A14) :1–9, 2010.
- [T20] R. S. Stoica, V. J. Martinez, J. Mateu, and E. Saar. Detection of cosmic filaments using the Candy model. *Astronomy and Astrophysics*, 434 :423–432, 2005.

- [T21] R. S. Stoica, V. J. Martinez, and E. Saar. A three dimensional object point process for detection of cosmic filaments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55 :189–205, 2007.
- [T22] R. S. Stoica, V. J. Martinez, and E. Saar. Filaments in observed and mock galaxy catalogues. *Astronomy and Astrophysics*, 510(A38) :1–12, 2010.
- [T23] R. S. Stoica. Marked point processes for statistical and morphological analysis of astronomical data. *The European Physical Journal Special Topics*, 186 :123–165, 2010.
- [T24] E. Tempel, R. S. Stoica, and E. Saar. Evidence for spin alignment of spiral and elliptical galaxies in filaments. *Monthly Notices of the Royal Astronomical Society*, 428 :1827–1836, 2013.
- [T25] E. Tempel, R. S. Stoica, E. Saar, V. J. Martinez, L. J. Liivamägi, and G. Castellán. Detecting filamentary pattern in the cosmic web : a catalogue of filaments for the SDSS. *Monthly Notices of the Royal Astronomical Society*, 438(4) :3465–3482, 2014.

**Première partie**

**Modélisation probabiliste**



## Qu'est-ce que la forme que l'on cherche ?

La réalisation d'un processus ponctuel marqué est un ensemble dont les éléments ont deux composantes : une position et une marque, la caractéristique de l'élément. Par exemple, une forêt peut être modélisée par un tel processus : des points aléatoires représentent les positions des arbres, alors que la marque en chaque point est donnée par un cercle de rayon aléatoire centré au point. Celle-ci peut représenter le diamètre de la couronne de chaque arbre. Naturellement, la marque peut être enrichie d'autres caractéristiques, telles que l'espèce ou l'âge de l'arbre.

L'un des nombreux domaines d'application où l'utilisation des processus ponctuels marqués a enregistré un franc succès est le traitement d'images [T1, 60, 61, 105, 189, 169, 170, 168, T13, T15]. Cette réussite tient non seulement à la qualité des solutions (approches complètement automatiques, robustes), mais aussi à la possibilité de formuler et de formaliser des réponses naturelles aux questions posées. Par exemple, pour détecter des objets dans l'image, la question posée n'est plus quels sont les pixels qui remplissent les critères définissant des objets. En modélisant les objets recherchés par un processus ponctuel marqué, il est possible de répondre directement à la question : où se trouvent les objets dans l'image et quelles sont leurs caractéristiques ?

Dans cette partie nous étendons les idées mises en oeuvre en traitement d'image, afin que cette utilisation des processus ponctuels marqués profite aussi à d'autres domaines d'application.

Posons comme principe que la forme que l'on cherche est une entité complexe construite à partir d'objets générateurs simples qui interagissent. Par exemple, il est naturel d'imaginer que les réseaux routiers sont formés à partir de segments connectés qui ont tendance à s'aligner.

Il en résulte que si l'on peut définir des objets générateurs et des interactions entre ces objets, la forme que l'on cherche peut être considérée comme la réalisation d'un processus ponctuel marqué. Il est tout à fait possible mathématiquement d'écrire la densité de probabilité d'un tel processus. Les paramètres de cette densité contrôlent le nombre et la marque des objets générateurs ainsi que l'intensité des différentes interactions. Ceci implique que les statistiques exhaustives du modèle pourraient être utilisées comme des descripteurs morphologiques de la forme ainsi construite.

Si le contexte le permet, il est possible de définir des interactions de portée finie entre objets. Le modèle ainsi construit aura un caractère markovien. Cette propriété est très intéressante. Du point de vue théorique c'est avec cette propriété que l'on brise la complexité de la forme. Du point de vue

pratique, elle peut réduire le coût calculatoire de la mise en oeuvre d'un tel modèle.

Cette partie présente des modèles de formes construits à partir de processus ponctuels marqués. A partir de cet axe, des ramifications se créent. Ainsi, nous considérons également des modèles construits à partir de champs aléatoires discrets ou des processus de type Arak.

Pour cette présentation nous avons suivi principalement [19, 36, 72, 190, 197, 141, 180]. Les références [45, 46, 157] traitent les processus ponctuels d'un point de vue plus théorique. Pour un point de vue plus orienté vers les applications en statistiques spatiales et analyse d'image nous recommandons [6, 60, 61, 68, 90]. Nous recommandons également [4, 5] pour une présentation théorique claire et précise en vue des applications dans le domaine des télécommunications. Des parties de cette section ont été publiées dans [T6, T7, T9, T10, T17, T16, T21, T22, T23, T11, T12].

## Chapitre 2

# Processus ponctuels marqués

### 2.1 Construction et définition

Soit  $\nu$  la mesure de Lebesgue dans  $\mathbb{R}^d$ ,  $W$  un sous-ensemble compact<sup>1</sup> de  $\mathbb{R}^d$ , et  $(W, \mathcal{B}_W, \nu)$  la restriction naturelle par rapport à  $W$  de  $(\mathbb{R}^d, \mathcal{B}, \nu)$ .

Pour chaque  $n \in \mathbb{N}$ , soit  $W_n$  l'ensemble de toutes les configurations  $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$  formées par  $n$  points  $w_i \in W$ . Les configurations ne sont pas ordonnées et les points ne sont pas obligatoirement distincts. La configuration vide  $W_0$  ne contient aucun point. L'espace des configurations peut être donné par  $\Omega_W = \bigcup_{n=0}^{\infty} W_n$ . A cet espace est associé la tribu  $\mathcal{F}_W$  générée par les applications de comptage

$$\mathcal{F}_W = \sigma(\{\mathbf{w} = \{w_1, \dots, w_n\} \in \Omega_W : n(\mathbf{w}_B) = n(\mathbf{w} \cap B) = m\})$$

où  $B \in \mathcal{B}$  et  $m \in \mathbb{N}$ .

**Définition 1** *Un processus ponctuel sur  $W$  est une application mesurable d'un espace de probabilité  $(\mathcal{S}, \mathcal{A})$  dans  $(\Omega_W, \mathcal{F}_W)$ . Sa loi  $\mathbb{P}$  est donnée par*

$$\mathbb{P}(X \in F) = \mathbb{P}\{\omega \in \mathcal{S} : X(\omega) \in F\},$$

avec  $F \in \mathcal{F}_W$ . Un processus ponctuel a comme réalisation un ensemble de points de  $W$ .

Il est possible de rajouter des caractéristiques à chaque point. Dans ce cas, nous avons affaire à un processus ponctuel marqué. Soit le processus ponctuel sur  $W \times M$  dont les réalisations sont données par les ensembles  $\{(w_i, m_i)\}_{i=1}^n$ . Ici les  $w_i \in W$  représentent les positions de chaque point et  $m_i \in M$  les

---

1. Nous avons choisi  $W$  compact pour être le plus proche possible des applications. Cependant, sous des conditions assez souples, la présentation de ce chapitre est applicable directement aux parties mesurables de  $W \subset \mathbb{R}^d$ .

marques correspondantes. L'espace des marques  $M$  est équipé d'une tribu  $\mathcal{M}$  et d'une mesure de probabilité  $\nu_M$ .

La construction d'un processus ponctuel marqué est similaire à celle d'un processus ponctuel. Pour chaque  $n \in \mathbb{N}$ ,  $\Xi_n$  est l'ensemble de toutes les configurations non-ordonnées  $\mathbf{x} = \{x_1, \dots, x_n\}$ . Ces configurations sont formées par les points marqués non nécessairement distincts  $x_i \in W \times M$ . La configuration vide est  $\Xi_0$ . L'espace des configurations est défini comme  $\Omega = \bigcup_{n=0}^{\infty} \Xi_n$  et il est équipé d'une  $\sigma$ -algèbre  $\mathcal{F}$ . La tribu  $\mathcal{F}$  est définie par les applications qui comptent le nombre de points dans des boréliens  $A \subseteq W \times M$ .

**Définition 2** *Un processus ponctuel sur  $W \times M$  est un processus ponctuel marqué si la loi marginale des positions des points est un processus ponctuel sur  $W$ . Un processus ponctuel marqué est une application mesurable d'un espace de probabilité dans  $(\Omega, \mathcal{F})$ .*

Les marques peuvent être par exemple des objets de forme géométrique simple : cercles, ellipses ou segments de droite. Dans ce cas, l'espace des marques représente le domaine de définition des paramètres des objets. Par exemple, il est possible de considérer des espaces des marques comme  $M_{\text{cercle}} = (0, \infty)$  ou  $M_{\text{segment}} = (0, \infty) \times [0, \pi]$ . Ces processus ponctuels marqués, avec les points indiquant la position de l'objet, et la marque précisant sa forme géométrique, sont connus sous les nom de processus *germe-graine*.

Un autre exemple de processus ponctuel marqué est le processus ponctuel avec plusieurs catégories de points. Dans ce cas,  $M = \{1, 2, \dots, k\}$  représente les différents types de points ou d'objets. Ceci est équivalent à un processus ponctuel multivarié de dimension  $k$ . Un tel processus peut être vu comme un  $k$ -tuple de processus ponctuels  $(X_1, X_2, \dots, X_k)$ , formé par les processus  $X_1, X_2, \dots, X_k$  qui correspondent chacun à un type d'objet.

## 2.2 Caractérisation des processus ponctuels

Soit  $X$  un processus ponctuel sur  $W$ , et soit  $N$  sa variable de comptage

$$N(B) = n(X_B), \quad B \in \mathcal{B}_W,$$

qui représente le nombre de points qui "tombent" dans  $B$ . Les ensembles de la forme

$$F_B = \{\mathbf{x} \in \Omega : n(\mathbf{x}_B) = 0\},$$

sont appelées *événements vides*.

Dans la suite, nous considérons que les processus ponctuels sont *simples*, c'est à dire  $N(\{w\}) \leq 1$  pour tout  $w \in W$ .

**Théorème 1** *La loi d'un processus ponctuel est caractérisée par la probabilité de ses événements vides*

$$v(B) = \mathbb{P}(N(B) = 0), \quad B \in \mathcal{B}_W.$$

La fonctionnelle génératrice d'un processus ponctuel  $X$  sur  $W$  est définie par

$$G_X(u) = \mathbb{E} \prod_{\xi \in X} u(\xi)$$

où  $u : W \rightarrow [0, 1]$ .

Soit  $u(\xi) = t^{\mathbf{1}\{\xi \in B\}}$  avec  $0 < t < 1$  et  $B \in \mathcal{B}_W$ . Alors

$$G_X(u) = \mathbb{E} t^{N(B)}$$

est la fonction génératrice de  $N(B)$ . Par conséquent, la loi de  $X$  est caractérisée par  $G_X(u)$ .

La moyenne et les moments d'ordre supérieur sont des quantités cruciales en théorie des probabilités. Il n'est pas évident de donner la définition d'une "forme moyenne". Cependant, nous pouvons restreindre notre attention à des configurations possédant certaines propriétés, en considérant des moments des quantités  $N(B)\mathbf{1}\{X \in F\}$ .

**Définition 3** *Soit  $X$  un processus ponctuel sur un espace, métrique, complet et séparable  $(W, d)$ . La mesure de Campbell de  $X$  est*

$$C(B \times F) = \mathbb{E}[N(B)\mathbf{1}\{X \in F\}],$$

pour tout  $B \in \mathcal{B}_W$  et  $F \in \mathcal{F}$ .

La mesure moment du premier ordre s'exprime en fonction des mesures de Campbell :

$$\mu_1(B) = \mathbb{E}[N(B)] = C(B \times \Omega).$$

D'une manière analogue, nous pouvons construire la mesure de Campbell d'ordre supérieur

$$C^{(2)}(B_1 \times B_2 \times F) = \mathbb{E}[N(B_1)N(B_2)\mathbf{1}\{X \in F\}],$$

qui nous donne la mesure moment de second ordre

$$\mu_2(B_1 \times B_2) = C^{(2)}(B_1 \times B_2 \times \Omega) = \mathbb{E}[N(B_1)N(B_2)].$$

Intuitivement, le mécanisme selon lequel une structure complexe est analysée, consiste d'abord en l'étude des caractéristiques de chaque objet, suivi d'une opération de synthèse des ces propriétés. Le résultat suivant, connu dans la littérature comme la formule de Campbell-Mecke, fixe le cadre d'une telle procédure.

**Théorème 2** *Soit  $h : W \times \Omega \rightarrow [0, \infty)$  une fonction mesurable. Alors*

$$\mathbb{E} \left[ \sum_{x \in X} h(x, X) \right] = \int_W \int_{\Omega} h(x, X) dC(x, X).$$

*Si la fonction  $h$  ne dépend pas du processus ponctuel  $X$ , la formule de Campbell-Mecke devient*

$$\mathbb{E} \left[ \sum_{x \in X} h(x) \right] = \int_W \int_{\Omega} h(x) d\mu_1(x).$$

## 2.3 Le processus ponctuel de Poisson

Ce processus est d'une importance fondamentale du point de vue théorique et appliqué. En plus des références citées, pour approfondir ce sujet nous recommandons [101, 107, 123, 132]. Ce processus est construit à partir de deux ingrédients : la densité et le processus ponctuel binomial.

Pour la densité, nous considérons une fonction  $\rho : W \rightarrow [0, \infty)$ . Cette fonction définit la *mesure intensité*

$$\Upsilon(B) = \int_B \rho(w) d\nu(w),$$

qui est *localement finie*<sup>2</sup> et *diffuse*<sup>3</sup>.

Le processus ponctuel binomial est décrit par la définition suivante.

**Définition 4** *Soit  $f(w)$  une densité de probabilité sur  $B \in \mathcal{B}_W$  et soit  $n \in \mathbb{N}$ . Un processus ponctuel  $X$  formé de  $n$  points indépendamment répartis dans  $B$  selon la densité  $f$  est un processus ponctuel binomial. Dans ce cas, nous écrivons  $X \sim \text{Binomial}(B, n, f)$ .*

---

2.  $\Upsilon(B) < \infty$  pour des ensembles  $B \subseteq W$  bornés

3.  $\Upsilon(\{w\}) = 0$  pour  $w \in W$

Le processus ponctuel binomial le plus simple est obtenu pour  $\nu(B) < \infty$  et  $f(w) = \frac{1}{\nu(B)}$ . La propriété d'indépendance et le nombre fixe d'objets, font que les processus binomiaux sont utilisés comme point de départ dans l'étude de la présence d'interaction entre objets composant une forme [90].

Nous pouvons définir maintenant le processus ponctuel de Poisson.

**Définition 5** *Un processus ponctuel  $X$  sur  $W$  est un processus ponctuel de Poisson de densité  $\rho$  si les propriétés suivantes sont vérifiées :*

(i) *pour tout  $B \subseteq W$  tel que  $\Upsilon(B) < \infty$ ,  $N(B)$  suit une loi de Poisson de paramètre  $\Upsilon(B)$ ,  $N(B) \sim \mathcal{P}(\Upsilon(B))$ <sup>4</sup>.*

(ii) *pour tout  $n \in \mathbb{N}$  et  $B \subseteq W$  tel que  $0 < \Upsilon(B) < \infty$ , conditionnellement à  $N(B) = n$ ,  $X_B \sim \text{binom}(B, n, f)$  avec  $f(w) = \frac{\rho(w)}{\Upsilon(B)}$ . Nous écrivons  $X \sim \text{Poisson}(W, \rho)$ .*

Pour tout ensemble borné  $B \subseteq W$ , la mesure intensité  $\Upsilon$  détermine le nombre moyen de points de  $B$ ,

$$\mathbb{E}N(B) = \Upsilon(B).$$

**Définition 6** *Un processus ponctuel  $X$  est stationnaire si sa distribution est invariante par translation. Le processus est dit isotrope si sa distribution est invariante par rotation.*

Si la fonction  $\rho$  est constante, nous parlons d'un processus de Poisson homogène. Sinon, le processus est appelé hétérogène. Si  $\rho = 1$ , alors le processus ponctuel  $\text{Poisson}(W, 1)$  est dit processus de Poisson standard ou processus de Poisson de densité unité.

**Proposition 1**

(i) *Le processus ponctuel  $X$  est  $\text{Poisson}(W, \rho)$  avec une mesure d'intensité  $\Upsilon(B) = \int_B \rho(w) d\nu(w) < \infty$ , si et seulement si pour tout  $B \subseteq W$  et  $F \in \mathcal{F}_W$ , nous avons*

$$\begin{aligned} \mathbb{P}(X_B \in F) &= \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\Upsilon(B)]}{n!} \int_B \cdots \int_B \mathbf{1}_F\{w_1, \dots, w_n\} \left[ \prod_{i=1}^n \rho(w_i) \right] d\nu(w_1) \dots d\nu(w_n) \end{aligned}$$

(ii) *Si  $X \sim \text{Poisson}(W, \rho)$ , alors pour les fonctions  $h : \Omega_W \rightarrow [0, \infty)$  et  $B \subseteq W$  avec  $\Upsilon(B) < \infty$ , nous avons :*

$$\begin{aligned} \mathbb{E}[h(X_B)] &= \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\Upsilon(B)]}{n!} \int_B \cdots \int_B h(\{w_1, \dots, w_n\}) \left[ \prod_{i=1}^n \rho(w_i) \right] d\nu(w_1) \dots d\nu(w_n) \end{aligned}$$

---

4. si  $\Upsilon(B) = 0$  alors  $N(B) = 0$

**Proposition 2** *Si  $X$  est un processus de Poisson sur  $W$ , et si  $B_1, B_2, \dots$  sont des ensembles deux à deux disjoints de  $W$ , alors les processus  $X_{B_1}, X_{B_2}, \dots$  sont indépendants.*

Si  $W$  est borné, l'existence du processus Poisson( $W, \rho$ ) découle directement de la Prop. 1. Le résultat suivant garantit l'existence d'un processus de Poisson même quand  $W$  n'est pas borné [141].

**Théorème 3** *Le processus  $X \sim \text{Poisson}(W, \rho)$  existe et sa loi est caractérisée par la probabilité des évènements vides*

$$v(B) = \exp[-\Upsilon(B)]$$

pour les ensembles  $B \subset W$  bornés.

**Proposition 3** *Si  $X \sim \text{Poisson}(W, \rho)$  alors*

$$G_X(u) = \exp \left[ - \int_W (1 - u(\xi)) \rho(\xi) d\nu(\xi) \right]$$

pour des fonctions  $u : W \rightarrow [0, 1]$ .

Les résultats suivants, connus sous le nom de théorèmes de Slyvniak-Mecke, constituent un outil précieux pour le calcul de quantités moyennes attachées aux objets issus d'un processus de Poisson.

**Théorème 4** *Si  $X \sim \text{Poisson}(W, \rho)$ , alors pour des fonctions  $h : W \times \Omega_W \rightarrow [0, \infty)$  nous avons*

$$\mathbb{E} \sum_{\xi \in X} h(\xi, X \setminus \{\xi\}) = \int_W \mathbb{E} h(\xi, X) \rho(\xi) d\nu(\xi), \quad (2.1)$$

(le membre de gauche étant fini si et seulement s'il en est de même du membre de droite).

**Théorème 5** *Si  $X \sim \text{Poisson}(W, \rho)$ , alors pour tout  $n \in \mathbb{N}$  et toute fonction  $h : W^n \times \Omega_W \rightarrow [0, \infty)$ , nous avons*

$$\begin{aligned} & \mathbb{E} \sum_{\xi_1, \dots, \xi_n \in X}^{\neq} h(\xi_1, \dots, \xi_n, X \setminus \{\xi_1, \dots, \xi_n\}) \\ &= \int_W \dots \int_W \mathbb{E} h(\xi_1, \dots, \xi_n, X) \prod_{i=1}^n \rho(\xi_i) d\nu(\xi_1) \dots d\nu(\xi_n), \end{aligned} \quad (2.2)$$

le terme  $\neq$  signifiant que la somme ne porte que sur des ensembles  $\{\xi_1, \dots, \xi_n\}$  de points distincts.



Dans ce qui suit, nous définissons deux opérations sur les processus ponctuels. Nous montrons ensuite que la classe des processus de Poisson est fermée sous ces opérations.

**Définition 7** *L'union disjointe  $\cup_{i=1}^{\infty} X_i$  de processus ponctuels  $X_1, X_2, \dots$  est appelée superposition.*

**Définition 8** *Soit la fonction  $q : W \rightarrow [0, 1]$  et soit  $X$  un processus ponctuel sur  $W$ . Le processus ponctuel  $X_{thin} \subset X$  obtenu en gardant les points  $\xi \in X$  dans  $X_{thin}$  avec la probabilité  $q(\xi)$ , indépendamment l'un de l'autre, est appelé amincissement indépendant de  $X$ . Plus précisément, nous avons*

$$X_{thin} = \{\xi \in X : R(\xi) \leq q(\xi)\},$$

où les  $R(\xi) \sim \mathcal{U}[0, 1]$ ,  $\xi \in W$  sont des variables aléatoires mutuellement indépendantes et indépendantes de  $X$ .

**Proposition 4** *Pour  $i = 1, 2$ , soient les processus indépendants*

$$X_i \sim \text{Poisson}(W, \rho_i),$$

de mesures d'intensité  $\Upsilon_i = \int \rho_i(\xi) d\nu(\xi)$ , respectivement. Les deux processus  $X_1$  et  $X_2$  sont presque sûrement disjoints, plus précisément

$$\mathbb{P}(X_1 \cap X_2 = \emptyset) = 1,$$

si et seulement si  $\Upsilon_1(\{w\})\Upsilon_2(\{w\}) = 0$  pour tout  $w \in W$ <sup>5</sup>.

**Proposition 5** *Si les processus  $X_i \sim \text{Poisson}(W, \rho_i)$ ,  $i = 1, 2, \dots$  sont mutuellement indépendants et si  $\rho = \sum \rho_i$  est localement intégrable, alors  $X = \cup_{i=1}^{\infty} X_i$  est presque sûrement une superposition et  $X \sim \text{Poisson}(W, \rho)$ .*

**Proposition 6** *Soit  $X \sim \text{Poisson}(W, \rho)$  sur lequel on applique une opération d'amincissement selon les probabilités  $q(\xi)$  avec  $\xi \in W$  et soit  $\rho_{thin} = q(\xi)\rho(\xi)$ . Alors,  $X_{thin}$  et  $X \setminus X_{thin}$  sont des processus de Poisson indépendants de densités respectives  $\rho_{thin}$  et  $\rho - \rho_{thin}$ .*

Une forme induite par un processus de Poisson préserve son caractère poissonnien si on lui applique les opérations de superposition et d'amincissement. Cette propriété est bien utile pour simuler des processus ponctuels plus complexes. Par exemple, un processus ponctuel de Poisson hétérogène peut être obtenu par l'amincissement d'un processus de Poisson homogène.

---

5. Il existe une manière alternative de présenter cette propriété des processus de Poisson, en utilisant la "Disjointness Lemma" [101, 102, 32]

**Corollaire 1** Soit  $X \sim \text{Poisson}(W, \rho)$  avec  $\rho$  majoré par une constante positive  $C$ . Alors  $X$  suit la loi du processus Poisson  $(W, C)$  aminci par les probabilités  $p(\xi) = \rho(\xi)/C$ .

La définition d'un processus ponctuel de Poisson peut être étendue au cadre des processus ponctuels marqués. Même si a priori cela ne pose pas de difficulté immédiate, cela doit être fait avec quelques précautions. En effet, la loi des marques peut affecter la propriété d'indépendance ainsi que les propriétés de stationnarité et d'isotropie du processus.

**Définition 9** Soit  $Y \sim \text{Poisson}(W, \rho)$  et soient les marques  $\{m_\xi : \xi \in Y\}$  choisies de façon conditionnellement indépendante à  $Y$ . Alors  $X = \{(\xi, m_\xi) : \xi \in Y\}$  est un processus de Poisson marqué. Si les marques sont de même loi  $\nu_M$  sur  $(M, \mathcal{M})$ , alors  $\nu_M$  est la loi des marques.

Il existe des processus d'objets géométriques aléatoires dont la position est poissonnienne, mais avec des caractéristiques dépendantes de la configuration des objets [49, 82, 43, 44]. Ces processus ne rentrent pas dans le cadre des processus de Poisson marqués.

**Exemple 1 : quelques processus de Poisson marqués connus.** Si  $M = \{1, 2, \dots, k\}$ , nous avons un processus de Poisson multitype. Si  $M = \{K \subseteq W : K \text{ compact}\}$ , alors  $X$  est un schéma booléen. Pour plus des détails sur les processus Booléens nous recommandons [107, 122, 132].

Soient  $X_1$  et  $X_2$  deux processus de Poisson sur  $W$ . La loi de  $X_1$  est absolument continue par rapport à celle de  $X_2$  si et seulement si  $\mathbb{P}(X_2 \in F) = 0$  implique  $\mathbb{P}(X_1 \in F) = 0$  quand  $F \in \mathcal{F}_W$ . Par le théorème de Radon-Nykodym, il existe une fonction  $f : \Omega_W \rightarrow [0, \infty)$  tel que

$$\mathbb{P}(X_1 \in F) = \mathbb{E}[\mathbf{1}\{X_2 \in F\}f(X_2)], \quad F \in \mathcal{F}_W.$$

La fonction  $f$  est appelée la *densité* du processus  $X_1$  par rapport à  $X_2$  [27, 141].

**Proposition 7**

Soient  $X_1$  et  $X_2$  deux processus de Poisson sur  $W$ , de densités  $\rho_1$  et  $\rho_2$ , respectivement.

- (i) Si les densités  $\rho_1$  et  $\rho_2$  sont des constantes et  $W = \mathbb{R}^d$ , alors  $X_1$  est absolument continu par rapport à  $X_2$  si et seulement si  $\rho_1 = \rho_2$ .
- (ii) Si les densités  $\rho_1$  et  $\rho_2$  sont des fonctions intégrables sur le domaine  $W$

borné, tels que  $\rho_2(\xi) > 0$  si  $\rho_1(\xi) > 0$ . Alors,  $X_1$  est absolument continu par rapport à  $X_2$ . De plus, sa dérivée de Radon-Nycodym vaut

$$f(\mathbf{x}) = \exp[\Upsilon_2(W) - \Upsilon_1(W)] \prod_{\xi \in \mathbf{x}} \rho_1(\xi)/\rho_2(\xi),$$

pour toute configuration finie de points  $\mathbf{x} \subset W$  (et en considérant  $0/0 = 0$ )<sup>6</sup>.

Le résultat précédent montre que deux processus de Poisson ne sont pas toujours absolument continus l'un par rapport à l'autre. Cependant, un processus de Poisson est toujours continu par rapport au processus de Poisson standard lorsque  $W$  est borné.

Les schémas Booléens sont l'un des premiers modèles mathématiques développés pour la construction de formes. La forme induite par ces modèles est considérée comme la plus simple, car il n'y a pas d'interaction entre objets. Cette absence d'interaction est garantie par la propriété d'indépendance des processus. Elle permet aussi d'en déduire des formules analytiques précises pour le calcul des caractéristiques de ces formes. L'application immédiate de ces résultats est la construction des tests statistiques permettant de rejeter l'hypothèse d'absence d'interaction entre les objets composant une structure [90, 190, 132, 141, 180].

## 2.4 Processus ponctuels avec interaction

Pour créer des structures, pour faire naître des relations de dépendance entre objets, plusieurs voies existent. Une voie beaucoup explorée dans la littérature est celle des processus de type Cox, Neymann-Scott et shot-noise. Ce cadre de modélisation suppose que les objets se placent dans un espace doté d'un "relief" aléatoire. Ce "relief" dirige la structure de la forme du processus en influençant le regroupement ou l'éloignement des objets. Cette voie a été ouverte au début de la deuxième moitié du siècle passé. Elle a permis la création des très beaux modèles qui possèdent des propriétés analytiques remarquables et connaissent d'importantes applications dans des domaines comme l'astronomie, la biologie, les sciences de la terre et de l'environnement [31, 30, 3, 116, 141, 146].

Une autre voie est celle des processus ponctuels avec interaction. La densité de probabilité de ces modèles est construite directement à partir des interactions entre objets. Une première différence avec les modèles de type Cox est que la forme résulte directement des interactions entre objets, et non pas de

---

6. Une forme plus générale de ce résultat peut être trouvée dans ([123], Thm. 1.7.12, pp. 61)

l'influence d'une "substance de fond" qui prédéfinit la forme ... La deuxième différence est moins métaphysique : la plupart des processus avec interaction ne peuvent être connus en totalité. Il n'est donc pas possible d'effectuer des calculs analytiques. Cependant, leur construction se fait d'une façon très simple et très naturelle. Cette direction de recherche a démarré à peu près à la même époque, mais a suscité moins d'enthousiasme parmi les chercheurs en raison du manque de formules analytiques. Ce manque d'intérêt a commencé à diminuer avec la croissance de la puissance de calcul des machines et le développement des techniques Monte Carlo. Ainsi, à partir de la fin des années '90 et au début des années 2000, les modèles basés sur les processus ponctuels avec interaction se sont largement développés et servent remarquablement des domaines comme l'analyse d'image, la reconnaissance des formes, les sciences de l'environnement et l'astronomie [T1, 60, 61, 68, 72, 105, 189, 190, 141, 169, 170, 168, T13, T15, T20, T16, T21, T22, T23].

C'est la voie des processus avec interaction que j'ai choisie pour mon activité de recherche. La rédaction de cette partie suit particulièrement [72, 190, 141].

### 2.4.1 Construction de la densité de probabilité

Soit  $X$  un processus ponctuel sur  $W$ , sous-ensemble compact de  $\mathbb{R}^d$ , de densité  $p$  par rapport au processus de Poisson standard  $\text{Poisson}(W, 1)$ . La densité  $p$  est concentrée sur l'espace  $\Omega_W$  des configurations à nombre fini de points. Cet espace est muni de la tribu  $\mathcal{F}_W$  construite à partir des mesures de comptage sur les boréliens  $B \subseteq W$ .

La Proposition 1 nous permet d'écrire pour tout  $F \in \mathcal{F}_W$  :

$$\begin{aligned} \mathbb{P}(X \in F) &= \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W \mathbf{1}(\{w_1, \dots, w_n\} \in F) p(\{w_1, \dots, w_n\}) \\ &\quad \times d\nu(w_1) \dots d\nu(w_n), \end{aligned}$$

lorsque  $n > 0$ . Dans le cas  $n = 0$ , on prend  $\exp[-\nu(W)] \mathbf{1}(\emptyset \in F) p(\emptyset)$ . Si  $\nu(W) = 0$ , alors  $P(X = \emptyset) = 1$ . Pour les applications nous supposons toujours  $\nu(W) > 0$ .

En général, une telle densité de probabilité est connue à une constante près  $p \propto h$ . Ceci est dû au fait que pour la plupart de ces modèles, il est difficile de calculer analytiquement la constante de normalisation

$$c = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W h(\{w_1, \dots, w_n\}) d\nu(w_1) \dots d\nu(w_n). \quad (2.3)$$

La quantité  $c$  est appelée en physique statistique *fonction de partition*.

**Exemple 2 : constante de normalisation pour le processus de Poisson.** Soit la densité de probabilité d'un processus de Poisson d'intensité  $\rho$

$$p(\mathbf{w}) \propto \prod_{w_i \in \mathbf{w}} \rho(w_i)$$

et soit  $\Upsilon(B) = \int_B \rho(w) d\nu(w)$  la mesure d'intensité associée. Nous supposons  $0 < \Upsilon(B) < \infty$  pour tout  $B \subseteq W$ .

En utilisant (2.3), nous obtenons

$$c = \exp[-\nu(W)] \sum_{n=0}^{\infty} \frac{\Upsilon(W)^n}{n!} = \exp[\Upsilon(W) - \nu(W)],$$

ce qui nous donne pour la densité dans sa forme complète

$$p(\mathbf{w}) = \exp[\nu(W) - \Upsilon(W)] \prod_{w_i \in \mathbf{w}} \rho(w_i)$$

Ici nous avons spécifié la densité d'un processus ponctuel par rapport au processus de Poisson standard. La théorie de la mesure permet de spécifier des densités de processus ponctuels par rapport à des processus de Poisson de densité quelconque [141]. Ceci peut s'avérer utile dans certaines applications en traitement d'image [60, 61].

Ce cadre de construction de densités de probabilité peut être étendu au cas de processus ponctuels marqués.

Soit  $X$  un processus ponctuel marqué sur  $W \times M$  de densité  $p$  par rapport au processus de Poisson marqué standard  $\text{Poisson}(W, \times M)$ . La densité  $p$  est concentrée sur l'espace  $\Omega$  des configurations à nombre fini de points marqués ou d'objets. Cet espace est muni de la tribu  $\sigma$ -algèbre  $\mathcal{F}$  construit à partir des mesures de comptage sur les boréliens  $B \subseteq W \times M$ . Le processus de Poisson marqué standard choisit un nombre d'objet selon une loi de Poisson de paramètre  $\nu(W)$ , puis il répartit les objets indépendamment et uniformément dans  $W$ . Finalement à chaque objet est associé une marque, choisie indépendamment selon la loi  $\nu_M$ .

Tout comme précédemment, la Proposition 1 nous permet d'écrire pour tout

$F \in \mathcal{F}$  :

$$\begin{aligned} \mathbb{P}(X \in F) &= \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_{W \times M} \cdots \int_{W \times M} \mathbf{1}(\{(w_1, m_1), \dots, (w_n, m_n)\} \in F) \\ &\quad \times p(\{(w_1, m_1), \dots, (w_n, m_n)\}) \\ &\quad \times d\nu(w_1) \dots d\nu(w_n) d\nu_M(m_1) \dots d\nu_M(m_n). \end{aligned}$$

Dans ce qui suit, nous notons  $\mu$  la mesure de référence du processus de Poisson marqué standard. L'expression de cette mesure s'obtient à partir de l'expression précédente en posant  $p(\{(w_1, m_1), \dots, (w_n, m_n)\}) = 1$  pour tout  $n$ . Par la suite, nous ne considérons que les densités de probabilité de processus ponctuels marqués construites par rapport à la mesure de référence  $\mu$ .

La densité de probabilité d'un processus ponctuel marqué dont les objets interagissent se construit en précisant les fonctions d'interaction

$$\phi(x_{i_1}, \dots, x_{i_k})^{(k)} : \Omega \rightarrow \mathbb{R}^+$$

pour tout  $k$ -tuple d'objets. Alors, la densité de probabilité est proportionnelle au produit de toutes les fonctions d'interaction

$$p(\mathbf{x}) \propto \prod_{x_i \in \mathbf{x}} \phi(x_i)^{(1)} \cdots \prod_{\{x_{i_1}, \dots, x_{i_k}\} \in \mathbf{x}} \phi(x_{i_1}, \dots, x_{i_k})^{(k)}$$

Les densités de probabilité (2.4.1) constituent un cadre très souple pour la modélisation pourvu qu'elles soient intégrables sur l'espace des configurations

$$\alpha = \int_{\Omega} p(\mathbf{x}) d\mu(\mathbf{x}) < \infty.$$

Les deux résultats suivants assurent l'intégrabilité de la densité de probabilité d'un processus ponctuel marqué. Ils sont connus dans la littérature sous le nom de conditions de stabilité de Ruelle [72, 141, 171].

**Définition 10** *Soit  $X$  un processus ponctuel marqué de densité de probabilité  $p$  par rapport à la mesure de référence  $\mu$ . Le processus  $X$  est stable au sens de Ruelle s'il existe un nombre  $\Lambda > 0$  tel que*

$$p(\mathbf{x}) \leq \Lambda^{n(\mathbf{x})}, \quad \forall \mathbf{x} \in \Omega. \quad (2.4)$$

**Proposition 8** *La densité de probabilité d'un processus ponctuel stable est intégrable.*

**Preuve:** La démonstration figure dans ([72], Cond. 1, pp. 102). L'intégrabilité de  $p(\mathbf{x})$  découle directement de la condition précédente :

$$\begin{aligned} \int_{\Omega} p(\mathbf{x})\mu(d\mathbf{x}) &\leq \int_{\Omega} \Lambda^{n(\mathbf{x})}\mu(d\mathbf{x}) \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)][\Lambda\nu(W)]^n}{n!} = \exp[\nu(W)(\Lambda - 1)]. \end{aligned}$$

□

**Définition 11** *Sous les mêmes hypothèses que la Proposition 8, un processus ponctuel marqué est dit localement stable s'il existe  $\Lambda > 0$  tel que*

$$p(\mathbf{x} \cup \{\eta\}) \leq \Lambda p(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \eta \in W \times M \setminus \mathbf{x} \quad (2.5)$$

**Proposition 9** *Un processus ponctuel localement stable est stable au sens de Ruelle.*

**Preuve:** La preuve figure aussi dans ([72], Cond. 2, pp. 103). Il est facile de montrer par récurrence

$$p(\mathbf{x}) = p(\emptyset)\Lambda^{n(\mathbf{x})}, \quad \forall \mathbf{x} \in \Omega.$$

La stabilité locale d'un processus ponctuel (2.5) implique donc son intégrabilité (2.4). □

La stabilité locale implique la condition dite *d'hérédité*

$$p(\mathbf{x}) = 0 \Rightarrow p(\mathbf{y}) = 0, \quad \text{si } \mathbf{x} \subseteq \mathbf{y}.$$

Elle permet la définition de l'intensité conditionnelle comme le rapport

$$\lambda(\eta; \mathbf{x}) = \frac{p(\mathbf{x} \cup \{\eta\})}{p(\mathbf{x})}, \quad \mathbf{x} \in \Omega, \eta \in W \times M \setminus \mathbf{x}, \quad (2.6)$$

moyennant la convention  $0/0 = 0$ . L'intensité conditionnelle est aussi connue dans la littérature sous le nom *d'intensité conditionnelle de Papangelou*.

L'intensité conditionnelle est un élément clef dans la modélisation à partir de processus ponctuels. Ainsi, les intensités conditionnelles bornées impliquent l'intégralité des processus ponctuels induits. Qui plus est, les intensités conditionnelles bornées sont nécessaires pour établir des propriétés de convergence des chaînes de Markov qui sont à la base des dynamiques Monte Carlo utilisées pour simuler ces modèles [21, 22, 72, 99, 141, 190, T7, T8].

L'intensité conditionnelle fournit également des informations concernant la morphologie de la structure induite par la réalisation d'un processus objet. Un processus  $X$  est dit attractif si  $\mathbf{x} \subseteq \mathbf{y}$  implique

$$\lambda(\eta; \mathbf{x}) \leq \lambda(\eta; \mathbf{y}), \quad (2.7)$$

et répulsif dans le cas contraire

$$\lambda(\eta; \mathbf{x}) \geq \lambda(\eta; \mathbf{y}), \quad (2.8)$$

La morphologie d'un processus répulsif manifeste une certaine régularité, au sens où les centres des objets ont tendance à se distancer. Pour un processus attractif, les objets ont tendance à former un ou plusieurs agrégats. Il existe des processus qui ne sont ni répulsifs ni attractifs. Les propriétés d'attraction (2.7) et de répulsion (2.8) sont aussi très utiles pour la construction d'algorithmes de simulation exacte [21, 22, 99, 141, 190, T8]. Dans ce contexte, nous parlons de monotonie et d'antimonotonie.

## 2.4.2 Processus ponctuels markoviens

L'intensité de Papangelou peut être interprétée comme la probabilité conditionnelle  $\lambda(\mathbf{x}, (w, m))d\nu(w)d\nu_M(m)$  de contenir un objet  $(w, m)$  dans le domaine infinitésimal  $d\nu(w)d\nu_M(m)$  sachant que le processus coïncide avec  $\mathbf{x}$  hors de ce domaine. Son expression est donnée par

$$\lambda(\eta; \mathbf{x}) = \phi(\eta)^{(1)} \prod_{x_i \in \mathbf{x}} \phi(x_i, \eta)^{(2)} \dots \prod_{\{x_{i_1}, \dots, x_{i_k}\} \in \mathbf{x}} \phi(x_{i_1}, \dots, x_{i_k}, \eta)^{(k+1)} \quad (2.9)$$

Cette interprétation est appelée dans [190, 141], conditionnement extérieur. Une façon complémentaire d'étudier une structure est le conditionnement intérieur : on étudie la forme connaissant un objet qui en fait partie. Cette autre manière de procéder mène aux mesures de Palm qui permettent la construction de statistiques descriptives très utilisées dans la pratique [190, 141, 180].

L'intensité conditionnelle dans sa forme (2.9) est difficile à manipuler. Pour simplifier l'étude d'une structure, il est naturel d'envisager un ordre limite d'interaction. On peut par exemple ne considérer que les interactions par paires d'objets. Une autre simplification naturelle consiste à supposer que les interactions entre objets sont *locales*, c'est-à-dire qu'un objet ne puisse interagir qu'avec ses "voisins proches". Toutes ces simplifications imposent un caractère markovien aux formes étudiées. Ceci est l'hypothèse clef pour le cadre de modélisation que l'on a choisi. Dans ce cas, l'intensité conditionnelle joue un rôle similaire à la probabilité conditionnelle en un site dans le



cas des champs de Markov discrets [113, 202].

Soit  $\sim$  une relation d'interaction symétrique et réflexive sur des objets de  $W \times M$ . Il s'agit typiquement d'une relation de voisinage basée sur une métrique (Euclidienne, Hausdorff, etc.) ou bien sur une intersection d'ensembles.

**Définition 12** Une clique est une configuration  $\mathbf{x} \in \Omega$  telle que  $\eta \sim \zeta$  pour tout  $\eta, \zeta \in \mathbf{x}$ . L'ensemble vide est une clique.

**Définition 13** Soit  $X$  un processus ponctuel marqué sur  $W \times M$  de densité  $p$  par rapport à la mesure de référence  $\mu$ . Le processus  $X$  est markovien si pour tout  $\mathbf{x} \in \Omega$  tel que  $p(\mathbf{x}) > 0$ , les conditions suivantes sont simultanément vérifiées :

- (i)  $p(\mathbf{z}) > 0$  pour tout  $\mathbf{z} \subseteq \mathbf{x}$  (hérédité)
- (ii)  $\frac{p(\mathbf{x} \cup \{\zeta\})}{p(\mathbf{x})}$  ne dépend que de  $\zeta$  et de  $\partial(\zeta) \cap \mathbf{x} = \{\eta \in \mathbf{x} : \eta \sim \zeta\}$ .

Ce type de processus est connu dans la littérature comme un processus de Markov au sens de Ripley-Kelly [190, 161].

**Exemple 3** Par rapport à la mesure de référence  $\mu$ , la densité de probabilité d'un processus de Poisson marqué sur  $W \times M$  d'intensité constante ( $\rho(\eta) = \beta > 0$ ) est

$$p(\mathbf{x}) = \beta^{n(\mathbf{x})} \exp[(1 - \beta)\nu(W)].$$

Il est clair que  $p(\mathbf{x}) > 0$  pour toute configuration d'objets  $\mathbf{x}$ . Son intensité de Papangelou vaut

$$\lambda(\eta; \mathbf{x}) = \beta \mathbf{1}\{\eta \notin \mathbf{x}\}.$$

Le processus de Poisson est donc markovien, indépendamment des fonctions d'interaction  $\phi^{(k)}$ . Ceci est bien conforme au choix du processus de Poisson comme structure complètement aléatoires [90, 190].

Le résultat suivant, connu sous le nom de *propriété de Markov spatiale*, généralise la propriété (ii) de la définition d'un processus ponctuel markovien. On peut aussi le voir comme l'équivalent spatial de la propriété de Markov temporelle [190, 141, 161].

**Théorème 6** Soit  $X$  un processus marqué markovien sur  $W \times M$  de densité  $p(\cdot)$ , et soit  $A$  un borélien de  $W \times M$ . Alors la loi conditionnelle de  $X \cap A$  sachant  $X \cap A^c$  ne dépend que de la partie de  $X$  située dans le voisinage

$$\partial(A) \cap A^c = \{u \in W \setminus A : u \sim a \text{ for some } a \in A\}.$$

Le résultat suivant est connu dans la littérature en tant que théorème de factorisation Hammersley-Clifford [11, 190, 141, 161].

**Théorème 7** *Le processus ponctuel de densité  $p : \Omega \rightarrow \mathbb{R}^+$  est markovien par rapport à la relation  $\sim$  si et seulement s'il existe une fonction mesurable  $\phi_c : \Omega \rightarrow \mathbb{R}^+$  telle que*

$$p(\mathbf{x}) = \prod_{\substack{\text{cliques} \\ \mathbf{z} \subseteq \mathbf{x}}} \phi_c(\mathbf{z}), \quad \alpha = \phi_c(\emptyset) \quad (2.10)$$

pour tout  $\mathbf{x} \in \Omega$ .

Ce théorème simplifie l'écriture de la densité de probabilité d'un processus ponctuel avec interaction (2.4.1) en ne considérant que le produit de fonctions d'interaction sur des cliques. Pour un processus ponctuel markovien, l'équivalence entre (2.10) et (2.4.1) est obtenue en posant  $\phi_c(\mathbf{z}) = 1$  lorsque  $\mathbf{z}$  n'est pas une clique.

Les processus ponctuels de Markov sont connus par la communauté de la physique statistique sous le nom de *processus ponctuels de Gibbs* [141, 180]. Les processus de Gibbs décrivent des systèmes à grand nombre de particules. La densité de probabilité d'un processus ponctuel de Gibbs est

$$p(\mathbf{x}) = \frac{1}{Z} \exp[-U(\mathbf{x})] = \frac{1}{Z} \exp \left[ - \sum_{\substack{\text{cliques} \\ \mathbf{z} \subseteq \mathbf{x}}} U_c(\mathbf{z}) \right],$$

où  $Z$  est la fonction de partition,  $U$  l'énergie du système et  $U_c = \log \phi_c$  le potentiel d'une clique. Tous les processus ponctuels de Markov peuvent être vus comme des processus de Gibbs. La réciproque n'est pas vraie.

Les exemples suivants analysent le caractère markovien des quelques processus ponctuels connus.

**Exemple 4** *Revenons sur le processus de Poisson de l'exemple précédent. Il a pour densité*

$$p(\mathbf{x}) = e^{(1-\beta)\nu(W)} \prod_{x \in \mathbf{x}} \beta.$$

*Ainsi, les fonctions d'interaction appliquées aux cliques sont*

$$\begin{aligned} \phi_c(\emptyset) &= e^{(1-\beta)\nu(W)} \\ \phi_c(\{\eta\}) &= \beta \end{aligned}$$

*et  $\phi_c \equiv 1$  si les cliques ont deux objets ou plus. Il en résulte que le potentiel des cliques à un objet est*

$$U_c(\eta) = -\log \beta,$$

alors qu'il vaut  $U_c = 0$  pour les cliques à plus d'un objet. Ceci confirme bien l'absence d'interaction dans le processus de Poisson, ainsi que le choix de ce processus pour modéliser des formes simples, sans structure morphologique particulière [90, 190].

**Exemple 5** *Modèle de Widom-Rowlinson.* Ce modèle de mélange pour des sphères superposables [200] est décrit par l'espace des marques  $M = \{1, 2\}$  et la densité

$$p(\mathbf{x}) = \alpha \prod_{(w,m) \in \mathbf{x}} \beta_m \prod_{(u,1),(v,2) \in \mathbf{x}} \mathbf{1}\{\|u - v\| > r\} \quad (2.11)$$

par rapport au processus de Poisson multitype standard sur  $W \times M$  avec  $\nu_M(1) = \nu_M(2)$ . Ici  $\alpha$  est la constante de normalisation. Ainsi, les sphères de type différent sont à distance au moins  $r > 0$  l'une de l'autre, tandis que les particules du même type s'agrègent dans les espaces laissés libres par les particules de l'autre type. Les paramètres  $\beta_1$  et  $\beta_2$  contrôlent l'intensité des particules de type 1 et 2, respectivement.

L'intensité conditionnelle qu'il faut pour ajouter l'objet  $(w, 1) \notin \mathbf{x}$  à la configuration  $\mathbf{x}$  est

$$\lambda((w, 1); \mathbf{x}) = \beta_1 \mathbf{1}\{\|u - w\| > r \text{ pour tous les } (u, 2) \in \mathbf{x}\};$$

une expression similaire est valable pour ajouter un objet de type 2. Le modèle Widom-Rowlinson est héréditaire et localement stable avec

$$\Lambda = \max\{\beta_1, \beta_2\}.$$

De plus,  $\lambda((w, m); \mathbf{x}') \geq \lambda((w, m); \mathbf{x})$  pour tout  $\mathbf{x}' \subseteq \mathbf{x}$  et  $(w, m) \in W \times M$ .

Les fonctions d'interaction appliquées sont

$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w, m)\}) &= \beta_m \\ \phi_c(\{(u, 1), (v, 2)\}) &= \mathbf{1}\{\|u - v\| > r\} \end{aligned}$$

ainsi que  $\phi_c \equiv 1$  si les cliques ont deux objets ou plus de même type.

**Exemple 6** *Processus multitype avec interactions par paires.* Nous prenons pour espace des marques  $M = \{1, \dots, I\}$  avec  $I \in \mathbb{N}$ , et pour  $\nu_M$  la distribution uniforme sur  $M$ . Les processus multitype avec interaction par paire [11,

161] sont définis par leur densité  $p$  par rapport au processus de Poisson multitype standard sur  $W \times M$ . La densité  $p$  a pour expression

$$p(\mathbf{x}) = \alpha \prod_{(w,m) \in \mathbf{x}} \beta_m \prod_{(u,i) \neq (v,j) \in \mathbf{x}} \gamma_{ij}(\|u - v\|), \quad (2.12)$$

le deuxième produit portant sur toutes les paires de points marqués différents. Les paramètres  $\beta_m > 0$ ,  $m \in M$  contrôlent l'intensité des points de type  $m$ . L'interaction entre chaque paire de types d'objet  $i, j \in M$  est décrite par la fonction mesurable  $\gamma_{ij} : [0, \infty) \rightarrow [0, 1]$ . Nous supposons que les interactions entre objets sont symétriques, c'est à dire  $\gamma_{ij} \equiv \gamma_{ji}$  pour tous les  $i, j \in M$ . Le modèle (2.12) est héréditaire et localement stable avec  $\Lambda = \max_{m \in M} \beta_m$ .

Pour  $(w, m) \notin \mathbf{x}$ , l'intensité conditionnelle de Papangelou

$$\lambda((w, m); \mathbf{x}) = \beta_m \prod_{(u,i) \in \mathbf{x}} \gamma_{im}(\|u - w\|)$$

est décroissante, par rapport à la relation d'ordre induite par l'inclusion.

Les fonctions d'interaction sont

$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w, m)\}) &= \beta_m \\ \phi_c(\{(u, i), (v, j)\}) &= \gamma_{ij}(\|u - v\|) \end{aligned}$$

avec  $\phi_c \equiv 1$  si les cliques ont trois objets ou plus.

## Chapitre 3

# Contributions

Notre problème est le suivant : nous observons des données spatialisées  $\mathbf{d}$  dans une fenêtre  $W$  et nous souhaitons trouver les structures présentes dans les observations. Par hypothèse, ces structures sont obtenues à partir d'objets simples qui interagissent. Notant  $M$  l'espace des paramètres de ces objets, nous supposons que la forme recherchée est la réalisation d'un processus ponctuel marqué sur  $W \times M$ .

Dans le cadre des processus de Gibbs, la densité de probabilité d'un tel processus s'écrit

$$p(\mathbf{x}|\theta) = \frac{\exp[-U(\mathbf{x}|\theta)]}{Z(\theta)} = \frac{\exp[-(U_{\mathbf{d}}(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta))]}{Z(\theta)}, \quad (3.1)$$

où  $\theta$  est le vecteur de paramètres et  $Z(\theta)$  la fonction de partition du modèle. L'énergie totale du système,  $U(\mathbf{x})$  est la somme de deux termes. Le premier terme,  $U_{\mathbf{d}}(\mathbf{x}|\theta)$  représente l'énergie d'attache aux données et concerne la position des objets dans  $W$ . Le deuxième terme,  $U_i(\mathbf{x}|\theta)$  représente l'énergie d'interaction et concerne l'interaction entre objets.

La densité (3.1) est définie par rapport à la mesure de référence  $\mu$ . Le processus décrit par cette équation peut être vu aussi comme la superposition d'un processus de Poisson hétérogènes et d'un processus d'interaction

$$p(\mathbf{x}|\theta) \propto p_{\mathbf{d}}(\mathbf{x}|\theta)p_i(\mathbf{x}|\theta).$$

Le processus de Poisson hétérogène est décrit par

$$p_{\mathbf{d}}(\mathbf{x}|\theta) = \frac{\exp[-U_{\mathbf{d}}(\mathbf{x}|\theta)]}{Z_{\mathbf{d}}(\theta)}, \quad (3.2)$$

avec  $U_{\mathbf{d}}(\mathbf{x}|\theta) = \sum_{x \in \mathbf{x}} u(x|\theta)$  et  $u$  une fonction mesurable, alors que le processus d'interaction est écrit sous la forme

$$p_i(\mathbf{x}|\theta) = \frac{\exp[-U_i(\mathbf{x}|\theta)]}{Z_i(\theta)}, \quad (3.3)$$

$Z_d(\theta)$  et  $Z_i(\theta)$  étant les constantes de normalisation associées respectivement à (3.2) et à (3.3) .

**Exemple 7 Rôle des composantes attache aux données et interaction.** *La Figure 3.1a montre une image satellite contenant un morceau du réseau routier. Le réseau forme la lettre “X”, le bras en bas à droite de l’image étant un peu diffus. Sur le bras en haut à gauche, on observe une route secondaire. Ce type de croisement est aussi observé sur le bras en bas à droite. La radiométrie ou la couleur de la route est plutôt homogène, et l’on peut dire que les routes dans cette image ont tendance à être “blanches”. Nous observons également des régions dans l’image où des structures filiformes peuvent être observées. Il est cependant difficile de les considérer comme faisant partie du réseau routier. Ce type d’objets représenterait plutôt une frontière entre champs.*

*Supposons le réseau formé par un ensemble de petits segments qui se connectent et qui s’alignent, tout au long de la route. Dans ce cas, le cadre des processus ponctuels permet approcher le réseau routier par la réalisation d’un processus marqué défini par une densité de la forme (3.1). Les milieux des segments représentent la position des objets cherchés, alors que leur longueurs et leur orientations représentent leur marques.*

*Sans entrer dans les détails, si le modèle utilisé n’était formé que d’un terme d’attache aux données (3.2), le résultat que l’on obtiendrait est présenté Figure 3.1b. Nous observons que l’on peut retrouver avec précision les grandes structures du réseau. Cependant, beaucoup de fausses alarmes sont également détectées. Le terme d’attache aux données localise assez bien les objets qui nous intéressent, mais n’élimine pas les objets indésirables, c’est-à-dire les objets qui pris ensemble ne forment pas un réseau. En fait, le problème de détection d’objets défini uniquement à partir de (3.2) est mal conditionné.*

*Si la densité ne comportait qu’un terme d’interaction entre les segments, alors le réseau que l’on obtiendrait ressemble à celui de la Figure 3.1c. Nous pouvons remarquer que le réseau qui se forme a bien les propriétés souhaitées. Les segments qui forment le réseau sont connectés et relativement bien alignés. Il y a très peu de petits segments isolés. Les croisements entre les différentes composantes du réseau sont autorisés. Cependant, le réseau obtenu est sans rapport avec le réseau que l’on souhaiterait détecter, l’énergie d’interaction ne dépendant pas des données.*

*C’est la combinaison de ces deux termes qui donne un résultat “proche” du réseau observé Figure ??d. Le terme d’attache aux données détecte les objets susceptibles de faire partie du réseau, selon des principes de filtrage morpho-*

*logique [175]. Le terme d'interaction régularise la solution, en ne gardant que les segments qui respectent au mieux les contraintes topologiques du réseau.*

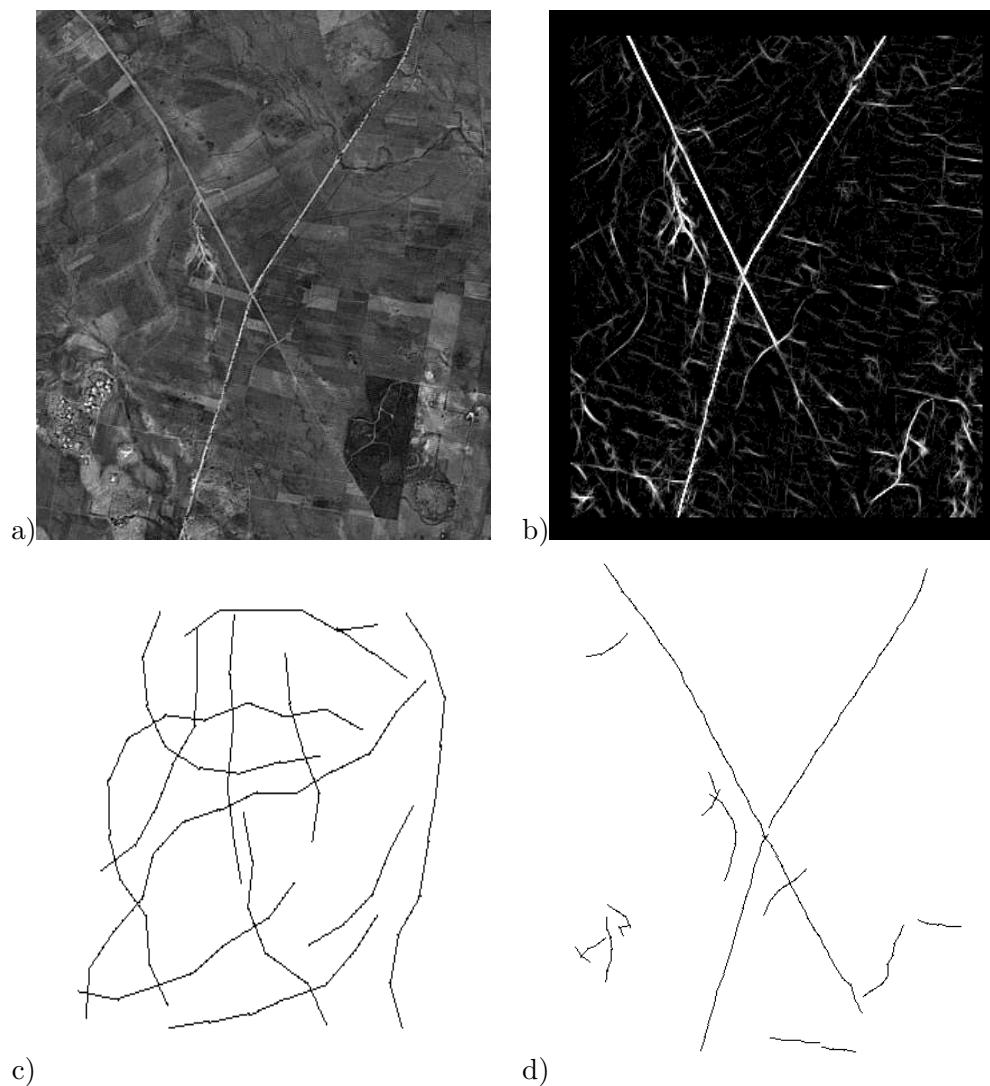


FIGURE 3.1 – Influence des différentes composantes d'un modèle pour la détection d'objets dans une image : a) image originelle SPOT, b) résultat obtenu en utilisant uniquement le terme d'attache aux données, c) résultat obtenu en utilisant uniquement le terme d'interaction, d) résultat obtenu en utilisant le modèle complet.

### 3.1 Modélisation des filaments galactiques : les processus Candy et Bisous

Le modèle Candy a été construit pendant la thèse. Ce modèle a été proposé comme solution possible au problème de l'extraction de réseaux linéiques dans des images satellitaires et aériennes [T2, 179, T14, T1, T15]. Ce processus ponctuel marqué a comme réalisations des configurations de segments en interaction : alignement, répulsion, connexion.

Après la thèse, le modèle Candy a été étudié du point de vue théorique [T7], tout en gardant le lien avec les applications [T3, T13, T20].

Sur cette base mathématique, une généralisation, appelée modèle Bisous, a été proposée ensuite [T4, T17]. Dans ce cas aussi, la motivation a été d'ordre pratique. La question intéresse les cosmologistes. Il est connu aujourd'hui que la distribution des galaxies dans notre univers n'est pas uniforme. Cette distribution répartit les galaxies en structures complexes comme des agrégats, des murs ou des filaments [116]. Le modèle Bisous est un processus ponctuel marqué capable de générer ce type de structures à partir d'objets simples en interaction. Une fois ce modèle construit, plusieurs travaux en coopération avec les cosmologistes ont été effectués. Ces travaux ont eu comme objectif la détection et la caractérisation morphostatistique de filaments galactiques [T21, T22, 184].

#### 3.1.1 Modélisation de l'interaction entre les objets

Pour construire une forme à partir d'objets en interaction dans le cadre des processus ponctuels marqués, il faut définir les objets, les interactions entre les objets ainsi que les fonctions d'interaction. Tout cela se formule au moyen d'une densité de probabilité.

Les objets que nous considérons sont de taille et de forme déterministes. Seules leur position et leur orientation sont aléatoires.

La marque représente l'orientation de l'objet. Si  $W \subset \mathbb{R}^2$ , l'espace des marques est  $M = [0, \pi)$ . Si  $W \subset \mathbb{R}^3$ , alors  $M = [0, 2\pi) \times [0, 1]$ . Dans ce cas, l'orientation est le vecteur

$$m(m_1, m_2) = (\sqrt{1 - m_2^2} \cos(m_1), \sqrt{1 - m_2^2} \sin(m_1), m_2).$$

Pour  $\nu_M$  nous choisissons la loi uniforme.



Il est commode d'attribuer des extrémités à tout objet. Il s'agit d'un ensemble de points rigides  $\{e_u, 1 \leq u \leq q\}$ . Ces points servent à construire une région d'attraction autour de l'objet.

**Définition 14** *La région d'attraction  $a(w)$  autour de l'objet  $x = (w, m)$  est définie comme une union disjointe de boules  $a(x) = \bigcup_{u=1}^q b(e_u, r_a)$ ,  $b(e_u, r_a)$  désignant la boule de rayon  $r_a$  centrée en  $e_u$ .*

Pour former des structures, les objets interagissent à travers leur région d'attraction. Une règle d'attraction  $\mathcal{C}$  entre deux objets  $x_i$  et  $x_j$  est une assertion logique portant sur l'ensemble des couples d'extrémités que l'on peut former à partir des deux objets. Par exemple, l'attraction peut être permise lorsque un couple d'extrémités est séparé d'une distance inférieure à  $r_a$ .

**Définition 15** *Soit deux objets  $x_1 = (w_1, m_1)$  et  $x_2 = (w_2, m_2)$ .*

*Les objets sont en attraction  $x_1 \sim_a x_2$ , si l'ensemble  $\{(u, v) : 1 \leq u, v \leq q, d(e_u(x_1), e_v(x_2)) \leq r_a\}$  contient un seul élément vérifiant  $\mathcal{C}$ .*

*Les objets sont en répulsion  $x_1 \sim_h x_2$ , si  $d(w_i, w_j) \leq 2r_h$  pour une valeur  $r_h > 0$  fixée.*

*Les objets sont alignés  $x_1 \sim_{\parallel} x_2$ , si*

$$m_1 \cdot m_2 \geq 1 - \tau$$

*où  $\tau \in (0, 1)$  est un paramètre de courbure fixé et  $\cdot$  est le produit scalaire des vecteurs d'orientation<sup>1</sup>.*

*Les objets sont connectés  $x_1 \sim_s x_2$ , si les trois propriétés suivantes sont respectées (simultanément)*

$$\begin{aligned} x_1 &\sim_a x_2 \\ x_1 &\sim_{\parallel} x_2 \\ x_1 &\not\sim_h x_2 \end{aligned}$$

**Définition 16** *Soit  $\mathcal{R}$  l'ensemble de tous les interactions par paires d'objets de  $W \times M$  qui sont à la fois locales<sup>2</sup>, symétriques et qui induisent une application mesurable sur  $\mathcal{F}$ .*

Toutes les relations qui ont été définies appartiennent à  $\mathcal{R}$ , hormis  $\sim_{\parallel}$  qui n'est pas locale.

---

1. Ici nous avons considéré que  $W \subseteq \mathbb{R}^3$ .

2. Une interaction locale est une interaction à portée finie.

**Définition 17** *Un objet est dit  $s$ -connexe s'il est connecté à travers exactement  $s$  de ses extrémités. Si  $s = 0$ , l'objet est dit libre.*

Deux objets sont dits de même type s'ils sont définis sur le même espace de paramètres, s'ils ont la même position et le même nombre d'extrémités, et s'ils suivent la même règle d'attraction. Jusqu'à présent, nous n'avons considéré que les structures formées par des objets de même type.

**Exemple 8 Structure linéaire.** *Un réseau linéaire peut être généré à partir de segments en interaction. La Figure 3.2a montre un segment de longueur  $2(r + r_a)$ , centré à l'origine. Il possède les  $q = 2$  extrémités suivantes*

$$e_u = (0, 0, (r + r_a) \sin \frac{(2u - 1)\pi}{2}), \quad 1 \leq u \leq q$$

*et son vecteur d'orientation est  $\omega = (0, 0, 1)$ .*

*Le rayon de répulsion est  $r_h = r$  alors que la règle d'attraction est  $\mathcal{C} = \{(1, 2), (2, 1)\}$ . Donc deux segments peuvent se connecter par extrémités d'indices différents. Tout segment de milieu et d'orientation aléatoires dans  $\mathbb{R}^3$  peut être obtenu à partir du segment de cet exemple, par une translation et deux rotations [88].*

**Structure planaire.** *La Figure 3.2b montre comment construire une structure planaire à partir d'un hexagone centré à l'origine. Son vecteur d'orientation est  $\omega = (0, 0, 1)$  et il possède  $q = 6$  extrémités données par*

$$e_u = ((r + r_a) \cos \frac{(2u - 1)\pi}{6}, (r + r_a) \sin \frac{(2u - 1)\pi}{6}, 0), \quad 1 \leq u \leq q.$$

*La région de répulsion est identique au cas précédent. La règle d'attraction est définie comme suit : le couple d'extrémités  $(e_u(x_i), e_v(x_j))$  accomplit les conditions d'attraction lorsque  $|u - v| = 3$ . Ceci donne les couples d'indices  $\mathcal{C} = \{(1, 4), (4, 1), (2, 5), (5, 2), (3, 6), (6, 3)\}$ .*

**Structure volumique.** *Pour former une structure volumique nous allons partir d'un cuboctaèdre comme élément générateur. La Figure 3.2c montre un cuboctaèdre centré, d'orientation  $\omega = (0, 0, 1)$ . Il possède  $q = 12$  points extrémaux qui s'écrivent*

$$e_u = ((r + r_a) \cos(\alpha_u + \beta_u), (r + r_a) \sin(\alpha_u + \beta_u), (r + r_a) \sin(\beta_u/2)) \quad 1 \leq q \leq 12,$$

*où  $\alpha_u = \frac{\pi}{4}(2u - 1)$  et  $\beta_u = \frac{\pi}{2} \left[ \frac{u-5}{4} \right]$ .*

*Autour de l'objet, la région de répulsion est construite comme dans les deux cas précédents. La règle  $\mathcal{C}$  d'attraction entre deux objets  $x_i$  et  $x_j$  s'établit*

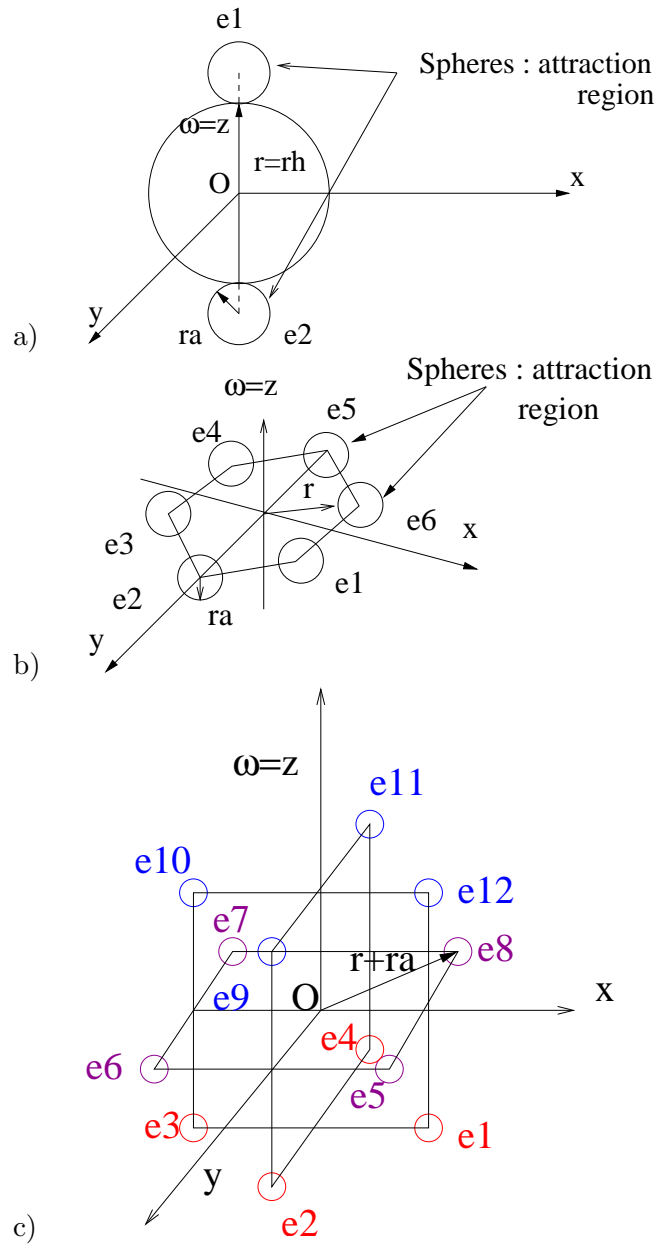


FIGURE 3.2 – Éléments générateurs : a) structure linéaire, b) structure planaire, c) structure volumique

par extrémités opposées, d'indices (1, 10), (2, 11), (3, 12), (4, 9), (5, 7), (6, 8) et leur symétrie.

**Lemme 1** *Le nombre  $n_s$  d'objets  $s$ -connexes d'une configuration d'objets de même type  $\mathbf{x} \in \Omega$ , est mesurable par rapport à  $\mathcal{F}$ .*

**Preuve:** Il s'agit d'une adaptation du résultat ([T7], Lemma 1, pp. 179). Ici, nous montrons la mesurabilité du nombre  $n_0$  d'objets libres. La mesurabilité des autres applications s'obtient par des arguments similaires.

Selon ([157], Section 3.1), nous devons montrer la mesurabilité des fonctions symétriques

$$f(x_1, \dots, x_n) : (W \times M)^n \rightarrow \sum_{i=1}^n \mathbf{1}\{x_i \text{ est libre}\}$$

pour tout  $n \in \mathbb{N}^*$ .

Pour tout  $i \neq j \in \{1, \dots, n\}$ , nous définissons  $f_{ij}$  tel que

$$f_{ij}(x_1, \dots, x_n) = \mathbf{1}\{x_i \not\sim_a x_j\} + \mathbf{1}\{x_i \sim_a x_j\} \mathbf{1}\{x_i \sim_h x_j\} + \mathbf{1}\{x_i \sim_a x_j\} \mathbf{1}\{x_i \not\sim_h x_j\} \mathbf{1}\{x_i \not\sim_{\parallel} x_j\}$$

La fonction  $f_{ij}$  est borélienne car c'est une application sur  $(W \times M)^n$ . Par conséquent, la fonction

$$\mathbf{1}\{x_i \text{ is free}\} = \prod_{i \neq j} f_{ij}(x_1, \dots, x_n)$$

est aussi borélienne et il en est de même pour la somme de ces fonctions indexées par  $i$ .  $\square$

Pour une configuration d'objets de même type  $\mathbf{x} = \{x_1, \dots, x_n\}$ , la densité de probabilité

$$p(\mathbf{x}) = p(\emptyset) \left[ \prod_{s=0}^q \gamma_s^{n_s(\mathbf{x})} \right] \prod_{\kappa \in \Gamma \subset \mathcal{R}} \gamma_{\kappa}^{n_{\kappa}(\mathbf{x})} \quad (3.4)$$

définit le modèle *Bisous* par rapport à la mesure de référence du processus de Poisson standard.

Les valeurs  $\gamma_s > 0, \gamma_{\kappa} \in [0, 1]$  représentent des paramètres du modèle et  $p(\emptyset)$  est la constante de normalisation. L'ensemble  $\Gamma$  contient les relations incluses dans  $\mathcal{R}$  qui sont prises en compte pour construire la forme  $\mathbf{x}$ . Si  $\gamma_{\kappa} = 0$ , l'interaction  $\sim_{\kappa}$  est dite hard-core (ici nous posons  $0^0 = 1$ ). Pour

chaque  $s$  et  $\kappa$  les statistiques suffisantes  $n_s(\mathbf{x})$  et  $n_\kappa(\mathbf{x})$  représentent respectivement, le nombre d'objets  $s$ -connexes et le nombre de paires d'objets ayant l'interaction  $\kappa$ , dans la configuration  $\mathbf{x}$ .

**Lemme 2** *Le modèle Bisous défini par (3.4) est localement stable.*

**Preuve:** Soit  $W$  un compact de  $\mathbb{R}^3$ . Si  $\sim_h \in \Gamma$  et si  $\gamma_h = 0$ , alors il existe un nombre maximum d'objets  $n_{\max}$  qui peut être contenu dans  $W$ .

Alors, l'intensité conditionnelle est bornée par la quantité

$$\lambda(\zeta; \mathbf{x}) \leq \prod_{s=0}^q \max\{\gamma_s, \gamma_s^{-1}\}^{n_{\max}} = \Lambda.$$

Le modèle est localement stable même si  $\Gamma$  ne contient pas d'interaction de type hard-core. En fait, le nombre maximal d'objets non-connectés qui peuvent se connecter à un autre objet est 12. Ce nombre est connu sous le nom de "kissing number"<sup>3</sup>. La borne pour l'intensité conditionnelle devient ainsi

$$\Lambda = \prod_{s=0}^q \max\{\gamma_s, \gamma_s^{-1}\}^{12}. \quad (3.5)$$

Comme la connexité suppose aussi les objets alignés, il est possible d'attribuer à l'intensité conditionnelle (3.5) des bornes encore plus resserrées qui dépendent du paramètre de courbure  $\tau$ .  $\square$

**Théorème 8** *La densité de probabilité du modèle Bisous (3.4) définit bien un processus ponctuel.*

**Preuve:** La  $(\Omega, \mathcal{F})$ -mesurabilité du modèle Bisous découle de la définition de  $\mathcal{R}$  et du Lemme 1. La stabilité locale implique la stabilité au sens de Ruelle (2.4), et donc l'intégrabilité du modèle.  $\square$

Le modèle Bisous est construit sur des interaction par paires, qui sont symétriques et locales, c'est-à-dire à portée finie. Par exemple, la  $s$ -connexité est symétrique et sa portée est de  $4r_h + 2r_a$  [T7, T17]. En utilisant ces arguments et en supposant les ensembles vides ainsi que les singletons comme des cliques, on prouve que les modèles Candy et Bisous sont markoviens au sens de (Ripley-Kelly) [161, 190, 141].

---

3. Le problème dont il est la réponse a été à l'origine une discussion mathématique entre Isaac Newton et David Gregory ([http://en.wikipedia.org/wiki/Kissing\\_number\\_problem](http://en.wikipedia.org/wiki/Kissing_number_problem)). L'appellation "kissing number" nous a inspiré également, pour le nom de notre modèle.

La Figure 3.3 montre des structures obtenues en utilisant le modèle Bisous. Nous observons que le modèle est capable de construire aussi bien des réseaux linéiques que des surfaces ou des agrégats.

### 3.1.2 Modélisation de l’attache aux données

Pour pouvoir localiser des structures dans un jeu de données, il faut rajouter au modèle, un terme d’attache aux données, comme expliqué plus haut dans ce chapitre.

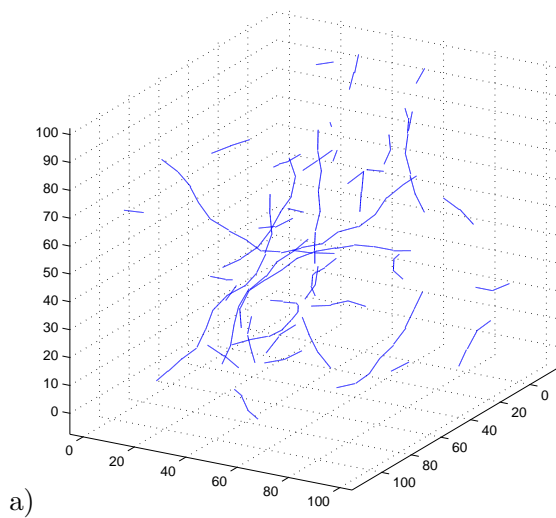
Il n’y a pas de méthode précise pour construire le terme d’attache aux données. En règle générale, il doit contenir la définition de ce que l’on cherche, il doit être facile à calculer et il doit garantir que le modèle complet est bien défini, c’est à dire intégrable. Dans ce qui suit, nous présentons les idées qui ont servi à construire le terme d’attache aux données pour la détection de filaments galactiques [T20, T21, T22].

Les données utilisées ici sont les positions d’un nombre fini de galaxies observées dans une fenêtre  $W$  de volume fini. Les filaments galactiques sont données par le réseau linéique induit par la position des galaxies. La Figure 3.4 montre un champ de galaxies observées dans un angle solide. On se limitera toutefois à un volume rectangulaire, afin que le champ de galaxies à l’intérieur préserve une certaine homogénéité : à cause de l’éloignement, seules les galaxies les plus brillantes sont observées. Plus on s’éloigne de l’observateur, plus la densité du champ de galaxies observées diminue.

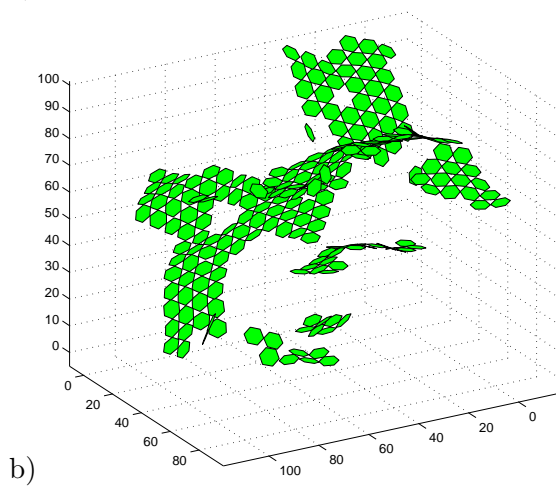
Pour la partie “interaction”, celle qui génère les filaments, nous avons utilisé le modèle Bisous pour des structures linéaires. Ainsi, un réseau est construit à partir de segments qui interagissent pour se connecter et s’aligner. La densité de probabilité de ce processus objet est donnée par

$$p_i(\mathbf{x}|\theta) = \gamma_\kappa^{n_\kappa(\mathbf{x})} \prod_{s=0}^2 \gamma_s^{n_s(\mathbf{B})},$$

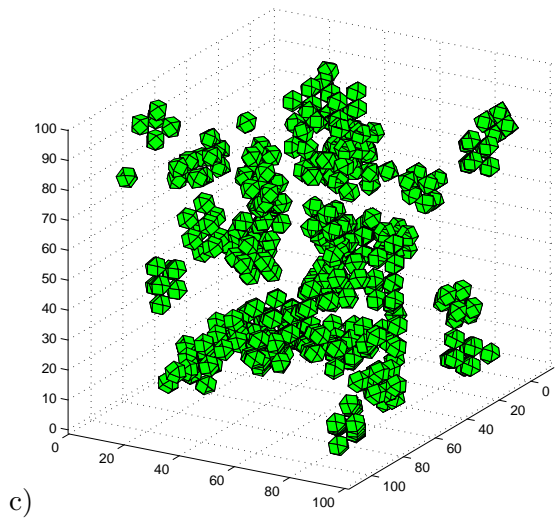
ce qui met en jeu les paramètres  $\gamma_s$  et les statistiques  $n_s$  définies comme plus haut. L’interaction  $\sim_\kappa$  combine la répulsion  $\sim_h$  avec un positionnement orthogonal  $\sim_\perp$  de deux objets. Deux objets sont dits orthogonaux  $x_1 \sim_\perp x_2$  si et seulement si  $|m_1 \cdot m_2| \leq \tau_\perp$ , avec  $\tau_\perp \in (0, 1)$  prédéfini. Par conséquent, pour deux objets  $x_1$  et  $x_2$ , nous pouvons écrire  $x_1 \sim_\kappa x_2$  si nous avons simultanément  $x_1 \sim_h x_2$  et  $x_1 \sim_\perp x_2$ . Pour que le modèle soit bien défini on doit prendre  $\gamma_\kappa \in [0, 1]$ . Les statistiques  $n_\kappa$  représentent le nombre de paires de segments en situation de répulsion et d’orthogonalité dans la configuration  $\mathbf{x}$ .



a)



b)



c)

FIGURE 3.3 – Formes générées à l'aide du modèle Bisous : a) structure linéaire, b) structure planaire, c) structure volumique

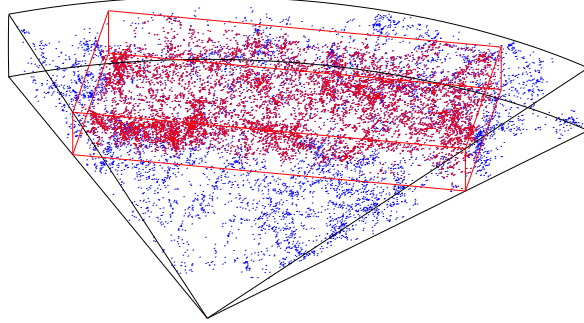


FIGURE 3.4 – Échantillon rectangulaire dans l’hémisphère nord du catalogue 2dF Redshift Survey. Le diamètre d’une galaxie est  $\sim 30 \times 3261.6$  années lumière.

L’énergie d’interaction du système s’écrit sous la forme suivante :

$$U_i(\mathbf{x}|\theta) = -n_\kappa \log \gamma_\kappa - \sum_{s=0}^2 n_s(\mathbf{x}) \log \gamma_s \quad (3.6)$$

Pour la partie “attache aux données”, celle qui positionne les filaments, nous procédons de la manière suivante. A chaque segment  $x$  du réseau linéique, nous lui attachons un cylindre  $s(x) = s(w, m)$  de rayon fixe  $r$  et d’axe de symétrie spécifié par les paramètres du segment, le centre  $w$  ; la longueur  $2r_h$  et le vecteur d’orientation  $m$ . A chaque cylindre est associé un extra-cylindre de paramètres identiques au précédent à l’exception du rayon qui est  $2r$ . Soit  $\tilde{s}(x)$  l’ombre de  $x$ , qui est la différence ensembliste entre l’extra-cylindre et le cylindre comme l’indique la Figure 3.5. Le cylindre  $s(x)$  est divisé en trois parties de volume égal, que nous allons noter  $s_1(x)$ ,  $s_2(x)$  et  $s_3(x)$ .

Un cylindre peut faire partie du réseau, s’il vérifie plusieurs conditions. Premièrement, la densité de galaxies à l’intérieur d’un cylindre doit être supérieure à la densité de galaxies à l’intérieur de son ombre. Nous écrivons cette condition de la manière suivante

$$\mathbf{1}\{\text{“densité”}\} = \mathbf{1}\{n(\mathbf{d} \cap s(x))\nu(\tilde{s}(x)) > n(\mathbf{d} \cap \tilde{s}(x))\nu(s(x))\},$$

$n(\mathbf{d} \cap s(x))$  et  $n(\mathbf{d} \cap \tilde{s}(x))$  désignant le nombre de galaxies couvertes par le cylindre et son ombre, de volumes respectifs  $\nu(s(x))$  et  $\nu(\tilde{s}(x))$ .

Ensuite, à l’intérieur du cylindre, les galaxies doivent être réparties d’une manière assez homogène. Ceci est formulé comme suit

$$\mathbf{1}\{\text{“homogénéité”}\} = \prod_{i=1}^3 \mathbf{1}\{n(\mathbf{d} \cap s_i(y)) > 1\},$$



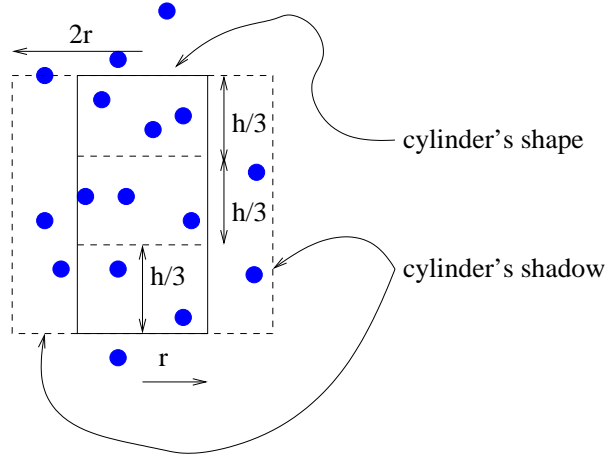


FIGURE 3.5 – La vue bidimensionnelle d’un cylindre et son ombre dans un champ de galaxies.

où  $n(\mathbf{d} \cap s_i(x))$  représente le nombre de galaxies situées dans  $s_i(y)$ .

Dans ce contexte, nous définissons le potentiel énergétique d’un cylindre comme la fonction

$$u(x) = \mathbf{1}\{\text{“densité”}\} \mathbf{1}\{\text{“homogénéité”}\} [n(\mathbf{d} \cap s(x)) - n(\mathbf{d} \cap \tilde{s}(x)) + u_{\max}] - u_{\max}, \quad (3.7)$$

où  $u_{\max}$  est une quantité positive préfixée qui est introduite pour donner une petite chance, à un cylindre qui ne remplit pas tous les critères de faire partie du réseau filamentaire. Ce terme vise à obtenir des réseaux un peu plus complets et il améliore les propriétés de mélange de la dynamique de simulation du modèle.

Le terme d’attache aux données est obtenu en sommant les potentiels (3.7) pour tous les cylindres d’une configuration

$$U_{\mathbf{d}}(\mathbf{x}|\theta) = - \sum_{x \in \mathbf{x}} u(x). \quad (3.8)$$

Comme la fenêtre d’observation  $W$  ne comporte qu’un nombre fini de galaxies, nous pouvons montrer qu’il existe une constante  $\Lambda_{\mathbf{d}} \geq \exp(-u_{\max})$  telle que

$$u(\eta) > -\log \Lambda_{\mathbf{d}}$$

pour tout  $\eta \in W \times M$ . Le modèle d’attache aux données et le modèle d’interaction sont tous deux localement stables. Ceci implique que le modèle complet (3.1) donné par (3.6) et (3.8) est bien défini.

Un résultat de détection des filaments obtenu en utilisant le modèle proposé est présenté Figure 3.6. Nous observons que les cylindres tendent à se connecter et qu'ils se positionnent dans des régions une inspection visuelle peut suggérer l'existence de filaments.

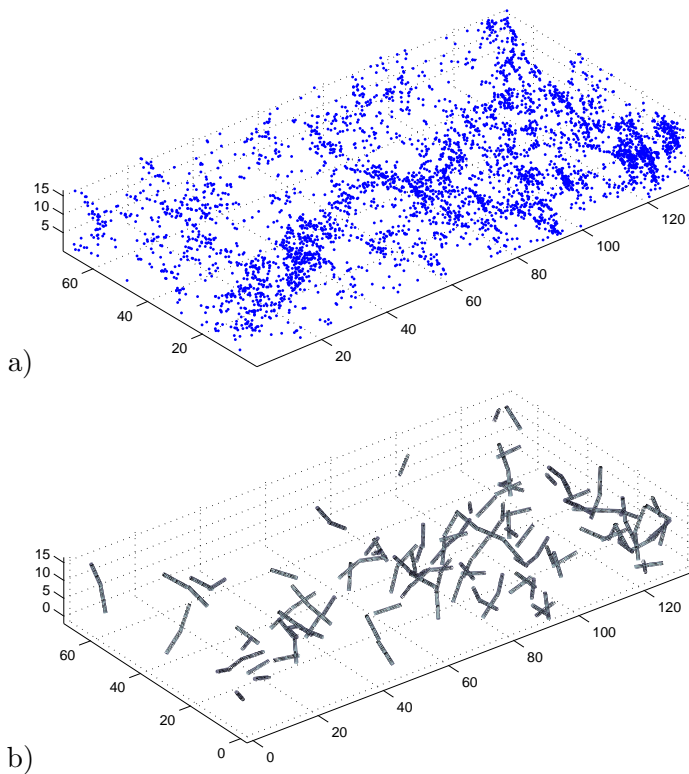


FIGURE 3.6 – a) Données : le catalogue de galaxies *NGP200*. b) Configuration de cylindres approximant les réseaux filamentaire.

La construction du terme d'attache aux données que nous avons présentée est très intuitive sans chercher une optimalité quelconque. Nous avons fait ce choix, en espérant que les cosmologistes vont réagir afin qu'en utilisant leur connaissances physiques, ils puissent donner une définition locale des propriétés d'un filament galactique qui puisse être intégrée au modèle. Cette stratégie a été fructueuse. Le terme d'attache aux données a été repris par les cosmologistes pour étudier la corrélation des vecteurs d'orientation des galaxies avec l'orientation du filament duquel elles font partie [184].

Actuellement, nous sommes en train de préciser les propriétés locales d'un filament sous la forme de tests d'hypothèses [T25]. Ceci ouvre la possibilité de relier notre modèle à des modèles de type Cox ou shot-noise. Sous les hypothèses du modèle, ce cadre permettrait la simulation des catalogues ga-

lactiques. Ce travail est en cours.

## 3.2 Modélisation d'agrégats en épidémiologie animale

A part les réseaux, d'autres structures morphologiques souvent rencontrées dans les applications, sont les agrégats ou les amas<sup>4</sup>. Nous en avons déjà parlé au paragraphe précédent en mentionnant la formation des amas formés par la position des centres des galaxies. C'était une des motivations pour la construction du modèle Bisous.

Cependant, le type des données impose des définitions différentes pour la structure que l'on cherche à détecter. En biométrie, sismologie ou en épidémiologie, on parle aussi d'agrégats. Mais dans ces domaines, plusieurs éléments d'un jeu de données peuvent avoir la même position, car les tremblements de terre ou les épidémies peuvent survenir plusieurs fois au même emplacement. De plus, par rapport à la structure des données du problème précédent, à la position d'un élément peut être associées certaines caractéristiques comme la magnitude d'un tremblement de terre ou la propension à une maladie [112, 111].

En Figure 3.7 un ensemble de points  $\mathbf{d}$  est observé dans une fenêtre rectangulaire  $W$ . Cet ensemble est composé de points bleus et blancs, formant une structure en trois agrégats. Chaque agrégat induit une région de  $W$ . Dans les régions délimitées par une ligne continue, il y a plus de points bleus que des points blancs. Cette situation est inversée dans la région délimitée par une ligne pointillée. Cet exemple illustre nos hypothèses relatives aux agrégats :

- agrégat : c'est un ensemble de points proches et de "mêmes" propriétés statistiques",
- structure d'agrégats : c'est un ensemble d'agrégats.

Par conséquent, détecter des agrégats signifie soit trouver l'ensemble de points dans  $\mathbf{d}$  qui forment la structure d'agrégats, soit trouver les caractéristiques géométriques des régions induites par les agrégats (positions, contours, périmètres, surfaces, etc.).

### 3.2.1 Présentation des données et formulation du problème

Le jeu de données analysées ici est une population de points. Chaque point représente la position d'une ferme en France, dans laquelle l'on élève des

---

4. En anglais, on utilise le mot "cluster".

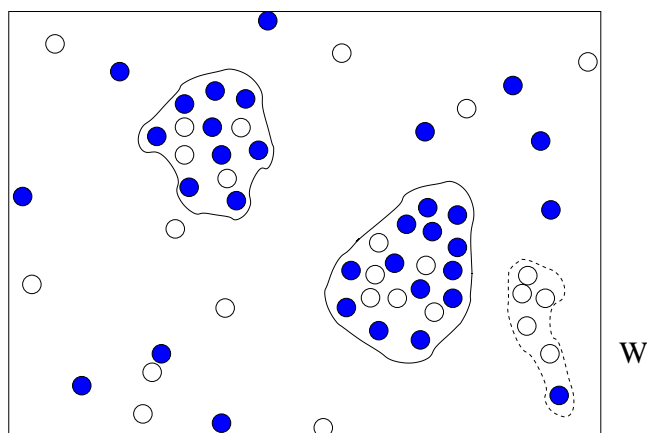


FIGURE 3.7 – Un champ de points dans la fenêtre d’observation  $W$ . Le champ observé comporte trois agrégats.

vaches laitières Holstein. La position d’une ferme est donnée par le centre de la commune à laquelle la ferme appartient. A chaque ferme est attachée une valeur numérique positive. Cette valeur est un indicateur annuel de la mastite bovine, pour les vaches de son élevage. Plus la valeur est grande, plus la ferme est “malade”. Nous avons analysé cinq jeux de données correspondant aux années 1996 jusqu’au 2000 [T16]. Typiquement, les données d’une année contiennent à peu près 30000 fermes enregistrées.

La mastite bovine est une maladie qui se transmet d’un animal à l’autre à l’intérieur d’un élevage, mais qui ne se transmet pas d’une ferme à une autre. La répartition “uniforme” de cette maladie sur tout le territoire observé est considérée comme un phénomène normal par les spécialistes [62]. L’apparition d’agrégats serait un indice de mauvaises pratiques et d’hygiène défectueuse dans les fermes.

La question que l’on se pose est dans quelle mesure le champ des observations possède une structure d’agrégats ?

Notre choix a porté sur la modélisation de régions là où les points forment des agrégats. Dans ce contexte, la région créée par un agrégat est supposée avoir les propriétés suivantes :

- localement, les points situés dans une telle région ont les “mêmes” propriétés statistiques,
- géométriquement, une telle région peut être approximée par une union de disques de centre et rayon aléatoires.

Ce point de vue est illustré Figure 3.8. Les agrégats d'une région sont approximatés par une configuration de disques aléatoires susceptibles de se superposer. Les disques ne sont situés que dans les régions de  $W$  où le nombre de points bleus prédomine.

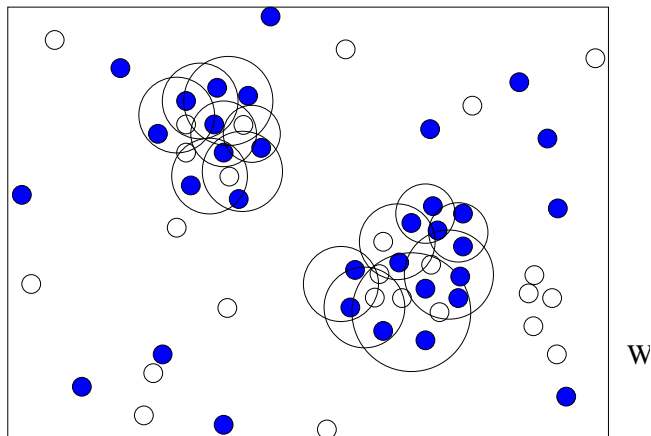


FIGURE 3.8 – Une configuration des disques aléatoires qui couvre une structure d'agrégats.

Nous faisons l'hypothèse que la configuration des disques qui approxime la structure d'agrégats est la réalisation d'un processus ponctuel marqué. La réalisation d'un tel processus est noté par  $\mathbf{x} = \{(w_i, m_i), i = 1, \dots, n\}$  où  $w_i \in W$  représente les positions des centres des disques et  $m_i \in [r_{\min}, r_{\max}]$  les rayons associées qui servent de marques. Le cadre de modélisation présenté précédemment peut s'appliquer. Le modèle proposé est un processus ponctuel de Gibbs dont la densité incorpore un terme d'attache aux données et un terme d'interaction. Dans ce qui suit, nous présentons les deux composantes du modèle.

### 3.2.2 Modélisation de l'attache aux données

Le terme d'attache aux données,  $U_{\mathbf{d}}(\mathbf{x}|\theta)$  spécifie la position des disques qui forment la structure des agrégats. Dans le contexte des mammites bovines, nous souhaitons que les disques soient placés dans les régions où l'on trouve beaucoup de fermes de score de maladie élevé.

Nous considérons qu'un disque  $x = (w, m)$  couvre un nombre suffisamment élevé de fermes si ce nombre, noté  $n_{\mathbf{d}}(x)$ , est plus grand qu'une valeur donnée  $n_0$ . Nous décidons que les fermes couvertes par un disque ont un score de

maladie élevé à l'aide d'un test statistique. Par rapport à une procédure de seuillage, un test statistique prend mieux en compte le comportement de la maladie et donc l'hétérogénéité des données. Ce test compare la moyenne des scores des fermes couvertes par un disque à une valeur seuil  $d_0$ . Soit  $W_{\mathbf{d}}(x, d_0, \alpha)$  la zone de rejet obtenue en utilisant le test de Student au niveau  $\alpha$ . Pour un jeu de données particulier, la valeur  $d_0$  retenue est la somme de sa moyenne et de son écart-type empiriques.

Sous ces hypothèses, le potentiel d'un disque s'écrit

$$v(x) = \mathbf{1}\{n_{\mathbf{d}}(c) > n_0\} \mathbf{1}\{\bar{\mathbf{d}}(c) \notin W_{\mathbf{d}}(x, d_0, \alpha)\} [\bar{\mathbf{d}}(x) - d_0 + v_{\max}] - v_{\max} \quad (3.9)$$

où  $\bar{\mathbf{d}}(x)$  est la moyenne des scores des fermes recouvertes par le disque  $x$  et  $v_{\max}$  une valeur préfixée. La définition de la fonction potentiel (3.9) fait que la densité du modèle va prendre des valeurs importantes quand les disques sont situées dans des régions où un grand nombre de fermes présente un score de maladie élevé. Comme pour les cylindres formant un réseau de filaments, le rôle du terme  $v_{\max}$  est de pénaliser les disques qui ne remplissent pas tous les critères, sans toutefois les interdire complètement. Ce choix de modélisation s'inscrit dans une logique de prévention forte, car toute région couverte par un disque a une probabilité non-nulle d'être considérée comme touchée par la maladie.

La somme des potentiels de tous les disques d'une configuration définit le terme d'énergie d'attache aux données de notre modèle, comme

$$U_{\mathbf{d}}(\mathbf{x}|\theta) = - \sum_{i=1}^{n(\mathbf{x})} v(x). \quad (3.10)$$

En prenant l'exponentielle de cette énergie (3.10) nous obtenons la première composante de notre modèle de structures d'agrégats. Cette composante définit un processus de Poisson inhomogène (3.2). Ce processus est localement stable.

### 3.2.3 Modélisation de l'interaction entre objets

Ainsi défini, le terme d'attache utilisé seul peut présenter des effets indésirables.

Premièrement, la maximisation de la densité poissonnienne (3.10) va placer tous les disques dans la région d'énergie d'attache maximale, provoquant la détection d'un seul agrégat. De plus, cette maximisation favorise l'apparition de configurations comportant un nombre moyen de disques de plus en plus grand. Cet effet est annulé en introduisant une énergie d'interaction qui

est une répulsion entre disques.

Deuxièmement, le terme d'attache aux données assigne la même probabilité à des disques de dimensions différentes. Ce phénomène est illustré Figure 3.9. Du point de vue mathématique, il produit un lissage du résultat final. Du point de vue applicatif, il élargit indûment les régions atteintes par la maladie. Pour annuler cet effet, le terme d'interaction doit agréger les disques tout en prenant en compte l'étendue de la structure d'agrégats.

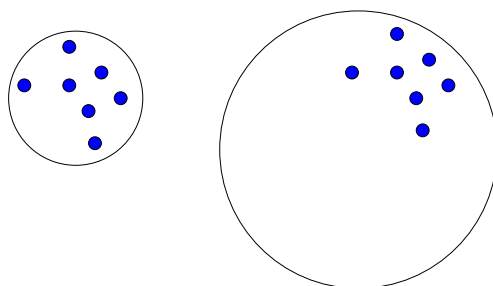


FIGURE 3.9 – Deux disques différents autour de la même configuration de points.

Dans ce contexte, l'énergie d'interaction d'une configuration des disques peut s'écrire sous la forme

$$U_i(\mathbf{x}|\theta) = -n_o(\mathbf{x}) \log \gamma_o + \nu[Z(\mathbf{x})] \log \gamma_a. \quad (3.11)$$

La première composante de (3.11) est un modèle d'interaction par paires d'objets, connu dans la littérature comme le modèle de Strauss [97, 141, 190, 180, 183]. Deux disques  $x_i = (w_i, m_i)$  et  $x_j = (w_j, m_j)$  se trouvent en interaction de type répulsion ou se chevauchent si  $d(w_i, w_j) \leq m_i + m_j$ . Dans ce cas, nous écrivons  $x_i \sim_o x_j$ . Le terme  $\log \gamma_o$  est le potentiel associé à une telle paire de disques. La notation  $n_o(\mathbf{x})$  désigne le nombre de paires de disques en répulsion dans la configuration  $\mathbf{x}$ . La Figure 3.10 montre trois réalisations du modèle de Strauss en fonction du paramètre d'interaction  $\gamma_o$ . Pour  $\log \gamma_o = 0$  le modèle est un processus de Poisson, qui permet aux disques de se chevaucher. Si  $\log \gamma_o = -0.3$ , le chevauchement des disques est pénalisé, ce qui réduit leur nombre. Finalement, si  $\log \gamma_o = -\infty$ , les disques ne peuvent pas se chevaucher. Dans ce cas nous avons affaire à un processus de type particules dures ou *hard-core*.

La deuxième composante de (3.11) est un modèle d'interaction par aires [17, 190]. Ici  $Z(\mathbf{x})$  est l'union des disques d'une configuration. Par conséquent,  $\nu[Z(\mathbf{x})]$  représente l'aire de cette union. Le paramètre  $\log \gamma_a$  contrôle la

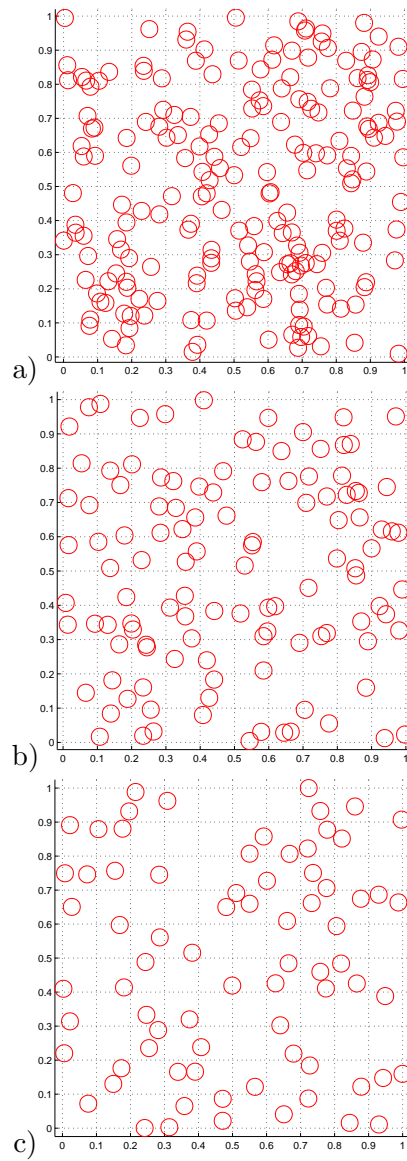


FIGURE 3.10 – Réalisations du processus de Strauss : a)  $\log \log \gamma_o = 0$ , b)  $\log \gamma_o = -0.3$  and c)  $\log \gamma_o = -\infty$ .



morphologie de  $Z(\mathbf{x})$ . Si  $\log \gamma_a < 0$ , alors les disques d'une configuration se repoussent et occupent plus d'espace dans la fenêtre d'observation  $W$ . Si  $\log \gamma_a > 0$ , alors les disques s'attirent et occupent moins d'espace. La Figure 3.11 montre des réalisations du modèle d'interaction par aires en fonction du paramètre  $\gamma_a$ . Comme pour le processus de Strauss si  $\log \gamma_a = 0$ , le modèle est un processus de Poisson. Si  $\log \gamma_a > 0$ , les disques se regroupent en occupant le moins de surface possible, en formant des agrégats. Inversement, si  $\log \gamma_a < 0$ , les disques ont tendance à occuper le plus d'aire possible.

Chacune de ces deux composantes est un processus ponctuel marqué markovien par rapport à la relation  $\sim_o$  et vérifie la propriété de stabilité locale [17, 97, 141, 190, 183]. Il en résulte que le modèle (3.11) est bien défini. Comme le terme d'attache aux données est également bien défini, il s'ensuit que le modèle complet - attache aux données et interaction - est également bien défini.

La Figure 3.12 montre une réalisation du modèle étudié. Les données sont représentées par des points bleus. Seule la position des fermes est visible, mais les scores de la maladie qui leur sont associés est bien pris en compte dans le terme d'attache aux données. Nous pouvons observer la formation d'agrégats importants au centre du territoire français. Quelques agrégats isolés apparaissent également dans d'autres régions.

Il est intéressant de savoir que Strauss a proposé son modèle pour créer des structures agrégées [183]. Kelly et Ripley ont déterminé pour quels paramètres le modèle est intégrable, prouvant que le modèle de Strauss ne pouvait simuler que des structures repulsives [97]. Cet inconvénient a été l'une des motivations de Baddeley et van Lieshout pour construire le modèle d'interaction par aires [17]. Dans cette application pratique, notre modèle synthétise les deux approches.

### 3.3 Modélisation du morcellement pour un environnement

Les processus ponctuels en interaction par paires de points peuvent être vues comme des limites de champs aléatoires définies sur une grille [23, 25]. Cela permet d'adapter le cadre mathématique utilisé à une situation pratique existante, et de passer du continu au discret et vice-versa, au moment opportun. Par exemple, il existe un équivalent de type processus ponctuel au bien connu modèle d'Ising [141].

Dans ce paragraphe nous montrons la construction d'un modèle similaire

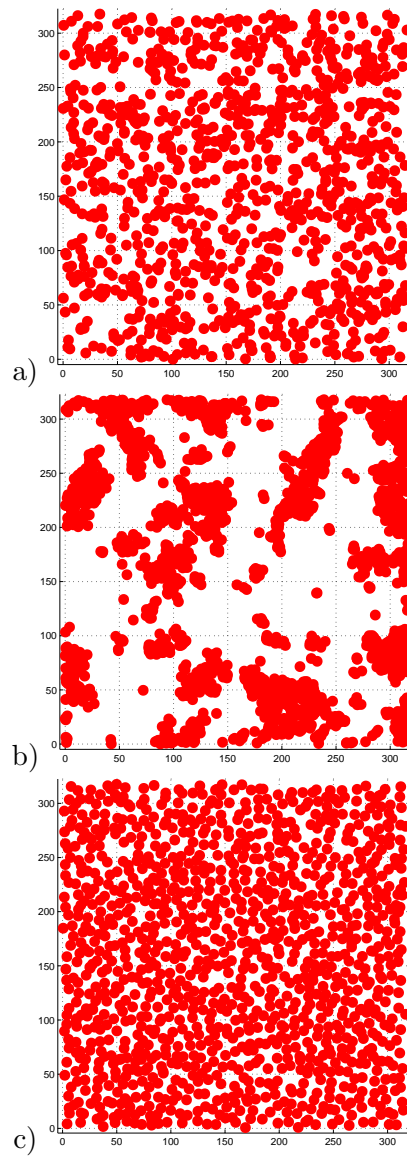


FIGURE 3.11 – Réalisations du processus d'interaction par aires : a)  $\log \gamma_a = 0$ , b)  $\log \gamma_a > 0$  and c)  $\log \gamma_a < 0$ .

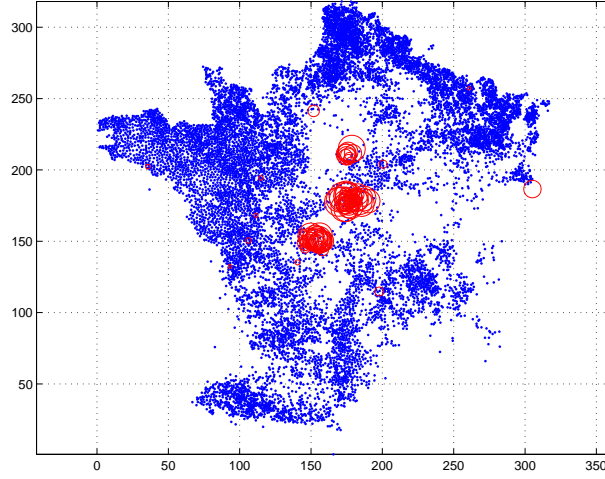


FIGURE 3.12 – Agrégats construits à partir de disques aléatoires (en rouge) avec les données (en bleu).

à celui d’Ising conditionné par un paramètre de localisation. Les structures obtenues par ce modèle ont été utilisées pour étudier l’influence du morcellement du paysage sur une dynamique de population dans des environnements périodiques [T10]. Les auteurs de l’article cité ont montré par simulation que moins le paysage est morcelé, plus l’espèce étudiée a de chance de survivre.

Soit  $W = [0, 1] \times [0, 1]$  un ensemble représentant une cellule d’un environnement périodique et soit  $W = \bigcup_{i=1}^{n_W} W_i$  une partition finie de  $W$ . Les cellules  $W_i$  sont toutes de même volume. Soit  $\Xi$  la grille construite à partir des centres  $x_i$  des cellules  $W_i$ , munie d’un système de voisinage  $V_\xi$ . Soient  $\mu^+$  et  $\mu^-$  deux réels tels que  $\mu^+ > \mu^-$ . Considérons maintenant  $\mathbf{x}$ , la réalisation d’un champ aléatoire défini sur la grille  $\Xi$ . Ce champ aléatoire attribue à chaque site  $\mathbf{x}_\xi$  une des valeurs  $\mu^+$  ou  $\mu^-$ , tandis que le nombre  $n_+ = \text{card}\{\mathbf{x}_\xi = \mu^+, \xi \in \Xi\}$ , prend une valeur préfixée. Dans ce contexte, soit  $\Omega$  l’espace des états possibles.

Le champ aléatoire  $\mathbf{x}$  représente un paysage dans lequel les sites où  $\mu(\mathbf{x}) = \mu^+$  sont interprétées comme “habitat” ; les autres pour lesquels  $\mu(\mathbf{x}) = \mu^-$  correspondent au “non-habitat”. Ainsi, caractérisons nous la *structure d’un paysage* par la structure du champ aléatoire  $\mathbf{x}$ .

La Figure 3.13 montre le système de voisinage d’ordre 4 considéré ici. Pour prendre en compte les environnements périodiques le domaine est enroulé sur un tore.

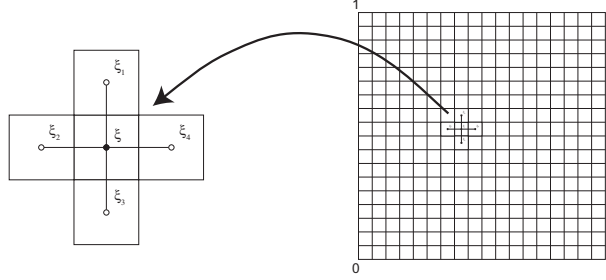


FIGURE 3.13 – Le système de voisinage d’ordre 4 : un élément  $\xi$  de  $\Xi$  et ses voisins  $\xi_i$ ,  $i = 1, \dots, 4$ .

Pour chaque  $\mathbf{x} \in \Omega$  soit

$$s(\mathbf{x}) = \frac{1}{2} \sum_{\eta, \xi \in \Xi, \xi \sim \eta} \mathbf{1}\{\mathbf{x}_\xi = \mathbf{x}_\eta\},$$

le nombre de paires de voisins  $\xi \sim \eta$  tel que  $\mathbf{x}_\xi = \mathbf{x}_\eta$ . La statistique du modèle,  $s(\mathbf{x})$ , est directement liée au morcellement de l’habitat : un paysage est agrégé ou morcelé selon que  $s(\mathbf{x})$  est grand ou petit.

*Premier modèle de paysage*

Soit le modèle décrit par la mesure de Gibbs  $\mathbb{P}$  définie sur  $\Omega$

$$P(X = \mathbf{x}) = \frac{1}{Z} \exp[-U(\mathbf{x})], \quad (3.12)$$

ou  $Z = \sum_{\mathbf{x} \in \Omega} \exp[-U(\mathbf{x})]$  est la constante de normalisation ou de partition, et  $U(\mathbf{x})$  est la fonction d’énergie définie par

$$U(\mathbf{x}) = -\frac{\beta}{2} \sum_{\eta, \xi \in \Xi, \xi \sim \eta} \mathbf{1}\{\mathbf{x}_\xi = \mathbf{x}_\eta\} = -\beta s(\mathbf{x})$$

avec  $\beta > 0$ .

La densité de probabilité du modèle (3.12) s’écrit

$$P(X = \mathbf{x}) = \frac{1}{Z} \exp[\beta s(\mathbf{x})], \quad (3.13)$$

ce qui montre que plus  $\beta$  est grand, plus le modèle favorise la même valeur entre sites voisins. Ainsi, le paramètre  $\beta$  contrôle la morphologie du paysage. Ceux qui sont familiers avec ce type de modélisation remarqueront que le processus (3.13) est un modèle d’Ising conditionné par le nombre de cellules de “spin” habitat et non-habitat.

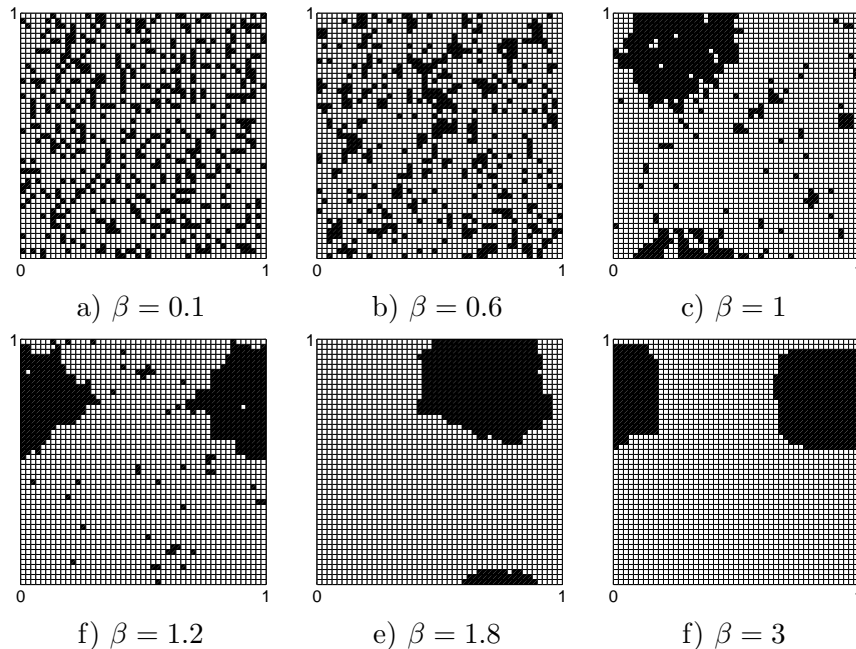


FIGURE 3.14 – Réalisations du premier modèle de paysage pour différentes valeurs du paramètre  $\beta$ .

Quelques réalisations du modèle (3.13) sont montrées en Figure 3.14. Le nombre total de cellules est de  $n_W = 50^2$  et l'abondance de l'habitat a été fixée à 20%, ce qui veut dire  $n_+ = 0.2n_w = 500$ . Comme prévu, nous observons que le paramètre  $\beta$  contrôle l'état d'agrégation du paysage.

La Figure 3.15 montre deux paysages présentant deux composantes identiques. Sous les hypothèses du modèle, les deux paysages ont la même probabilité, alors que les composantes de la Figure 3.15a sont plus rapprochées que celles dans la Figure 3.15b. D'un point de vue plus réaliste, il est souhaitable qu'un modèle puisse différencier ces deux situations [173].

#### *Deuxième modèle de paysage*

Suite à la remarque précédente, un deuxième modèle est envisagé. Soit  $\kappa$  un point aléatoire choisi uniformément dans  $W$ . Le modèle est défini, conditionnellement au point  $\kappa$ , par la mesure de Gibbs

$$P(X = \mathbf{x}|\kappa) = \frac{1}{Z_\kappa} \exp[-U_\kappa(\mathbf{x})], \quad (3.14)$$

la fonction d'énergie s'écrivant

$$U_\kappa(\mathbf{x}) = -\beta s(\mathbf{x}) + \frac{1}{\sigma^2} \tau_\kappa(\omega),$$

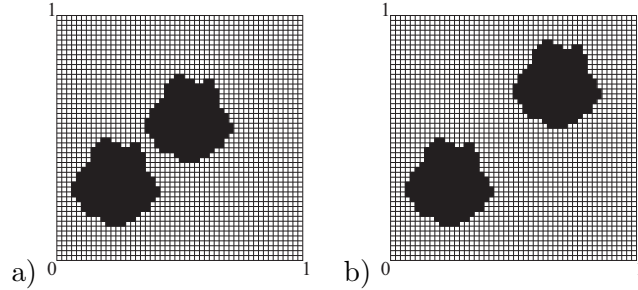


FIGURE 3.15 – Deux structures de paysage construites avec deux composantes identiques.

avec  $\sigma > 0$  et

$$\tau_\kappa(\mathbf{x}) = \sum_{\eta \in \Xi, \mathbf{x}_\eta = \mu^+} d(\eta, \kappa)^2.$$

avec  $d(\eta, \kappa)$  la distance euclidienne sur  $W$  muni de la symétrie toroïdale.

Le terme  $\frac{1}{\sigma^2} \tau_\kappa(\mathbf{x})$  agit comme un champ externe. Ce champ a la forme d'une gaussienne tronquée centrée en  $\kappa$  et de variance  $\sigma^2$ . Ce choix se justifie par le fait que nous savons qu'au moins dans le cas où la différence  $|\mu^+ - \mu^-|$  n'est pas trop petite et que l'habitat est de 20%, la configuration optimale d'un paysage tend à devenir circulaire [167]. Ici, la configuration optimale d'un paysage est celle qui donne le plus de chances de survie à l'espèce étudiée.

Quelques réalisations du modèle (3.14) sont montrées Figure 3.16. Ici, on a repris les mêmes valeurs qu'auparavant pour  $n_W$  et  $n_+$ . On a aussi choisi  $\sigma = 0.4$ . Aux fortes valeurs de  $\beta$  le paysage tend à se structurer autour d'un centre.

### 3.4 Combinaison des étiquettes dans des modèles de champ aléatoire

Ce problème est apparu lors du travail effectué sur la simulation exacte et l'étude de la discrétisation de processus ponctuels. La partie concernant simulation exacte est présentée dans un autre chapitre de ce mémoire.

La segmentation d'images est la partition d'images en zones de textures homogènes, c'est à dire en régions où les pixels présentent des propriétés statistiques voisines. Cette opération se fait en assignant à chaque pixel une étiquette correspondant à un certain type de texture. Les champs de Markov sont un des outils les plus utilisés pour résoudre ce problème [113, 202].

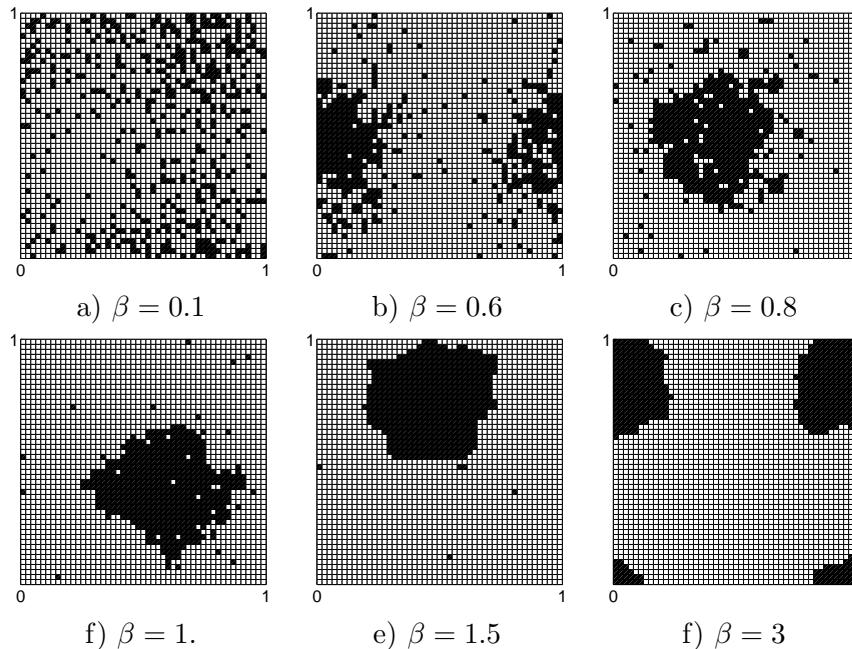


FIGURE 3.16 – Réalisations du deuxième modèle de paysage pour différentes valeurs de  $\beta$ .

En général, le nombre de textures n'est pas connu. Pour y remédier, des approches bayésiennes traitent ce nombre comme une variable aléatoire du modèle [59, 77].

La difficulté rencontrée en utilisant ces approches est la même que pour les processus ponctuels : la plupart des lois de champs de Markov n'ont pas de constante de normalisation disponible analytiquement. Ainsi, en changeant le nombre d'étiquettes, les constantes de normalisation ne se simplifient plus quand on met en place une méthode de type Monte Carlo. Un problème encore plus important est que la structure d'interaction peut changer d'une manière significative quand l'on groupe deux étiquettes ensemble. En d'autres termes, le groupement d'étiquettes ne préserve pas le caractère markovien. Il les transforme en modèles assez peu naturels pour la segmentation non-supervisée. En revanche, nous montrons que la classe des champs de Markov à composantes connexes proposée par [140] reste stable par cette opération. Elle peut donc fournir des modèles appropriés pour la segmentation d'images avec un nombre inconnu de textures.

Les résultats que nous présentons ci-dessous ont été publiés dans [T9] et correspondent à des résultats obtenus pour les processus ponctuels [35, 191].

## Les champs aléatoires $C$ -markoviens <sup>5</sup>

Soit  $S = (s_1, \dots, s_m)$  un ensemble de sites (par exemple une région finie dans  $\mathbb{Z}^2$ ) et soit  $\sim$  une relation symétrique et réflexive sur  $S$ . Dans le langage de la théorie des graphes, les sites représentent les sommets, et une arête connecte les sites  $s$  et  $r$  si et seulement si  $s \sim r$ . A chaque site correspond une étiquette ou une couleur appartenant à l'ensemble  $\mathbb{L} = \{1, \dots, q\}$ ,  $q \geq 2$ , choisie aléatoirement. Le champ aléatoire qui en résulte est  $X = (X_1, \dots, X_m)$ ,  $X_i$  indiquant l'étiquette au site  $s_i$ .

Un champ aléatoire  $X$  de loi  $P(\cdot)$  est dit markovien par rapport à  $\sim$  si pour tout  $i = 1, \dots, m$ , la probabilité conditionnelle

$$P(X_i = x_i \mid X_j = x_j, j \neq i) = P(X_i = x_i \mid X_j = x_j, s_j \sim s_i, j \neq i) \quad (3.15)$$

depend seulement de  $x_i$  et les étiquettes des sites connectés à  $s_i$ . L'ensemble  $\partial(s_i)$  de ces sites est appelé le voisinage du site  $i$ , et les lois conditionnelles dans (3.15) portent le nom de *caractéristiques locales*.

Pour caractériser un champ de Markov, nous avons besoin de la définition suivante. Une *clique* est un sous-ensemble  $C \subseteq S$  saturé en arêtes, c'est à dire un sous-ensemble pour lequel  $s \sim t$  pour tout  $s, t \in C$ . Les singletons et l'ensemble vide sont aussi des cliques. Soit  $\mathcal{C}$  l'ensemble de toutes les cliques dans  $S$ . Sous l'hypothèse de positivité de la loi multivariable, d'après le théorème de Hammersley-Clifford [23, 113, 202] le champ aléatoire  $X$  est markovien si et seulement si sa loi de probabilité peut être écrite comme suit

$$P(X_1 = x_1, \dots, X_m = x_m) = \prod_{C \in \mathcal{C}} \varphi_C(x_c, c \in C), \quad (3.16)$$

ce qui met en jeu des fonctions d'interaction  $\varphi_C : \mathbb{L}^C \rightarrow \mathbb{R}^+$  définies pour chaque clique  $C \in \mathcal{C}$ . L'équation (3.16) permet aussi établir l'équivalence avec les champs de Gibbs, en écrivant

$$P(X_1 = x_1, \dots, X_m = x_m) = \sum_{C \in \mathcal{C}} V_C(x_c, c \in C) = \sum_{C \in \mathcal{C}} \log \varphi_C(x_c, c \in C),$$

où  $V_C(\cdot) = \log \varphi_C(\cdot)$  pour des ensembles non vides  $C$  sont les potentiels d'interaction.

Le sous-ensemble  $K \subset S$  est une composante connexe si pour tout  $s, r \in K$ , il existe un chemin  $s = s_1 \sim \dots \sim s_n = r$  avec  $s_i \in K$ . Les *champs de Markov à composantes connexes* ou  *$C$ -markoviens* introduits par [140] sont

---

5. Ce terme est la traduction de l'anglais : Markov connected component random fields. Cette traduction nous a été suggérée par C. Lantuéjoul.



caractérisés par une factorisation de la forme

$$P(X_1 = x_1, \dots, X_m = x_m) = \frac{1}{Z} \prod_{K \in \mathcal{K}(\mathbf{x})} \Psi(K, l(\mathbf{x}_K)), \quad (3.17)$$

le produit portant sur l'ensemble des composantes connexes maximales  $K$  de même étiquette  $l(\mathbf{x}_K)$ , et  $\Psi(\cdot, \cdot)$  étant une fonction positive définie sur l'espace produit  $\mathcal{K} \times \mathbb{L}$ , des composantes  $\sim$ -connectées et des étiquettes. En général, une définition de dépendance locale n'existe pas, hormis le cas où une des couleurs peut être vue comme couleur de fond.

Les classes de champs de Markov et de champs  $C$ -markoviens ne sont pas comparables, au sens où aucune des ces deux classe ne contient l'autre. Pour plus des détails, nous recommandons [140].

**Exemple 9** *Le modèle de Potts*

Soient  $S = (s_1, \dots, s_m)$  un ensemble fini de sites,  $\sim$  une relation symétrique et réflexive sur  $S$  et  $\mathbb{L} = \{1, \dots, q\}$  un ensemble fini d'étiquettes. Le modèle de Potts est la variable aléatoire sur  $\mathbb{L}^S$  définie par la loi jointe

$$\pi(x_1, \dots, x_m) = \frac{1}{Z} \exp \left[ -\beta \sum_{s_i \sim s_j, i < j} \mathbf{1}_{x_i \neq x_j} \right]. \quad (3.18)$$

Le paramètre  $\beta \in \mathbb{R}$  est aussi appelé température inverse,  $Z$  est la constante de normalisation qui assure que  $\sum_{\mathbf{x} \in \mathbb{L}^S} \pi(\mathbf{x}) = 1$ . Dans le cas particulier  $q = 2$ , nous retrouvons le modèle d'Ising. Ainsi (3.18) est de la forme (3.4) avec  $V_C(x_i, x_j) = -\beta \mathbf{1}_{x_i \neq x_j}$  pour les ensembles  $C = \{s_i, s_j\}$  with  $s_i \sim s_j$ , et  $V_A \equiv 0$  sinon.

Les caractéristiques locales du modèle du Potts satisfont

$$\frac{P(X_i = \ell \mid X_j = x_j, j \neq i)}{P(X_i = 1 \mid X_j = x_j, j \neq i)} = \frac{\exp \left[ -\beta \sum_{s_i \sim s_j} \mathbf{1}_{\{x_j \neq \ell\}} \right]}{\exp \left[ -\beta \sum_{s_i \sim s_j} \mathbf{1}_{\{x_j \neq 1\}} \right]}$$

pour tout  $\ell \in \mathbb{L}$  et  $x_j \in \mathbb{L}$ ,  $j \neq i \in \{1, \dots, m\}$ . Pour résumer,  $X$  est markovien pour tout nombre de couleurs. Si  $\beta > 0$ , le modèle est dit "ferromagnétique" et pénalise les paires de sites de couleurs différentes, en formant des "grandes" régions de la même couleur. Si  $\beta < 0$ , le modèle est dit "anti-ferromagnétique" et encourage les paires de sites de couleurs différentes. Dans ce cas, les régions d'une seule couleur ne constituent pas une structure dont l'apparition est souhaitée. Elles ne font que refléter le caractère aléatoire du modèle. Les fonctions d'interaction sont  $\varphi_\emptyset = 1/Z$ ,

$\varphi_C(x_i, x_j) = \exp[-\beta \mathbf{1}\{x_i \neq x_j\}]$  pour des cliques  $C = \{s_i, s_j\}$  formées par des paires de voisins  $s_i \sim s_j$ , et  $\varphi_C \equiv 1$  sinon.

Comme

$$\pi(x_1, \dots, x_m) = \frac{1}{Z} \exp \left[ -\frac{\beta}{2} \sum_{K \in \mathcal{K}(\mathbf{x})} \sum_{s_i \in K} |\{s_j \notin K : s_i \sim s_j\}| \right],$$

le modèle de Potts est aussi un champ  $C$ -markovien.

### Fusion des étiquettes

La classe des processus ponctuels marqués  $C$ -markoviens est fermée par rapport à la superposition indépendante [11, 35]. Les résultats suivants montrent que la même chose est vraie pour les champs  $C$ -markoviens quand l'on effectue la fusion des étiquettes.

**Théorème 9** Soit  $S = (s_1, \dots, s_m)$  un ensemble fini de sites,  $\sim$  une relation symétrique et réflexive sur  $S$  et  $\mathbb{L} = \{1, \dots, q\}$  un ensemble fini de  $q \geq 2$  étiquettes. Soit  $X$  un champ  $C$ -markovien sur  $S$ , ayant  $q$  couleurs, muni de la relation  $\sim$  et décrit par (3.17) et soit aussi le champ aléatoire  $Y$  à valeurs dans  $\{0, \dots, q-2\}$  défini par

$$Y_i = X_i \mathbf{1}\{X_i \leq q-2\}, \quad i = 1, \dots, m.$$

Alors,  $Y$  est un champ  $C$ -markovien à  $q-1$  couleurs muni de la relation  $\sim$ .

**Preuve:** Soit  $\mathbf{y} = (y_1, \dots, y_m) \in \{0, \dots, q-2\}^S$ . Pour  $\mathbf{x} \in \Omega_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{L}^S : y_i = x_i \mathbf{1}\{x_i \leq q-2\}, i = 1, \dots, m\}$  et  $j \in \mathbb{L}$ , nous considérons  $\mathcal{K}_j(\mathbf{x})$  l'ensemble de composantes connexes maximales dans  $\mathbf{x}$  étiquetées  $j$ . Premièrement, nous remarquons que  $\mathbf{x}$  et  $\mathbf{y}$  ont les mêmes composantes connexes maximales étiquetés de 1 à  $q-2$ . Ensuite, chaque composante étiquetée  $q$  ou  $q-1$  est une partie d'une seule composante maximale étiquetée 0 dans  $\mathbf{y}$ . Soit  $\mathcal{K}_0(\mathbf{y})$  la famille des composantes étiquetées 0. La loi de  $Y$  est donnée par

$$\begin{aligned} \pi_Y(\mathbf{y}) &= \sum_{\mathbf{x} \in \Omega_{\mathbf{y}}} \pi(\mathbf{x}) = & (3.19) \\ & \frac{1}{Z} \prod_{j=1}^{q-2} \prod_{K \in \mathcal{K}_j(\mathbf{y})} \Psi(K, j) \times \prod_{K \in \mathcal{K}_0(\mathbf{y})} \left\{ \sum_{\mathbf{x} \in \{q-1, q\}^K} \prod_{j=q-1}^q \prod_{L \in \mathcal{K}_j(\mathbf{x})} \Psi(L, j) \right\} \end{aligned}$$

ce qui prouve le résultat.  $\square$

Il est à remarquer que la constante de normalisation  $Z$  de  $\pi_Y$  est la même que celle de  $\pi$ , la loi de probabilité de  $X$ . La factorisation (3.19) doit être comparée avec celle des processus ponctuels [35, 190]. Un résultat similaire n'existe pas pour les champs de Markov. Un bilan de cette situation concernant les processus ponctuels de Markov se trouve dans [190, 191].

**Exemple 10** *Le modèle de Potts (suite)*

Soit le modèle de Potts  $X$  défini par (3.18), et pour  $i = 1, \dots, m$ , soit  $Y_i = X_i \mathbf{1}\{X_i \leq q - 2\}$ . Alors la loi de probabilité de  $Y$  est

$$\begin{aligned} \pi_Y(\mathbf{y}) &= \sum_{\mathbf{x} \in \Omega_{\mathbf{y}}} \frac{1}{Z} \exp \left[ -\beta \sum_{s_i \sim s_j, i < j} \mathbf{1}\{x_i \neq x_j\} \right] = \\ &= \frac{1}{Z} \exp \left[ -\beta \sum_{s_i \sim s_j, i < j} \mathbf{1}\{y_i \neq y_j\} \right] \times \\ &= \prod_{K \in \mathcal{K}_0(\mathbf{y})} \left\{ \sum_{\mathbf{z} \in \{q-1, q\}^K} \exp \left[ -\beta \sum_{s_i \sim s_j \text{ in } K, i < j} \mathbf{1}z_i \neq z_j \right] \right\}. \end{aligned} \quad (3.20)$$

Ainsi, (3.20) est proportionnelle au produit d'une loi de probabilité d'un modèle de Potts à  $q - 1$  couleurs et celle d'un champ de Markov à composantes connexes sur la composante maximale étiquetée 0 prise dans  $\mathcal{K}_0(\mathbf{y})$ . La Figure 3.17 montre une réalisation du champ de Potts  $X$  à  $q = 5$  couleurs et les champs à composantes connectés  $Y$ , obtenus en fusionnant les étiquettes une à une.

En conclusion, l'opération de fusion d'étiquettes d'un modèle de Potts amène à un champ à composantes connexes. Cependant, comme nous l'avons dit auparavant, ceci n'est pas vrai pour les champs de Markov généraux. Les auteurs montrent dans [T9] par un exemple, que si  $X$  est Markov, le champ  $Y$  obtenu par fusion d'étiquettes ne l'est pas forcément.

Dans ce qui suit, nous portons notre attention sur la classe des champs aléatoires à la fois markoviens et  $C$ -markoviens. Ces champs sont définis par une densité de probabilité qui se factorise comme suit

$$\pi(x_1, \dots, x_m) = \frac{1}{Z} \prod_{K \in \mathcal{K}(\mathbf{x})} \prod_{\emptyset \neq C \in \mathcal{C}, C \subseteq K} \Phi(C, l(\mathbf{x}_K)). \quad (3.21)$$

Par exemple, le modèle de Potts (3.18) est un tel champ si l'on considère  $C = \{c\}$  avec  $c \in K$ ,  $\Phi(C, l(\mathbf{x}_K)) = \exp \left[ -\frac{\beta}{2} |\{s_j \notin K : c \sim s_j\}| \right]$  et  $\Phi \equiv 1$

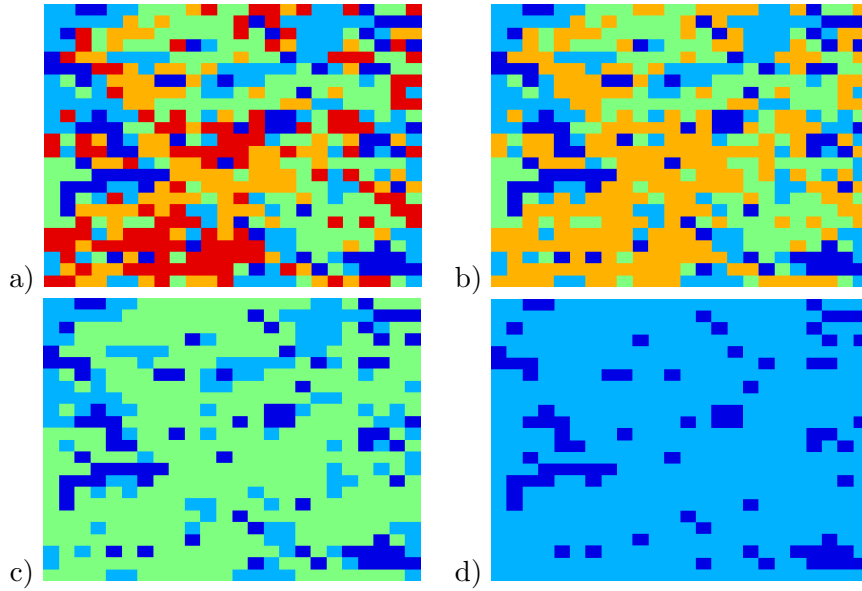


FIGURE 3.17 – Réalisation d'un modèle de Potts à  $q = 5$  couleurs, relation de voisinage  $s_i \sim s_j \Leftrightarrow \|s_i - s_j\| \leq 1$  et paramètre  $\beta = 1$  (a) et les champs à composantes **connexes** obtenus après fusion des étiquettes (b-c).

sinon.

Le résultat suivant est un corollaire du Théorème 9.

**Corollaire 2** Soit  $S = (s_1, \dots, s_m)$  un ensemble fini de sites,  $\sim$  une relation symétrique et réflexive sur  $S$  et  $\mathbb{L} = \{1, \dots, q\}$ ,  $q \geq 2$ , un ensemble fini d'étiquettes. Soit  $X$  un champ  $C$ -markovien à  $q$ -couleurs, muni de la relation  $\sim$  et décrit par la loi de probabilité (3.21), et soit également le champ  $Y$  à valeurs dans  $\{0, \dots, q-2\}$  défini par

$$Y_i = X_i \mathbf{1}\{X_i \leq q-2\}, \quad i = 1, \dots, m.$$

Alors  $Y$  est un champ  $C$ -markovien à  $(q-1)$ -couleurs, muni de la relation  $\sim$  et avec

$$\Psi(K, 0) = \sum_{\mathbf{x} \in \{q-1, q\}^K} \prod_{j=q-1}^q \prod_{L \in K_j(\mathbf{x})} \prod_{\emptyset \neq C \in \mathcal{C}, C \subseteq L} \Phi(C, j).$$

Comme nous l'avons déjà mentionné,  $Y$  n'est pas forcément un champ de Markov.

Pour tout champ de Markov, nous avons le résultat suivant.

**Théorème 10** Soit  $S = (s_1, \dots, s_m)$  un ensemble fini de sites,  $\sim$  une relation symétrique et réflexive sur  $S$  et  $\mathbb{L} = \{1, \dots, q\}$ ,  $q \geq 2$ , un ensemble fini d'étiquettes. Soit  $X$  un champ de Markov à  $q$ -couleurs, muni de la relation  $\sim$  et décrit par la loi de probabilité (3.16), et soit également le champ  $Y$  à valeurs dans  $\{0, \dots, q-2\}$  défini par

$$Y_i = X_i \mathbf{1}\{X_i \leq q-2\}, \quad i = 1, \dots, m.$$

Alors  $Y$  a comme loi de probabilité de  $\pi_Y$  qui est donnée par

$$\prod_{\mathcal{C} \ni \mathcal{C} \subseteq (\cup \mathcal{K}_0(\mathbf{y}))^c} \varphi_{\mathcal{C}}(y_c, c \in \mathcal{C}) \prod_{K \in \mathcal{K}_0(\mathbf{y})} \left\{ \sum_{\mathbf{x} \in \{q-1, q\}^K} \prod_{\mathcal{C} \ni \mathcal{C} \cap K \neq \emptyset} \varphi_{\mathcal{C}}(\mathbf{x}_c, c \in \mathcal{C} \cap K; y_d, d \in \mathcal{C} \setminus K) \right\}.$$

**Preuve:** Soient  $\mathbf{y} = (y_1, \dots, y_m) \in \{0, \dots, q-2\}^S$  et  $\Omega_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{L}^S : y_i = x_i \mathbf{1}x_i \leq q-2, i = 1, \dots, m\}$ . Alors la loi de probabilité de  $Y$  est donnée par

$$\pi_Y(\mathbf{y}) = \sum_{\mathbf{x} \in \Omega_{\mathbf{y}}} \pi(\mathbf{x}) = \prod_{\mathcal{C} \ni \mathcal{C} \subseteq (\cup \mathcal{K}_0)^c} \varphi_{\mathcal{C}}(y_c, c \in \mathcal{C}) \times \sum_{\mathbf{x} \in \{q-1, q\}^{\mathcal{K}_0(\mathbf{y})}} \prod_{\mathcal{C} \ni \mathcal{C} \cap \mathcal{K}_0(\mathbf{y}) \neq \emptyset} \varphi_{\mathcal{C}}(\mathbf{x}_c, c \in \mathcal{C} \cap \mathcal{K}_0(\mathbf{y}); y_d, d \in \mathcal{C} \setminus \mathcal{K}_0(\mathbf{y})).$$

L'observation importante est que si  $y_i = 0 = y_j$  et  $s_i \sim s_j$ , alors  $s_i$  et  $s_j$  doivent appartenir à la même composante maximale étiquetée 0, donc aucune clique ne peut contenir des points de composantes distinctes de  $\mathcal{K}_0(\mathbf{y})$ . Par conséquent,  $\pi_Y$  a l'expression souhaitée.  $\square$

En conclusion,  $Y$  n'est pas forcément un champ  $C$ -markovien.

Pour les processus ponctuels marqués, nous pouvons dire qu'il est presque toujours possible d'ajouter ou enlever un objet à une configuration, sans que cela modifie la loi de probabilité de la forme induite par le processus. Cela permet entre autres de construire des dynamiques naturelles de type Monte Carlo, pour simuler et utiliser ces modèles en pratique. Dans le cas des champs aléatoires à dimension de grille fixe, des opérations naturelles consistent à ajouter ou enlever une couleur. Par conséquent, il faut savoir comment la loi de la structure induite par ce processus évolue. Les résultats de cette partie montrent que les champs  $C$ -markoviens ont les bonnes propriétés pour la simulation et les applications pratiques.

## 3.5 Modélisation de paysages à l'aide de pavages en $T$

Le développement de la génétique a encouragé les études de propagation du pollen dans différentes structures du paysage. Pour ce faire, on s'intéresse à la modélisation de paysages de parcellaire agricole. Ici nous proposons un modèle construit à partir des pavages en  $T$ . Les différentes caractéristiques d'un paysage sont décrites par les propriétés géométriques des pavages. Ces propriétés sont probabilisées, à la manière de celles qui décrivent les processus ponctuels marqués. Cela pose des problèmes mathématiques très intéressants, car les pavages en  $T$  sont des objets géométriques assez contraints. Notre travail a été publié dans [T6].

### 3.5.1 L'espace des $T$ -pavages

Seuls sont considérés ici les pavages définis sur un domaine compact  $W \subset \mathbb{R}^2$ . Ce domaine est supposé convexe et polygonal. Soient  $a(W)$ ,  $l(W)$ ,  $n_e(W)$  et  $n_v(W)$  la surface, le périmètre, le nombre d'arêtes et de sommets de  $W$ .

Un *pavage* polygonale de  $W$  est une partition finie de  $W$  en polygones. Un tel polygone est appelé *cellule*. Les sommets du pavage sont les sommets des cellules. Ses *arêtes* sont des portions de frontière de cellules délimitées par sommets adjacents. L'ensemble des arêtes d'un pavage peut être vu comme un ensemble de segments, dont certains sont des unions d'arêtes alignées et jointives.

**Définition 18** *Un sommet du pavage est un  $T$ -sommet s'il est l'intersection d'exactly trois arêtes dont deux sont alignées. Le pavage de  $W$  est appelé  $T$ -pavage si tous ses sommets, à l'exception éventuelle de ceux de  $W$ , sont des  $T$ -sommets, et qu'il n'y ait pas de paire de segments situés sur la même droite.*

La Figure 3.18 montre un exemple de  $T$ -pavage.

L'espace des  $T$ -pavages sur le domaine  $W$  est noté  $\mathcal{T}$ . Pour tout  $T$ -pavage  $T \in \mathcal{T}$  soient  $C(T)$  l'ensemble de ses cellules,  $V(T)$  l'ensemble de ses sommets,  $E(T)$  l'ensemble de ses arêtes et  $S(T)$  l'ensemble de ses segments. Les cardinaux de ces ensembles sont respectivement désignés par  $n_c(T)$ ,  $n_v(T)$ ,  $n_e(T)$  et  $n_s(T)$ . Les sommets, les arêtes et les segments sont dits *internes* quand ils ne sont pas totalement contenus dans le bord de  $W$ . Les ensembles et les mesures correspondant aux objets internes sont notés par le même symbole surmonté d'un cercle. Par exemple  $\overset{\circ}{V}(T)$  et  $\overset{\circ}{n}_v(T)$  désignent l'ensemble

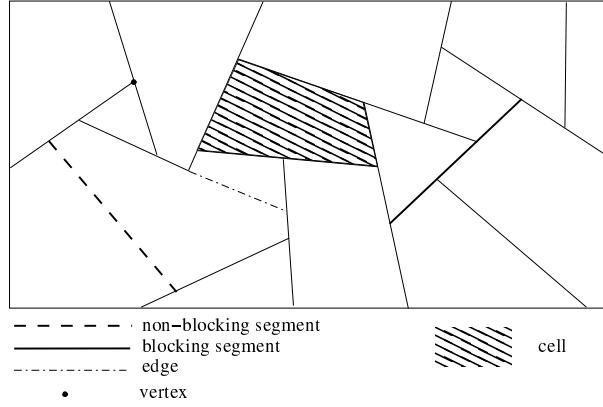


FIGURE 3.18 – Exemple d'un  $T$ -pavage avec ses composantes.

et le nombre de sommets internes.

Une tessellation induit une configuration de droites  $L(T)$ , qui est l'ensemble de droites supportant les segments. Réciproquement, une configuration finie de droites touchant  $W$  engendre un ensemble de tessellations  $\mathcal{T}(L)$ , telles que  $L(T) = L$  pour tout  $T \in \mathcal{T}(L)$ . La cardinalité de cet ensemble est noté  $n_t(L)$ . La majoration du nombre de droites de  $L$

$$n_t(L) \leq C^k \left( \frac{k}{(\log k)^{1-\epsilon}} \right)^{k-k/\log k}, \quad (3.22)$$

avec  $\epsilon > 0$ ,  $k \geq 2$  et  $C$  dépendant de  $\epsilon$ , sera utile pour prouver l'intégrabilité des modèles proposés pour les  $T$ -pavages. Ce résultat a été obtenu par [96].

**Définition 19** Une fonction  $\phi : \mathcal{T} \rightarrow \mathbb{R}^+$  est dite stable, s'il existe une constante positive réelle  $K$ , telle que

$$\phi(T) \leq K^{\hat{n}_s(T)},$$

pour tout  $T \in \mathcal{T}$ .

La définition précédente fait un premier lien avec la première condition de stabilité des processus ponctuels [72, 171]. Cependant, l'analogie avec la stabilité locale des processus ponctuels ne se fait pas trivialement. Ceci tient au fait que l'ajout ou la suppression d'un segment doit préserver la structure d'un  $T$ -pavage.

Dans ce qui suit, nous montrons qu'il est possible de transformer un  $T$ -pavage dans un autre  $T$ -pavage, grâce à trois opérations, que nous appelons divi-

sion, fusion et flip<sup>6</sup>.

Une division est l'opération qui coupe une cellule en deux (cf. Figures 3.19a et 3.19b). Une division  $S$  de  $T$  est représentée par  $S = (c, l)$  où  $c$  est une cellule et  $l$  une droite qui touche l'intérieur de  $c$ . La division de  $T$  par  $S$  entraîne l'ajout de l'arête  $c \cap l$ . Le nouveau pavage obtenu est  $ST = T \cup (c \cap l)$ .

Une fusion est l'opération qui enlève un segment contenant une seule arête (cf. Figures 3.19a et 3.19c). Les segments qu'une fusion peut affecter sont appelés non-bloquants, et les autres bloquants. Une fusion dans le pavage  $T$  s'identifie à un segment intérieur non-bloquant. Pour tout  $T$ -pavage, le nombre de fusions possibles est  $\overset{\circ}{n}_{s,nb}(T)$ , le nombre de segments intérieurs non-bloquants dans  $T$ . Le pavage après la fusion est notée  $MT$ .

Un flip est l'opération qui enlève une arête finale d'un segment bloquant, puis ajoute une nouvelle arête. Cette nouvelle arête prolonge le segment anciennement bloqué (cf. Figure 3.19d). Un flip s'identifie à une arête finale d'un segment interne bloquant. Pour une tessellation, il y a  $2\overset{\circ}{n}_{s,b}(T)$  flips possibles. Le pavage obtenu après un flip est notée  $FT$ .

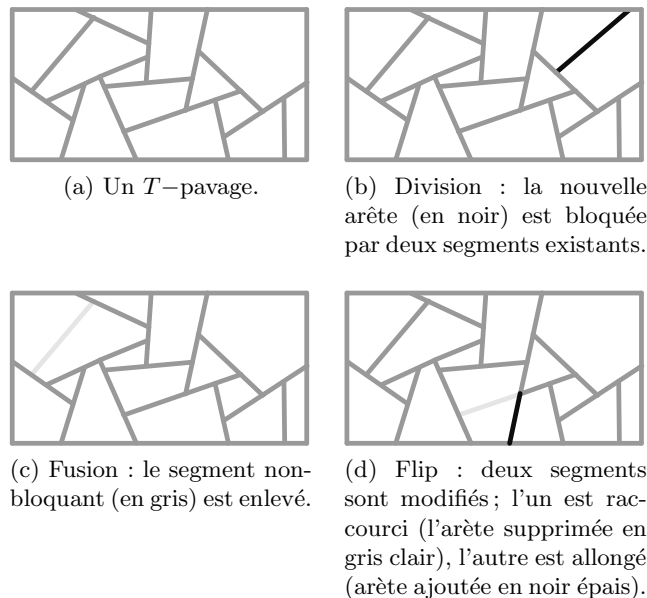


FIGURE 3.19 – Division, fusions et flips sont des transformations locales d'un  $T$ -pavage.

6. Ces trois termes sont traduits de l'anglais : split, merge et flip. Nous avons préservé le nom tel quel pour flip.



L'effet d'une division peut être compensée par une fusion, et vice-versa. Pour tout  $S \in \mathcal{S}_T$ ,  $S^{-1}$  est la fusion qui compense  $S$ . Toute opération flip  $F$  peut être compensée par une autre opération flip  $F^{-1}$ . Le pavage vide  $T_W$  est le  $T$ -pavage qui a comme unique cellule le domaine  $W$ .

**Proposition 10** *Soit  $T$  un  $T$ -pavage non vide. Il existe une séquence finie de divisions, fusions et flips qui transforme  $T$  en tessellation vide. Réciproquement, il existe une séquence finie de divisions et flips qui transforme la tessellation vide en  $T$ .*

**Corollaire 3** *Etant données deux  $T$ -pavages distincts, il existe une séquence finie d'opérations division, fusion et flip qui transforme l'un en l'autre.*

Soient  $\mathcal{S}_T$ ,  $\mathcal{M}_T$  et  $\mathcal{F}_T$  l'ensemble de toutes les divisions, fusions et flips qui peuvent être appliqués à  $T$ . Nous allons construire des mesures de probabilité sur chacun de ces espaces. Pour  $\mathcal{M}_T$  cela se fait d'une manière naturelle, car  $\mathcal{M}_T$  est un ensemble discret et fini. Par conséquent, quand cet ensemble est non-vide, on peut lui associer une loi uniforme. S'il n'est pas possible d'effectuer une fusion, la probabilité d'une fusion sera nulle. Pour les mêmes raisons, pour l'ensemble  $\mathcal{F}_T$ , nous procédons d'une manière analogue. Pour la mesure de probabilité sur  $\mathcal{S}_T$ , nous procédons en deux étapes, car cet ensemble n'est pas discret. Premièrement, une cellule  $c$  du pavage  $T$  est sélectionnée avec un poids proportionnel à son périmètre, ensuite on génère une droite uniforme rencontrant  $c$ . Cette construction permet de choisir uniformément une division dans l'ensemble  $\mathcal{S}_T$ . Tous les détails de ces constructions sont donnés dans [T6].

### 3.5.2 Le $T$ -pavage complètement aléatoire

L'espace  $\mathcal{T}$  est équipé avec le  $\sigma$ -algèbre  $\sigma(\mathcal{T})$  généré par les événements

$$\left\{ T \in \mathcal{T} : \left( \bigcup_{e \in E(T)} e \right) \cap K \neq \emptyset \right\}$$

avec  $K$  un sous-ensemble compact de  $W$  [122]. Un  $T$ -pavage est vu comme un ensemble fermé aléatoire.

Notre construction du  $T$ -pavage complètement aléatoire s'appuie sur le processus des droites poissonniennes, noté ici  $\mathbf{L}$ . Ce processus est supposé de densité unité. Ceci signifie que la longueur moyenne de la configuration de droites observées à travers une unité de surface, est un. Ici, nous considérons les  $T$ -pavages d'un domaine borné  $W$ . Ainsi, seulement la restriction de  $\mathbf{L}$

au  $W$  est relevante pour notre construction. Pour la simplicité, nous notons cette restriction  $\mathbf{L}$ . Tous les détails sont donnés dans [T6]. Pour la définition du processus des droites poissoniennes, nous recommandons [36, 180].

Soit la mesure de probabilité  $\mu$  définie par

$$\mu(A) = Z^{-1} \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \mathbf{1}_A(T), \quad A \in \sigma(\mathcal{T}), \quad (3.23)$$

où  $Z = \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \mathbf{1}_{\mathcal{T}(\mathbf{L})}(T)$  est la constante de normalisation. Pour que  $\mu$  soit bien défini, il faut vérifier que  $Z$  est finie. Ceci est une conséquence immédiate du résultat suivant.

**Théorème 11** *Soit  $\mathbf{L}$  la restriction au domaine  $W$  du processus de droites poissoniennes d'intensité unité et soit  $\phi : \mathcal{T} \rightarrow \mathbb{R}^+$  une fonction stable. Alors :*

$$\mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha$$

est finie pour tout  $\alpha \in \mathbb{R}$ .

**Preuve:** Pour  $\mathbf{T} \sim \mu$  et  $\phi$ , il existe un réel  $K$  tel que

$$\mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha \leq \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} K^{\alpha n_s(T)}.$$

Le nombre de segments internes de  $T$  est égal au nombre  $\mathbf{k}$  de droites dans  $\mathbf{L}$ . Il résulte que

$$\mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha \leq \mathbb{E} n_t(\mathbf{L}) K^{\alpha \mathbf{k}},$$

et en utilisant (3.22), nous obtenons

$$\mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha \leq c_1 + \mathbb{E} \mathbf{1}\{\mathbf{k} \geq 2\} C^{\mathbf{k}} \left( \frac{\mathbf{k}}{(\log \mathbf{k})^{1-\epsilon}} \right)^{k-k/\log k} K^{\alpha \mathbf{k}}.$$

La constante  $c_1$  est due aux termes de la somme pour lesquels le nombre de droites  $\mathbf{k}$  est plus petit que 2. Comme  $\mathbf{L}$  est la réalisation d'un processus de Poisson de droites d'intensité unité, alors  $\mathbf{k}$  suit une loi de Poisson de paramètre  $l(W)/\pi$ . Soit  $c_2$  la constante de normalisation de cette distribution. Nous avons

$$\mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha \leq c_1 + c_2 \sum_{k \geq 2} \left( \frac{l(W)}{\pi} \right)^k \frac{1}{k!} C^k \left( \frac{k}{(\log k)^{1-\epsilon}} \right)^{k-k/\log k} K^{\alpha k}.$$

En utilisant la formule de Stirling<sup>7</sup> et en posant  $\epsilon = 1/2$ , il s'ensuit le résultat

$$\begin{aligned} \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \phi(T)^\alpha &\leq c_1 + \frac{c_2}{\sqrt{2\pi}} \sum_{k \geq 2} \left( \frac{CK^\alpha l(W)e}{\pi \sqrt{\log k}} \right)^k \frac{1}{\sqrt{k}} \left( \frac{\sqrt{\log k}}{k} \right)^{k/\log k}, \\ &< +\infty. \end{aligned}$$

□

Par l'application du théorème 11 avec  $\phi = 1$ , nous obtenons que la mesure  $\mu$  définie par (3.23) est finie. La constante de normalisation  $Z$  n'a pas d'expression analytique connue. Une autre conséquence de ce théorème est que pour  $\mathbf{T} \sim \mu$  et  $\phi$  une fonction stable, tous les moments de  $\phi(\mathbf{T})$  sont finis.

Pour caractériser la mesure  $\mu$ , nous allons utiliser les mesures de Campbell et les noyaux de Papangelou. Utilisés, avec les processus ponctuels, ces outils nous permettent d'étudier le comportement d'un point par rapport au reste de la configuration observée. L'extension de ces outils pour les  $T$ -pavages n'est pas trivial. Ceci est du au fait que les composantes des  $T$ -tesselations points, arêtes et segments sont beaucoup plus contraints qu'un point dans une configuration de points.

Pour un  $T$ -pavage, les composantes qui peuvent être facilement ajoutées ou enlevées sont les segments non-bloquants. Dans la suite, toute division représentée par la paire  $(c, l)$  formée par une cellule  $c$  touchée par une droite  $l$  est identifiée avec le segment  $c \cap l$ . Par conséquent  $\mathcal{S}_T$  est identifié avec l'espace de ces segments et  $dS$  la mesure uniforme sur  $\mathcal{S}_T$  induit une mesure sur l'espace de segments non-bloquants, notée dans la suite  $ds$ . D'une manière analogue,  $\mathcal{M}_T$  est identifié avec le nombre de segments qui peuvent être enlevés de la tessellation  $T$ .

**Définition 20** Soit  $\mathcal{C}_s$  l'espace

$$\mathcal{C}_s = \{(s, T) : s \in \mathcal{S}_T, T \in \mathcal{c}T, \}.$$

La mesure (réduite) de Campbell des divisions d'un  $T$ -pavage aléatoire est définie par

$$C_s(\phi) = \mathbb{E} \sum_{m \in \mathcal{M}_T} \phi(m, \mathbf{T} \setminus \{m\}), \quad \phi : \mathcal{C}_s \rightarrow \mathbb{R}. \quad (3.24)$$

Le noyau de Papangelou division est le noyau  $P_s$  caractérisé par l'identité

$$C_s(\phi) = \mathbb{E} \int_{\mathcal{S}_T} \phi(s, \mathbf{T}) P_s(\mathbf{T}, ds), \quad \phi : \mathcal{C}_s \rightarrow \mathbb{R}. \quad (3.25)$$

---

7. La formule de Stirling utilisée est  $:k! = \frac{k^k}{e^k} \sqrt{2\pi k} [1 + O(\frac{1}{n})]$

**Proposition 11** Soit  $\mathbf{T} \sim \mu$ , alors l'identité suivante est vérifiée

$$C_s(\phi) = \mathbb{E} \int_{\mathcal{S}_{\mathbf{T}}} \phi(s, \mathbf{T}) ds, \quad \phi : \mathcal{C}_s \rightarrow \mathbb{R}. \quad (3.26)$$

**Preuve:** Les définitions de la mesure  $\mu$  et de la mesure de Campbell des divisions, données par (3.23) et la (3.24), respectivement, impliquent que cette dernière peut être écrite sous la forme

$$C_s(\phi) = \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \sum_{m \in \mathcal{M}_T} \phi(m, T \setminus \{m\}).$$

La somme appliquée aux segments non-bloquants est égale à la somme appliquée aux droites  $l \in \mathbf{L}$  combinées avec une fonction indicatrice. Soit  $s(l, T)$  le segment de  $T$  appartenant à la droite  $l \in \mathbf{L}$ . Cela nous donne

$$C_s(\phi) = \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \sum_{l \in \mathbf{L}} \mathbf{1}_{\{s(l, T) \text{ est non-bloquant}\}} \phi(s(l, T), T \setminus \{s(l, T)\}).$$

Pour  $l \in \mathbf{L}$  avec  $\mathbf{L}$  une configuration de droites donnée, soit la correspondance  $T \mapsto \tilde{T} = T \setminus s(l, T)$ . Clairement, si  $s(l, T)$  est non-bloquant, alors  $\tilde{T}$  est toujours un  $T$ -pavage. Ainsi, nous avons

$$\begin{aligned} & \sum_{T \in \mathcal{T}(\mathbf{L})} \mathbf{1}_{\{s(l, T) \text{ est non-bloquant}\}} \phi(s(l, T), T \setminus \{s(l, T)\}) \\ &= \sum_{\tilde{T} \in \mathcal{T}(\mathbf{L} \setminus l)} \sum_{c \in C(\tilde{T})} \mathbf{1}_{\{c \cap l \neq \emptyset\}} \phi(c \cap l, \tilde{T}). \end{aligned}$$

Il résulte

$$C_s(\phi) = \mathbb{E} \sum_{l \in \mathbf{L}} \sum_{T \in \mathcal{T}(\mathbf{L} \setminus l)} \sum_{c \in C(T)} \mathbf{1}_{\{c \cap l \neq \emptyset\}} \phi(c \cap l, T),$$

avec le terme à droite qui peut s'écrire

$$\mathbb{E} \sum_{l \in \mathbf{L}} \psi(l, \mathbf{L} \setminus l).$$

Comme  $\mathbf{L}$  est la réalisation d'un processus de Poisson d'intensité unité dans l'espace des droites, sa mesure de Campbell se décompose selon un noyau de Papangelou égal à la fonction d'intensité

$$\mathbb{E} \sum_{l \in \mathbf{L}} \psi(l, \mathbf{L} \setminus l) = \mathbb{E} \int \psi(l, \mathbf{L}) dl.$$

Ainsi, la mesure de Campbell des divisions pour  $\mathbf{T}$  s'écrit

$$\begin{aligned} C_s(\phi) &= \mathbb{E} \int \sum_{T \in \mathcal{T}(\mathbf{L})} \sum_{c \in C(T)} \mathbf{1}_{\{c \cap l \neq \emptyset\}} \phi(c \cap l, T) dl \\ &= \mathbb{E} \int \sum_{c \in C(\mathbf{T})} \mathbf{1}_{\{c \cap l \neq \emptyset\}} \phi(c \cap l, \mathbf{T}) dl. \end{aligned}$$

avec la dernière partie droite qui est obtenue grâce à (3.23).

Comme dans [T6], la mesure  $dS$  sur  $\mathcal{S}_T$  est définie par

$$ds = \sum_{c \in \mathcal{C}(T)} \mathbf{1}_{\{c \cap l \neq \emptyset\}} dl,$$

nous obtenons le résultat

$$C_s(\phi) = \mathbb{E} \int_{\mathcal{S}_T} \phi(S, \mathbf{T}) dS.$$

□

Comme conséquence directe de ce résultat, nous obtenons le corollaire suivant.

**Corollaire 4** *Le noyau de Papangelou des divisions d'un  $T$ -pavage  $\mathbf{T} \sim \mu$  est*

$$P_s(T, ds) = ds. \quad (3.27)$$

Si nous comparons ces résultats, avec ceux analogues relatifs aux processus ponctuels de Poisson homogènes, une possible interprétation intuitive serait que le coût de l'ajout d'un segment non-bloquant ne dépend pas de la tessellation existante, suggérant ainsi l'absence d'interaction dans un  $T$ -pavage distribué selon  $\mu$ .

Dans la suite, le raisonnement appliqué aux transformations division et fusion est étendu aux transformations flip.

**Définition 21** *Soit l'espace*

$$\mathcal{C}_f = \{(F, T) : F \in \mathcal{F}_T, T \in \mathcal{T}\}.$$

*La mesure de Campbell des flips d'un  $T$ -pavage aléatoire est définie par*

$$C_f(\phi) = \mathbb{E} \sum_{F \in \mathcal{F}_T} \phi(F^{-1}, F\mathbf{T}), \quad \phi : \mathcal{C}_f \rightarrow \mathbb{R}. \quad (3.28)$$

*Le noyau de Papangelou flip est le noyau caractérisé par l'identité*

$$C_f(\phi) = \mathbb{E} \sum_{F \in \mathcal{F}_T} \phi(F, \mathbf{T}) P_f(\mathbf{T}, F), \quad \phi : \mathcal{C}_f \rightarrow \mathbb{R}. \quad (3.29)$$

**Proposition 12** *Soit  $\mathbf{T} \sim \mu$ , alors l'identité suivante est vérifiée*

$$C_f(\phi) = \mathbb{E} \sum_{F \in \mathcal{F}_{\mathbf{T}}} \phi(F, \mathbf{T}), \quad \phi : \mathcal{C}_f \rightarrow \mathbb{R}. \quad (3.30)$$

**Preuve:** Les définitions de la mesure  $\mu$  et de la mesure de Campbell des flips, données par (3.23) et 21, respectivement, impliquent que cette dernière peut être écrite sous la forme

$$C_f(\phi) = \mathbb{E} \sum_{T \in \mathcal{T}(\mathbf{L})} \sum_{F \in \mathcal{F}_T} \phi(F^{-1}, FT).$$

La double somme peut être re-écrite en utilisant le changement de variables suivant

$$(F, T) \mapsto (\tilde{F} = F^{-1}, \tilde{T} = FT).$$

Une opération flip ne modifie pas l'ensemble de droites qui est le support d'un  $T$ -pavage :

$$L(T) = L(\tilde{T}), \quad \text{i.e.} \quad T \in \mathcal{T}(L) \Leftrightarrow \tilde{T} \in \mathcal{T}(L).$$

De plus, nous pouvons appliquer  $\tilde{F}$  à la tessellation  $FT = \tilde{T} : \tilde{F} \in \mathcal{F}_{\tilde{T}}$ . Il résulte que la mesure de Campbell des flips s'écrit

$$C_f(\phi) = \mathbb{E} \sum_{\tilde{T} \in \mathcal{T}(\mathbf{L})} \sum_{\tilde{F} \in \mathcal{F}_{\tilde{T}}} \phi(\tilde{F}, \tilde{T}) = \mathbb{E} \sum_{F \in \mathcal{F}_{\mathbf{T}}} \phi(F, \mathbf{T}).$$

□

**Corollaire 5** *Le noyau de Papangelou flip d'un  $T$ -pavage  $\mathbf{T} \sim \mu$  est*

$$P_f(T, F) \equiv 1. \quad (3.31)$$

En continuant, l'analogie avec les processus ponctuels de Poisson stationnaire, le noyau de Papangelou flip d'une  $T$ -tessellation distribué selon  $\mu$  montre que le coût de la transformation associée est constant et ne dépend pas du  $T$ -pavage initial.

Tous ces résultats, nous amènent à considérer le  $T$ -pavage généré selon  $\mu$  comme une structure complètement aléatoire ayant une dépendance spatiale minimale. C'est pour ces raisons que ce modèle est appelé dans [T6] le CRTT (Completely Random  $T$  Tessellation) modèle.

### 3.5.3 Modèles de Gibbs pour les $T$ -pavages

Nous présentons dans cette partie quelques modèles pour les  $T$ -pavages. Ces modèles sont décrits par des densités de probabilités construites par rapport à la mesure de référence  $\mu$ . Ils permettent construire des  $T$ -pavages ayant des caractéristiques géométriques structurées, plus proches de ce que l'on pourrait rencontrer dans les applications.

**Définition 22** Soit  $h : \mathcal{T} \rightarrow \mathbb{R}^+$  une fonction stable. Un  $T$ -pavage aléatoire de Gibbs est décrit par la densité non-normalisée  $h$  et il a comme distribution

$$\mathbb{P}(dT) \propto h(T) \mu(dT), \quad (3.32)$$

avec  $\int_{\mathcal{T}} \mathbb{P}(dT) = 1$ .

L'intégrabilité de la densité de probabilité  $h$  est assurée par la condition de stabilité. En utilisant la terminologie utilisée pour les processus de Gibbs,  $-\log h$  est appelée fonction d'énergie.

Les résultats obtenus pour les noyaux de Papangelou des  $T$ -pavages complètement aléatoires, peuvent être étendus aux  $T$ -pavages de Gibbs, sous une hypothèse similaire à l'hérédité pour les processus ponctuels.

**Définition 23** Un  $T$ -pavage de Gibbs avec densité de probabilité  $h$  est héréditaire si les conditions suivantes sont respectées :

- pour tout  $(T, S)$  tel que  $T \in \mathcal{T}$  et  $S \in \mathcal{S}_T$ ,  $h(ST) = 0 \Rightarrow h(T) = 0$ .
- pour tout  $(T, F)$  tel  $T \in \mathcal{T}$  et  $F \in \mathcal{F}_T$ ,  $h(FT) = 0 \Rightarrow h(T) = 0$ .

**Proposition 13** Soit  $\mathbf{T}$  un  $T$ -pavage de Gibbs héréditaire. Alors les noyaux de Papangelou division et flip sont

$$P_s(T, ds) = \frac{h(T \cup \{s\})}{h(T)} ds, \quad P_f(T, F) = \frac{h(FT)}{h(T)}, \quad (3.33)$$

et les deux formules de type Georgii-Nguyen-Zessin sont vérifiées

$$\mathbb{E} \sum_{m \in \mathcal{M}_{\mathbf{T}}} \phi(m, \mathbf{T} \setminus \{m\}) = \mathbb{E} \int_{s \in \mathcal{S}_{\mathbf{T}}} \phi(s, \mathbf{T}) \frac{h(\mathbf{T} \cup \{s\})}{h(\mathbf{T})} ds, \quad (3.34)$$

$$\mathbb{E} \sum_{F \in \mathcal{F}_{\mathbf{T}}} \phi(F^{-1}, F\mathbf{T}) = \mathbb{E} \sum_{F \in \mathcal{F}_{\mathbf{T}}} \phi(F, \mathbf{T}) \frac{h(F\mathbf{T})}{h(\mathbf{T})}, \quad (3.35)$$

en considérant  $0/0 = 0$ .

**Exemple 11** Soit la fonction d'énergie

$$-\log h(T) = -\overset{\circ}{n}_s(T) \log \tau, \quad (3.36)$$

qui a chaque segment interne de la configuration associe le potentiel  $\log \tau$ . Le paramètre  $\tau$  contrôle la densité de segments internes ou de droites sur laquelle le  $T$ -pavage est construit. Si  $\tau = 1$ , nous avons le modèle CRTT. D'une manière indirecte, le paramètre  $\tau$  peut être utilisé pour contrôler aussi le nombre de cellules dans un  $T$ -pavage. Les noyaux de Papangelou associés à la fonction d'énergie (3.36) sont

$$P_s(T, ds) = \tau ds, \quad P_f(T, F) \equiv 1.$$

Une réalisation du modèle est montrée dans la Figure 3.20a.

**Exemple 12** La fonction d'énergie

$$-\log h(T) = \frac{\tau}{\pi} l(T) + \overset{\circ}{n}_v(T) \log 2 - \overset{\circ}{n}_s(T) \log \tau \quad (3.37)$$

est le modèle du  $T$ -pavage proposé dans [1]. D'après le nom des auteurs de l'article, Arak, Clifford et Surgailis, ce modèle est appelé l'ACS modèle. Une réalisation de l'ACS modèle est montrée dans la Figure 3.20b. Le paramètre  $\tau$  a été ajusté de telle manière que le nombre moyen de cellules soit proche de celui obtenu par le CRTT modèle dans la Figure 3.20a.

Le modèle introduit [1] a des propriétés remarquables. Il est capable d'engendrer en fonction de ses paramètres, des différents types d'ensembles aléatoires comme les segments, les lignes brisées, les polygones ou les  $T$ -pavages aléatoires. La constante de normalisation de la densité de probabilité  $h$  a une expression analytique précise. Le paramètre  $\tau$  est égal à l'intensité linéaire de ce  $T$ -pavage. L'ACS modèle est markovien dans un sens analogue à celui exprimé par le Théorème 6. Un algorithme de simulation de ce modèle est également donnée dans [1]. Les distributions des longueurs des segments et des arêtes, et les distributions des surfaces et des périmètres des cellules ont été étudiées par [129].

Le noyaux de Papangelou division et flip pour l'ACS modèle peuvent être calculées à partir des relations (3.33) et (3.37). Pour la division avec un segment  $s$ , nous obtenons

$$\log \frac{P_s(T, ds)}{ds} = -\frac{\tau}{\pi} l(s) - 2 \log 2 + \log \tau,$$

avec  $l(s)$  la longueur du segment. Par rapport au CRTT, l'ACS modèle tend à diviser les cellules par des segments courts, en produisant des cellules de petite surface. Cet aspect a été vérifié numériquement par simulation. Le résultat de cette comparaison est montré dans la Figure 3.21.



**Exemple 13** Pour obtenir des  $T$ -pavages ayant des cellules avec des aires moins dispersées, nous pouvons considérer la fonction d'énergie suivante

$$-\log h(T) = -\overset{\circ}{n}_s(T) \log \tau + \alpha a^2(T), \quad (3.38)$$

où  $a^2(T)$  est la somme du carré des aires des cellules de la tessellation. La statistique  $a^2(T)$  est minimale pour des cellules de même taille. En remarquant que  $a^2(T) \leq a(W)^2$ , il résulte facilement que la fonction d'énergie (3.38) est stable, donc le modèle est bien défini. La Figure 3.20c montre une réalisation du modèle. Les paramètres du modèle ont été réglés afin qu'approximativement, le nombre moyen de cellules soit le même que pour les réalisations des modèles CRTT et ACS, montrées dans les Figures 3.20a et 3.20b, respectivement. La Figure 3.21 montre le résultat d'une étude par simulation, qui confirme qu'avec ces paramètres, ce modèle produit un nombre de cellules petites beaucoup moins important que les modèles CRTT et ACS.

**Exemple 14** Pour contrôler les angles entre deux segments qui s'intersectent, la fonction d'énergie suivante est considérée

$$-\log h(T) = -\overset{\circ}{n}_s(T) \log \tau - \beta \sum_{v \in V(T)} \phi(v), \quad (3.39)$$

avec  $\phi(v)$  le plus petit angle entre deux segments se touchant dans le sommet  $v$ . Avec  $\beta > 0$  nous favorisons les configurations ayant des angles  $\phi(v)$  proches de  $\pi/2$ . Comme  $\phi(v)$  est une grandeur finie, nous pouvons facilement prouver que la fonction d'énergie (3.39) est stable, donc que le modèle est bien défini. Une réalisation de ce modèle est montrée dans la Figure 3.20d. Par simulation, nous avons comparé les angles des  $T$ -pavages obtenus de ce modèle avec ceux obtenus avec le CRTT modèle. Les histogrammes de la Figure 3.21 indiquent que les angles dans le modèle CRTT ont tendance à être plus aigus.

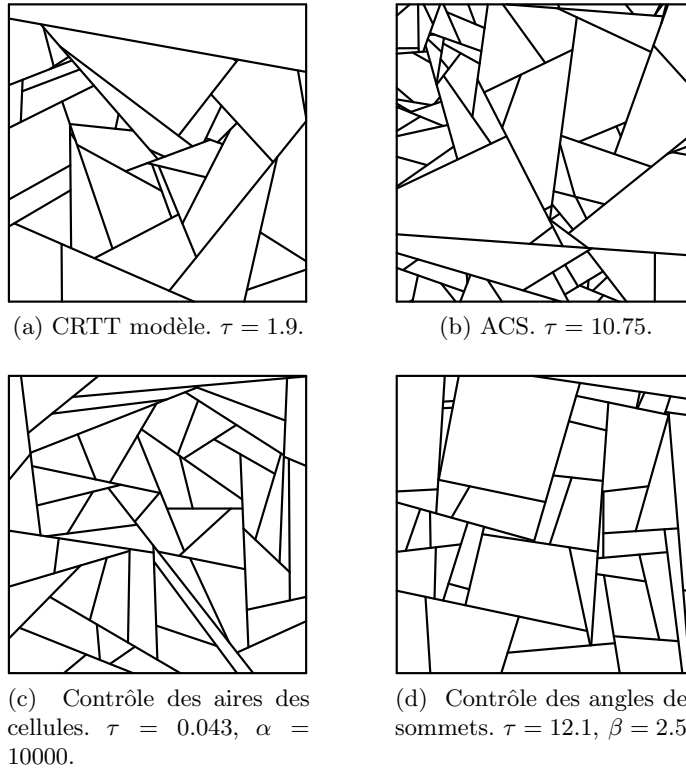


FIGURE 3.20 – Réalisations des quatre modèles présentés comme exemples.

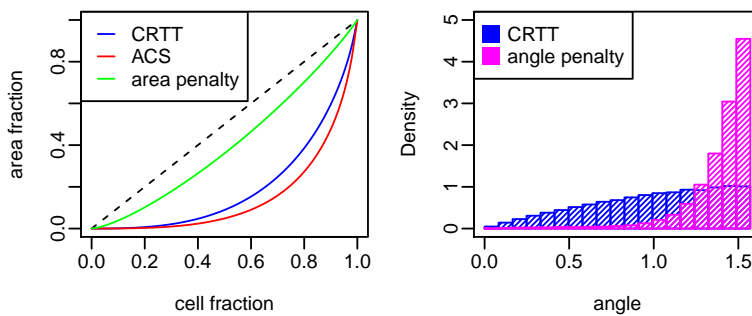


FIGURE 3.21 – Comparaison statistique des modèles présentés. A gauche : utilisations des courbes de Lorenz, pour étudier les distributions de l'aire des cellules pour les modèles CRTT, ACS et de contrôle des aires ( $Ox$  : pourcentage de petites cellules,  $Oy$  : taux de recouvrement de  $W$ , pointillé : courbe théorique pour des cellules avec des aires uniformément distribuées). A droite : histogrammes des angles aigus pour les modèles CRTT et de contrôle des angles.

**Deuxième partie**

**Simulation de type MCMC**

## Mise en oeuvre de la forme

Les densités de probabilité des processus ponctuels marqués avec interaction et d'autres modèles de structures spatiales possèdent des constantes de normalisation qui ne sont pas calculables analytiquement. Par conséquent, pour simuler la loi de tels processus, il faut recourir à des techniques de type MCMC (Markov chain Monte Carlo). Le principe de ces techniques réside dans la simulation d'une chaîne de Markov qui a pour loi d'équilibre la loi du processus d'intérêt.

Cette partie du mémoire est dédiée au développement et à l'analyse de techniques MCMC en vue de simuler les modèles de formes que nous avons construits. Deux grandes familles de méthodes MCMC vont être considérées : les dynamiques MCMC adaptées et les techniques de simulation exacte.

Les dynamiques adaptées sont des techniques MCMC qui permettent la construction des noyaux de transition qui accélèrent la convergence vers la loi stationnaire. Par exemple, nous pouvons proposer de connecter deux objets à l'aide d'un noyau de transition spécialement conçu dans ce but. Ce noyau complète le noyau "habituel" qui "jette" les objets uniformément dans le domaine jusqu'à ce qu'un objet "réalise" la connexion. De façon intuitive, une telle stratégie va améliorer les propriétés de la chaîne. Mais pour être rigoureux, la construction d'une telle dynamique doit être accompagnée de l'établissement théorique de ses propriétés (irréductibilité, récurrence et ergodicité).

Le grand inconvénient de la majorité des méthodes MCMC est que l'on ne sait pas avec précision quand la convergence de la chaîne est effectivement atteinte. La simulation exacte ou parfaite, est le nom générique des méthodes MCMC qui indiquent quand la chaîne a atteint son régime d'équilibre. Des méthodes de simulation exacte de processus ponctuels marqués comportant un nombre raisonnable d'objets en interaction simple peuvent être mises en oeuvre.

Pour cette présentation nous avons principalement suivi [2, 73, 72, 99, 107, 190, 128, 141, 148, 163, 186, 187]. Des parties de ce chapitre ont été publiées dans [T6, T7, T8, T17, T23].

## Chapitre 4

# Chaînes de Markov : notions, propriétés et algorithmes

### 4.1 Cadre théorique pour la simulation MCMC “classique”

Une chaîne de Markov peut être vue comme une suite de variables aléatoires indexées par un temps discret, telle que la réalisation de la variable à l’instant  $t$  ne dépend que de celle à l’instant  $t - 1$ . Un algorithme très général de simulation d’une chaîne de Markov pendant le temps  $T$  est le suivant :

#### Algorithme 1

$\mathbf{x}$  = Mon premier échantillonneur MCMC ( $T$ )

1. choisir une condition initiale  $\mathbf{x}_0$
2. pour  $i = 1$  à  $T$ , faire
  - {
  - $\mathbf{x}_i = \text{mettreAJour}(\mathbf{x}_{i-1})$
  - }
3. sortir  $\mathbf{x}_T$ .

D’un point de vue mathématique, la procédure `mettreAJour` est décrite par le noyau de transition de la chaîne que l’on souhaite simuler. Les propriétés demandées au noyau de transition ou à la chaîne sont l’irréductibilité, la récurrence et l’ergodicité. Ces propriétés garantissent que l’itération infinie du noyau de transition à partir de n’importe quelle condition initiale converge en variation totale vers l’unique distribution invariante de la chaîne. Il est évident que l’Algorithme 1 ne peut pas être utilisé un temps infini. Cependant, pour des valeurs de  $T$  suffisamment grandes, les échantillons obtenus suivent quasiment la loi invariante de la chaîne, permettant ainsi

l'utilisation de résultats importants comme la loi des grands nombres et le théorème de la limite centrale. Ceci justifie le succès de cette approche, pour peu que le mécanisme de mise à jour soit suffisamment simple par rapport à la loi que l'on souhaite simuler.

Ce chapitre décrit des outils mathématiques nécessaires pour la construction d'algorithmes MCMC en vue de la simulation de processus ponctuels ou d'autres modèles de structures géométriques aléatoires.

Soit  $(\Omega, \mathcal{F}, \mu)$  un espace de probabilité et  $X_0, X_1, X_2, \dots, X_n, \dots$  une suite de variables aléatoires définies dans cet espace.

**Définition 24** *Un noyau de transition est une fonction  $P : \Omega \times \mathcal{F} \rightarrow [0, 1]$  telle que*

(i)  $\forall \mathbf{x} \in \Omega, P(\mathbf{x}, \cdot)$  est une mesure de probabilité

(ii)  $\forall A \in \mathcal{F}, P(\cdot, A)$  est mesurable

Si l'espace  $\Omega$  est discret, le noyau de transition est donné par la matrice de transition  $P$

$$P_{\mathbf{x}\mathbf{y}} = \mathbb{P}(X_n = \mathbf{y} | X_{n-1} = \mathbf{x}).$$

Si  $\Omega$  est continu, le noyau de transition est donné par la densité conditionnelle  $P(\mathbf{x}, \mathbf{x}')$  du noyau de transition  $P(\mathbf{x}, \cdot)$ , c'est à dire

$$\mathbb{P}(X \in A | \mathbf{x}) = \int_A P(\mathbf{x}, d\mathbf{x}').$$

**Définition 25** *Soit  $P$  un noyau de transition. La suite de variables aléatoires  $X_0, X_1, \dots, X_n, \dots$  est une chaîne de Markov notée  $(X_n)$ , si pour tout  $t$  la loi conditionnelle de  $X_t$  sachant  $\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_0$  est la même que celle de  $X_t$  sachant  $\mathbf{x}_{t-1}$  :*

$$\mathbb{P}(X_{k+1} \in A | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k) = \mathbb{P}(X_{k+1} \in A | \mathbf{x}_k) = \int_A P(\mathbf{x}_k, d\mathbf{x}).$$

Une chaîne de Markov est dite *homogène* si

$$\mathbb{P}(X_{t_1}, X_{t_2}, \dots, X_{t_n} | \mathbf{x}_{t_0}) = \mathbb{P}(X_{t_1-t_0}, X_{t_2-t_0}, \dots, X_{t_n-t_0} | \mathbf{x}_0)$$

pour tout  $n \in \mathbb{N}^*$ .

Tout comme dans le cas discret, la probabilité de passer de  $\mathbf{x}$  en  $A$  en  $n$  pas est définie d'une manière itérative :

$$P^n(\mathbf{x}, A) = \int_{\Omega} P(\mathbf{x}, d\mathbf{y}) P^{n-1}(\mathbf{y}, A) \quad \mathbf{x} \in \Omega, A \in \mathcal{F}.$$

On pose aussi  $P^0(\mathbf{x}, A) = \delta_{\mathbf{x}}(A)$  qui est la mesure de Dirac en  $\mathbf{x}$  :

$$\delta_{\mathbf{x}}(A) = \begin{cases} 1 & \mathbf{x} \in A \\ 0 & \mathbf{x} \notin A. \end{cases}$$

**Théorème 12 Les équations de Chapman-Kolmogorov.** *On a pour tout couple d'entiers  $m$  et  $n$*

$$P^{m+n}(\mathbf{x}, A) = \int_{\Omega} P^m(\mathbf{x}, d\mathbf{y}) P^n(\mathbf{y}, A), \quad \mathbf{x} \in \Omega, A \in \mathcal{F}.$$

Intuitivement, ce résultat indique que pour passer de  $\mathbf{x}$  à  $A$  en  $m + n$  pas, la chaîne doit transiter en un certain état  $\mathbf{y}$  au pas  $m$ .

Dans le cas discret, les équations de Chapman-Kolmogorov peuvent être interprétées comme un produit matriciel. Dans le cas général, le noyau  $P^n$  est un opérateur qui agit à la fois sur des fonctions mesurables bornées et sur des mesures  $\sigma$ -finies sur  $\mathcal{F}$  via les formules

$$P^n f(\mathbf{x}) = \int_{\Omega} P^n(\mathbf{x}, d\mathbf{y}) f(\mathbf{y}), \quad \mu P^n(A) = \int_{\Omega} \mu(d\mathbf{x}) P^n(\mathbf{x}, A).$$

Ainsi, nous pouvons également écrire

$$P^n f(\mathbf{x}) = \int_{\Omega} P^n(\mathbf{x}, d\mathbf{y}) f(\mathbf{y}) = \delta_{\mathbf{x}} P^n f$$

**Définition 26** *Une mesure  $\sigma$ -finie sur  $(\Omega, \mathcal{F})$  est invariante si*

$$\pi(A) = \int_{\Omega} \pi(d\mathbf{x}) P(\mathbf{x}, A), \quad A \in \mathcal{F}.$$

La mesure invariante garantit la stationnarité de la chaîne de Markov comme suit : si un état initial est choisi selon une loi invariante  $\pi$ , alors l'action du noyau de transition est de produire un nouvel état de même loi  $\pi$ . Cet aspect est illustré par le calcul suivant :

$$\begin{aligned} \pi(A) &= \int_{\Omega} \pi(d\mathbf{x}) P(\mathbf{x}, A) \\ &= \int_{\Omega} \left[ \int_{\Omega} \pi(d\mathbf{y}) P(\mathbf{y}, d\mathbf{x}) \right] P(\mathbf{x}, A) \\ &= \int_{\Omega} \pi(d\mathbf{y}) \left[ \int_{\Omega} P(\mathbf{y}, d\mathbf{x}) P(\mathbf{x}, A) \right] \\ &= \int_{\Omega} \pi(d\mathbf{y}) P^2(\mathbf{y}, A) \\ &\quad \vdots \\ &= \int_{\Omega} \pi(d\mathbf{y}) P^n(\mathbf{y}, A), \end{aligned}$$

pour tout  $n \in \mathbb{N}^*$  et  $A \in \mathcal{F}$ .

Soit  $\mu$  une loi initiale quelconque. S'il existe une mesure  $\gamma(\mu)$  vérifiant

$$\gamma_\mu(A) = \lim_{n \rightarrow \infty} \int_{\Omega} \mu(d\mathbf{x}) P^{(n)}(\mathbf{x}, A)$$

pour une certaine topologie définie sur l'espace des mesures de probabilité et pour tout  $A \in \mathcal{F}$ , alors

$$\begin{aligned} \gamma_\mu(A) &= \lim_{n \rightarrow \infty} \int_{\Omega} \mu(d\mathbf{x}) \int_{\Omega} P^{n-1}(\mathbf{x}, d\mathbf{y}) P(\mathbf{y}, A) \\ &= \int_{\Omega} \gamma_\mu(d\mathbf{y}) P(\mathbf{y}, A). \end{aligned}$$

Il en résulte que  $\gamma_\mu$  est une mesure de probabilité invariante.

Par conséquent, si la limite  $\gamma_\mu$  existe, il s'agit d'une mesure invariante. Si de plus, la chaîne possède une unique mesure invariante, dans ce cas  $\gamma_\mu$  ne dépend pas de  $\mu$  et constitue la loi d'équilibre de la chaîne.

**Définition 27** *Un noyau de transition  $P(\mathbf{x}, A)$  est réversible par rapport à la mesure  $\pi$ , si l'intégrale*

$$\int_A \int_B \mathbf{1}_A(\mathbf{x}) \mathbf{1}_B(\mathbf{y}) \pi(d\mathbf{x}) P(\mathbf{x}, d\mathbf{y}) \quad (4.1)$$

*est symétrique en interchangeant  $A$  et  $B$ .*

En appliquant la définition précédente à la formule (4.1), nous obtenons

$$\int_A \pi(d\mathbf{x}) P(\mathbf{x}, B) = \int_B \pi(d\mathbf{x}) P(\mathbf{x}, A),$$

qui signifie que les probabilités de transiter de  $A$  vers  $B$  ou bien de  $B$  vers  $A$  sont les mêmes.

Prenons  $B = \Omega$ . Comme  $P(\mathbf{x}, \Omega) = 1$ , il en résulte immédiatement

$$\pi(A) = \int_{\Omega} \pi(d\mathbf{x}) P(\mathbf{x}, A),$$

Il apparaît ainsi qu'un noyau  $\pi$ -réversible admet  $\pi$  comme loi invariante.

De point de vue pratique et informatique, la procédure `mettreAJour` représente et simule un noyau de transition réversible par rapport à la distribution d'intérêt. Une telle procédure ne possédant que cette propriété, ne garantit la convergence de l'algorithme seulement que si les conditions initiales sont choisies selon la loi d'équilibre. Mais comme c'est précisément cette loi



que l'on souhaite simuler, il est nécessaire que la procédure `mettreAJour` "pousse" la chaîne simulée vers le régime d'équilibre. Cette condition est accomplie, si la chaîne construite est irréductible, apériodique et récurrente.

Dans ce qui suit, nous détaillons ces notions.

**Définition 28** *Pour tout ensemble  $A \in \mathcal{F}$ , la variable*

$$\tau_A = \min\{n \geq 1 : X_n \in A\}$$

*est appelée temps du premier retour de la chaîne en  $A$ . Si  $X_n \notin A$  pour tout  $n$ , on pose par convention  $\tau_A = +\infty$ .*

Pour les chaînes de Markov à espace d'états discret, l'irréductibilité signifie que tous les états communiquent, c'est à dire

$$\mathbb{P}_{\mathbf{x}}(\tau_{\mathbf{y}} < \infty) > 0 \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Mais si l'espace d'états est continu on a toujours  $\mathbb{P}_{\mathbf{x}}(\tau_{\mathbf{y}}) = 0$ . Pour définir correctement l'irréductibilité pour des chaînes sur un espace d'états général, nous avons besoin d'une mesure auxiliaire  $\phi$ .

**Définition 29** *Soit  $\phi$  une mesure sur  $\mathcal{F}$ . La chaîne de Markov  $(X_n)$  est dite  $\phi$ -irréductible si  $\phi(A) > 0$  implique  $\mathbb{P}_{\mathbf{x}}(\tau_A < \infty) > 0$  pour tout  $\mathbf{x} \in \Omega$ .*

Une manière équivalente pour de définir la  $\phi$ -irréductibilité est de dire qu'il existe  $n > 0$  tel que  $P^n(\mathbf{x}, A) > 0$ . L'irréductibilité permet donc à tout ensemble  $A \in \mathcal{F}$  suffisamment grand pour la mesure  $\phi$  d'être atteint à partir de n'importe quel état initial  $\mathbf{x} \in \Omega$ .

**Définition 30** *Un ensemble  $C$  est petit<sup>1</sup> s'il existe  $m \in \mathbb{N}^*$  et une mesure non-nulle positive  $\nu_m$  telle que*

$$P^m(\mathbf{x}, A) \geq \nu_m(A), \tag{4.2}$$

*pour tout  $\mathbf{x} \in C$  et  $A \in \mathcal{F}$ . L'ensemble  $C$  qui vérifie (4.2) est appelé  $\nu_m$ -petit.*

Le résultat suivant montre le lien entre la  $\phi$ -irréductibilité et les ensembles petits.

---

1. Les auteurs dans [128] utilisent les mots "small" et "petite" pour des ensembles de propriétés différentes.

**Théorème 13** *Soit  $(X_n)$  une chaîne  $\phi$ -irréductible. Pour tout  $A \in \mathcal{F}$  tel que  $\phi(A) > 0$ , il existe  $m \in \mathbb{N}^*$  et un ensemble petit  $C \subset A$  tel que  $\nu_m(C) > 0$ . De plus,  $\Omega$  est une union dénombrable d'ensembles petits.*

Ainsi, les ensembles petits peuvent être vues comme un outil pour la vérification de la non-irréductibilité.

L'irréductibilité, ou l'accessibilité de tout état à partir de tout autre état, est une propriété désirable quand on construit des chaînes de Markov pour simuler des lois de probabilité. Cependant, pour ce même objectif, le comportement cyclique ou périodique des chaînes est une propriété indésirable. Un tel comportement ne garantit pas la convergence vers la distribution d'équilibre.

**Définition 31** *Soit  $P(\cdot, \cdot)$  le noyau de transition d'une chaîne de Markov ayant une mesure invariante  $\pi$ . Supposons qu'il existe un ensemble  $A \in \mathcal{F}$ , une mesure de probabilité  $\nu$  telle que  $\nu(A) = 1$ , une constante  $\epsilon > 0$  et un entier  $n_0 \geq 1$  tel que*

$$P^{n_0}(\mathbf{x}, \cdot) \geq \epsilon \nu(\cdot) \quad \forall \mathbf{x} \in A.$$

*La chaîne de Markov associée est apériodique si*

$$\text{pgcd}\{m : \exists \epsilon_m > 0 \text{ such that } P^m(\mathbf{x}, \cdot) \geq \epsilon_m \nu(\cdot)\} = 1.$$

En particulier, nous simulons une chaîne de Markov apériodique si la procédure *mettreAJour* permet à la chaîne de rester dans le même état avec une probabilité positive.

**Théorème 14** *Soit  $(X_n)$  une chaîne  $\pi$ -irréductible et apériodique, qui admet  $\pi$  comme unique mesure invariante. Alors, il existe un ensemble  $\Omega' \subseteq \Omega$  tel que  $\pi(\Omega') = 1$  et*

$$\sup_{A \in \mathcal{F}} |P^n(\mathbf{x}, A) - \pi(A)| \rightarrow 0 \tag{4.3}$$

*pour tout  $\mathbf{x} \in \Omega'$ .*

Le résultat précédent permet la construction d'une chaîne de Markov qui converge vers sa loi d'équilibre  $\pi$  à partir de presque tout état initial. En effet, selon le théorème 14, la chaîne peut se trouver dans un état  $\mathbf{x} \in \Omega'$  avec  $\pi(\Omega') = 1$ , pour lequel la limite (4.3) est non-nulle.

La récurrence est la propriété qui prémunit contre l'existence d'un tel ensemble de mesure nulle  $\Omega''$ . Elle garantit ainsi l'indépendance des conditions

initiales pour un algorithme de type MCMC. La récurrence nous assure également que tout ensemble  $A \in \mathcal{F}$  est suffisamment visité à partir de tout état initial.

**Définition 32** *Pour tout ensemble  $A \in \mathcal{F}$ , le nombre de passages de  $(X_n)$  en  $A$  après le temps zero*

$$\eta_A = \sum_{n=1}^{\infty} \mathbf{1}\{X_n \in A\}$$

*est appelé temps d'occupation de l'ensemble  $A$ .*

**Définition 33** *Un ensemble  $A$  est Harris récurrent si  $\mathbb{P}_{\mathbf{x}}(\eta_A = \infty) = 1$  pour tout  $\mathbf{x} \in A$ . La chaîne  $(X_n)$  est Harris récurrente si elle est  $\phi$ -irréductible et si tout ensemble  $A \in \mathcal{F}$  tel que  $\phi(A) > 0$  est Harris récurrent.*

**Définition 34** *Soit*

$${}_A P^n(\mathbf{x}, B) = \mathbb{P}_{\mathbf{x}}(X_n \in B, \tau_A \geq n), \quad \mathbf{x} \in \Omega, \quad A, B \in \mathcal{F},$$

*la probabilité d'arriver dans l'ensemble  $B$  en  $n$  pas, en évitant l'ensemble  $A^2$ .*

Dans ce qui suit, nous utilisons la notation

$$U_A(\mathbf{x}, B) = \sum_{n=1}^{\infty} {}_A P^n(\mathbf{x}, B), \quad \mathbf{x} \in \Omega, \quad A, B \in \mathcal{F}$$

et **remarquons** que

$$U_A(\mathbf{x}, A) = \mathbb{P}_{\mathbf{x}}(\tau_A < \infty).$$

**Proposition 14** *Si pour tout ensemble  $A \in \mathcal{F}$ ,  $\mathbb{P}_{\mathbf{x}}(\tau_A < \infty) = 1$  pour tout  $\mathbf{x} \in A$ , alors  $\mathbb{P}_{\mathbf{x}}(\eta_A = \infty) = \mathbb{P}_{\mathbf{x}}(\tau_A < \infty)$  pour tout  $\mathbf{x} \in \Omega$  et  $(X_n)$  est Harris récurrent.*

Le résultat précédent montre que la Harris récurrence est une propriété plus forte que la  $\phi$ -irréductibilité. La Harris récurrence impose  $\mathbb{P}_{\mathbf{x}}(\tau_A < \infty) = 1$ , alors que la  $\phi$ -irréductibilité ne demande que  $\mathbb{P}_{\mathbf{x}}(\tau_A < \infty) > 0$ .

Souvent, les mécanismes **mettreAJour** qui sont  $\phi$ -irréductibles sont aussi Harris récurrents. Le résultat suivant donne un outil pour établir la Harris récurrence. Ce résultat utilise la condition de dérive.

---

2. En anglais : the  $n$ -step taboo probability.

**Théorème 15** Soit  $(X_n)$  une chaîne de Markov  $\phi$ -irréductible. Supposons qu'il existe un ensemble petit  $C \in \mathcal{F}$  et une fonction  $V : \Omega \rightarrow (0, \infty)$  telles que les ensembles de niveau

$$C_V(\alpha) = \{\mathbf{x} \in \Omega : V(\mathbf{x}) \leq \alpha\}$$

soient petits. La chaîne est Harris récurrente si la dérive  $\Delta V(\mathbf{x}) = PV(\mathbf{x}) - V(\mathbf{x})$  est négative pour tout  $\mathbf{x} \notin C$ , c'est à dire :

$$PV(\mathbf{x}) = \mathbb{E}[V(X_{t+1}) | X_t = \mathbf{x}] = \int_{\Omega} P(\mathbf{x}, d\mathbf{y})V(\mathbf{y}) \leq V(\mathbf{x}), \quad \mathbf{x} \notin C. \quad (4.4)$$

L'intérêt principal des algorithmes MCMC est que les échantillons obtenus peuvent être ensuite utilisés dans des résultats forts comme la loi des grands nombres ou le théorème de la limite centrale. Un algorithme qui simule une chaîne Harris récurrente ne permet pas directement une telle application. Il faudrait en effet attendre que le régime d'équilibre soit établi. Cela est impossible en pratique. Ce problème est toutefois résolu lorsque la chaîne est également ergodique.

Une chaîne de Markov  $(X_n)$  est ergodique si elle est à la fois Harris récurrente et apériodique. Dans ce cas, la limite (4.3) converge indépendamment des conditions initiales. De plus, nous connaissons la vitesse de convergence de cette limite, c'est à dire la rapidité avec laquelle la chaîne atteint son régime d'équilibre. Ainsi, même si comme précédemment, nous ne disposons que d'un nombre limité d'échantillons, nous avons l'espoir légitime d'obtenir des résultats d'autant plus précis que leur nombre est élevé.

**Définition 35** La norme en variation totale d'une mesure bornée et signée  $\nu$  sur  $(\Omega, \mathcal{F})$  est définie par

$$\|\nu\| = \sup_{A \in \mathcal{F}} \nu(A) - \inf_{A \in \mathcal{F}} \nu(A).$$

La distance en variation totale entre deux mesures  $\nu_1$  et  $\nu_2$  est  $\|\nu_1 - \nu_2\|$ .

**Définition 36** Soit  $(X_n)$  une chaîne ergodique de distribution invariante  $\pi$ . La chaîne est dite géométriquement ergodique s'il existe une fonction  $M : \Omega \rightarrow \mathbb{R}^+$  avec  $\pi\|M\| < \infty$  et une constante  $\rho \in (0, 1)$  telles que

$$\|P^n(\mathbf{x}, \cdot) - \pi\| \leq M(\mathbf{x})\rho^n$$

pour tout  $\mathbf{x} \in \Omega$ . La chaîne est dite uniformément ergodique s'il existe des constantes  $M > 0$  et  $\rho \in (0, 1)$  telles que

$$\sup_{\mathbf{x} \in \Omega} \|P^n(\mathbf{x}, \cdot) - \pi\| \leq M\rho^n.$$

L'ergodicité géométrique signifie que l'itération répétée du noyau de transition approche la distribution d'équilibre de la chaîne à une vitesse géométrique. Le résultat suivant permet d'établir cette propriété.

**Théorème 16** *Soit  $(X_n)$  une chaîne  $\phi$ -irréductible et apériodique. La chaîne est géométriquement ergodique s'il existe une fonction  $V : \Omega \rightarrow [1, \infty)$ , des constantes  $a < 1$  et  $b < \infty$ , et un ensemble petit  $C \in \mathcal{F}$  telles que*

$$PV(\mathbf{x}) \leq aV(\mathbf{x}) + b\mathbf{1}_C(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega. \quad (4.5)$$

Il est possible de montrer que si  $V$  satisfait (4.5), alors pour tout  $\alpha > 0$ , les ensembles de niveau  $C_V = \{\mathbf{x} \in \Omega : V(\mathbf{x}) \leq \alpha\}$  sont petits. Ceci implique que la condition de dérive pour l'ergodicité (4.5) implique la condition de dérive pour la récurrence (4.4). Il en résulte que si une chaîne est géométriquement ergodique, alors elle est Harris récurrente ([72] pp. 106).

## 4.2 Algorithme de type Metropolis-Hastings pour les processus ponctuels marqués

### 4.2.1 Construction de l'algorithme

Du point de vue historique, l'algorithme de Metropolis-Hastings a été introduit pour simuler des systèmes physiques à nombre fixé de particules [127]. La généralisation de cette approche aux chaînes de Markov à états discrets a été accomplie par [87]. L'adaptation de ce cadre pour les chaînes de Markov à états quelconques a été faite par [76, 79]. La simulation des processus ponctuels en utilisant un algorithme de type Metropolis-Hastings a été proposée par [73].

Pour échantillonner un processus ponctuel marqué, nous construisons une procédure `mettreAJour` qui simule une chaîne de Markov de loi d'équilibre

$$\pi(A) = \int_A p(\mathbf{x})\mu(d\mathbf{x}), \quad A \in \mathcal{F},$$

où  $p(\mathbf{x})$  est la densité de probabilité du processus et  $\mu$  la mesure de référence (2.4). Le processus ponctuel est supposé héréditaire.

Dans le cadre des algorithmes Metropolis-Hastings, le principe d'une telle procédure est le suivant : à partir d'une configuration initiale  $\mathbf{x}$ , une autre configuration  $\mathbf{y} \in A$  est proposée. Cet autre état est soit accepté comme

**nouvel** état initial, soit rejeté. Si l'état  $\mathbf{y}$  est rejeté, la chaîne reste dans son état initial  $\mathbf{x}$ . Cette idée est décrite par **le noyau** de transition

$$P(\mathbf{x}, A) = \int_A \alpha(\mathbf{x}, \mathbf{y})Q(\mathbf{x}, d\mathbf{y}), \quad \mathbf{x} \in \Omega, A \in \mathcal{F}, \mathbf{x} \notin A \quad (4.6)$$

et

$$P(\mathbf{x}, \{\mathbf{x}\}) = \int_{\mathbf{y} \neq \mathbf{x}} [1 - \alpha(\mathbf{x}, \mathbf{y})]Q(\mathbf{x}, d\mathbf{y}),$$

où  $Q$  est le noyau de proposition ou d'instrumentation<sup>3</sup> et  $\alpha(\cdot, \cdot)$  est la probabilité d'acceptation.

Le noyau de transition (4.6) est construit à partir de trois transformations ou mouvements appliqués à une configuration d'objets initiale  $\mathbf{x}$ . Ces transformations sont : ajout d'un objet, suppression d'un objet et remplacement d'un objet.

Un objet est ajouté, supprimé ou remplacé selon les probabilités respectives  $p_b, p_d$  et  $p_c$ . Un objet à ajouter  $\eta$  est simulé dans  $W \times M$  selon la densité  $b(\mathbf{x}, \eta)$ . Un objet à supprimer  $\eta$  est sélectionné dans la configuration  $\mathbf{x}$  selon la probabilité  $d(\mathbf{x}, \eta)$ . Un objet à modifier  $\eta$  est sélectionné dans  $\mathbf{x}$  puis remplacé par un nouvel objet  $\zeta$  selon la densité  $c(\mathbf{x}, \eta, \zeta)$ .

Un noyau de transition qui effectue ces transformations est

$$\begin{aligned} P(\mathbf{x}, A) = & p_b \int_K b(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{y} := \mathbf{x} \cup \{\eta\}) \mathbf{1}\{\mathbf{y} \in A\} d\sigma(\eta) \\ & + p_d \sum_{\eta \in \mathbf{x}} d(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{y} := \mathbf{x} \setminus \{\eta\}) \mathbf{1}\{\mathbf{y} \in A\} \\ & + p_c \sum_{\eta \in \mathbf{x}} q(\mathbf{x}, \eta) \int_K c(\mathbf{x}, \eta, \zeta) \alpha(\mathbf{x}, \mathbf{y} := \mathbf{x} \setminus \{\eta\} \cup \{\zeta\}) \mathbf{1}\{\mathbf{y} \in A\} d\sigma(\zeta) \\ & + \mathbf{1}\{\mathbf{x} \in A\} \left[ 1 - p_b \int_K b(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) d\sigma(\eta) \right. \\ & - p_d \sum_{\eta \in \mathbf{x}} d(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\}) \\ & \left. - p_c \sum_{\eta \in \mathbf{x}} q(\mathbf{x}, \eta) \int_K c(\mathbf{x}, \eta, \zeta) \alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\} \cup \{\zeta\}) d\sigma(\zeta) \right], \quad (4.7) \end{aligned}$$

avec  $K = W \times M$ ,  $d\sigma(\eta) = d\sigma((w, m)) = d\nu(w) \times d\nu_M(m)$  et  $0 < p_b + p_d + p_c \leq 1$ .

---

3. En anglais : proposal kernel.

Si l'état initial est l'ensemble vide, le noyau de transition ne peut pas proposer "enlever" ou "modifier" un objet. Dans ce cas, le noyau (4.7) s'écrit

$$P(\emptyset, A) = p_b \int_K b(\emptyset, \eta) \alpha(\emptyset, \{\eta\}) \mathbf{1}\{\{\eta\} \in A\} d\sigma(\eta) + \mathbf{1}\{\emptyset \in A\} \left[ 1 - p_b \int_K b(\emptyset, \eta) \alpha(\emptyset, \{\eta\}) d\sigma(\eta) \right]. \quad (4.8)$$

Nous observons que le noyau de transition (4.7) est une combinaison linéaire de deux noyaux. Le premier "ajoute" ou "supprime" un objet à la configuration initiale. Ce noyau se concentre sur les sous-espaces  $\Omega_{n-1} \cup \Omega_n \cup \Omega_{n+1}$  (ou  $\Omega_0 \cup \Omega_1$  si  $n = 0$ ). Le deuxième noyau "modifie" les caractéristiques d'un objet dans une configuration. Ce noyau se limite au sous-espace  $\Omega_n$ , car cette transformation ne change pas le nombre d'objets d'une configuration.

La loi  $p(\mathbf{x})$  est la loi invariante de la chaîne si le noyau de transition associé est réversible. Il n'est pas difficile de vérifier que si ces deux noyaux sont réversibles, alors il en est de même de leur combinaison linéaire [72].

Etablir la réversibilité d'une chaîne de Markov d'une manière directe peut s'avérer difficile. En général, il est plus simple de choisir la loi d'instrumentation et la probabilité d'acceptation d'un noyau afin que la condition d'équilibre local<sup>4</sup> soit respectée, et puis dans ce contexte, prouver la réversibilité de la chaîne.

Supposons que la chaîne se trouve dans l'état  $\mathbf{x} \in A \subseteq \Omega_n$  et qu'un "ajout" est proposé de sorte que la chaîne peut aller dans l'état  $\mathbf{x} \cup \{\eta\} \in B \subseteq \Omega_{n+1}$ . Le seul mouvement qui compense l'ajout est la "suppression" d'un objet. Dans ce cas, la condition d'équilibre local s'écrit

$$p(\mathbf{x}) p_b b(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) = p(\mathbf{x} \cup \{\eta\}) p_d d(\mathbf{x} \cup \{\eta\}, \eta) \alpha(\mathbf{x} \cup \{\eta\}, \mathbf{x}). \quad (4.9)$$

Dans ce qui suit, nous montrons comment la condition (4.9) implique la réversibilité du noyau de transition construit à partir des mouvements "ajouter" et "suppression".

En utilisant (2.1), (4.7) et (4.1), la probabilité d'être en  $A$  puis en  $B$  à l'itération suivante vaut

$$\frac{\exp[-\nu(W)]}{n!} \int_{(W \times M)^{(n+1)}} p_b p(\mathbf{x}) b(\mathbf{x}, \eta) \alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) \times \mathbf{1}\{\mathbf{x} \in A, \mathbf{x} \cup \{\eta\} \in B\} d\sigma(x_1) \dots d\sigma(x_n) d\sigma(\eta), \quad (4.10)$$

tandis que celle d'être en  $B$  puis en  $A$  à l'itération suivante est égale à

$$\frac{\exp[-\nu(W)]}{(n+1)!} \int_{(W \times M)^{(n+1)}} \sum_{\mathbf{y}_i \in \mathbf{y}} p_d p(\mathbf{y}) d(\mathbf{y}, y_i) \alpha(\mathbf{y}, \mathbf{y} \setminus \{y_i\}) \times \mathbf{1}\{\mathbf{y} \in B, \mathbf{y} \setminus \{y_i\} \in A\} d\sigma(y_1) \dots d\sigma(y_n) d\sigma(y_{n+1}). \quad (4.11)$$

---

4. En anglais : detailed balance condition.

En permutant les signes somme et intégrale de la dernière équation, nous obtenons  $(n + 1)$  intégrales car l'ordre des objets n'importe pas dans une configuration, et ceci nous permet d'écrire (4.11) de la manière suivante

$$\frac{\exp[-\nu(W)]}{n!} \int_{(W \times M)^{(n+1)}} p_d p(\mathbf{y}) d(\mathbf{y}, y_1) \alpha(\mathbf{y}, \mathbf{y} \setminus \{y_1\}) \times \mathbf{1}\{\mathbf{y} \in B, \mathbf{y} \setminus \{y_1\} \in A\} d\sigma(y_1) d\sigma(y_2) \dots d\sigma(y_{n+1}).$$

Posons  $\mathbf{x} = \mathbf{y} \setminus \{y_1\}$  et  $\eta = y_1$  dans (4.10). La condition (4.9) fournit alors la réversibilité, c'est à dire l'égalité des probabilités (4.10) et (4.11).

La probabilité d'acceptation peut être obtenue directement à partir des conditions d'équilibre détaillé (4.9). Plusieurs solutions existent [18, 127]. Pour les processus ponctuels marqués, nous avons opté pour le choix classique de la littérature [73, 72, 190, 141].

Soit  $r(\mathbf{x}, \eta)$ , le taux d'acceptation

$$r(\mathbf{x}, \eta) = \frac{p_d d(\mathbf{x} \cup \{\eta\}, \eta) p(\mathbf{x} \cup \{\eta\})}{p_b b(\mathbf{x}, \eta) p(\mathbf{x})} \quad (4.12)$$

et supposons qu'il soit toujours bien défini, c'est à dire que les quantités  $d(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$ ,  $p(\cdot)$  sont toujours strictement positives.

La probabilité d'acceptation d'un "ajouter" s'écrit alors

$$\alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) = \min\{1, r(\mathbf{x}, \eta)\}, \quad (4.13)$$

tandis que la probabilité d'acceptation d'une "suppression" s'obtient comme l'inverse du taux d'acceptation (4.12), ce qui donne

$$\alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\}) = \min\{1, 1/r(\mathbf{x} \setminus \{\eta\}, \eta)\}. \quad (4.14)$$

Pour éviter toute situation pathologique, on suppose  $\alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) = 0$  lorsque  $p(\mathbf{x} \cup \{\eta\}) = 0$ .

La probabilité d'acceptation d'une "modification" se déduit d'une manière analogue. La condition d'équilibre local pour ce type de transition s'écrit

$$p(\mathbf{x}) q(\mathbf{x}, \eta) c(\mathbf{x}, \eta, \zeta) \alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\} \cup \{\zeta\}) = p(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}) q(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}, \zeta) c(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}, \zeta, \eta) \alpha(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}, \mathbf{x}).$$

Sous ces hypothèses, la réversibilité du noyau de transition associé s'obtient en introduisant (4.7) et (2.1) dans (4.1).



Comme auparavant, l'élimination de cas pathologiques se fait à l'aide d'hypothèses très faibles. Il en résulte que la probabilité d'acceptation pour une "modification" est

$$\begin{aligned} \alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\} \cup \{\zeta\}) &= & (4.15) \\ &= \min \left\{ 1, \frac{q(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}, \zeta) c(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\}, \zeta, \eta) p(\mathbf{x} \setminus \{\eta\} \cup \{\zeta\})}{q(\mathbf{x}, \eta) c(\mathbf{x}, \eta, \zeta) p(\mathbf{x})} \right\}. \end{aligned}$$

L'algorithme suivant est un exemple de procédure `miseAJour` qui peut être implémenté pour simuler des processus ponctuels marqués.

### Algorithme 2

$\mathbf{y} = \text{miseAJour}(\mathbf{x})$

1. Choisir un mouvement selon les probabilités  $p_b, p_d$  and  $p_c$ .
2. Si un mouvement "ajouter" a été choisi, alors générer un nouveau objet  $\eta$  à l'aide de la densité  $b(\mathbf{x}, \eta)$ . Accepter la nouvelle configuration  $\mathbf{y} = \mathbf{x} \cup \{\eta\}$  avec la probabilité  $\alpha(\mathbf{x}, \mathbf{y})$  donnée par (4.13).
3. Si un mouvement "supprimer" est choisi, alors sélectionner un objet  $\eta \in \mathbf{x}$  pour être éliminé avec la probabilité  $d(\mathbf{x}, \eta)$ . Accepter la nouvelle configuration  $\mathbf{y} = \mathbf{x} \setminus \{\eta\}$  avec la probabilité  $\alpha(\mathbf{x}, \mathbf{y})$  donnée par (4.14).
4. Si un mouvement "remplacer" est choisi, alors sélectionner un objet  $\eta \in \mathbf{x}$  avec la probabilité  $q(\mathbf{x}, \eta)$  et lui appliquer une transformation à l'aide de la densité de probabilité  $c(\mathbf{x}, \eta, \zeta)$ . Accepter la nouvelle configuration  $\mathbf{y} = \mathbf{x} \setminus \{\eta\} \cup \{\zeta\}$  avec la probabilité  $\alpha(\mathbf{x}, \mathbf{y})$  donnée par (4.15).
5. Si la nouvelle configuration  $\mathbf{y}$  est acceptée, alors retourner  $\mathbf{y}$ . Sinon, retourner  $\mathbf{x}$ .

Le type de mouvement est choisi selon le mécanisme suivant. Premièrement, on génère une valeur uniforme sur  $[0, 1]$ . Les mouvements "ajout", "suppression" ou "modification" sont effectués selon que  $u \leq p_b$ ,  $p_b < u \leq p_b + p_d$  ou bien  $p_b + p_d < u \leq p_b + p_d + p_c$ . De même, l'acceptation d'une nouvelle configuration  $\mathbf{y}$  à partir de  $\mathbf{x}$  se fait si  $w \leq \alpha(\mathbf{x}, \mathbf{y})$  où  $w$  est une valeur uniforme  $w$  sur  $[0, 1]$ . Dans le cas contraire, la chaîne reste dans la configuration initiale  $\mathbf{x}$ .

### 4.2.2 Propriétés de convergence

L'algorithme précédent permet un large choix des lois d'instrumentation.

Par exemple, nous pouvons choisir pour  $p_b, p_d, p_c$  n'importe quelles valeurs positives pourvu que  $0 < p_b + p_d + p_c \leq 1$ . Ici, ces probabilités sont constantes. Il est possible de les choisir en fonction de la configuration courante de la

chaîne tout en préservant ses propriétés de convergence [72, 76]. Par ailleurs, les densités  $b(\mathbf{x}, \eta)$ ,  $d(\mathbf{x}, \eta)$ ,  $q(\mathbf{x}, \eta)$  et  $c(\mathbf{x}, \eta, \zeta)$  doivent être choisies strictement positives pour éviter les cas pathologiques [T7, T15, T17].

Pour ajouter un nouveau objet à la configuration courante, le choix le plus simple est la densité uniforme

$$b(\mathbf{x}, \eta) = \frac{1}{\nu(W)} \quad (4.16)$$

par rapport à  $d\sigma(\eta)$ . Par ce choix, l'objet proposé a une position uniformément distribué dans  $W$  alors que sa marque est choisie d'une façon indépendante selon  $\nu_M$ .

De même, pour supprimer un objet, le choix le plus simple est

$$d(\mathbf{x}, \eta) = \frac{1}{n(\mathbf{x})} \quad (4.17)$$

avec  $\mathbf{x} \neq \emptyset$ . Par ce choix uniforme, tout objet a la même chance d'être proposé pour être éliminé. Si  $\mathbf{x} = \emptyset$  aucun objet ne peut être proposé.

Le choix proposé par (4.16) et (4.17) n'est pas le seul. Pour l'instant, nous souhaitons montrer que ce simple choix mène à des algorithmes convergents pour une large famille de modèles. Disons toutefois, d'une façon intuitive, qu'une dynamique basée sur de telles lois instrumentales uniformes est plutôt adaptée aux modèles à interactions faibles.

Pour modifier les caractéristiques d'un objet d'une configuration, nous avons plusieurs manières de faire [126, 165, 166]. Ici, nous appliquons à nouveau une stratégie simple qui consiste à choisir uniformément l'objet à modifier

$$q(\mathbf{x}, \eta) = \frac{1}{n(\mathbf{x})}, \quad \mathbf{x} \neq \emptyset.$$

Une fois l'objet sélectionné, ses propriétés sont modifiées à l'aide de la densité  $c(\mathbf{x}, \eta, (w, m))$ . Soit  $C(\eta) = C_W(w_\eta) \times C_M(m_\eta)$  un voisinage du point marqué  $\eta = (w_\eta, m_\eta)$  tel que  $\nu(C_W(w_\eta))$  et  $\nu_M(C_M(m_\eta))$  soient strictement positives. Alors, la densité  $c(\mathbf{x}, \eta, (w, m))$  vaut

$$c(\mathbf{x}, \eta, (w, m)) = \frac{\mathbf{1}\{w \in C_W(w_\eta)\}}{\nu(C_W(w_\eta))} \times \frac{\mathbf{1}\{m \in C_M(m_\eta)\}}{\nu_M(C_M(m_\eta))}.$$

Pour garantir la réversibilité, nous devons avoir  $\zeta \in C(\eta)$  si et seulement si  $\eta \in C(\zeta)$ . Typiquement,  $C(\eta)$  est un voisinage relativement petit qui donne un caractère "local" à la modification d'un objet. De telles modifications "locales" ont un taux d'acceptation (4.15) proche de 1, ce qui fait que la

transition proposée est souvent acceptée. Ceci est très positif. Cependant, cette propriété introduit une corrélation importante entre échantillons. Du point de vue pratique, il y a un compromis à régler entre le résultat souhaité, le résultat obtenu et le coût calculatoire.

Les choix présentés et accompagnés des conditions d'équilibre local nous assurent que la chaîne de Markov simulée est réversible. Le résultat suivant nous montre que la chaîne simulée est irréductible, Harris récurrente et géométriquement ergodique [73, 72, 141, 190, T7].

**Théorème 17** *Soient les fonctions  $b, d, q$  et  $c$  décrites antérieurement. Supposons que  $b(\mathbf{x}, \eta)$  et  $d(\mathbf{x}, \eta)$  soient strictement positives sur leur domaines de définition respectives et que*

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \left[ \sup_{\eta \in W \times M, \mathbf{x} \in \Xi_n} \frac{d(\mathbf{x} \cup \{\eta\}, \eta)}{b(\mathbf{x}, \eta)} \right] \rightarrow 0.$$

*Posons également que  $p_b, p_d, p_c \in (0, 1)$  avec  $p_b + p_d + p_c = 1$  et que  $p(\mathbf{x})$  est la densité de probabilité d'un processus ponctuel marqué sur  $W \times M$ . Le processus ponctuel est localement stable et  $p(\mathbf{x})$  est construite par rapport à la mesure de référence  $\mu$ , représentée par le processus de Poisson marqué standard. Alors, l'échantillonneur Metropolis-Hastings pour la densité  $p$  défini par l'algorithme 2 simule une chaîne de Markov qui est  $\phi$ -irréductible, Harris récurrente et géométriquement ergodique.*

**Preuve:**

Soient  $\Lambda > 0$  la borne supérieure du rapport des vraisemblances  $p(\mathbf{x} \cup \{\eta\})/p(\mathbf{x})$  et  $V(\mathbf{x}) = A^{n(\mathbf{x})}$  avec  $A > 1$ .

Un objet généré  $\eta \notin \mathbf{x}$  est ajouté à la configuration  $\mathbf{x}$  avec la probabilité (4.13)

$$\min \left\{ 1, \frac{p_d d(\mathbf{x} \cup \{\eta\}, \eta) p(\mathbf{x} \cup \{\eta\})}{p_b b(\mathbf{x}, \eta) p(\mathbf{x})} \right\} \leq \frac{p_d \Lambda}{p_b} \times u_{n(\mathbf{x})} \leq \epsilon.$$

En effet, comme  $u_{n(\mathbf{x})} \rightarrow 0$ , quand  $n(\mathbf{x})$  est suffisamment grand la séquence ne peut pas dépasser une certaine quantité pre-fixée.

De manière analogue, un objet sélectionné  $\eta \notin \mathbf{x}$  est supprimé de la configuration  $\mathbf{x} \cup \{\eta\}$  avec la probabilité

$$\begin{aligned} & \min \left\{ 1, \frac{p_b b(\mathbf{x}, \eta) p(\mathbf{x})}{p_d d(\mathbf{x} \cup \{\eta\}, \eta) p(\mathbf{x} \cup \{\eta\})} \right\} \\ & \geq \min \left\{ 1, \inf \left[ \frac{b(\mathbf{x}, \eta)}{d(\mathbf{x} \cup \{\eta\}, \eta)} \right] \times \frac{p_b}{p_d \Lambda} \right\} \geq 1, \end{aligned}$$

car selon l'hypothèse nous avons

$$\inf_{\eta \in W \times M, \mathbf{x} \in \Xi_n} \frac{b(\mathbf{x}, \eta)}{d(\mathbf{x} \cup \{\eta\}, \eta)} \rightarrow \infty$$

quand  $n(\mathbf{x}) \rightarrow \infty$ .

Il résulte que pour des configurations  $\mathbf{x}$  à grand nombre d'objets  $n(\mathbf{x}) > N_\epsilon$ , on a  $\alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) \leq \epsilon$  et  $\alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\}) = 1$ . Par suite, en appliquant à la fonction  $V(\mathbf{x})$  le noyau de transition (4.7), nous obtenons

$$\begin{aligned} PV(\mathbf{x}) &= \int_{\Omega} P(\mathbf{x}, d\mathbf{y})V(\mathbf{y}) \\ &= p_b A^{n(\mathbf{x})} \int_{W \times M} b(\mathbf{x}, \eta)(A-1)\alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\})d\sigma(\eta) \\ &+ p_d A^{n(\mathbf{x})} \sum_{\eta \in \mathbf{x}} d(\mathbf{x}, \eta)(A^{-1}-1)\alpha(\mathbf{x}, \mathbf{x} \setminus \{\eta\}) + A^{n(\mathbf{x})} \\ &\leq [p_b(A-1)\epsilon + p_d(A^{-1}-1) + 1] V(\mathbf{x}). \end{aligned} \quad (4.18)$$

En passant à la limite dans (4.18), il résulte que

$$p_d(A^{-1}-1) + 1 = p_b + p_c + \frac{p_d}{A} < 1,$$

quand  $\epsilon$  tend vers zero. Il est donc possible de choisir  $\epsilon$  tel que

$$PV(\mathbf{x}) = \int_{\Omega} P(\mathbf{x}, d\mathbf{y})V(\mathbf{y}) \leq aV(\mathbf{x}) \quad (4.19)$$

pour une valeur  $a < 1$ .

Soit  $C$  l'ensemble des configurations comportant moins de  $N_\epsilon$  objets, c'est-à-dire

$$C = \{\mathbf{x} \in \Omega : n(\mathbf{x}) \leq N_\epsilon\}.$$

Il s'agit d'un petit ensemble. Pour vérifier cette affirmation, nous remarquons d'abord que la probabilité d'acceptation pour enlever un objet est toujours plus grande que

$$\Delta := \min \left\{ \frac{p_b}{u_n p_d \Delta} : n(\mathbf{x}) \leq N_\epsilon \right\}.$$

Evidemment, nous avons  $\Delta < 1$ . Puis en utilisant les équations Chapman-Kolmogorov et aussi le fait que  $P(\emptyset, \Xi_0) \geq p_d$ , il s'ensuit que

$$P^{N_\epsilon}(\mathbf{x}, \Xi_0) \geq P^{n(\mathbf{x})}(\mathbf{x}, \Xi_0)P^{N_\epsilon - n(\mathbf{x})}(\emptyset, \Xi_0) \geq (p_d \Delta)^{N_\epsilon}, \quad (4.20)$$

pour toute configuration  $\mathbf{x}$  contenant au plus  $N_\epsilon$  objets. L'ensemble  $C$  est **donc** petit par rapport à la mesure non-nulle  $(p_d \Delta)^{N_\epsilon} \delta_{\mathbf{x}}(\emptyset)$ .

Ayant une probabilité positive de rester dans le même état, la chaîne est apériodique. Elle est aussi  $\phi$ -irréductible en conséquence de (4.20). La condition de dérive (4.19) entraîne alors qu'elle est récurrente au sens de Harris. Définissant la fonction  $V(\mathbf{x})$  et l'ensemble  $C$  comme précédemment, et posant  $b = A^{N_\epsilon+1}$ , on voit que la condition de dérive pour l'ergodicité géométrique (4.5) est valable pour tout  $\mathbf{x} \in \Omega$ .  $\square$

### 4.3 Cadre théorique pour la simulation "parfaite"

Nous présentons ici des outils mathématiques nécessaires à la construction d'algorithmes de simulation "parfaite" pour les processus ponctuels. Ces algorithmes de simulation "parfaite" ou "exacte" sont des algorithmes de simulation markovienne qui indiquent par eux mêmes quand la convergence vers la loi d'équilibre a lieu. Le très fameux article de Prop et Wilson [155] a été le précurseur de ce domaine de la simulation.

Pour illustrer les idées principales de cette section, supposons que nous souhaitons simuler une loi  $\pi$  définie sur un espace d'états fini  $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ , par une approche markovienne. La chaîne de Markov  $(X_n)$  est représentée par sa fonction de transition  $\phi(\cdot, \cdot)$  sous la forme

$$X_{n+1} = \phi(X_n, V_n), \quad (4.21)$$

où  $V_n$  des variables aléatoires i.i.d.

Considérons  $m$  chaînes  $(X_n(\omega_i))$ , toutes initialisées d'un état différent  $\Omega$ , et qui évoluent du temps  $-T < 0$  jusqu'au temps 0. Ces chaînes sont couplées, au sens où elles utilisent les mêmes variables  $V_n$ . Si à un certain moment  $n \in -T, \dots, 0$  toutes les chaînes se trouvent dans le même état<sup>5</sup>, c'est à dire  $X_n \equiv \mathbf{x}$ , alors elles vont continuer d'évoluer toutes dans le même état, jusqu'au temps 0. L'influence des conditions initiales a été effacée. Nous pouvons commencer la simulation bien avant le temps  $-T$ , à l'infini, les chaînes vont se trouver toutes dans le même état au même moment. Il résulte donc que  $X_0$  est un échantillon "parfait" de la loi d'équilibre  $\pi$ .

Il n'est pas toujours pratique de lancer beaucoup de chaînes en parallèle. Cependant, si l'espace d'états est ordonné, avec un maximum  $\omega_{\max}$  et un minimum  $\omega_{\min}$ , et si en plus la fonction de transition respecte cette relation d'ordre, au sens où  $\omega \leq \omega'$  implique  $\phi(\omega) \leq \phi(\omega')$ , alors il n'est pas nécessaire de considérer tous les états initiaux possibles. Seuls les états  $\omega_{\min}$  et  $\omega_{\max}$  suffisent, car le comportement de tous les autres chaînes va être

---

5. Ce phénomène est appelé coalescence.

encadré par celui des chaînes extrémales  $X_n(\omega_{\min})$  et  $X_n(\omega_{\max})$ .

L'idée précédemment énoncée est connue sous le nom de couplage du passé<sup>6</sup>. L'algorithme suivant illustre ce principe. La procédure `mettreAJour` décrit la fonction de transition (4.21). Elle dépend de l'état de la chaîne et d'une variable aléatoire, par laquelle le couplage est réalisé.

### Algorithme 3

$(\mathbf{x}, T) = \text{Le premier échantillonneur parfait}$

1. initialiser  $T = 1$
2. initialiser  $V_n, n = -1, -2, \dots$  utilisant la loi uniforme sur  $[0, 1]$  d'une manière i.i.d
3. faire
  - {
  - $U_{-T}(-T) = \omega_{\max}, L_{-T}(-T) = \omega_{\min}$
  - for  $n = -T$  to  $-1$  do
  - {
  - $U_{-T}(n+1) = \text{mettreAJour}(U_{-T}(n), V_n)$
  - $L_{-T}(n+1) = \text{mettreAJour}(L_{-T}(n), V_n)$
  - }
  - if  $(U_{-T}(0) = L_{-T}(0))$  then  $\mathbf{x} = U_{-T}(0) = L_{-T}(0)$
  - else  $T = 2T$
  - }
  - tant que  $(U_{-T}(0) \neq L_{-T}(0))$
4. return  $\mathbf{x}$  et  $T$

L'exemple qui suit montre pourquoi il est nécessaire que le couplage se fasse du passé jusqu'en 0, et non pas vers le futur, c'est à dire de 0 jusqu'à un certain temps  $T$ .

**Exemple 15** Soit  $\Omega = \{1, 2\}$  et soit la chaîne de Markov définie sur  $\Omega$  par les probabilités de transition suivantes  $p_{11} = 1/2$ ,  $p_{12} = 1/2$ ,  $p_{21} = 1$  et  $p_{22} = 0$ . La coalescence a toujours lieu en l'état 1. Cependant, la mesure invariante est  $\pi_1 = 2/3$  et  $\pi_2 = 1/3$ . En faisant le couplage à partir de 0 vers le futur, nous introduisons donc un biais. Alors que, en faisant le couplage de  $-T$  vers 0, nous sommes sûrs que les deux états 1 et 2 vont être retournés, assortis des bonnes fréquences.

---

6. En anglais : Coupling From The Past (CFTP).

Le résultat suivant donne les conditions suffisantes pour que la sortie de l'Algorithme 3 constitue un échantillon non biaisé de la loi d'équilibre.

**Théorème 18** *Soit  $\pi$  une probabilité strictement positive sur l'espace muni d'une relation ordre  $(\Omega, \leq)$  et qui possède un état minimum  $\omega_{\min}$  et un état maximum  $\omega_{\max}$ . Soit  $\phi$  la fonction de transition d'une chaîne de Markov irréductible, apériodique avec  $\pi$  comme distribution invariante. La fonction  $\phi$  respecte la relation d'ordre sur  $\Omega$ . Si l'Algorithme C finit presque sûrement, alors il retourne un échantillon non biaisé de  $\pi$ .*

Toutes les fonctions de transition ne garantissent pas la coalescence presque sûre des chaînes.

**Exemple 16** *Comme dans l'exemple précédent, soit  $\Omega = \{1, 2\}$  et soit la chaîne de Markov définie par la matrice de transition  $p_{ij} = 1/2, i = 1, 2$ . La chaîne est irréductible, apériodique et elle a comme distribution invariante  $\pi = (1/2, 1/2)$ . Cette chaîne peut être simulée à l'aide de la fonction de transition*

$$\phi(X_n, (V_{n,1}, V_{n,2})) = \begin{cases} 1 & \text{si } V_{n,X_n} > \frac{1}{2} \\ 2 & \text{si } V_{n,X_n} \leq \frac{1}{2} \end{cases}$$

les marginales du vecteur  $(V_{n,1}, V_{n,2})$  étant uniformes sur  $[0, 1]$ , mais leur loi jointe arbitraire.

*Dans un premier temps, supposons  $V_{n,2} = 1 - V_{n,1}$ . Dans ce cas, la chaîne ne peut pas coalescer. Si  $V_{n,1} > 1/2$ , nous avons  $V_{n,2} < 1/2$  et donc  $\phi(1, (V_{n,1}, V_{n,2})) = 1$  ainsi que  $\phi(2, (V_{n,1}, V_{n,2})) = 2$ . Si  $V_{n,1} < 1/2$ , nous avons  $V_{n,2} > 1/2$  et donc  $\phi(1, (V_{n,1}, V_{n,2})) = 2$  ainsi que  $\phi(2, (V_{n,1}, V_{n,2})) = 1$ .*

*Considérons maintenant, le cas plus naturel  $V_{n,1} = V_{n,2}$ . Dans ce cas, la coalescence est immédiate. Si  $V_{n,1} = V_{n,2} < 1/2$ , nous avons  $\phi(1, (V_{n,1}, V_{n,2})) = \phi(2, (V_{n,1}, V_{n,2})) = 2$ . Si  $V_{n,1} = V_{n,2} > 1/2$ , nous avons  $\phi(1, (V_{n,1}, V_{n,2})) = \phi(2, (V_{n,1}, V_{n,2})) = 1$ .*

*Dans le cas où les composantes du vecteur sont indépendantes, l'Algorithme 3 se termine presque sûrement, mais ce choix n'est pas optimal en terme de temps de coalescence.*

## 4.4 Algorithme CFTP pour les processus ponctuels marqués

### 4.4.1 Processus spatiaux de naissance et de mort

Le recours à ces processus a peut-être constitué la technique la plus ancienne de simulation des processus ponctuels [97, 190, 141, 159].

Un processus spatial de naissance et de mort est un processus stochastique à temps continu qui à chaque instant  $t$  est la configuration  $X_t = \mathbf{x}$  d'un processus ponctuel sur  $\Omega$ . Ce processus évolue par sauts, un saut consistant en l'ajout ou la suppression d'un point. Le nombre de sauts dans un intervalle de temps borné est fini.

La loi de  $X_t$  est caractérisé par les fonctions  $b(\mathbf{x}, \eta)$  et  $d(\mathbf{x}, \eta)$ , qui représentent le taux de naissance et le taux de mort, respectivement. Le taux complet de naissance<sup>7</sup> est

$$B(\mathbf{x}) = \int_{W \times M} b(\mathbf{x}, \eta) d\sigma(\eta), \quad (4.22)$$

alors que le taux complet de mort<sup>8</sup> est

$$D(\mathbf{x}) = \sum_{\eta \in \mathbf{x}} d(\mathbf{x}, \eta). \quad (4.23)$$

L'algorithme suivant simule un processus de naissance-mort spatial.

#### Algorithme 4

$\mathbf{x} =$  *Processus Naissance-Mort* ( $T$ )

1. *initialiser*  $T_{pnm} = 0$ ,  $X^{(0)} = \mathbf{x}_0$  avec  $\mathbf{x}_0 \in \Omega$  et  $n = 0$
2. *tant que* ( $T_{pnm} < T$ )

{

·  $\mathbf{x} = X^{(n)}$

· *calculer*  $D(\mathbf{x})$  et  $B(\mathbf{x})$

· *choisir le temps de séjour*  $T^{(n)}$  *distribuée selon*  $\exp(1/(D(\mathbf{x}) + B(\mathbf{x})))$  *indépendamment des autres temps de séjour ou des états passés*

· *avec probabilité*  $\frac{B(\mathbf{x})}{B(\mathbf{x}) + D(\mathbf{x})}$  *la transition suivante est naissance ; un nouveau objet*  $\eta$  *est généré à l'aide de la densité de probabilité*

$$\frac{b(\mathbf{x}, \eta)}{B(\mathbf{x})}$$

---

7. En anglais : total birth rate

8. En anglais : total death rate



par rapport à la mesure de référence  $\sigma$  et ensuite il est rajouté à la configuration initiale  $\mathbf{x} := \mathbf{x} \cup \{\eta\}$

· avec probabilité  $\frac{D(\mathbf{x})}{B(\mathbf{x})+D(\mathbf{x})}$  la transition suivante est mort; l'objet  $\eta$  est enlevé de la configuration avec la probabilité

$$\frac{d(\mathbf{x}, \eta)}{D(\mathbf{x})},$$

et la nouvelle configuration devient  $\mathbf{x} := \mathbf{x} \setminus \{\eta\}$

·  $T_{nmp} = T_{nmp} + T^{(n)}$ ,  $n = n + 1$ ,  $X^{(n)} = \mathbf{x}$

}

Par rapport à cet algorithme, la dynamique Metropolis-Hastings que nous avons déjà présentée peut être vue comme une "discrétisation". Ceci tient à ce qu'on s'intéresse plutôt à ce qui se passe aux instants où a lieu un saut. Dans les deux algorithmes, les transitions sont des naissances et des morts, c'est à dire des ajouts ou des suppressions de point à la configuration courante. Cependant, dans le cas des processus naissance et mort, les transitions sont toujours acceptées.

Comme pour l'algorithme Metropolis-Hastings, nous avons plusieurs choix possibles pour les taux de naissance et de mort. Ces choix doivent garantir que l'algorithme converge et qu'il n'explose pas. Ceci veut dire que le processus de naissance et de mort a une mesure invariante et que le nombre de sauts reste fini dans un intervalle de temps borné. La loi du processus de naissance et de mort induit une loi sur le processus ponctuels que nous observons aux moments de sauts. Nous souhaitons pouvoir contrôler la loi de ce processus ponctuel.

Soit donc  $p(\mathbf{x})$ , la densité par rapport au processus de Poisson standard d'un processus ponctuel que l'on souhaite simuler. Supposons ici que la condition d'équilibre détaillé

$$b(\mathbf{x}, \eta)p(\mathbf{x}) = p(\mathbf{x} \cup \{\eta\})d(\mathbf{x} \cap \{\eta\}, \eta) \quad (4.24)$$

est respectée.

**Proposition 15** Si (4.24) est vérifiée et si  $\mathbb{E}B(X) < \infty$  avec  $X \sim \pi = \int p(\mathbf{x})d\mu(\mathbf{x})$ , alors  $X_t, t \geq 0$  est réversible par rapport à  $\pi$ .

Cela signifie que si  $X_0 \sim \pi$ , alors pour tout  $t > 0$ , le processus  $X_s, 0 \leq s \leq t$  et  $X_{t-s}, 0 \leq s \leq t$  sont identiquement distribués.

**Théorème 19** Soient  $n \in \mathbb{N}$ ,  $\beta_n = \sup_{n(\mathbf{x})=n} B(\mathbf{x})$  et  $\delta_n = \inf_{n(\mathbf{x})=n} D(\mathbf{x})$ . On suppose l'équilibre local (4.24) satisfait ainsi que  $\delta_n > 0$  pour tout  $n > 1$ . On suppose aussi que l'une des conditions suivantes est respectée  
(C1)  $\beta_n = 0$  pour  $n > 0$  suffisamment grand,  
(C2)  $\beta_n > 0$  pour  $n > 1$  et

$$\sum_{n=2}^{\infty} \frac{\beta_1 \cdots \beta_{n-1}}{\delta_1 \cdots \delta_n} < \infty, \quad \sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_{n-1}}{\beta_1 \cdots \beta_n} = \infty.$$

Alors il existe un unique processus spatial de naissance et de mort défini par les taux de transition  $b(\cdot, \cdot)$  et  $d(\cdot, \cdot)$  qui possède une unique loi d'équilibre vers laquelle il converge indépendamment des conditions initiales. Si la condition (4.24) découle de la loi d'un processus ponctuel, i.e.  $\pi = \int p(\mathbf{x}) d\mu(\mathbf{x})$ , nous avons

$$\pi(F) = \lim_{t \rightarrow \infty} \mathbb{P}(X_t \in F | X_0 = \mathbf{x}).$$

Soit  $p(\cdot)$  la densité d'un processus ponctuel localement stable, c'est à dire  $\lambda(\eta; \mathbf{x}) < \Lambda$  avec  $\Lambda > 0$ . Le processus spatial de naissance et de mort construit avec  $b(\mathbf{x}, \eta) = \lambda(\eta; \mathbf{x})$  et  $d(\mathbf{x}, \eta) = 1$  vérifie la condition d'équilibre détaillé (4.24). Le taux total de naissance est

$$B(\mathbf{x}) = \int_{W \times M} b(\mathbf{x}, \eta) d\sigma(\eta) \leq \Lambda \nu(W)$$

avec  $\beta_n \leq \Lambda \nu(W)$ , et le taux total de mort est  $D(\mathbf{x}) = n$  avec  $\delta_n = n$ .

Si  $\beta_n > 0$  pour tout  $n > 1$ , alors la deuxième condition du résultat précédent est vérifiée car

$$\sum_{n=2}^{\infty} \frac{\beta_1 \cdots \beta_{n-1}}{\delta_1 \cdots \delta_n} \leq \sum_{n=2}^{\infty} \frac{[\Lambda \nu(W)]^{n-1}}{n!} < \infty,$$

et

$$\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_{n-1}}{\beta_1 \cdots \beta_n} \geq \sum_{n=1}^{\infty} \frac{n!}{[\lambda \nu(W)]^{n-1}} = \infty.$$

Dans ces conditions, le théorème 19 dit qu'il existe un unique processus spatial de naissance et de mort défini par les taux de transition  $b(\mathbf{x}, \eta) = \lambda(\eta; \mathbf{x})$  et  $d(\mathbf{x}, \eta) = 1$ , et la loi de ce processus converge vers  $\pi = \int p d\mu$  à partir de toute configuration initiale.

La chaîne de Markov simulée par l'Algorithme 4 est uniformément ergodique ([141] Prop. G.7, pp. 276). Ceci a été un argument supplémentaire utilisé par la communauté pour simuler des processus ponctuels à partir de processus de naissance et de mort. Cependant, l'utilisation de cet algorithme est limitée par la nécessité de pouvoir calculer le taux de naissance total  $B(\mathbf{x})$ . A priori, ce problème peut être contourné par une technique

d'amincissement<sup>9</sup>. Son principe est le suivant : un processus de naissance et de mort de taux complet de naissance  $B(\mathbf{x}) = \Lambda\nu(W)$  est simulé, et la transition  $\mathbf{x} \rightarrow \mathbf{x} \cup \{\eta\}$  est acceptée avec probabilité  $\lambda(\mathbf{x}; \eta)/\Lambda$ . Cette technique fonctionne très bien quand les interactions entre points sont faibles. Dans le cas contraire,  $\Lambda$  peut être très grand, ce qui conduit à un très faible taux d'acceptation des naissances et limite l'utilisation pratique de l'algorithme [190].

#### 4.4.2 Algorithme CFTP pour les processus ponctuels

Les principes des algorithmes CFTP peuvent être appliqués pour simuler des processus ponctuels marqués par l'intermédiaire de processus de naissance et de mort. Soit  $p(\cdot)$  la densité d'un processus ponctuel  $X$  sur  $W$ . L'espace des configurations  $\Omega$  est ordonné par la relation d'inclusion. L'ensemble vide est l'élément minimum, mais il n'existe pas d'élément maximum. Il est remplacé par un maximum stochastique.

Le processus de naissance et de mort est un processus à temps continu, alors que l'algorithme CFTP est un algorithme à temps discret. Le rôle de la fonction de transition  $\phi(\cdot, \cdot)$  dans (4.21) est alors tenu par les taux de transition  $B(\mathbf{x})$  et  $D(\mathbf{x})$  respectivement donnés par (4.22) et (4.23).

Supposons le processus ponctuel  $X$  localement stable, c'est à dire qu'il existe  $\Lambda > 0$  tel que

$$0 \leq \lambda(\eta; \mathbf{x}) = \frac{p(\mathbf{x} \cup \{\eta\})}{p(\mathbf{x})} \leq \Lambda,$$

pour tout  $\eta \in W$  et tout  $\mathbf{x} \in \Omega$  tel que  $p(\mathbf{x}) > 0$ . Posons également que la densité conditionnelle est croissante pour l'inclusion, i.e.

$$\lambda(\eta; \mathbf{x}) \leq \lambda(\eta; \mathbf{x}') \quad \text{si} \quad \mathbf{x} \subseteq \mathbf{x}'. \quad (4.25)$$

Les processus ponctuels qui vérifient (4.25) sont appelés monotones ou attractifs.

Sous ces hypothèses, l'application des idées du couplage dans le passé aux processus ponctuels est décrite par l'algorithme suivant [99, 190, 141, T8].

#### Algorithme 5

$(\mathbf{x}, T) = \text{CFTP}$  pour les processus ponctuels marqués

1. Soit  $V_t, t \leq 0$  une famille de v.a. i.i.d de loi uniforme sur  $[0, 1]$

---

9. En anglais : thinning.

2. Soit  $D(0)$  la réalisation d'un processus de Poisson de densité  $\Lambda$  sur  $W \times M$ . Les points sont répartis uniformément dans  $W$  et les marques correspondantes sont générées indépendamment selon  $\nu_M$
3. Initialiser  $T = 1$
4. Le processus  $D(\cdot)$  est simulé de 0 vers  $-T$  en utilisant un processus spatial de naissance et de mort de taux de naissance  $b(\mathbf{x}, \eta) = \Lambda$  et de taux de mort  $d(\mathbf{x}, \eta) = 1$
5. Générer  $L^{-T}(\cdot)$  (processus inférieur) et  $U^{-T}(\cdot)$  (processus supérieur) de  $-T$  à 0 comme suit :
  - $L^{-T}(-T) = \emptyset$  et  $U^{-T}(-T) = D(-T)$  ;
  - si  $D(\cdot)$  a vécu une naissance dans le passé, i.e.  $D(t-) = D(t) \cup \{(w, m)\}$  où  $D(t-)$  est l'état juste avant le temps  $t$ , alors enlever le point  $(w, m)$  de  $L^{-T}(t-)$  et  $U^{-T}(t-)$  ;
  - si  $D(\cdot)$  a vécu une mort dans le passé, i.e.  $D(t-) = D(t) \setminus \{(w, m)\}$ , alors ajouter le point  $(w, m)$  à  $L^{-T}(t-)$  si
 
$$V_t \leq \alpha_{\min}(U^{-T}(t-), L^{-T}(t-), (w, m)),$$
 et à  $U^{-T}(t-)$  si
 
$$V_t \leq \alpha_{\max}(U^{-T}(t-), L^{-T}(t-), (w, m)).$$
6. Si  $U^{-T}(0) = L^{-T}(0)$  stop. Sinon, on pose  $T = 2T$  et on recommence.
7. Sortir  $U^{-T}(0)$  et  $T$ .

Les probabilités d'acceptation des naissances sont données par

$$\begin{aligned} \alpha_{\min}(U, L, (w, m)) &:= \min \{ \lambda((w, m); \mathbf{x}) / \Lambda : L \subseteq \mathbf{x} \subseteq U \} \\ \text{et} & \\ \alpha_{\max}(U, L, (w, m)) &:= \max \{ \lambda((w, m); \mathbf{x}) / \Lambda : L \subseteq \mathbf{x} \subseteq U \}. \end{aligned} \quad (4.26)$$

Il faut préciser que les valeurs des processus aléatoires générées sont réutilisées quand le temps est doublé.

Lorsque le processus à simuler est attractif, alors les probabilités (4.26) deviennent

$$\begin{aligned} \alpha_{\min}(U, L, (w, m)) &:= \lambda((w, m); L) / \Lambda \\ \text{et} & \\ \alpha_{\max}(U, L, (w, m)) &:= \lambda((w, m); U) / \Lambda. \end{aligned}$$

Les processus ponctuels qui vérifient la condition suivante sont dits antimotones ou répulsifs,

$$\lambda(\eta; \mathbf{x}) \leq \lambda(\eta; \mathbf{x}') \quad \text{si} \quad \mathbf{x}' \subseteq \mathbf{x}.$$

L'algorithme précédent peut s'utiliser pour ces processus également, en posant

$$\begin{aligned} \alpha_{\min}(U, L, (w, m)) &:= \lambda((w, m); U) / \Lambda \\ \text{et} & \\ \alpha_{\max}(U, L, (w, m)) &:= \lambda((w, m); L) / \Lambda. \end{aligned}$$

## Convergence

**Théorème 20** *Soit  $X$  un processus ponctuel marqué localement stable sur  $W \times M$  défini par la probabilité  $\pi = \int p d\mu$ ,  $\mu$  étant la mesure de référence poissonnienne unité et de marques i.i.d. selon  $\nu_M$ . L'Algorithme 5 finit presque sûrement et il retourne un échantillon non biaisé de densité  $p(\cdot)$ .*

**Preuve:** La preuve figure dans ([99] Thm. 2.1 pp. 847, [190] Thm. 3.5 pp. 94, [141] pp. 231).

Le processus dominant  $D(\cdot)$  est un processus de Poisson stationnaire de densité  $\Lambda$ . En tout  $T > 0$ , nous avons

$$\emptyset = L_{-T}(-T) \subseteq U_{-T}(-T) = D(-T).$$

Par construction, les fonctions  $\alpha_{min}$  et  $\alpha_{max}$  respectent l'ordre induit par l'inclusion, c'est-à-dire

$$L_{-T}(t) \subseteq U_{-T}(t), \quad (4.27)$$

pour tout  $t \in [-T, 0]$ . De plus, les processus  $L$  et  $U$  ont la structure d'un entonnoir<sup>10</sup>, i.e.

$$L_{-T}(t) \subseteq L_{-S}(t) \subseteq U_{-S}(t) \subseteq U_{-T}(t), \quad (4.28)$$

si  $-S \leq -T \leq t \leq 0$ . La première inclusion peut être vérifiée en remarquant que  $L_{-T}(t) = \emptyset \subseteq L_{-S}(t)$  et que la définition des transitions respectent l'inclusion. En ce qui concerne la dernière inclusion, nous avons  $U_{-T}(t) = D(-T) \supseteq U_{-S}(-T)$ . Si  $L_{-T}(t_0) = U_{-T}(t_0)$ , alors  $L_{-T}(t) = U_{-T}(t)$  pour tout  $t \in [t_0, 0]$  car les processus sont couplés.

Posons  $X_{-T} = \emptyset$  et définissons le processus  $X_{-T}(\cdot)$  sur  $[-T, 0]$  de la même manière que les processus  $L$  et  $U$ , à la différence qu'une naissance est acceptée si  $V_t \leq \lambda(\eta; X(t-))/\Lambda$ . Le processus  $X_{-T}(\cdot)$  est un processus spatial de naissance et de mort de taux de naissance  $\lambda(\cdot, \cdot)$  et de taux de mort unité. Par hypothèse  $\lambda(\eta; \mathbf{x}) \leq \Lambda$ , ce qui fait que le taux total de naissance est borné par  $\Lambda\nu(W)$ . En vertu du Thm. 19, le processus  $X_{-T}(\cdot)$  admet  $\pi$  comme distribution d'équilibre.

Par la propriété de monotonie de la construction, on a  $L_{-T}(0) \subseteq X_{-T}(0) \subseteq U_{-T}(0)$ , et en supposant pour l'instant que l'algorithme finit presque sûrement, alors la limite  $\lim_{T \rightarrow \infty} X_{-T}(0)$  est bien définie avec probabilité 1. Comme  $D$  est en équilibre et réversible,  $X_{-T}(0)$  a la même loi que son extension vers le futur, toujours en couplage avec le processus  $D$ , pendant une période de temps de longueur  $T$ . Il en résulte que  $L_{-T}(0)$  est un échantillon non biaisé

---

10. En anglais : funneling property.

de  $p(\cdot)$ .

Il reste à prouver que la coalescence se produit presque sûrement. Nous savons que  $\pi_\Lambda(\emptyset) = \exp[-\Lambda\nu(W)] > 0$  ou  $\pi_\Lambda$  est la loi d'un processus de Poisson de densité  $\Lambda$ . Pour  $n \in \mathbb{N}^*$ , définissons  $E_n = \mathbf{1}\{D(-n) \neq \emptyset\}$ . Il se trouve que  $(E_n)$  est une chaîne de Markov irréductible et apériodique sur  $\{0, 1\}$ , donc récurrente puisque discrète ([65], Prop 3.10, pp.14). La probabilité de l'état 0 est  $\pi_\Lambda(\emptyset)$  qui est strictement positive. Par conséquent, l'état 0 est atteint avec la probabilité unité. Ceci signifie que le processus  $D(t)_{t \leq 0}$  va se "vider" presque sûrement, auquel cas, le couplage implique que l'algorithme finit presque sûrement et la preuve est complète.  $\square$

# Chapitre 5

## Contributions

### 5.1 Construction des dynamiques adaptées

La convergence des dynamiques de type Metropolis-Hastings est influencée par le choix des lois d'instrumentation. Dans le cas de la simulation de processus ponctuels marqués, nous avons montré que le choix des lois d'instrumentation uniformes  $b(\mathbf{x}, \eta) = \frac{1}{\nu(W)}$  et  $d(\mathbf{x}, \eta) = \frac{1}{n(\mathbf{x})}$  garantit des propriétés remarquables de la dynamique ainsi construite. Cependant, un tel choix paraît moins naturel si le processus que l'on simule exhibe une forme plus structurée, avec des interactions fortes entre objets. Dans le cas de processus comme Candy ou Bisous, il est légitime de penser à des lois d'instrumentation qui "aident" le modèle, par exemple, en proposant des objets connectés à la structure déjà existante.

Un premier essai en ce sens a été fait pour simuler le modèle Candy dans [179, T15]. Les auteurs y construisent une dynamique Monte Carlo à sauts réversibles adaptée et des résultats numériques montrent les performances de l'algorithme proposé, sans toutefois donner la preuve de sa convergence théorique. Ce problème a été ultérieurement traité et résolu dans le cadre des dynamiques Metropolis-Hastings par [T7]. Depuis, la solution proposée a été généralisée à la simulation du modèle Bisous [T17].

L'idée de la solution consiste en la construction d'une densité non uniforme pour ajouter un objet à la configuration. Son expression est donnée par le mélange suivant

$$b(\mathbf{x}, \eta) = \frac{p_1}{\nu(W)} + p_2 b_a(\mathbf{x}, \eta), \quad (5.1)$$

avec  $p_1 + p_2 = 1$  et  $b_a(\mathbf{x}, \eta)$  une densité de probabilité proposant un objet en attraction et en alignement avec un autre objet de la configuration.

Cette stratégie est illustrée Figure 5.1. Les cercles blancs représentent les

extrémités autour desquelles l'on souhaite connecter d'autres segments pour prolonger le réseau existant. Les triangles rouges représentent les extrémités déjà connectées. De par la définition du modèle, une proposition d'extension dans ces régions a de fortes chances d'être refusée. Les cercles bleus représentent les extrémités libres, mais trop proches du bord du domaine  $W$ . Les segments connectés dans ces régions ne feront très probablement pas partie du domaine observé. Par conséquent, nous proposons de prolonger le réseau de préférence dans les régions où la connexité est permise. Une telle stratégie doit toujours avoir en vue de possibles effets de bord [179, T7, T17]. Des techniques d'échantillonnage sont proposées dans [7] pour éviter ces écueils.

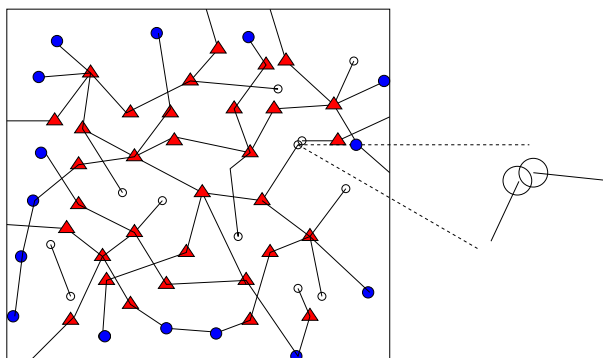


FIGURE 5.1 – Une configuration de segments aléatoires et régions d'attraction autour des segments favorables ou non à la connexité.

L'expression de la densité  $b_a(\mathbf{x}, \eta)$  est donnée par

$$b_a(\mathbf{x}, \eta) = \frac{1}{n(A(\mathbf{x}))} \sum_{x \in A(\mathbf{x})} \tilde{b}(x, \eta)$$

où  $A(\mathbf{x})$  est l'ensemble des objets de la configuration  $\mathbf{x}$  qui n'ont pas toutes leurs  $q$  extrémités connectées. Après avoir choisi uniformément un objet  $x$  dans l'ensemble  $A(\mathbf{x})$ , un nouvel objet  $\eta = (w_\eta, m_\eta)$  de paramètres  $(w, m_1, m_2)$  est proposé selon la densité

$$\tilde{b}(w, m_1, m_2) = f(w|x)g(m_1)h(m_2). \quad (5.2)$$

Dans cette densité (5.2) les termes suivants sont utilisés :

$$f(w|x) = \frac{\mathbf{1}\{w \in \tilde{a}(x)\}}{\nu(\tilde{a}(x) \cap W)} \quad (5.3)$$

$$g(m_1) = 1 \quad (5.4)$$

$$h(m_2) = \frac{2}{\tau} \mathbf{1}\{m_2 \in (1 - \tau, 1)\}, \quad (5.5)$$



où  $\tilde{a}(x) = \bigcup_{u=1}^{\tilde{q}} b(e_u(x), r_a)$  désigne l'union de toutes les sphères d'attraction de  $x$  qui ne contiennent aucune extrémité d'autre objet de  $\mathbf{x}$ . En utilisant  $m_1$  et  $m_2$ , le vecteur  $m_\eta$  est obtenu par une translation et deux rotations, telles que les conditions d'alignement soient respectées par rapport à  $m_x$ . Par translation suivant la règle d'attraction  $\mathcal{C}$ , le point  $w$  devient le centre du nouvel objet  $w_\eta$ .

L'utilisation de la densité (5.1) pour ajouter des objets et d'une loi uniforme pour supprimer des objets (4.17) mènent à la construction d'un algorithme convergent. Nous avons prouvé que la chaîne de Markov ainsi simulée est  $\phi$ -irréductible, Harris récurrente et géométriquement ergodique pourvu que

$$\sup_{\eta \in \Omega, n(\mathbf{x})=n} \frac{d(\mathbf{x} \cup \{\eta\}, \eta)}{b(\mathbf{x}, \eta)} \rightarrow 0$$

quand  $n \rightarrow \infty$  [T7, T17].

### 5.1.1 Etude par simulation de la dynamique adaptée au modèle Candy

Cette section présente une étude par simulation de la dynamique adaptée au modèle Candy [T7]. Ce modèle génère des configurations aléatoires de segments qui peuvent se connecter en formant un réseau. Il peut être vu aussi comme une application bidimensionnelle du modèle Bisous, pour générer des structures linéaires [T17]. Une réalisation du modèle ainsi que les paramètres d'interaction de la densité de probabilité sont donnés Figure 7.4. Le processus est défini dans la fenêtre  $W = [0, 256] \times [0, 256]$ . Les segments sont de longueur uniforme entre  $[30, 40]$  et leur orientation uniforme dans  $[0, \pi)$ . Le rayon d'attraction est  $r_a = 1/\sqrt{\pi}$ . Les valeurs seuil définissant les interactions par répulsion et alignement sont respectivement de  $\delta = 0.05\pi$  et de  $\tau = 0.2\pi$ .

Dans une première expérience, l'algorithme de Metropolis-Hastings défini par (4.7) et les probabilités  $p_b = 0.6$ ,  $p_d = 0.2$  et  $p_c = 0.2$  a été lancé à partir d'une configuration vide, pendant  $2 \times 10^7$  itérations. Des statistiques exhaustives ont été prélevées toutes les  $10^3$  itérations. La densité d'ajout d'un nouvel objet à la configuration  $b(\mathbf{x}, \eta)$  est un mélange de la forme (5.1), où  $p_1 = 0.2$  et  $p_2 = 0.8$ . Pour une configuration de segments  $\mathbf{x}$ , l'ensemble  $A(\mathbf{x})$  est l'union de toutes les sphères d'attraction des extrémités non connectées, à une distance supérieure à  $\frac{1}{2}l_{\max} + r_a$  du bord de  $W$ .

La probabilité de supprimer un objet est donnée par (4.17). Un objet est modifié de deux manières. Avec la probabilité  $p_{c1} = 0.1$  un objet est sélectionné au hasard, oté de la configuration, puis de nouvelles valeurs sont attribuées

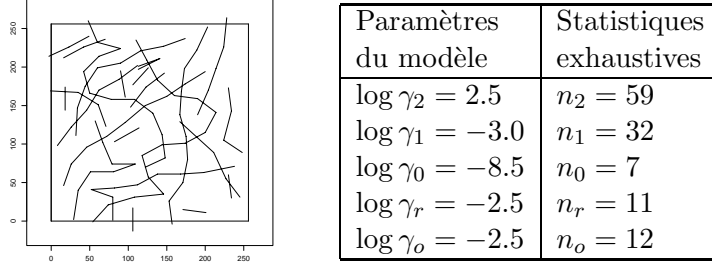
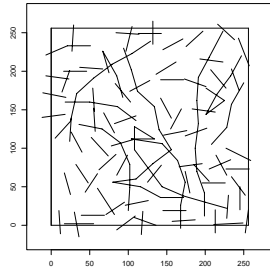


FIGURE 5.2 – Réalisation du Candy modèle avec les valeurs des paramètres d’interaction et les valeurs observées des statistiques exhaustives.

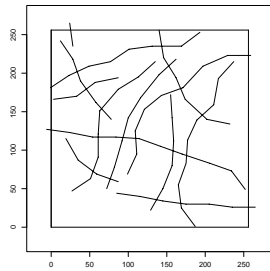
aux paramètres de cet objet selon la loi  $c(\mathbf{x}, x_i, \eta) = b(\mathbf{x} \setminus \{x_i\}, \eta)$ . Avec la probabilité  $p_{c2} = 0.1$  un objet est sélectionné selon le même schéma, puis une modification uniforme sur  $[0, \pi)$  de son orientation est proposée.

La Figure 5.3 illustre la dépendance de la morphologie d’une configuration de segments vis-à-vis des paramètres du modèle. La connexité du réseau est influencée par les paramètres  $\gamma_0, \gamma_1$  et  $\gamma_2$ , alors que sa courbure est plutôt contrôlée par les paramètres  $\gamma_o$  et  $\gamma_r$ .

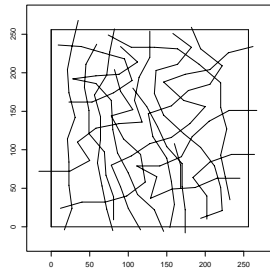
La deuxième expérience étudie les performances de l’algorithme Metropolis-Hastings en fonction de la configuration initiale et de l’influence des différents mouvements sur la vitesse de convergence. Les Figures 5.4 et 5.5 montrent deux réalisations du modèle Candy de référence (ses paramètres sont donnés Figure 7.4), obtenues par l’algorithme précédent, avec pour configuration initiale la réalisation d’un processus binomial et d’un réseau aléatoire. La réalisation du processus binomial est obtenue pour 200 segments. Le réseau aléatoire est obtenu à l’aide d’un algorithme Metropolis-Hastings qui ne fait que des modifications. L’algorithme est défini par un noyau de transition (4.7) où  $p_b = p_d = 0$  et  $p_{c1} = p_{c2} = 0.5$ , et les lois d’instrumentation  $c(\mathbf{x}, x_i, \eta)$  et  $c_\theta(\mathbf{x}, x_i, \eta)$  définies comme plus haut. L’état initial de l’algorithme est une réalisation d’un processus binomial de 200 segments. Comme pour la première expérience les algorithmes ont tourné pendant  $2 \times 10^7$  itérations et des statistiques exhaustives du modèle ont été prélevées toutes les  $10^3$  itérations. Les valeurs des moyennes des statistiques sont proches et leur évolution pendant la simulation ne donnent pas d’indications contraires à la convergence.



Paramètres du modèle	Statistiques exhaustives
$\log \gamma_2 = 2.5$	$n_2 = 39$
$\log \gamma_1 = -3.0$	$n_1 = 20$
$\log \gamma_0 = -5.0$	$n_0 = 68$
$\log \gamma_r = -2.5$	$n_r = 15$
$\log \gamma_o = -2.5$	$n_o = 4$

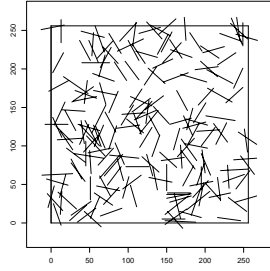


Paramètres du modèle	Statistiques exhaustives
$\log \gamma_2 = 2.5$	$n_2 = 47$
$\log \gamma_1 = -3.0$	$n_1 = 24$
$\log \gamma_0 = -8.5$	$n_0 = 1$
$\log \gamma_r = -6.5$	$n_r = 0$
$\log \gamma_o = -6.5$	$n_o = 0$

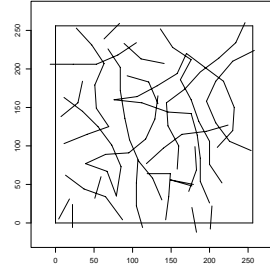


Paramètres du modèle	Statistiques exhaustives
$\log \gamma_2 = 4.0$	$n_2 = 107$
$\log \gamma_1 = -3.0$	$n_1 = 30$
$\log \gamma_0 = -8.5$	$n_0 = 0$
$\log \gamma_r = -2.5$	$n_r = 14$
$\log \gamma_o = -2.5$	$n_o = 23$

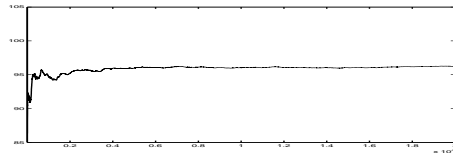
FIGURE 5.3 – Réalisations du modèle Candy, avec ses valeurs des paramètres d'interaction et les statistiques exhaustives observées.



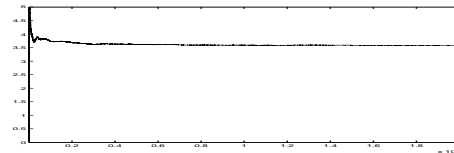
Configuration initiale



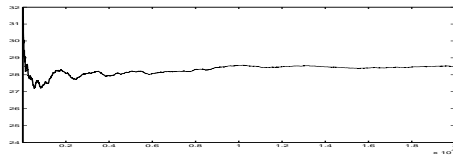
Etat final



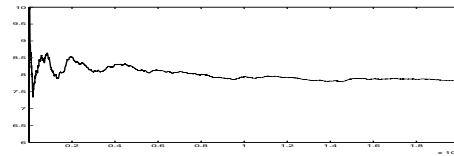
$$\bar{n}_t = \bar{n}_2 + \bar{n}_1 + \bar{n}_0 = 96.23$$



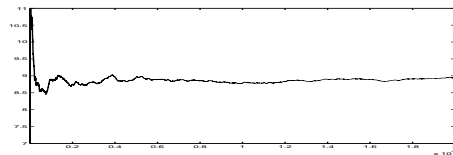
$$\bar{n}_0 = 3.58$$



$$\bar{n}_1 = 28.48$$

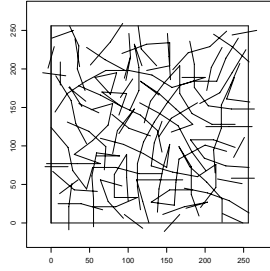


$$\bar{n}_r = 7.84$$

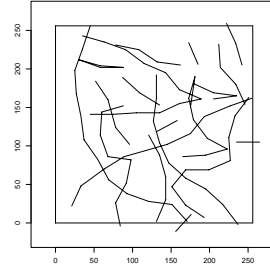


$$\bar{n}_o = 8.94$$

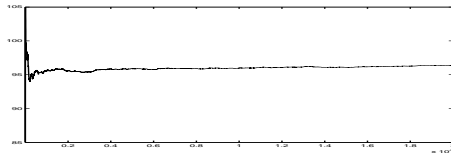
FIGURE 5.4 – Séries temporelles des moyennes cumulées des statistiques exhaustives durant la simulation du modèle Candy par l’algorithme Metropolis-Hastings décrit dans le texte. La configuration initiale (réalisation d’un processus binomial de 200 segments) est montrée en haut à gauche, la configuration finale en haut à droite.



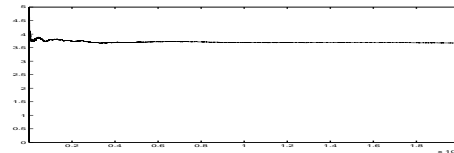
Configuration initiale



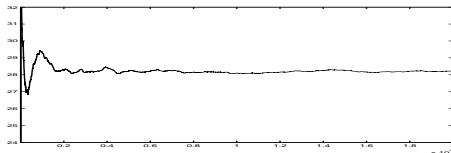
Etat final



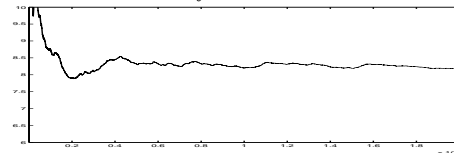
$$\bar{n}_t = \bar{n}_2 + \bar{n}_1 + \bar{n}_0 = 96.39$$



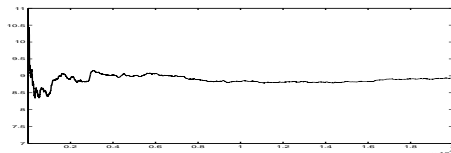
$$\bar{n}_f = 3.67$$



$$\bar{n}_c = 28.21$$



$$\bar{n}_r = 8.19$$



$$\bar{n}_o = 8.93$$

FIGURE 5.5 – Séries temporelles des moyennes cumulées des statistiques exhaustives durant la simulation du modèle Candy par l’algorithme Metropolis-Hastings décrit dans le texte. La configuration initiale (réalisation d’un processus binomial de 200 segments) est montrée en haut à gauche, la configuration finale en haut à droite.

Nous avons ensuite fait varier les poids des différents mouvements. La Figure 5.6 montre une réalisation et les séries temporelles des moyennes cumulées des statistiques pour les poids  $p_b = 0.45$ ,  $p_d = 0.15$ ,  $p_{c1} = 0.3$ ,  $p_{c2} = 0.1$ . Pour la Figure 5.7 ces poids sont  $p_b = 0.7$ ,  $p_d = 0.1$ ,  $p_{c1} = 0.1$  et  $p_{c2} = 0.1$ . Dans les deux cas, nous avons  $p_1 = 0.2, p_2 = 0.8$ . ainsi que  $p_M = p_{\tilde{M}} = 0.5$ .

Les résultats indiquent que ni la configuration initiale, ni les poids des mouvements du noyau de transition influent sur les résultats de la simulation avec les paramètres considérés. Cependant, la probabilité  $p_2$  ne doit pas être trop petite pour ne pas complètement exclure des mouvements adaptés. La Figure 5.8 montre une simulation où nous avons utilisé seulement des naissances uniformes, c'est-à-dire  $p_b = 0.75, p_d = 0.25$ ,  $p_1 = 1.0$  et  $p_2 = p_c = 0$ . Les courbes montrent qu'après un grand nombre d'itérations un réseau connecté se forme. Mais l'évolution des statistiques exhaustives révèle une certaine absence de stationnarité, ce qui contraste avec les autres exemples.

## 5.2 Dynamique Monte Carlo pour la simulation d'un modèle de pavage en $T$

Pour la construction d'un  $T$ -pavage, nous avons mis en place les idées utilisées pour les dynamiques adaptées aux processus ponctuels avec interaction. Cette fois-ci, nous avons du prendre en compte que le passage d'une configuration à une autre se faisait en utilisant trois transformations qui conservent la structure de  $T$ -pavage. Cest trois transformations sont la division, la fusion et le flip. Nous présentons ici que les points clef de la construction de cette dynamique. Tous les détails se trouvent dans l'article [T6].

L'algorithme fonctionne selon la procédure suivante :

### Algorithme 6

$\mathbf{T}_{n+1} = \text{miseAJour-}T\text{-Tesselation}(\mathbf{T}_n)$

1. La configuration initiale est la  $T$ -tesselation  $\mathbf{T}_n$ .
2. Choisir une transformation division, fusion ou flip, avec les probabilités respectives  $p_s, p_m, p_f$ .
3. Si la transformation choisie n'est pas applicable, retourner  $\mathbf{T}_n$ .
4. Selon la transformation choisie, proposer une nouvelle tessellation  $\mathbf{T}_{n+1}$ , en utilisant le noyau d'instrumentation approprié ( $q_s(T, \cdot), q_m(T, \cdot)$  ou  $q_f(T, \cdot)$ ).

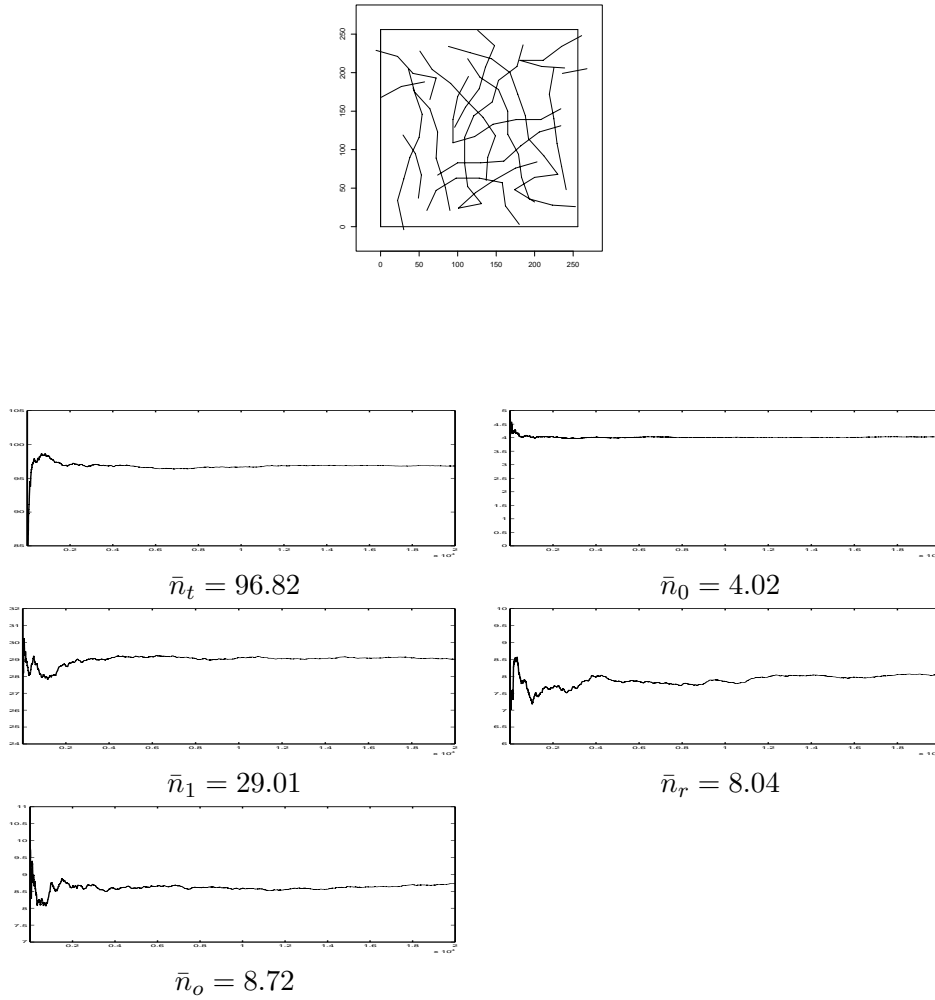


FIGURE 5.6 – Séries temporelles des moyennes cumulées des statistiques exhaustives obtenues par l’algorithme Metropolis-Hastings avec les paramètres  $p_b = 0.45$ ,  $p_d = 0.15$ ,  $p_{c1} = 0.3$ ,  $p_{c2} = 0.1$ . La configuration initiale est l’ensemble vide. La configuration est montrée en haut de la figure.

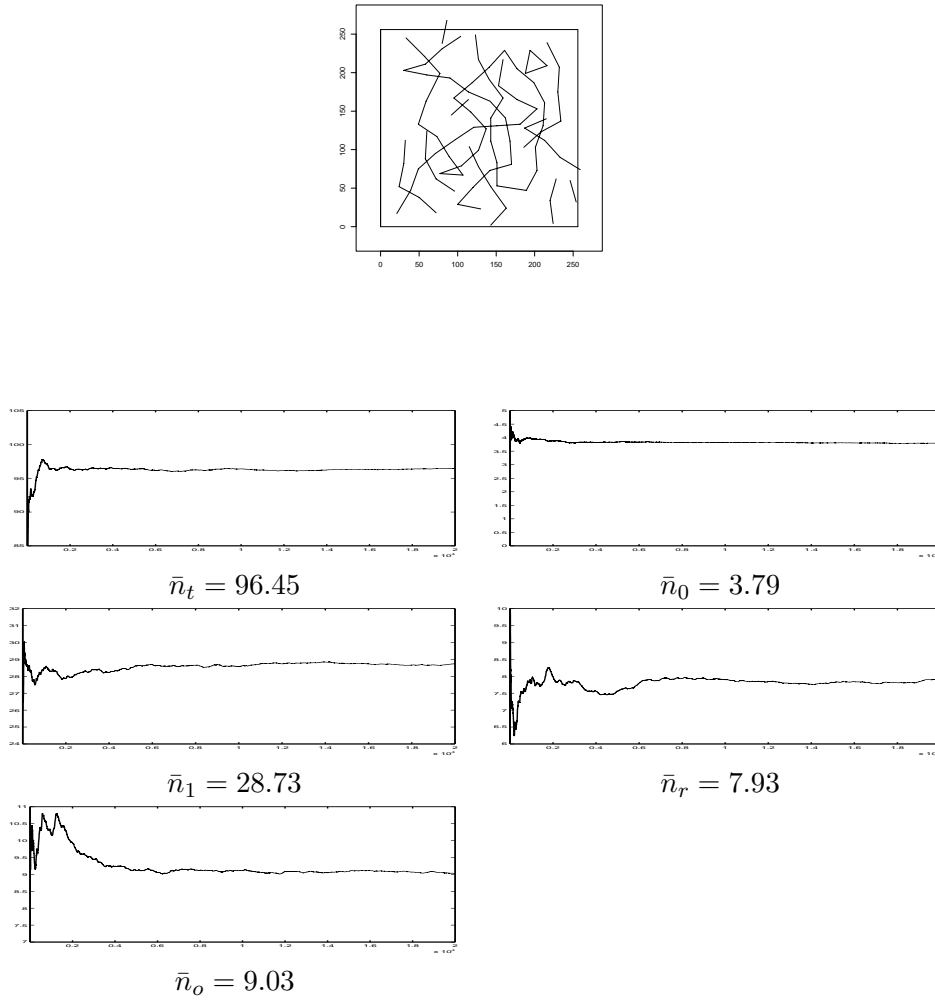


FIGURE 5.7 – Séries temporelles des moyennes cumulées des statistiques exhaustives obtenues par l’algorithme Metropolis-Hastings avec les paramètres  $p_b = 0.7$ ,  $p_d = 0.1$ ,  $p_{c1} = 0.1$ ,  $p_{c2} = 0.1$ . La configuration initiale est l’ensemble vide. La configuration est montrée en haut de la figure.



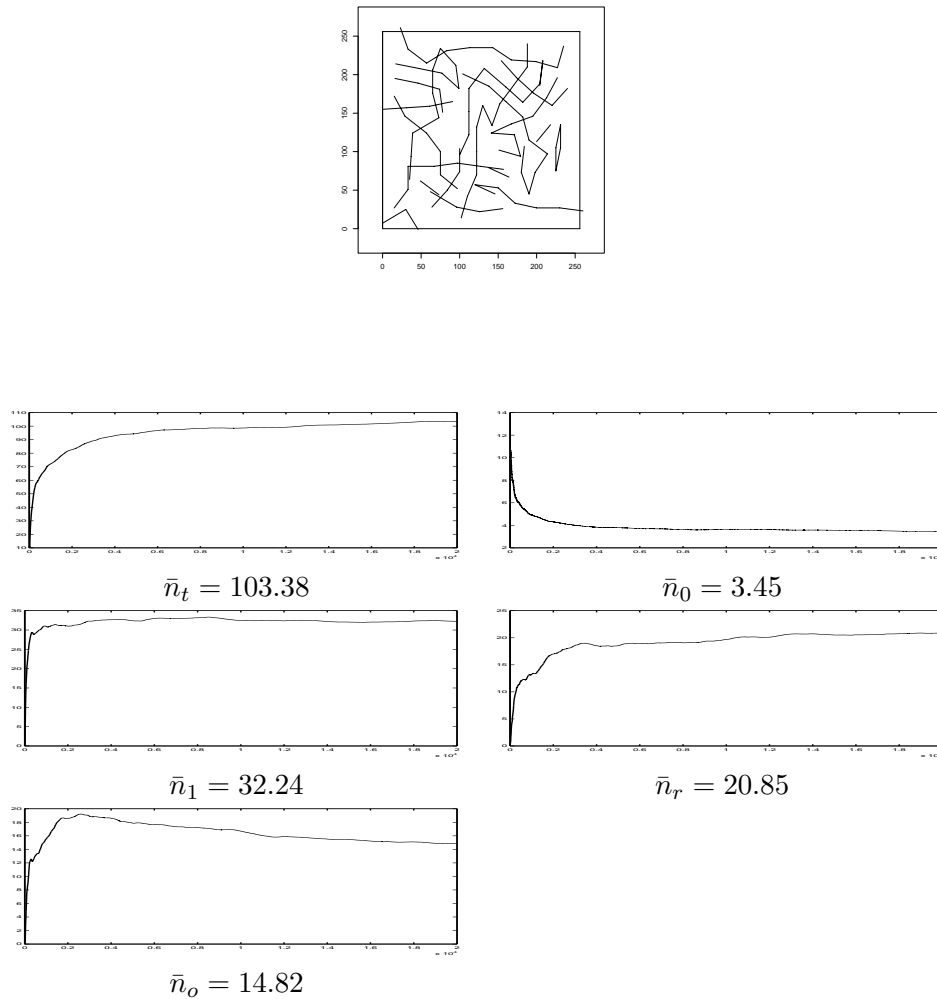


FIGURE 5.8 – Séries temporelles des moyennes cumulées des statistiques suffisantes obtenues par l’algorithme Metropolis-Hastings avec les paramètres  $p_b = 0.75$ ,  $p_d = 0.25$ ,  $p_1 = 1$ ,  $p_2 = p_c = 0$ . La configuration initiale est l’ensemble vide. La configuration est montrée en haut de la figure.

5. Calculer le taux d'acceptation correspondant  $r_t(\mathbf{T}_n, \mathbf{T}_{n+1})$  avec  $t = \{s, m, f\}$  représentant la transformation choisie.
6. Accepter la nouvelle configuration  $\mathbf{T}_{n+1}$  avec la probabilité
$$\alpha = \min\{1, r_t(\mathbf{T}_n, \mathbf{T}_{n+1})\}.$$
7. Si la nouvelle configuration  $\mathbf{T}_{n+1}$  est acceptée, alors retourner  $\mathbf{T}_{n+1}$ . Sinon, retourner  $\mathbf{T}_n$ .

Une version simple de l'algorithme est obtenue en choisissant les probabilités  $p_s$ ,  $p_m$  et  $p_t$  indépendamment de la tessellation courante  $\mathbf{T}_n$  et en considérant des lois d'instrumentation uniforme suivantes :

$$\begin{aligned} q_s(T, S) &= \frac{\pi}{2l(T) - l(D)}, \\ q_m(T, M) &= \frac{1}{\overset{\circ}{n}_{s, nb}(T)}, \\ q_f(T, F) &= \frac{1}{2\overset{\circ}{n}_{s, b}(T)}. \end{aligned}$$

Le taux d'acceptation d'une division  $S$  est

$$r_s(T, S) = \frac{h(ST) p_m}{h(T) p_s} \frac{2l(T) - l(D)}{\pi(\overset{\circ}{n}_{s, nb}(T) + 1 - \xi)}$$

où  $\xi \in \{0, 1, 2\}$  désigne le nombre de segments internes non bloquants de  $T$  formés par le nouveau segment produisant la division d'une cellule.

De même pour une fusion, nous obtenons

$$r_m(T, M) = \frac{h(MT) p_s}{h(T) p_m} \frac{\pi \overset{\circ}{n}_{s, nb}(T)}{2(l(T) - l(M)) - l(D)},$$

où  $l(M)$  est la longueur du segment non bloquant supprimé par la fusion  $M$ .

Finalement, pour un flip  $F$ , nous avons

$$r_f(T, F) = \frac{h(FT) \overset{\circ}{n}_{s, b}(T)}{h(T) \overset{\circ}{n}_{s, b}(FT)},$$

$\overset{\circ}{n}_{s, b}(FT) - \overset{\circ}{n}_{s, b}(T)$  variant entre  $-2$  et  $2$ .

Dans [T6], les auteurs prouvent que la chaîne ainsi construite est apériodique,  $\phi$ -irréductible et qu'elle converge vers la distribution d'équilibre  $\mathbb{P}(dT) = \int h(T)\mu(dT)$ . Les preuves de convergence ont été développées dans le cadre des dynamiques MCMC à sauts réversibles [76]<sup>1</sup>. Une étude par simulation de la dynamique a été également faite. Ceci ouvre des perspectives intéressantes de travail en vue de l'application du modèle à des données réelles.

---

1. En anglais : Reversible Jump Monte Carlo Markov Chains (RJMCMC)

## 5.3 Analyse des algorithmes de simulation exacte

Le CFTP n'est pas le seul algorithme pour obtenir des échantillons "parfaits" d'un processus ponctuel marqué. Dans [T8], les auteurs ont effectué l'étude théorique, l'implémentation informatique et la comparaison des performances de différentes méthodes de simulation parfaite pour les processus ponctuels marqués.

### 5.3.1 Clan des ancêtres

Dans le cas où la loi du processus à simuler possède une structure d'ordre partielle, l'algorithme CFTP est particulièrement efficace. Cependant, beaucoup d'objets  $y$  sont générés inutilement. Cette observation a incité [63] à construire un algorithme qui évite la naissance de ces objets, ou d'une façon plus générale d'individus d'une certaine classe. L'utilisation de cet algorithme pour la simulation de processus ponctuels peut être trouvée dans [21, 63, 141]. Son adaptation aux processus ponctuels marqués est montrée dans [T8].

Soit  $r$  le rayon d'interaction de la densité  $p$  d'un processus ponctuels marqué, au sens où

$$\lambda((w, m); \mathbf{x}) = \lambda((w, m); \mathbf{x} \cap B(w, r)),$$

et  $B(w, r) = \{(u, i) \in W \times M, \|u - w\| \leq r\}$ . L'équation précédente est bien sûr satisfaite pour  $r$  égale au diamètre de  $W$ . Cependant, l'algorithme qui va être décrit est efficace pour des valeurs de  $r$  assez faibles par rapport à  $\Lambda^2$ .

#### Algorithme 7

$(\mathbf{x}, T) = \text{Clan des ancêtres}$

*Soit  $D(0)$  la réalisation d'un processus ponctuel marqué de Poisson sur  $W \times M$ . Les points sont répartis selon la densité  $\Lambda$  et les marques correspondantes sont choisies indépendamment selon la loi  $\nu_M$ .  $A = D(0)$  est la valeur initiale du clan des ancêtres de la configuration que l'on souhaite obtenir.*

1. *Le processus  $D(\cdot)$  est simulé dans le passé en utilisant un processus spatial de naissance et de mort avec un taux de naissance  $b(\mathbf{x}, \eta) = \Lambda$  et un taux de mort  $d(\mathbf{x}, \eta) = 1$ . La simulation a lieu seulement si  $A_t = A \cap D(t) \neq \emptyset$ , sinon elle s'arrête. Pour chaque mort vers le passé, c'est-à-dire  $D(t-) = D(t) \cup \{(w, m)\}$  pour  $t < 0$ , si  $(w, m) \in A \cap D(t)$*

---

2. Voir ([T7], Section 3.2, pp.683) pour une explication plus détaillée

alors ajouter les points  $B(w, r) \cap D(t)$  au clan des ancêtres  $A$ . Le temps d'arrêt obtenu est noté  $-T$ .

2. Générer  $X(\cdot)$  dans le futur comme il suit :

- $X(-T) = \emptyset$  ;
- si une naissance apparait dans le passé, i.e.  $D(t-) = D(t) \cup \{(w, m)\}$ , alors enlever  $(w, m)$  de  $X(t-)$
- si une mort survient dans le passé, i.e.  $D(t-) = D(t) \setminus \{(w, m)\}$ , et si  $(w, m) \in A_t$ , alors ajouter  $(w, m)$  à  $X(t-)$  avec la probabilité  $\lambda((w, m); X(t-))/\Lambda$

L'Algorithme 7 s'appuie lui aussi sur la simulation de processus spatiaux de naissance et de mort, tout comme l'Algorithme 5. Cependant, l'Algorithme 7 ne requiert pas la monotonie du processus ponctuel, et l'obtention de la coalescence n'est pas nécessaire. Un aller vers le passé suivi d'un retour vers le futur suffisent à construire le clan des ancêtres et effectuer l'amincissement du processus dominant. La structure de l'interaction a une importance capitale. Ici nous avons utilisé une interaction de portée fixe. On peut lui substituer un *index d'incompatibilité*

$$I((w, m), (u, n)) := \mathbf{1}\{\sup_{\mathbf{x}} |\lambda((w, m); \mathbf{x}) - \lambda((w, m); \mathbf{x} \cup \{(u, n)\})| > 0\}. \quad (5.6)$$

### 5.3.2 Convergence de l'algorithme clan des ancêtres

La preuve de convergence de l'Algorithme 7 est similaire à celle des processus ponctuels non marqués [63, 21, 141].

Pour vérifier que l'Algorithme 7 finit bien, il faut remarquer que le processus  $D$  est réversible et en équilibre. Les hypothèses impliquent qu'en allant vers le passé le processus  $D$  se vide presque sûrement après un temps fini  $T$ , moment auquel  $A_T = \emptyset$  également. Durant la simulation vers le futur, les naissances dépendent de l'intensité conditionnelle. Ceci garantit que le processus résultant respecte les conditions d'équilibre détaillé et qu'il admet bien  $p(\cdot)$  comme loi d'équilibre.

En regardant un peu plus en détail les naissances de la simulation vers le futur, on peut remarquer qu'un point ajouté à  $D(t)$  est également ajouté à  $Y(t)$  en fonction de la densité conditionnelle qui ne dépend que de ses voisins à distance au plus  $r$ . Par conséquent, les points qui ne se trouvent pas dans  $A_t$  peuvent être ignorés.

De plus, cet algorithme peut être utilisé pour simuler des processus ponctuels définis sur des espaces infinis dans une fenêtre bornée [63].

### 5.3.3 Metropolis-Hastings exact

Un algorithme exact de type Metropolis-Hastings pour simuler des processus ponctuels localement stables a été construit par [99]. Dans cet algorithme, on a utilisé une discrétisation de l'espace  $W$  et des probabilités d'instrumentation qui dépendent de la configuration courante. Une version plus simple qui utilise des probabilités d'instrumentation fixes et qui est adaptée aux processus ponctuels marqués a été proposée par [T8].

L'algorithme de Metropolis-Hastings a été présenté dans une section précédente (v. Algorithme 2). Nous rappelons ici que cette méthode est une technique qui fait des propositions de mises à jour qui peuvent être acceptées ou pas [73, 72, 190, 141]. Si l'état courant est une configuration d'objets  $\mathbf{x}$  de densité positive  $p(\mathbf{x})$ , alors une mise à jour standard est de choisir une naissance (ajouter un objet) selon la probabilité  $p_b \in (0, 1)$ , ou une mort (supprimer un objet) selon la probabilité  $p_d = 1 - p_b$ . Pour une naissance, un nouvel objet  $\eta$  est proposé selon la densité  $\nu \times \nu_M / \nu(W)$  et accepté avec la probabilité

$$\min \left\{ 1, \frac{p_d \nu(W) \lambda(\eta; \mathbf{x})}{p_b (1 + n(\mathbf{x}))} \right\}. \quad (5.7)$$

Pour une mort, si la configuration  $\mathbf{x}$  est vide, alors l'état  $\mathbf{x}$  reste inchangé. Sinon, un objet  $\eta$  est choisi uniformément dans  $\mathbf{x}$  et supprimé avec la probabilité

$$\min \left\{ 1, \frac{p_b n(\mathbf{y})}{p_d \nu(W) \lambda(\eta; \mathbf{x} \setminus \{\eta\})} \right\}. \quad (5.8)$$

Cette dynamique converge vers un échantillon non biaisé de  $p$  [73, 72, 190, T7].

L'adaptation de ces mises à jour à un algorithme de simulation exacte demande beaucoup d'attention, tout spécialement pour les transitions de mort. L'idée des méthodes CFTP pour les processus ponctuels localement stables s'appuie sur la simulation d'un couple de processus qui peuvent coalescer. Ces processus sont couplés à un processus dominant. Le processus dominant est censé respecter la relation d'ordre basée sur l'inclusion. Pour cela il serait convenable que la probabilité d'acceptation d'une mort soit toujours égale à 1, indépendamment de la configuration initiale. Le taux d'acceptation (5.8) est clairement plus grand que  $p_b / (p_d \nu(W) \Lambda)$ , mais il est peut être plus petit que 1. Pour lever cette difficulté, nous discrétisons  $W$  à la manière de [99] et nous construisons des mises à jour par bandes.

Soit  $\cup_{i=1}^{n_W} W_i$  une partition finie de  $W$  telle que  $\nu(W_i) > 0$  pour toutes les cellules  $W_i$ ,  $i = 1, \dots, n_W$ . Si la configuration courante est  $\mathbf{x}$  et si la bande visitée est  $W_i \times M$ , alors avec probabilité  $p_b^i$  on choisit d'y ajouter un nouveau objet, ou bien, avec probabilité  $p_d^i = 1 - p_b^i$  on choisit d'enlever un.

Supposons la dernière égalité vérifiée pour tous les  $i = 1, \dots, n_W$ . Dans le cas d'une naissance, un nouvel objet est généré. La position de l'objet est uniforme sur  $W_i$  et sa marque est distribué selon  $\nu_M$ . Dans le cas d'une mort, alors un objet est choisi uniformément parmi ceux de  $W_i$ , quand cela est possible. Des taux de Metropolis-Hastings similaires aux (5.7) et (5.8) sont ensuite calculés pour accepter ou pas les transitions proposées. Cette procédure est bien définie sur l'ensemble de configurations de densité positive, et elle converge vers  $p$ . Nous en donnons la preuve dans le paragraphe suivant.

Pour décrire le couplage basé sur cette dynamique, nous commençons par définir le processus dominant  $D$ . Ce processus accepte toutes les transitions qu'on lui propose. La chaîne  $D$  visite les bandes aléatoirement avec la même probabilité. Si la bande  $W_i \times M$  est visitée, alors un nouvel objet est généré avec la probabilité  $p_b^i$ . L'objet est uniformément dans  $W_i$  et sa marque est choisie selon  $\nu_M$ . Avec la probabilité complémentaire  $p_d^i = 1 - p_b^i$  un objet est choisi uniformément parmi ceux de  $W_i$  et effacé de la configuration courante. Cette opération a lieu pourvu que la bande  $W_i \times M$  soit non vide. Sinon, le processus  $D$  reste inchangé.

Faisons l'hypothèse que  $0 < p_b^i/p_d^i < 1$ , c'est-à-dire  $0 < p_b^i < 1/2$ , et soit  $\pi_i(n) = \mathbb{P}(n \text{ points in } W_i \times M)$ . Les conditions d'équilibre local pour le nombre d'objets dans chaque bande  $i$

$$\pi_i(n) p_b^i = \pi_i(n+1) p_d^i, \quad n = 0, 1, 2, \dots,$$

donnent l'unique solution

$$\pi_i(n) = \left( \frac{p_b^i}{p_d^i} \right)^n \left( 1 - \frac{p_b^i}{p_d^i} \right). \quad (5.9)$$

Par conséquent, le nombre d'objets de  $W_i \times M$  suit une loi géométrique décalée de paramètre  $1 - p_b^i/p_d^i$ . Sachant que l'on a  $n$  objets dans une bande  $W_i \times M$ , à l'équilibre ils sont i.i.d avec les positions distribués selon  $\nu(\cdot)/\nu(W_i)$  et les marques choisies selon  $\nu_M$ . La chaîne est réversible et les configurations dans les différentes bandes  $W_i \times M$  sont indépendantes.

Construisons maintenant la dynamique cible. Le taux de Metropolis-Hastings pour une transition de l'état  $\mathbf{x} \neq \emptyset$  à l'état  $\mathbf{x} \setminus \{x_j\}$  avec  $x_j \in (W_j \times M) \cap \mathbf{x}$  est donné par

$$\frac{p_b^i n(\mathbf{x} \cap (W_i \times M))}{p_d^i \nu(K_i) \lambda(x_j; \mathbf{x} \setminus \{x_j\})}.$$

En considérant la restriction de la chaîne sur l'ensemble  $\{\mathbf{x} : p(\mathbf{x}) > 0\}$  et en utilisant le fait que  $p$  est héréditaire, il apparait que le taux Metropolis-Hastings est bien défini. Une condition suffisante pour qu'il soit supérieur à

1 est

$$\frac{p_b^i}{p_d^i} \geq \Lambda\nu(W_i)$$

où  $\Lambda$  est la borne supérieure de l'intensité conditionnelle. Sous cette condition, ces transitions sont toujours acceptées.

Soit  $L \subseteq U$  deux configurations finies de points marqués, et soient

$$\begin{aligned} \alpha_{\min}(U, L, (w, m), i) &:= \min \left\{ \frac{p_d^i \nu(W_i) \lambda((w, m); \mathbf{x})}{p_b^i (1 + n(\mathbf{x} \cap (W_i \times M)))} : L \subseteq \mathbf{x} \subseteq U \right\} \\ \text{et} & \\ \alpha_{\max}(U, L, (w, m), i) &:= \max \left\{ \frac{p_d^i \nu(W_i) \lambda((w, m); \mathbf{x})}{p_b^i (1 + n(\mathbf{x} \cap (W_i \times M)))} : L \subseteq \mathbf{x} \subseteq U \right\} \end{aligned} \quad (5.10)$$

les bornes des taux de Metropolis-Hastings pour la naissance d'un objet  $(w, m) \notin U$  dans la bande  $i$  pour les configurations d'objets qui se situent entre  $L$  et  $U$  par inclusion.

Nous avons maintenant tous les ingrédients pour proposer l'algorithme suivant. Comme pour l'Algorithme 5, les variables aléatoires déjà construites sont réutilisées quand le temps est doublé.

**Algorithme 8**  $(\mathbf{x}, T) = \text{Metropolis Hastings Exact}$

*Soit une partition de  $W$  telle que  $0 < \Lambda\nu(W_i) < 1$  et choisissez les probabilités  $p_b^i$  et  $p_d^i$  telles que  $p_b^i/p_d^i \in [\Lambda\nu(W_i), 1)$  pour tout  $i = 1, \dots, n_W$ . Soit  $V_t$ ,  $t = -1, -2, \dots$  une famille de v.a. indépendantes uniformément distribuées sur  $\{1, \dots, n_W\}$  et soit  $U_t$  une famille de v.a. indépendantes sur  $(0, 1)$ . Initialise  $T = 1$ , et soit  $D(0)$  une réalisation du processus ponctuel marqué avec les points indépendamment distribués par bandes comme il suit : pour chaque bande un nombre  $n$  de points est choisi par rapport à (5.9), puis les positions des objets sont choisies indépendamment dans  $W_i$  selon la loi  $\nu$ -uniforme et finalement à chaque point une marque indépendante lui est associée avec la distribution  $\nu_M$ .*

1. *Simuler  $D(\cdot)$  vers le temps passé  $-T$  comme il suit : avec probabilité  $p_d^{V_t}$  enlever un objet choisi aléatoirement dans  $W_{V_t} \times M$  ; sinon ajouter un objet  $\xi_t$  avec une position  $\nu$ -uniforme dans  $W_{V_t}$  et une marque donnée par  $\nu_M$ , indépendamment des autres variables.*
2. *Générer  $L^{-T}(\cdot)$  et  $U^{-T}(\cdot)$  vers le futur de la manière suivante :*
  - *initialiser  $L^{-T}(-T) = \emptyset$  et  $U^{-T}(-T) = D(-T)$  ;*
  - *si  $D(\cdot)$  a vécu une naissance dans le passé, i.e.  $D(t) = D(t+1) \cup \{(w, m)\}$ , alors un point est enlevé de  $L^{-T}(t)$  et  $U^{-T}(t)$  en utilisant le mécanisme de permutations dans [99]*

- si  $D(\cdot)$  a vécu une mort dans le passé, i.e.  $D(t) = D(t+1) \setminus \{(w, m)\}$ , l'objet  $(w, m)$  est ajouté au  $L^{-T}(t)$  si

$$U_t \leq \alpha_{\min}(U^{-T}(t), L^{-T}(t), (w, m), V_t)$$

et au  $U^{-T}(t)$  si

$$U_t \leq \alpha_{\max}(U^{-T}(t), L^{-T}(t), (w, m), V_t)$$

3. Si  $U^{-T}(0) = L^{-T}(0)$  arrêter. Sinon poser  $T = 2T$  et recommencer.
4. Return  $\mathbf{x} = U^{-T}(0)$  et  $T$ .

Si la densité  $p$  est répulsive, les équations (5.10) deviennent

$$\alpha_{\min}(U, L, (w, m), i) = (p_d^i \nu(W_i) \lambda((w, m); U)) / p_b^i (1 + n(U \cap (W_i \times M)))$$

et respectivement

$$\alpha_{\max}(U, L, (w, m), i) = (p_d^i \nu(W_i) \lambda((w, m); L)) / p_b^i (1 + n(L \cap (W_i \times M))).$$

Dans le cas attractif, une telle simplification n'est pas possible. Cependant, les bornes suivantes peuvent être utilisés à la place

$$\alpha_{\min}(U, L, (w, m), i) = (p_d^i \nu(W_i) \lambda((w, m); L)) / p_b^i (1 + n(U \cap (W_i \times M)))$$

et

$$\alpha_{\max}(U, L, (w, m), i) = (p_d^i \nu(W_i) \lambda((w, m); U)) / p_b^i (1 + n(L \cap (W_i \times M))).$$

### 5.3.4 Convergence de l'algorithme Metropolis-Hastings exact

Pour prouver la convergence de l'algorithme nous devons vérifier les conditions du Théorème 20.

Premièrement, nous devons préciser la dynamique du processus dominant ainsi que les variables aléatoires associées. A part  $V_t$  et  $U_t$ , nous associons à chaque  $D(t)$  une permutation  $\Sigma(t)$  des points marqués qui se situent dans la bande  $V_t$ . Cette permutation est utilisée pour spécifier l'ordre dans laquelle les points peuvent mourir dans la simulation vers le futur dans le processus dominant ainsi que dans les sous-processus correspondants. Ainsi, quand une mort est proposée au temps  $t$ , si  $D(t) \cap (W_{V_t} \times M) = \mathbf{x} = \{x_1, \dots, x_n\}$  et  $\Sigma(t) = (x_{j_1}, \dots, x_{j_n})$  pour une permutation  $(j_1, \dots, j_n)$  de l'ensemble  $\{1, \dots, n\}$ , l'objet enlever du processus dominant est  $x_{j_1}$ . La loi du processus  $\Sigma$  est définie de la manière conditionnelle suivante. Si  $D(t+1)$  est obtenu à partir de  $D(t)$  par une mort dans la bande indexée par  $V_t$ , le rang 1 est assigné à l'objet à enlever, alors que les rangs des objets situés dans la bande  $V_t$  du processus  $D(t+1)$  sont donnés par une permutation uniforme,



indépendante des autres variables. Sinon,  $D(t+1)$  est obtenu à partir de  $D(t)$  par une naissance dans la bande indexée par  $V_t$ . Dans ce cas,  $\Sigma(t)$  est une permutation uniforme des points marqués dans  $D(t) \cap (W_{V_t} \times M)$ . Par conséquent,  $\Sigma(t)$  dépend de  $V_t$ ,  $D(t)$  et  $D(t+1)$ . La réversibilité du processus  $D(\cdot)$ , l'indépendance de  $U_t$  et  $V_t$  et la construction de  $\Sigma$  impliquent la stationnarité conjointe du processus

$$\{(D(t), V_t, \Sigma(t), U_t) : -\infty < t < \infty\}.$$

Il est à remarquer que pour  $D(t) = \mathbf{x}$  et  $V_t = i$  fixés, la permutation  $\Sigma$  de  $\{1, \dots, n(\mathbf{x} \cap (W_i \times M))\}$  a comme probabilité marginale conditionnelle

$$\begin{aligned} \mathbb{P}(\sigma | \mathbf{x}, i) &= \\ &= p_b^i \frac{1}{n(\mathbf{x} \cap (W_i \times M))!} + p_d^i \frac{1}{n(\mathbf{x} \cap (W_i \times M))} \frac{1}{(n(\mathbf{x} \cap (W_i \times M)) - 1)!} \\ &= \frac{1}{n(\mathbf{x} \cap (W_i \times M))!} \quad , \end{aligned}$$

c'est à dire une distribution uniforme si  $n(\mathbf{x} \cap (W_i \times M)) > 0$  et une distribution dégénérée sinon.

Le processus dominant peut être simulé comme suit. En  $D(0)$ , la bande  $i$  contient un nombre d'objets de loi géométrique décalée (5.9) de paramètre  $1 - p_b^i/p_d^i$ . Les objets sont indépendants. Les points sont répartis uniformément dans  $W_i$  et les marques sont choisies selon la loi  $\nu_M$ . Le nombre aléatoire  $U_0$  est uniforme sur  $(0, 1)$ , alors que  $V_0$  est uniforme sur  $\{1, \dots, n_W\}$ , indépendamment de toute autre source d'aléatoire. Grâce à la réversibilité,  $D$  peut être simulé de 0 à  $-T$ , et de 0 à  $T$ . Il en est de même pour les quantités indépendantes  $U_t$  et  $V_t$ . Conditionnellement aux trajectoires de  $D$  et  $V$ , les réalisations de  $\Sigma(-T), \dots, \Sigma(-1), \Sigma(0), \dots, \Sigma(T-1)$  sont générées comme expliqué plus haut.

Lors d'une mort, l'objet de plus faible rang attribué par  $\sigma$  est éliminé des processus  $U$  et  $L$ . Nous devons vérifier que ce choix équivaut à enlever uniformément un point situé dans la bande sélectionnée pour être mise à jour. Pour cela, soit  $\Sigma$  une permutation uniforme de  $A = \{x_1, \dots, x_n\}$  et soit sa restriction au sous-ensemble  $B = \{x_{i_1}, \dots, x_{i_m}\}$  de  $A$ . Nous avons  $n!$  permutations possibles pour l'ensemble  $A$ . Pour la restriction au sous-ensemble  $B$ , le nombre de possibilités se calcule comme suit. Premièrement, nous choisissons  $m$  indices, ce qui fait  $\binom{n}{m}$  possibilités. Puis les  $n-m$  places libres donne lieu à  $n-m$  permutations possibles. La probabilité de toute permutation attaché à  $B$  étant le rapport entre le nombre de permutations favorables et le nombre de permutations possibles, nous arrivons à

$$\mathbb{P}(\sigma(B)) = \frac{\binom{n}{m} \times (n-m)!}{n!} = \frac{1}{m!}$$

comme il se doit. Il est en effet facile de prouver que sélectionner le point d'une configuration de plus faible rang après une permutation uniforme revient à choisir directement ce point uniformément dans la configuration. De plus, les transitions de type mort respectent l'ordre par inclusion : enlèvement du point de plus faible  $\Sigma$ -rang, l'ensemble  $B$  modifié est un sous-ensemble de  $A$  modifié.

Dans [98, 99], les auteurs proposent une procédure appelée *MHDeath* pour ne calculer que la partie des permutations nécessaire à la simulation.

### Algorithme 9

$$\eta_{\mathbf{y}} = \text{MHDeath}(\mathbf{x}, \mathbf{y}, \eta_x, \eta_y)$$

Soient  $\mathbf{x} \subseteq \mathbf{y} \subseteq \mathbf{z}$  trois configurations d'objets, et soient  $\{\eta_x\} \subseteq \mathbf{x}$ ,  $\{\eta_y\} \subseteq \mathbf{y}$  et  $\{\eta_z\} \subseteq \mathbf{z}$  les candidats à l'enlèvement respectifs. Si  $\{\eta_w\} = \emptyset$ , ceci signifie qu'il n'y a pas de candidat pour la mort retenu dans la "population"  $w$  (avec  $w = \mathbf{x}, \mathbf{y}$  ou  $\mathbf{z}$ ). Supposons que

$$\eta_z \in \mathbf{y} \Rightarrow \eta_y = \eta_z \text{ si } \eta_z \text{ existe,}$$

$$\eta_y \in \mathbf{x} \Rightarrow \eta_x = \eta_y \text{ si } \eta_y \text{ existe,}$$

$\eta_z$  est choisi uniformément sur  $\mathbf{z}$ ,

une mort est effectué dans  $\mathbf{x}$ , alors  $\eta_x$  est choisi uniformément sur  $\mathbf{x}$ .

Conditionnellement aux valeurs  $(\mathbf{x}, \mathbf{y}, \eta_x, \eta_z)$  et au fait qu'une mort soit proposée pour  $\mathbf{y}$ , le candidat  $\eta_y$  est choisi uniformément dans  $\mathbf{y}$  par le mécanisme suivant :

$$\eta_y = \begin{cases} \eta_z & \text{si } \eta_z \in \mathbf{y} \\ \eta_x & \text{si } \eta_z \notin \mathbf{y} \text{ et } \omega \leq n(\mathbf{x})/n(\mathbf{y}) \\ \eta'_y & \text{sinon} \end{cases}$$

avec  $\omega$  uniforme sur  $[0, 1]$ ,  $\eta'_y$  choisi uniformément sur  $\mathbf{y} \setminus \mathbf{x}$  et la convention  $n(\mathbf{x})/n(\mathbf{y}) = 0$  si  $n(\mathbf{y}) = 0$ .

**Preuve:** Nous montrons que sous les hypothèses précédentes, l'Algorithme 9 choisi  $\eta_y$  uniformément sur  $\mathbf{y}$  [192]. Pour cela soit

$$\begin{aligned} \mathbb{P}(\eta_y = y_0) &= \\ &= \mathbb{P}(\eta_y = y_0 | \eta_z \in \mathbf{y}) \mathbb{P}(\eta_z \in \mathbf{y}) + \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}) \mathbb{P}(\eta_z \notin \mathbf{y}) \end{aligned}$$

Pour le premier terme de l'équation à droite, nous avons

$$\mathbb{P}(\eta_y = y_0 | \eta_z \in \mathbf{y}) \mathbb{P}(\eta_z \in \mathbf{y}) = \mathbb{P}(\eta_z = y_0) = \frac{1}{n(\mathbf{z})},$$

alors pour le deuxième terme nous obtenons

$$\begin{aligned}
& \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}) \mathbb{P}(\eta_z \notin \mathbf{y}) = \\
& = \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega \leq n(\mathbf{x})/n(\mathbf{y})) \mathbb{P}(\omega \leq n(\mathbf{x})/n(\mathbf{y}) | \eta_z \notin \mathbf{y}) \mathbb{P}(\eta_z \notin \mathbf{y}) + \\
& + \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega > n(\mathbf{x})/n(\mathbf{y})) \mathbb{P}(\omega > n(\mathbf{x})/n(\mathbf{y}) | \eta_z \notin \mathbf{y}) \mathbb{P}(\eta_z \notin \mathbf{y}) \\
& = \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega \leq n(\mathbf{x})/n(\mathbf{y})) \frac{n(\mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})} + \\
& + \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega > n(\mathbf{x})/n(\mathbf{y})) \frac{n(\mathbf{y} \setminus \mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})}.
\end{aligned}$$

Maintenant, si  $y_0 \in \mathbf{x}$  nous avons

$$\mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega > n(\mathbf{x})/n(\mathbf{y})) = 0,$$

et donc

$$\begin{aligned}
& \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega \leq n(\mathbf{x})/n(\mathbf{y})) \frac{n(\mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})} \\
& = \frac{1}{n(\mathbf{x})} \frac{n(\mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})} \\
& = \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})n(\mathbf{y})}.
\end{aligned}$$

Et pour le cas  $y_0 \notin \mathbf{x}$

$$\mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega \leq n(\mathbf{x})/n(\mathbf{y})) = 0,$$

ce qui donne

$$\begin{aligned}
& \mathbb{P}(\eta_y = y_0 | \eta_z \notin \mathbf{y}, \omega > n(\mathbf{x})/n(\mathbf{y})) \frac{n(\mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})} = \\
& = \frac{1}{n(\mathbf{y} \setminus \mathbf{x})} \frac{n(\mathbf{y} \setminus \mathbf{x})}{n(\mathbf{y})} \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})} \\
& = \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})n(\mathbf{y})}.
\end{aligned}$$

Nous sommes tous les termes et nous obtenons comme résultat

$$\mathbb{P}(\eta_y = y_0) = \frac{1}{n(\mathbf{z})} + \frac{n(\mathbf{z} \setminus \mathbf{y})}{n(\mathbf{z})n(\mathbf{y})} = \frac{1}{n(\mathbf{y})}.$$

□

En conclusion, les sous-processus  $U^{-T}(\cdot)$  et  $L^{-T}(\cdot)$  sont des fonctionnelles adaptées du processus marqué dominant  $\{(D(t), V_t, \Sigma(t), U_t) : -\infty < t < \infty\}$ . Par construction, les propriétés d'inclusion et d'entonnoir [99, (2.4)–(2.6)] sont respectées. Ces propriétés sont tout à fait similaires à celles

que l'on demande à la simulation CFTP par les processus de naissance et de mort, décrites par (4.27) et (4.28).

En second lieu, comme la loi géométrique translatée attribue une probabilité positive au 0, et comme le processus dominant est à l'équilibre, l'extension du processus vers le passé atteint l'état  $\emptyset$  presque sûrement, d'où l'arrêt presque sûr de l'Algorithme 8.

Il reste à s'assurer de la convergence du processus cible  $X^{-T}(\cdot)$  défini comme suit. Nous posons  $X^{-T}(-T) = \emptyset$  et nous lui appliquons la même dynamique qu'à  $U^{-T}(\cdot)$  et  $L^{-T}(\cdot)$ , à la différence près que si  $X^{-T}(t) = \mathbf{x}$  et si la naissance d'un objet  $(w, m)$  est proposé dans la bande indexée par  $V_t$ , alors elle est acceptée quand

$$U_t \leq \frac{p_d^{V_t} \nu(W_{V_t}) \lambda((w, m); \mathbf{x})}{p_b^{V_t} (1 + n(\mathbf{x} \cap (W_{V_t} \times M)))}.$$

Bien évidemment,  $L^{-T}(t) \subseteq X^{-T}(t) \subseteq U^{-T}(t)$  pour tout entier  $0 \geq t \geq -T$ .

Le processus  $X(t)$  avec  $t = 0, 1, 2, \dots$  et  $X(0) = \emptyset$  est défini de manière analogue, à partir d'une simulation vers le futur du processus dominant et ses variables aléatoires associées. Le processus  $X(t)$  évolue selon une dynamique de type Metropolis-Hastings. Si l'état courant est  $\mathbf{x}$ , une bande  $W_i \times M$  est choisie uniformément. Avec la probabilité  $p_b^i$  un nouveau point marqué  $\eta$  est généré dans  $W_i \times M$  selon la loi  $\nu \times \nu_M / \nu(W_i)$  et accepté avec la probabilité  $(p_d^i \nu(W_i) \lambda(\xi; \mathbf{x})) / (p_b^i n(\mathbf{y} \cap K_i \times M)) \leq 1$ . Avec la probabilité complémentaire  $p_d^i = 1 - p_b^i$  un point uniforme de  $W_i \times M$  est enlevé, sauf si la bande  $i$  est vide, auquel cas la configuration reste inchangée. La procédure est bien définie sur l'ensemble de configurations de densité positive.

Par construction,  $p$  est une densité invariante du noyau de transition. Comme les morts sont toujours acceptées et comme  $p_d^i > 1/2$ , la probabilité d'atteindre la configuration vide à partir de toute configuration  $\mathbf{x}$  est au moins

$$(2n_W)^{-n(\mathbf{x})} > 0.$$

Il en résulte que la dynamique de Metropolis-Hastings est  $\delta_0$ -irréductible, où  $\delta_0$  désigne la mesure de Dirac de  $\emptyset$ . Si l'état courant est vide, la probabilité de le rester est au moins

$$\sum_i p_d^i / n_W > 1/2 \geq (2n_W)^{-1}.$$

Par conséquent, les ensembles de la forme  $\{\mathbf{x} : n(\mathbf{x}) \leq n\}$ , ainsi que leur sous-ensembles mesurables, sont petits par rapport  $(2n_W)^{-n} \delta_0$ . Comme la

probabilité de rester dans le même état est positive, la chaîne simulée est apériodique. Il en résulte que  $X(t)$  converge faiblement en variation totale vers la loi  $p$ .

La convergence est assurée pour tous les états  $p$ -positifs, mais il faut se rappeler que  $X(0) = \emptyset$  est un atome de la loi limite. Pour éliminer l'influence des ensembles de mesure nulle, nous pouvons utiliser la condition de dérive. Soit la fonction  $V$  définie par

$$V(\mathbf{x}) = c(\mathbf{x})A^{m(\mathbf{x})}$$

où  $m(\mathbf{x}) = \max_i n(\mathbf{x} \cap (W_i \times M))$ ,  $c(\mathbf{x})$  étant la multiplicité du maximum et  $A > n_W$ . Nous pouvons remarquer que  $V$  est strictement positive et que ses ensembles de niveau  $\{\mathbf{x} : V(\mathbf{x}) \leq c\}$  pour  $c > 0$  sont des ensembles petits. Ceci tient au fait que  $V(\mathbf{x}) \leq c$  implique  $n(\mathbf{x}) \leq n_W \log(c)/\log(A)$  ou bien 0, selon que  $c > A$  ou non. Maintenant, pour tout  $\epsilon > 0$  nous pouvons choisir un entier  $N_\epsilon > 0$  tel que la probabilité d'acceptation d'une naissance dans une bande  $i$  satisfaisant  $n(\mathbf{x} \cap (W_i \times M)) \geq m(\mathbf{x}) - 1$  ne dépasse pas  $\epsilon$  pour tout  $\mathbf{x}$  tel que  $n(\mathbf{x}) \geq N_\epsilon$ . Les modifications dans des bandes à  $m(\mathbf{x}) - 1$  objets n'affectent pas  $V(\cdot)$ . Pour la configuration  $\mathbf{x}$  avec  $n(\mathbf{x}) \geq N_\epsilon$ , nous avons

$$\begin{aligned} \mathbb{E}[V(X(t+1))|X(t) = \mathbf{x}] - V(\mathbf{x}) \leq \\ \frac{\max_i p_b^i}{n_W} (A + n_W - 2)\epsilon V(\mathbf{x}) - \frac{\min_i p_d^i}{n_W} \left(1 - \frac{n_W}{A}\right) V(\mathbf{x}). \end{aligned}$$

Il est possible de choisir  $\epsilon$  tel que  $0 < \max_i p_b^i (A + n_W - 2)\epsilon < \min_i p_d^i (1 - n_W/A)$ , de telle manière que l'accroissement attendu de  $V$  soit négatif. La chaîne induite par  $X$  est récurrente au sens de Harris, ce qui complète la preuve.

### 5.3.5 L'échantillonneur de Gibbs exact

L'échantillonneur de Gibbs est une technique de simulation de lois multivariées. Le principe de cette technique est la simulation, à tour de rôle, de la loi d'une variable conditionnellement aux autres. Les densités des processus ponctuels marqués ne rentrent pas forcément dans le cadre des lois multivariées. Par exemple, l'étude des processus ponctuels multiobjets montre que simuler la loi marginale d'une seule composante est tout aussi difficile que simuler la loi jointe de toutes les composantes [190]<sup>3</sup>. Toutefois, l'approximation d'un processus ponctuel par un processus sur une grille est une chose assez commune, notamment dans le cadre de l'estimation des paramètres par pseudo-vraisemblance [25, 75]. Dans ce contexte, l'échantillonneur de Gibbs

---

3. Il existe tout de même une exception avec le modèle Widom-Rowlinson [85]

s'applique naturellement.

### Discrétisation d'un processus ponctuel marqué

Comme **da**u paragraphe 5.3.3, soit  $W = \bigcup_{i=1}^{n_W} W_i$  une partition finie de l'espace des positions telle que  $0 < \nu(W_i) < \infty$  pour tout  $0 < \nu(W_i) < \infty$ . Supposons aussi l'espace des marques partitionné, i.e.  $M = \bigcup_{j=1}^{n_M} M_j$  avec  $0 < \nu_M(M_j) \leq 1$  pour tout  $j = 1, \dots, n_M$ . Alors, la partition induite sur l'espace produit est  $W \times M = \bigcup_{i=1}^{n_W} \bigcup_{j=1}^{n_M} (W_i \times M_j)$ . A chaque cellule  $W_i \times M_j$  est associé un point représentatif  $\xi_{ij}$ , par exemple son centre. On désignera par  $\Xi$  l'ensemble de tous les points représentatifs.

En suivant [25, 75], nous approximations la densité  $p$  de  $X$  par une loi autologistique. Plus précisément, soit  $N$  un champ aléatoire sur la grille induite par  $\Xi$  en prenant des valeurs binaire comme ci-dessous. L'espace d'états est fini et il est constitué des vecteurs  $(n_{ij})$ ,  $i = 1, \dots, n_W$ ,  $j = 1 \dots n_M$  tels que  $n_{ij} \in \{0, 1\}$  et  $\sum_{j=1}^{n_M} n_{ij} \leq 1$  pour tout  $i$ . Si nous interprétons  $n_{ij}$  comme un indicateur de présence/absence d'un point marqué dans  $W_i \times M_j$ , la condition sur  $\sum_j n_{ij}$  nous assure que chaque bande  $W_i \times M$  contient au maximum un seul point marqué. Ainsi, presque sûrement, le vecteur aléatoire  $N_i = (N_{i1}, \dots, N_{in_M})$  prend ses valeurs dans  $S := \{\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_{n_M}\}$ , l'ensemble contenant le vecteur nul et les vecteurs unités.

Pour préciser la fonction de probabilité  $\mathbb{P}$  du champ aléatoire  $N$ , nous rappelons que  $p$  est héréditaire par hypothèse et que par le Théorème 7 (Hammersley-Clifford), la densité peut-être factorisée sous la forme

$$p(\mathbf{x}) = \prod_{\mathbf{z} \subseteq \mathbf{x}} \varphi(\mathbf{x}) \quad (5.11)$$

où les fonctions  $\varphi$  sont les fonctions d'interaction.

La loi de probabilité de  $N$  est obtenue à partir de (5.11) en limitant le produit à des sous-ensembles de  $\Xi$  et en corrigeant par rapport à la masse d'une cellule [75]. Il en résulte

$$\mathbb{P}(\mathbf{n}) \propto \prod_{\{i,j\}} [\nu(W_i)\nu_M(M_j)]^{n_{ij}} \prod_{\emptyset \neq \mathbf{z} \subseteq \Xi} \varphi(\mathbf{z})^{\eta(\mathbf{n};\mathbf{z})}$$

où  $\mathbf{n}$  est dans l'espace d'états et  $\eta(\mathbf{n}; \mathbf{z}) = \prod_{\xi_{ij} \in \mathbf{z}} n_{ij}$  prend des valeurs 0 ou 1. Par convention,  $0^0 = 1$ .

## Caractéristiques locales

Le comportement local d'un champ aléatoire est spécifié par les lois d'un point du champ conditionnellement au reste. Nous appelons ces lois les caractéristiques locales. Soit  $\mathbf{n}_{-i}$  le vecteur formé par les  $n_{i'j}$  avec  $i' \neq i$ . Les caractéristiques locales de la bande  $i$  sont données par

$$\mathbb{P}(N_i = \mathbf{e}_j | \mathbf{n}_{-i}) = \frac{\nu(K_i) \nu_M(M_j) \prod_{\mathbf{z}} \varphi(\mathbf{z} \cup \{\xi_{ij}\})^{\eta(\mathbf{n}_{-i}; \mathbf{z})}}{1 + \nu(K_i) \sum_{j'=1}^{n_M} \nu_M(M_{j'}) \prod_{\mathbf{z}} \varphi(\mathbf{z} \cup \{\xi_{ij'}\})^{\eta(\mathbf{n}_{-i}; \mathbf{z})}} \quad (5.12)$$

quand  $j \neq 0$ , pourvu que l'événement par rapport auquel on conditionne soit de masse strictement positive sous  $\mathbb{P}$ . Le produit s'effectue sur les ensembles  $\mathbf{z} \subseteq \Xi \setminus \{\xi_{ij'} : j' = 1, \dots, n_M\}$ , et  $\eta(\mathbf{n}_{-i}; \mathbf{z})$  est défini comme précédemment. Si les processus ponctuels que l'on discrétise sont markoviens, alors l'équation (5.12) n'utilise qu'un sous-ensemble réduit de points autour de  $\xi_{ij}$ . Dans ce cas, l'ensemble des points  $\xi_{i'j}$  utilisés est appelé le voisinage du  $\xi_{ij}$  et les caractéristiques données par (5.12) sont vraiment locales.

Le champ aléatoire  $N$  est dit monotone par rapport à une relation d'ordre donnée  $\leq$  sur  $S$ , si pour tout  $s \in S$ ,  $\mathbb{P}(N_i \leq s | \mathbf{n}_{-i}) \geq \mathbb{P}(N_i \leq s | \tilde{\mathbf{n}}_{-i})$  quand  $\mathbf{n}_{-i} \leq \tilde{\mathbf{n}}_{-i}$  a lieu bande par bande,  $\mathbb{P}(\mathbf{n}_{-i}) > 0$  et  $\mathbb{P}(\tilde{\mathbf{n}}_{-i}) > 0$ . Le modèle de mélange Widom-Rowlinson est monotone par rapport à la relation d'ordre sur  $S$  donnée par  $\mathbf{e}_1 < \mathbf{0} < \mathbf{e}_2$  [84]. De même pour le processus ponctuel bivarié avec interaction par paires, pour lequel  $\gamma_{ii} \equiv 1$  et  $\gamma_{ij}(t) > 0$  pour tout  $t > 0$ .

Pour vérifier cette dernière assertion, considérons un processus bivarié avec interaction par paires décrit par la densité (2.12). Supposons  $\gamma_{ii} \equiv 1$  pour  $i \in \{1, 2\}$  et  $\gamma_{12}(t) > 0$  pour tout  $t > 0$ . Après avoir discrétisé  $W$ , pour prouver la monotonie par rapport à la relation d'ordre  $\mathbf{e}_1 < \mathbf{0} < \mathbf{e}_2$  on  $S$ , nous devons établir

$$\begin{aligned} \mathbb{P}(N_i = \mathbf{e}_1 | \mathbf{n}_{-i}) &\geq \mathbb{P}(N_i = \mathbf{e}_1 | \tilde{\mathbf{n}}_{-i}) \\ \text{et} \\ \mathbb{P}(N_i = \mathbf{e}_2 | \mathbf{n}_{-i}) &\leq \mathbb{P}(N_i = \mathbf{e}_2 | \tilde{\mathbf{n}}_{-i}) \end{aligned}$$

pour tout  $i$  et toutes les paires de configuration pour lesquelles  $\mathbf{n}_{-i} \leq \tilde{\mathbf{n}}_{-i}$ . En posant  $\xi_{i1} = (w_i, 1)$ ,  $\xi_{i2} = (w_i, 2)$  pour les points représentatifs de la cellule  $W_i$  et  $\alpha_i = \nu(W_i)/2$ , alors la probabilité  $\mathbb{P}(N_i = \mathbf{e}_1 | \mathbf{n}_{-i})$  s'écrit

$$\frac{\alpha_i \beta_1 \prod_{w \in n_{-i}(2)} \gamma_{12}(\|w_i - w\|)}{1 + \alpha_i \beta_1 \prod_{w \in n_{-i}(2)} \gamma_{12}(\|w_i - w\|) + \alpha_i \beta_2 \prod_{w \in n_{-i}(1)} \gamma_{12}(\|w_i - w\|)} = \frac{1}{\left[ \frac{1}{\alpha_i \beta_1} + \frac{\beta_2}{\beta_1} \prod_{w \in n_{-i}(1)} \gamma_{12}(\|w_i - w\|) \right] \prod_{w \in n_{-i}(2)} \gamma_{12}(\|w_i - w\|)^{-1} + 1}.$$

Ici nous avons noté par  $n_{-i}(m)$  les positions des points représentatifs  $\xi_{i'j}$ ,  $i' \neq i$ , qui ont la marque  $m \in \{1, 2\}$  dans  $\mathbf{n}$ . Comme par hypothèse  $\gamma_{12}(t)$  prend des valeurs dans  $(0, 1]$  et comme  $\tilde{n}_{-i}(1) \subset n_{-i}(1)$  tandis que  $\tilde{n}_{-i}(2) \supset n_{-i}(2)$ , les caractéristiques locales satisfont bien la relation d'ordre requise. Le même raisonnement s'applique aux probabilités conditionnelles d'étiquette 2.

La discrétisation présentée ci-dessus s'applique à la lettre aux modèles Candy et Bisous. Cependant, nous n'avons pas pu ordonner les caractéristiques locales [T8].

### L'échantillonneur de Gibbs pour des processus sur grilles

L'échantillonneur de Gibbs exact de Häggström and Nelander [84] pour des champs aléatoires sur grilles finies s'appuie sur les principes du CFTP [155]. Pour éviter le couplage simultané de trop de chaînes, les auteurs proposent de garder la trace de toutes les valeurs possibles à chaque moment dans chaque cellule de la grille. L'espace d'états de la chaîne de Markov obtenue est construit à partir de sous-ensembles de l'ensemble de valeurs de la loi du champ aléatoire à simuler. Dans notre cas, la grille est représentée par  $\Xi$ , et chaque  $n_{ij}$  prend des valeurs dans  $\{\mathbf{0}, 1\}$ .

Une implémentation directe de cette méthode risque de manquer d'efficacité si l'on n'exploite pas la contrainte  $\sum_{j=1}^{n_M} n_{ij} \leq 1$  pour tout  $i$ . Pour cette raison, nous proposons de mettre en oeuvre des mises à jour par bandes. Plus précisément, supposons que l'état courant est  $\mathbf{n} = (\mathbf{n}_1, \dots, \mathbf{n}_{n_W})$  où  $\mathbf{n}_i \in S$  est le vecteur de la bande  $i$  de composantes  $n_{ij}$ ,  $j = 1, \dots, n_M$ . Naturellement, chaque  $\mathbf{n}$  appartient à l'espace produit  $S^{\otimes n_W}$ . A chaque instant  $t$ , une bande  $W_i \times M$  est sélectionnée au hasard et sa valeur courante  $\mathbf{n}_i$  est remplacée par un échantillon de caractéristiques locales (5.12), les valeurs des autres bandes restant inchangées. Comme  $S$  est fini, un seul nombre pseudo-aléatoire  $u$  suffit à effectuer la mise à jour une fois la bande sélectionnée. Notant  $\psi$  le noyau de transition, le nouvel état est  $\psi(\mathbf{n}, u, i)$ .

Si l'on applique le CFTP à toutes les chaînes initialisées par tous les états de probabilité positive possibles, et si l'on garde la trace de l'union des états possibles  $\mathbf{en}$  chaque point de la grille au cours du temps plutôt que l'évolution des trajectoires individuelles de chaque chaîne, nous obtenons l'algorithme suivant. Comme auparavant, les variables aléatoires produites sont réutilisées quand le temps est doublé.

#### Algorithme 10



$(\mathbf{x}, T) = \text{Gibbs Exact}$

Soit  $V_t$ ,  $t = -1, -2, \dots$  une famille de variables aléatoires indépendantes uniformes sur  $\{1, \dots, n_W\}$  et soit  $U_t$  une famille de variables aléatoires indépendantes uniformes sur  $(0, 1)$ . Supposons que  $\psi : S^{\otimes n_W} \times (0, 1) \times \{1, \dots, n_W\}$  est un noyau de transition pour l'échantillonneur de Gibbs. Ceci signifie que  $\psi(\mathbf{n}, U, v)$  modifie l'état  $\mathbf{n}$  dans la bande  $v$  de telle manière que le nouvel état suit la loi  $\mathbb{P}$ -conditionnelle sachant  $\mathbf{n}_{-v}$ . Initialiser  $T = 1$ .

1. Soit  $\mathcal{X}_i^{-T}(-T) = S$  pour tout  $i = 1, \dots, n_W$ .
2. Simuler vers le futur à partir de  $-T$  en posant

$$\mathcal{X}_i^{-T}(t+1) = \{s \in S : \exists \chi \in \mathcal{X}^{-T}(t) : \psi(\chi, U_t, V_t) = s\}$$

et en laissant inchangés les ensembles des autres bandes.

3. Si pour toutes les bandes  $W_i \times M$ , le cardinal de  $\mathcal{X}_i^{-T}(0)$  vaut 1, alors arrêter. Sinon, poser  $T = 2T$  et recommencer.
4. Retourner  $\mathbf{x} = \mathcal{X}^{-T}(0)$  et  $T$ .

### 5.3.6 Convergence de l'échantillonneur de Gibbs

Le résultat ([84] Thm. 1 pp. 400) implique que si la probabilité que  $N_i = \mathbf{0}$  sachant les valeurs des autres bandes est uniformément minorée par une valeur  $\delta > 0$ , alors l'algorithme finit presque sûrement et retourne un échantillon non-biaisé de  $\mathbb{P}$ . Sous l'hypothèse que (5.11) est localement stable, pour tous les  $\mathbf{n}_{-i}$  tels que  $\mathbb{P}(\mathbf{n}_{-i}) > 0$  nous avons

$$\begin{aligned} \mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i}) &= \frac{1}{1 + \nu(W_i) \sum_{j=1}^{n_M} \nu_M(M_j) \prod_{\mathbf{z}} \varphi(\mathbf{z} \cup \{\xi_{ij}\})^{\eta(\mathbf{n}_{-i}; \mathbf{z})}} \\ &\geq \frac{1}{1 + \nu(W_i) \sum_{j=1}^{n_M} \nu_M(M_j) \Lambda} \geq \frac{1}{1 + W^* \Lambda} > 0 \end{aligned}$$

où  $W^* = \max\{\nu(W_i) : i = 1, \dots, n_W\}$ , et la condition sur la borne inférieure est respectée.

ce qui suit, nous allons faire quelques remarques d'implémentation. Tout comme l'Algorithme 5, l'Algorithme 10 est particulièrement intéressant si une monotonie par rapport à une relation d'ordre convenablement choisie peut être établie [155, 83]. Dans ce cas, si le noyau de transition respecte cet ordre, par exemple

$$\psi(\mathbf{n}, u, i) = \max\{j = 0, \dots, n_M : \mathbb{P}(N_i \geq \mathbf{e}_j | \mathbf{n}_{-i}) \geq u\},$$

l'Algorithme 10 n'est autre que l'algorithme original de Propp et Wilson dans [155], ou encore, selon notre présentation, l'Algorithme 3, la fonction

*mettre à jour* étant donnée par  $\psi$ .

De plus, il est souvent possible de gagner du temps si un nouvel état  $\mathcal{X}_{V_t}^{-T}(t+1)$  n'est calculé explicitement que si l'on sait d'avance qu'il s'agira d'un singleton, ou bien poser  $\mathcal{X}_{V_t}^{-T}(t+1) = S$  sinon [84, 100, 145]. En effet, comme

$$\mathbb{P}(N_i = \mathbf{0} | \mathbf{n}_{-i}) \geq \frac{1}{1 + \Lambda \max_i \nu(W_i)} > 0$$

est vrai pour tout  $i$  et tout  $\mathbf{n}_{-i}$ , nous pouvons choisir le noyau de transition de façon à ce que  $\psi(\chi, u, v) = \mathbf{0}$  à chaque fois que  $u \leq 1/(1 + \Lambda \max_i \nu(W_i))$ . Ainsi, pour de petites valeurs de  $U_t$ , la valeur  $\mathbf{0}$  est attribuée à la bande  $V_t$ , sans tenir compte de  $\mathcal{X}^{-T}(t)$ . Lorsque  $U_t > 1/(1 + \Lambda \max_i \nu(W_i))$ , si un ensemble  $\mathcal{X}_v^{-T}(t)$  d'un voisin de  $V_t$  (au sens où les caractéristiques locales dans  $V_t$  dépendent de  $v$ ) n'est pas un singleton, alors aucune mise à jour n'est faite, c'est à dire que l'on pose  $\mathcal{X}_{V_t}^{-T}(t+1) = S$ . Les articles précédemment cités donnent différents noms à cette approche. Pour la suite nous ferons référence en tant que la technique de couplage "multigamma".

### 5.3.7 Etude par simulation

Dans cette partie nous présentons une étude par simulation pour mettre en évidence l'applicabilité et l'efficacité des algorithmes de simulation exacte étudiés. Les modèles que nous considérons ici sont le Widom-Rowlinson (2.11) et Strauss bivarié (2.12). L'étude complète qui inclut aussi les modèles de Cauchy et Candy figure dans [T8].

Pour chacun des quatre algorithmes considérés, nous générons 25 échantillons indépendants et affichons l'évolution du temps de convergence ou plus précisément du nombre de "sauts" en fonction des paramètres des modèles.

Pour le modèle de Widom-Rowlinson, les densités vérifient  $\beta_1 = \beta_2$ . Même chose pour le modèle de Strauss bivarié. Ses fonctions d'interaction valent  $\gamma_{11}(t) = \gamma_{22}(t) = 1$  et  $\gamma_{12}(t) = 0.5\mathbf{1}\{t \leq r\}$ . Pour ces deux modèles les densités varient de 1 par pas de 1 jusqu'à un maximum dépendant de l'algorithme. Le paramètre de portée  $r$  des deux modèles évolue de 0.005 par pas de 0.005, sauf pour l'échantillonneur de Gibbs pour lequel la discrétisation est un tout petit peu différente.

Les algorithmes ont été programmé en C++ et sont disponibles sur simple demande [178].

## CFTP

Pour l’Algorithme 5, un saut est défini comme une naissance ou une mort dans la chaîne dominante. Il convient de remarquer que le nombre d’opérations est en vérité plus grand, en raison du schéma de dédoublement du temps. En effet, si le temps de coalescence est  $T = -4$ , les sauts entre  $T = -4$  et  $T = -2$  sont amincis une fois, ceux entre  $T = -2$  et  $T = -1$  deux fois et ceux après  $T = -1$  trois fois.

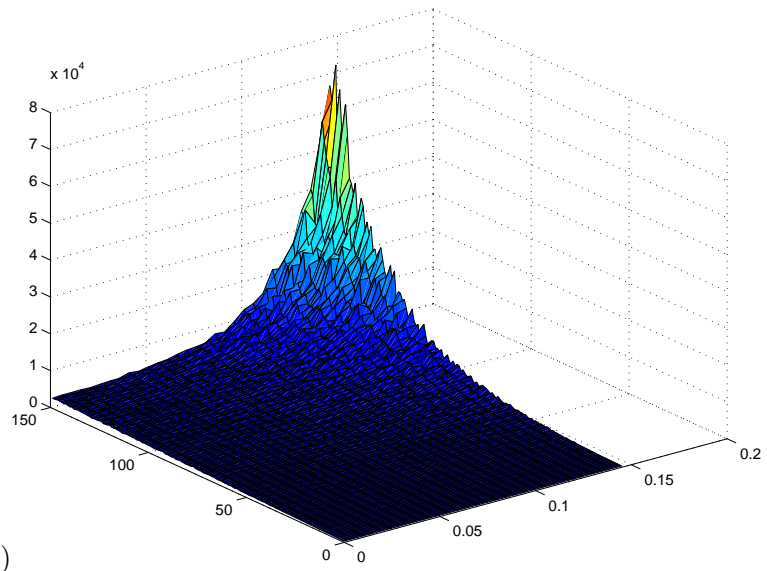
La Figure 5.9 montre l’évolution du nombre moyen de sauts par rapport aux densités et à la portée. Les bornes supérieures sont 150 et 0.15, respectivement. Nous pouvons remarquer que plus l’interaction est forte, plus long est le temps de convergence de l’algorithme. Effectivement, pour le modèle Widom-Rowlinson l’interaction est en tout ou rien ; elle est dite dure ou de type hard-core. Pour le modèle de Strauss, l’interaction est spécifiée par une constante positive ; elle est dite molle ou de type soft-core.

## Clan des ancêtres

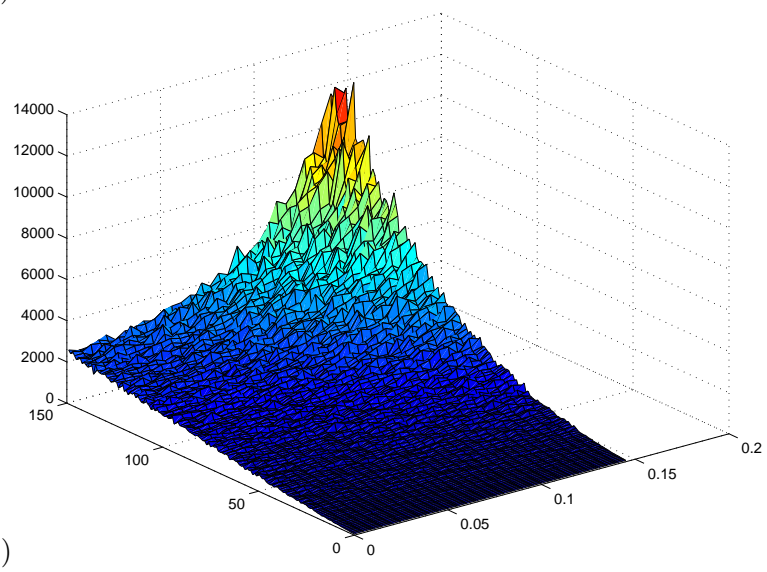
Pour l’Algorithme 7, un saut est défini comme une naissance ou une mort dans la simulation vers le passé. Le nombre total de sauts représente bien la charge de calcul de cet algorithme car il n’y a pas doublement du temps.

Le nombre moyen de sauts est présenté Figure 5.10(a) en fonction des paramètres du modèle. Ici, la densité maximale est de 100 et la portée de 0.08. L’évolution vers le passé n’est déterminée que par  $\Lambda$  et  $r$ . En conséquence, la structure de l’interaction n’est pas pertinente et tous les modèles obtenus par amincissement du même processus dominant nécessitent le même temps de simulation. Sur la plage de paramètres considérée, le clan des ancêtres est plus rapide que le couplage du passé. Cependant, le couplage du passé s’applique à une plus grande famille de paramètres.

Nous avons ensuite utilisé l’index d’incompatibilité (5.6) qui décrit mieux la structure de voisinage qu’une boule de rayon fixé. Effectivement, pour le modèle de Widom-Rowlinson ainsi que pour le Strauss bivarié, la densité conditionnelle d’un point  $(w, m)$  ne dépend que des points de type différent de  $m$ , situés à distance inférieure à  $r$  de  $w$ . Si nous prenons à nouveau 25 échantillons indépendants pour approximer le nombre moyen de sauts pour chaque couple de paramètres, nous obtenons le graphe de la Figure 5.10(b). L’évident gain d’efficacité est dû au fait que le nombre des ancêtres est inférieur à celui des voisins pour un système de voisinage à taille fixe.



a)



b)

FIGURE 5.9 – Nombre moyen de sauts en fonction des densités et de la portée, pour 25 échantillons indépendants obtenus avec l’algorithme CFTP. Les modèles sont : a) Widom-Rowlinson, b) Strauss bi-varié.

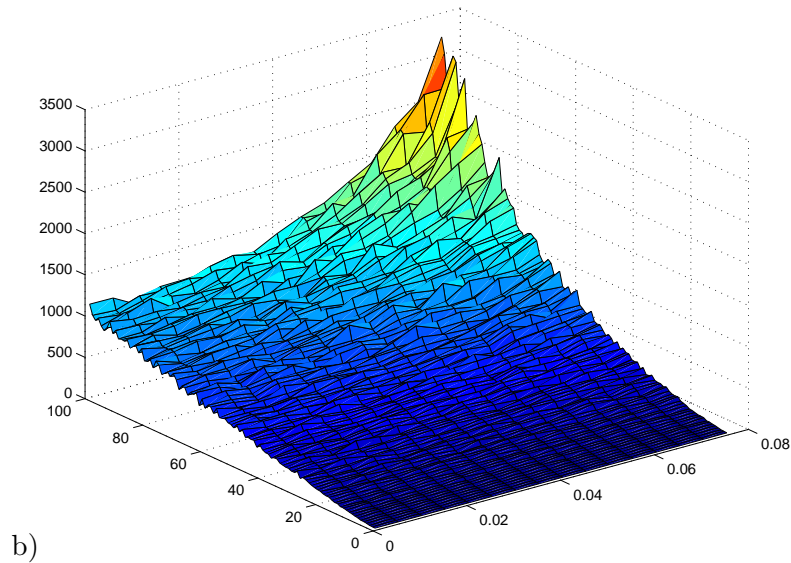
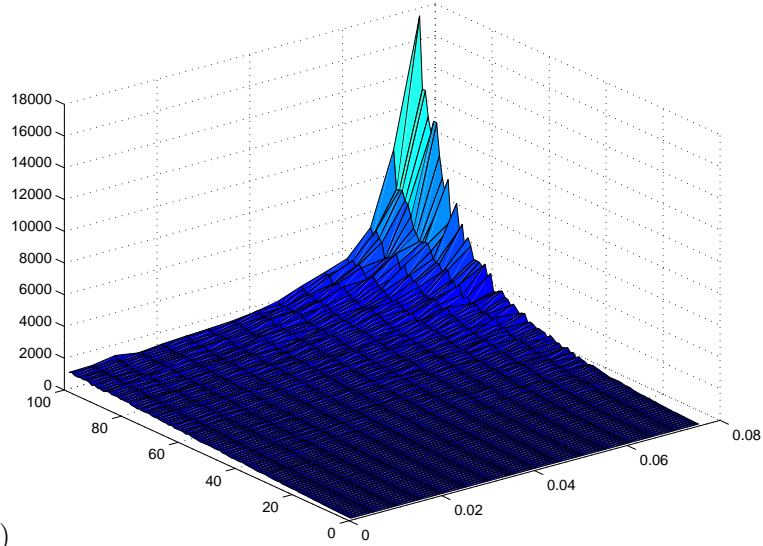


FIGURE 5.10 – Nombre moyen de sauts en fonction des densités et de la portée, pour 25 échantillons indépendants obtenus avec l’algorithme clan des ancêtres. Les modèles sont le Widom-Rowlinson et le Strauss bi-varié. a) l’algorithme utilise un voisinage de dimension fixe, b) l’algorithme utilise l’index d’incompatibilité.

## Metropolis–Hastings

L’Algorithme 8 laisse un peu de liberté pour le choix de la partition et des probabilités de naissances. L’effet de ces choix est analysé dans les expériences suivantes. Nous supposons que toutes les probabilités de naissance sont égales  $p_b^i \equiv p_b$  et que la partition du domaine  $W = [0, 1] \times [0, 1]$  se fait en  $m \times m$  cellules de même volume  $1/m^2$ .

Dans la première expérience, la taille  $m$  de la cellule est adaptée à  $p_b$  en prenant  $\lceil \sqrt{(\Lambda/p)} \rceil$  où  $p = p_b/(1 - p_b)$ . En utilisant l’Algorithme 8, nous simulons un processus ponctuel de Poisson de densité  $\Lambda = 50$  pour différentes valeurs de  $p_b$ . Nous observons que si les marques sont indépendantes, le choix de leur distribution n’affecte pas le temps de convergence de l’algorithme. Pour chaque valeur de  $p_b$  nous avons simulé 500 réalisations indépendantes et calculé le temps moyen de coalescence. Nous avons également prélevé le nombre de sauts en écartant les suppressions dans des bandes vide. Les résultats sont consignés Tableau 5.1. Nous y observons que les petites et les grandes valeurs de  $p_b$  augmentent le temps de coalescence, tandis que des valeurs intermédiaires ( $p_b = 0.10$  ou  $0.25$ ) donnent plutôt de bons résultats.

$p_b$	Temps de coalescence	Nombre de sauts
0.01	34471.94	690.80
0.10	4816.90	965.64
0.25	4048.90	2023.12
0.45	24707.07	22233.11

TABLE 5.1 – Evolution du temps moyen de coalescence et nombre moyen de sauts en fonction de  $p_b$ , obtenue sur 500 échantillons simulés à partir d’un processus de Poisson de densité  $\beta = 50$ .

L’influence de la taille de la cellule  $m$  est également étudiée en prenant  $p_b = 0.1$ . Les résultats obtenus donnés au Tableau 5.2. Comme précédemment, le temps moyen de coalescence et le nombre moyen de sauts ont été calculés sur 500 échantillons d’un processus de Poisson de densité 50. Pour  $p_b = 0.10$ , nous obtenons  $\lceil \sqrt{(\Lambda/p)} \rceil = 22$ . Le Tableau 5.2 montre que le raffinement de la grille augmente le temps de coalescence.

La Figure 5.11 donne le temps moyen de coalescence en fonction de la densité et de la portée. Il a été calculé sur 25 simulations obtenues avec l’Algorithme 8. Le coût de calcul est augmenté par le dédoublement de temps. Les bornes supérieures de la densité et de la portée sont respectivement de 150 et de 0.15. La probabilité de naissance est  $p_b = 0.1$ . Le domaine  $W$  est découpé en  $m^2$  cellules carrées de dimension  $m = \lceil \sqrt{(\Lambda/p)} \rceil$ . Les

Grille	Temps de coalescence	Nombre de sauts
25	6520.83	1304.47
30	9437.18	1890.16
35	13828.10	2768.14
40	17580.03	3518.75

TABLE 5.2 – Evolution du temps moyen de coalescence et du nombre moyen de sauts en fonction de la taille de la grille, obtenue, avec  $p_b = 0.1$ , par simulation de 500 échantillons d’un processus de Poisson de densité  $\beta = 50$ .

résultats obtenus sont comparables à ceux obtenus par l’Algorithme 5 : une interaction forte nécessite des temps de coalescence plus grands. Cependant, l’algorithme exact de Metropolis-Hastings est plus lent que le couplage du passé basé sur les processus spatiaux de naissance et de mort.

### Echantillonneur de Gibbs

A chaque instant, l’état courant est mis à jour par l’échantillonneur de Gibbs. Pour cette raison, la Figure 5.12 montre le temps moyen de coalescence en fonction de la densité et de la portée. Comme pour le couplage du passé, le temps calcul est plus important à cause du dédoublement du temps. Pour les modèles Widom-Rowlinson et Strauss bivarié, des relations de monotonie sont utilisées pour les mises à jour.

Tous les modèles ont été simulés sur une grille  $W$  de taille  $20 \times 20$ . En raison de la discrétisation, seules quelques valeurs de portée peuvent être considérés. Ici, nous avons pris  $r = 0.005, 0.055, 0.075, 0.105, 0.115$  et  $0.155$ . La borne supérieure pour l’intensité est 100. Pour les modèles de Widom-Rowlinson et de Strauss bivarié, l’espace des marques est discret.

Pour les modèles étudiés, la Figure 5.12 montre des résultats comparables à ceux obtenus par le couplage du passé : plus l’interaction est forte, plus long est le temps de coalescence. Les auteurs de [T8] ont étudié également la simulation parfaite sur une version du modèle Candy en utilisant le couplage multigamma. Il apparaît que les algorithmes de Gibbs qui peuvent exploiter les propriétés de monotonie sont beaucoup plus rapides.

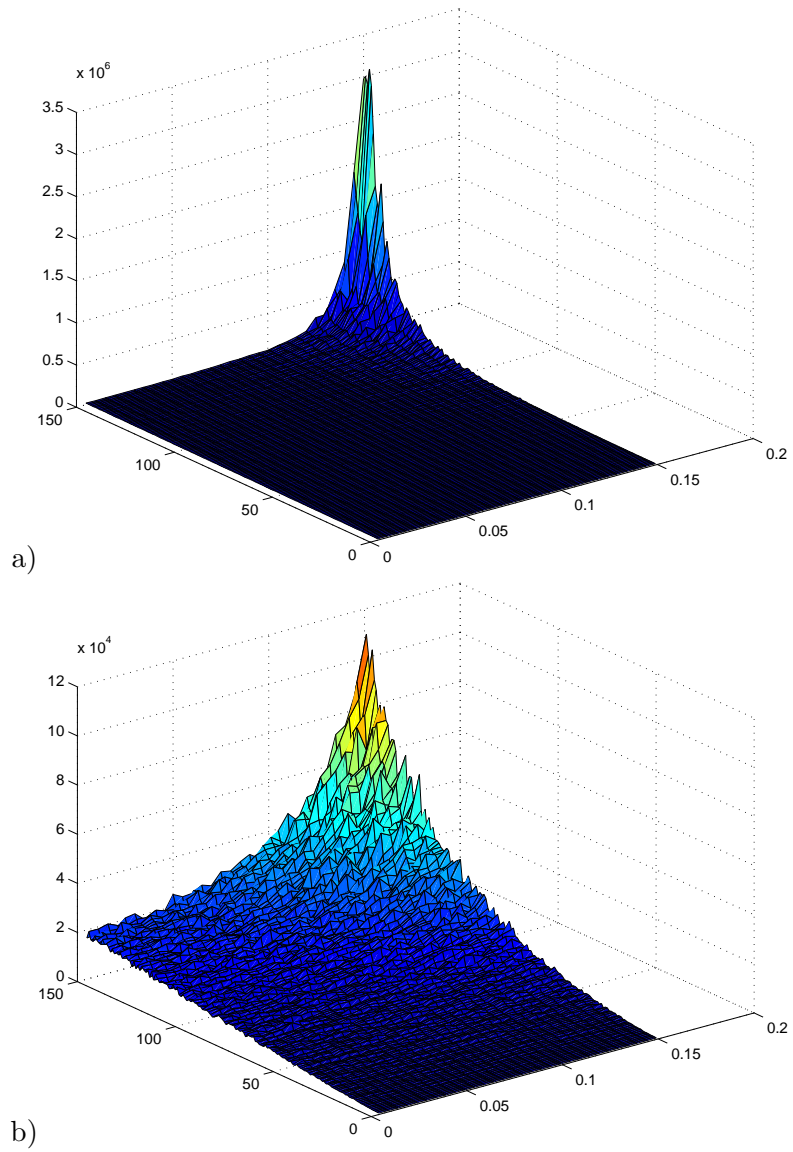


FIGURE 5.11 – Temps moyen de coalescence en fonction de la densité et de la portée, obtenu sur 25 échantillons indépendants simulés par l’algorithme Metropolis-Hastings. Les modèles sont : a) Widom-Rowlinson, b) Strauss bivarié. Dans les deux cas, la grille a été adaptée à la probabilité de naissance  $p_b = 0.1$ .



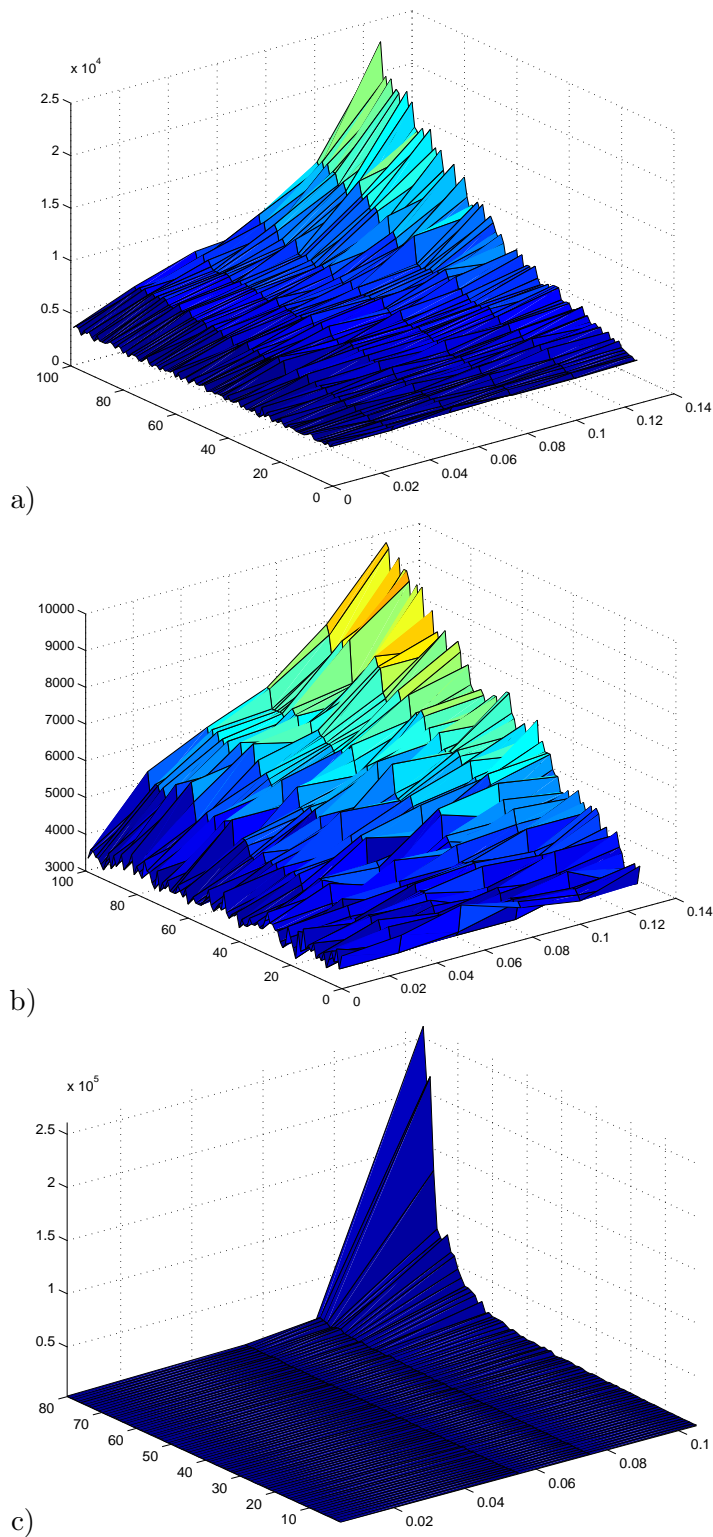


FIGURE 5.12 – Temps moyen de coalescence en fonction de la densité et de la portée, obtenu sur 25 échantillons indépendants simulés par l'algorithme de Gibbs. Les modèles sont : a) Widom-Rowlinson, b) Strauss bi-varié, c) Candy.

**Troisième partie**  
**Inférence statistique**

## Questions sur la forme

Nous disposons d'un modèle de forme et d'un moyen de le simuler. Se posent alors les questions suivantes :

- quelle est la forme qui se cache dans les données ?
- une forme étant observée, quels sont les paramètres du modèle qui permettent de la reproduire ?
- la forme détectée existe-t-elle vraiment ou bien s'agit il d'un effet du hasard ?

Dans cette partie du mémoire, des réponses à ces questions sont proposées.

La première question traite les données comme un champ externe. Pour modéliser la forme qui s'y trouve, il faut prendre en compte deux termes : l'attache aux données et le terme d'interaction. Le premier terme localise les objets qui composent la forme et le deuxième régit leurs interactions. Le degré de connaissance des paramètres du modèle détermine la façon de répondre à cette question. Il se peut que les paramètres soient connus d'avance ou bien qu'ils doivent être modélisés par des connaissances a priori sur leur comportement. Une fois ce choix fait, une loi de probabilité qui porte à la fois sur le champ de données, les paramètres du modèle et la forme, peut être construite. La réponse à la question posée est fournie par la configuration d'objets qui maximise cette loi de probabilité. Cette maximisation peut se faire à l'aide d'un algorithme de type recuit simulé. Cet algorithme sera basé sur une dynamique MCMC spécialement conçue pour simuler le modèle proposé.

La deuxième question est posée à partir d'une réalisation de la forme. Les descripteurs statistiques de la forme sont alors considérés comme des statistiques exhaustives. Ces statistiques peuvent être observées dans leur totalité ou partiellement. Dans la première situation, nous sommes dans le cas de données complètes. Il est possible alors d'estimer les paramètres du modèle par le maximum de vraisemblance calculé par méthodes Monte Carlo (MCMCML). Si les statistiques ne sont observées que partiellement, il est possible de répondre à la question posée en utilisant des algorithmes de type EM.

La dernière question de notre projet vise à développer des outils qui permettent de valider les résultats obtenus.

Sous certaines hypothèses, quand il s'agit de l'estimation des paramètres il

est possible de calculer des erreurs d'estimation. Dans un premier temps, il est possible d'approximer l'erreur standard, qui est la différence entre l'estimateur de maximum de vraisemblance et la valeur vraie des paramètres. Nous pouvons ensuite calculer l'erreur standard de type Monte Carlo qui approxime la différence entre le maximum de vraisemblance calculé par des méthodes de type Monte Carlo et son estimateur. Le calcul de ces quantités sert à évaluer la précision de l'estimation.

Une fois la forme détectée dans un champ de données, il faut vérifier si le résultat obtenu est dû à la présence de la forme ou bien au choix des paramètres. Des stratégies comme des permutations circulaires ou bien les calculs de caractéristiques moyennes de forme, peuvent être mises en oeuvre. La première nous permet de calculer une p-valeur du résultat obtenu, c'est à dire de valider l'existence d'une structure dans les données. La deuxième nous permet d'inférer la forme à partir de quantités moyennes et non pas à partir d'une réalisation.

Pour cette présentation nous avons suivi principalement [72, 190, 141, 143]. Des parties de cette section ont été publiées dans [T5, T7, T17, T16, T21, T22, T23, T11, T12].

## Chapitre 6

# Outils mathématiques pour l'inférence statistique de processus ponctuels marqués

### 6.1 Détection de la forme cachée dans les données

Soit  $W \subset \mathbb{R}^d$  une fenêtre d'observation de volume fini  $0 < \nu(W) < \infty$ . Dans cette fenêtre, nous observons un champ de données spatialisées  $\mathbf{d} = \{d_1, d_2, \dots, d_m\}$  avec  $m$  entier positif. Ces données peuvent représenter les positions et les valeurs de pixels dans une image numérique, les emplacements d'élevages laitiers bovins en France et un score indiquant la susceptibilité à une certaine maladie, les positions des centres des galaxies de notre Univers et les différentes caractéristiques qui leur sont associées (masse, forme, luminosité, etc.). Les images peuvent contenir des objets : cellules biologiques, réseaux routiers et hydrographiques, parcellaire agricole, etc. En épidémiologie, la structure d'intérêt consiste en des agrégats, c'est à dire les régions où la maladie apparaît plus souvent que d'habitude. En cosmologie, les formes et les structures étudiées sont des filaments, des murs et des agrégats. Ces structures apparaissent "dessinées" par la position des centres des galaxies. L'étude de la répartition spatiale des galaxies de notre Univers est un problème d'intérêt majeur pour les cosmologistes.

Tous les modèles envisagés pour étudier ces structures s'appuient sur l'hypothèse que les formes recherchées sont des structures complexes composées d'objets simples en interaction. Les processus ponctuels marqués sont adaptés à ce cadre. De façon naturelle, la forme d'intérêt est vue comme la réalisation d'un processus ponctuel marqué, caractérisé par la densité de probabilité (3.1). Son expression est

$$p(\mathbf{x}|\theta) = \frac{\exp[-U(\mathbf{x}|\theta)]}{Z(\theta)} = \frac{\exp[-(U_d(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta))]}{Z(\theta)},$$

où  $\theta$  désigne le vecteur paramètre et  $Z(\theta)$  la fonction de partition du modèle. L'énergie totale du système  $U(\mathbf{x})$  comporte deux composantes, l'attache aux données,  $U_d(\mathbf{x}|\theta)$  et l'énergie d'interaction,  $U_i(\mathbf{y}|\theta)$ .

La densité d'un processus ponctuel marqué est caractérisée par trois ensembles de paramètres. Le premier est lié à la loi des marques des objets, le deuxième aux interactions entre objets et le troisième à l'attache aux données. Selon le contexte, certains ensembles peuvent être fusionnés [179, T7].

Notre objectif ici est de déterminer les paramètres de position et de marque des objets composant la forme cachée dans les données. Pour ce faire, nous devons introduire des connaissances a priori sur les autres paramètres du modèle.

La manière la plus simple d'introduire cette connaissance est de supposer ces paramètres connus. Cette hypothèse est très forte, mais elle peut être étayée par des analyses exploratoires, des tests empiriques ou bien par les connaissances des spécialistes travaillant sur ces données. Dans ce cas, la configuration d'objets qui se cachent dans les données peut-être approximée par l'estimateur suivant :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \Omega} p(\mathbf{x}|\theta). \quad (6.1)$$

Une deuxième stratégie consiste à adopter un cadre bayésien. Au lieu de synthétiser la connaissance sur les paramètres du modèle en une seule valeur numérique, il est beaucoup plus naturel de lui associer une loi a priori  $\tilde{p}(\theta)$ . L'estimateur que nous obtenons dans ce cas est de la forme suivante :

$$(\hat{\mathbf{x}}, \hat{\theta}) = \arg \max_{(\mathbf{x}, \theta) \in \Omega \times \Theta} p(\mathbf{x}|\theta)\tilde{p}(\theta),$$

$\Theta$  représentant l'ensemble de définition des paramètres  $\theta$ .

Dans [179, T11, T12] nous avons montré sur un exemple comment introduire cette connaissance a priori. Supposons que nous utilisons le modèle Candy ou Bisous pour détecter un réseau filamenteux dans un jeu de données quelconque. La Figure 6.1 montre deux configurations très semblables hormis le segment  $s_3$  qui a une orientation très différente. Dans le premier cas, la connexité n'empêche pas une certaine tortuosité, comme on pourrait voir sur un réseau hydrographique. Dans le deuxième cas, l'alignement des segments est favorisé, suggérant plutôt un réseau routier.

Ces stratégies de détection de formes nécessitent un algorithme d'optimisation. La méthode généralement adoptée est le recuit simulé, car il s'agit

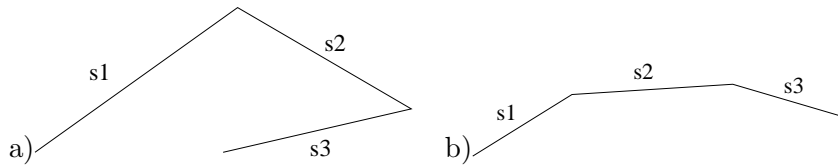


FIGURE 6.1 – a) Configuration de segments pour laquelle la connectivité est privilégiée par rapport à l’alignement, b) Configuration de segments pour laquelle la connectivité et l’alignement ont des potentiels équivalents.

d’une technique d’optimisation globale. Son principe est le suivant : la loi  $p(\mathbf{x}|\theta)^{1/T_n}$  est simulée tandis que le paramètre de température  $T_n$  converge lentement vers 0. Pour simuler la loi à température donnée, l’algorithme utilise des dynamiques de simulation Monte Carlo. Selon la dynamique utilisée, des résultats montrent l’existence d’un schéma de refroidissement pour la température qui garantit la convergence vers un optimum global de l’algorithme [69, 189, T17]. Par ce schéma de refroidissement, l’algorithme converge en loi vers la distribution uniforme sur l’ensemble des configurations qui maximisent  $p(\mathbf{x}|\theta)$ . Ceci n’est pas inconvenient pour les praticiens. Le plus grand obstacle à l’utilisation de cet algorithme est son temps de calcul. Malgré tout, la qualité des solutions obtenues fait qu’il est souvent adopté pour résoudre ce type de problème.

## 6.2 Estimation des paramètres de la forme

La deuxième question qui nous intéresse est la suivante. Soit  $W$  une fenêtre bornée dans laquelle nous observons une configuration d’objets formant une structure  $\mathbf{x}$ . Nous supposons que cette structure est la réalisation d’un processus ponctuel marqué. Quels sont les paramètres  $\theta$  du processus donnant lieu à une telle configuration d’objets ?

### 6.2.1 Maximum de la pseudo-vraisemblance

Historiquement, le concept de pseudo-vraisemblance a été introduit par Julian Besag pour estimer les paramètres d’abord de champs aléatoires discrets [23, 24], puis de processus ponctuels avec interaction par paires [25]. Il a ensuite été étendu à des classes plus larges de processus ponctuels [160, 92]. Schématiquement, l’idée principale est la suivante : à partir d’une observation  $\mathbf{x} = \{x_1, \dots, x_n\}$ , une pseudo-vraisemblance est construite comme produit des pseudo-vraisemblances conditionnelles de chaque variable connaissant les autres. Pour plus de détails, nous recommandons [141].

Soit  $p(\mathbf{x}|\theta)$  la densité d'un processus ponctuel marqué (3.1) que l'on peut écrire sous la forme

$$p(\mathbf{x}|\theta) = \frac{\exp\langle t(\mathbf{x}), \theta \rangle}{Z(\theta)} \quad (6.2)$$

où  $t(\mathbf{x})$  est le vecteur des statistiques exhaustives,  $\theta \in \Theta$  le paramètre vectoriel des d'interaction et  $\Theta$  l'espace de ces paramètres, sous-ensemble compact de  $\mathbb{R}^d$ . La dimension  $d$  correspond au nombre de paramètres scalaires à estimer.

Lorsque, moyennant un changement de variables, la loi du processus de Gibbs vu en (3.1) peut se mettre sous cette forme, on dit que cette loi décrit une famille exponentielle de paramètres canoniques  $\theta \in \Theta$  et de statistiques exhaustives le vecteur  $t(\mathbf{x})$ . Ces statistiques concentrent toute l'information disponible sur les paramètres. Cela signifie qu'à partir d'une forme observée, il suffit de calculer le vecteur des statistiques  $t(\mathbf{x})$  pour estimer les paramètres  $\theta$  du modèle (6.2). Les caractéristiques morphologiques et statistiques d'une forme sont donc synthétisées en un vecteur à valeurs réelles.

La construction de la pseudo-vraisemblance pour les processus ponctuels repose sur l'intensité conditionnelle (2.6). Intuitivement, l'intensité conditionnelle  $\lambda(\eta; \mathbf{x})$  peut être interprétée comme la probabilité conditionnelle que l'objet  $\eta$  appartienne au processus  $\mathbb{X}$  sachant que le reste du processus coïncide avec la configuration  $\mathbf{x}$ .

Pour un processus ponctuel défini selon (6.2), l'intensité conditionnelle s'écrit

$$\lambda_\theta(\eta; \mathbf{x}) = \frac{p(\mathbf{x} \cup \{\eta\}|\theta)}{p(\mathbf{x}|\theta)} = \exp \langle t(\mathbf{x} \cup \{\eta\}) - t(\mathbf{x}), \theta \rangle,$$

lorsque  $\eta \notin \mathbf{x}$ , alors que pour tout  $x_i \in \mathbf{x}$  nous avons

$$\lambda_\theta(x_i; \mathbf{x}) = \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x} \setminus \{x_i\}|\theta)} = \exp \langle t(\mathbf{x}) - t(\mathbf{x} \setminus \{x_i\}), \theta \rangle.$$

Notons que cette intensité conditionnelle est a priori discontinue aux points  $x_i \in \mathbf{x}$ , et qu'elle ne dépend pas de la fonction de normalisation  $\theta \in \Theta \mapsto Z(\theta)$ .

Par analogie avec le cas discret, la pseudo-vraisemblance d'un processus ponctuel marqué est définie comme suit :

$$PV_W(\theta; \mathbf{x}) = \lim_{\delta \rightarrow 0} \prod_{i=1}^{m(\delta)} p(\mathbf{x}_{W_i} | \mathbf{x}_{W \setminus W_i}, \theta),$$



où  $(W_i)_{i=1,\dots,m(\delta)}$  est une famille d'ensembles deux à deux disjoints qui forme une partition de  $W$  telle que  $\delta = \max_{1 \leq i \leq m} \nu(W_i)$  et  $m(\delta)\delta^2 \xrightarrow{\delta \rightarrow 0} 0$ .

**Proposition 16** *Soit  $W$  un domaine borné dans lequel nous observons la configuration d'objets  $\mathbf{x}$ . Nous supposons que cette configuration est la réalisation d'un processus ponctuel marqué de densité de probabilité (6.2) et d'intensité conditionnelle  $\lambda_\theta(u; \mathbf{x})$ . Alors, sa pseudo-vraisemblance vaut*

$$\begin{aligned} PV_W(\theta; \mathbf{x}) &= \tag{6.3} \\ &= \left[ \prod_{x_i \in \mathbf{x}} \lambda_\theta(x_i; \mathbf{x}) \right] \exp \left[ - \int_{W \times M} \lambda_\theta((w, m); \mathbf{x}) \nu(dw) \nu_M(dm) \right]. \end{aligned}$$

L'estimateur du maximum de pseudo-vraisemblance  $\hat{\theta}$  est obtenu en maximisant (6.3). Les processus ponctuels qui admettent leur densité dans la famille exponentielle, ont leur pseudo-vraisemblance concave [92]. La consistance de l'estimateur a été étudiée par [92, 117]. La normalité asymptotique de l'estimateur a été traitée dans [91, 118, 29].

Tous les ingrédients sont réunis pour construire un schéma d'optimisation de type descente de gradient pour maximiser la pseudo-vraisemblance. Le point délicat d'une telle méthode reste quand même le calcul de l'intégrale de (6.3). Différents schémas d'intégration ont été proposés par [20, 16, 121]. Aujourd'hui, la librairie **R spatstat** construite par A. J. Baddeley et R. Turner, conjointement avec des contributeurs de la communauté scientifique, propose des procédures pour le calcul de l'estimateur de la pseudo-vraisemblance pour plusieurs modèles de processus ponctuels.

### 6.2.2 Maximum de vraisemblance

Rappelons brièvement notre problème. Dans un domaine borné  $W$ , une configuration d'objets  $\mathbf{x}$  est observée et considérée comme la réalisation d'un processus ponctuel marqué. Ce processus est caractérisé par sa densité  $p(\mathbf{x}|\theta)$  par rapport à la mesure de référence (2.4). Cette densité est supposée relever de la famille des modèles exponentiels (6.2).

Dans ce contexte, la fonction de log-vraisemblance s'écrit :

$$l(\theta) = \langle t(\mathbf{x}), \theta - \psi \rangle - \log \frac{Z(\theta)}{Z(\psi)}$$

où  $\psi \in \Theta$  est une valeur connue quelconque.

Le calcul du rapport des constantes de **normalisation**,

$$\begin{aligned}
\frac{Z(\theta)}{Z(\psi)} &= \frac{1}{Z(\psi)} \int_{\Omega} p(\mathbf{x}|\theta) d\mu(\mathbf{x}) \\
&= \frac{1}{Z(\psi)} \int_{\Omega} p(\mathbf{x}|\theta) \frac{p(\mathbf{x}|\psi)}{p(\mathbf{x}|\psi)} d\mu(\mathbf{x}) \\
&= \int_{\Omega} \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\psi)} \frac{p(\mathbf{x}|\psi)}{Z(\psi)} d\mu(\mathbf{x}) \\
&= \mathbb{E} \left[ \frac{p(\mathbb{X}|\theta)}{p(\mathbb{X}|\psi)} \right]
\end{aligned}$$

nous donne

$$\frac{Z(\theta)}{Z(\psi)} = \mathbb{E} [\exp\langle t(\mathbb{X}), \theta - \psi \rangle], \quad (6.4)$$

expression qui n'a pas toujours d'expression analytique.

Pendant longtemps, ce problème a été l'obstacle principal à l'adoption du maximum de vraisemblance. Il peut être contourné en utilisant les méthodes Monte Carlo. Aujourd'hui, l'approximation du maximum de vraisemblance par des techniques de type Monte Carlo est devenu un outil classique en statistique inférentielle [74, 71, 72, 190, T7, 141, 149, 150, 151, T11, T12].

Considérons une approximation de type Monte Carlo de (6.4)

$$\frac{Z(\theta)}{Z(\psi)} \approx \frac{1}{n} \sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle, \quad (6.5)$$

où  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$  sont des échantillons simulés selon  $p(\mathbf{x}|\psi)$ . Alors, en utilisant (6.5) pour calculer la log-vraisemblance, nous obtenons

$$l_n(\theta) = \langle t(\mathbf{x}), \theta - \psi \rangle - \log \left( \frac{1}{n} \sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle \right).$$

Si la chaîne de Markov qui génère les échantillons  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$  est irréductible, alors  $\frac{1}{n} \sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle$  converge vers  $Z(\theta)/Z(\psi)$  presque sûrement. Ceci implique que  $l_n(\theta)$  converge vers  $l(\theta)$  presque sûrement également. Soient  $\hat{\theta}$  et  $\hat{\theta}_n$  les maxima respectifs de  $l(\theta)$  et de  $l_n(\theta)$ . Alors, sous l'hypothèse que la densité de notre processus ponctuel relève de la famille exponentielle lorsque  $\theta \in \Theta$ ,  $\hat{\theta}_n$  converge presque sûrement vers  $\hat{\theta}$  [74, 72, 141, 143].

La fonction de vraisemblance  $l(\theta)$  des modèles exponentiels est convexe [143]. En principe, pour calculer le maximum d'une fonction convexe nous pouvons recourir à une méthode d'optimisation locale. La convergence presque sûre

de  $l_n(\theta)$  vers  $l(\theta)$ , nous incite à calculer le gradient de la fonction de log-vraisemblance approchée

$$\nabla l_n(\theta) = t(\mathbf{x}) - \mathbb{E}[n, \theta, \psi] t(\mathbb{X}) \quad (6.6)$$

où l'on a posé

$$\mathbb{E}[n, \theta, \psi] t(\mathbb{X}) = \frac{\sum_{i=1}^n t(\mathbb{X}_i) \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle}{\sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle}$$

qui est l'approximation de Monte Carlo par échantillonnage pondéré de  $\mathbb{E}_{\theta} t(\mathbb{X})$ .

De la même façon, nous calculons le hessien de la fonction de log-vraisemblance approximée

$$-\nabla^2 l_n(\theta) = \text{Var}_{n, \theta, \psi}[t(\mathbb{X})] \quad (6.7)$$

où

$$\text{Var}_{n, \theta, \psi}[t(\mathbb{X})] = \mathbb{E}[n, \theta, \psi] t(\mathbb{X}) t(\mathbb{X})^t - \mathbb{E}[n, \theta, \psi] t(\mathbb{X}) \mathbb{E}[n, \theta, \psi] t(\mathbb{X})^t.$$

Les estimateurs du gradient et du hessien ainsi obtenus sont consistants. Nous observons que le gradient et le hessien de la log-vraisemblance ont les mêmes expressions que leur estimateurs respectifs lorsque l'on remplace dans (6.6) et (6.7) les espérances Monte Carlo par les espérances exactes.

Nous disposons maintenant de tous les ingrédients pour construire des méthodes itératives d'optimisation locale afin de trouver  $\hat{\theta}_n$ . Par la méthode Newton-Raphson de type Monte Carlo, on obtient

$$\theta_{k+1} = \theta_k - [\nabla^2 l_n(\theta_k)]^{-1} \nabla l_n(\theta_k) \quad (6.8)$$

pour  $k = 1, 2, \dots$ ,  $l_n(\cdot)$  étant calculée à partir de  $n$  échantillons de la loi  $p_{\mathbf{d}}(\mathbf{y}|\theta_0)$ . Les calculs de l'inverse du hessien et du gradient peuvent s'avérer coûteux et parfois numériquement instables. Une alternative possible au (6.8) est la méthode du gradient stochastique [144, 203] :

$$\theta_{k+1} = \theta_k + \epsilon_k [t(\mathbf{x}) - t(\mathbb{X}_k)]$$

où  $\epsilon_k > 0$  est une séquence décroissante et  $\mathbb{X}_k$  un échantillon tiré selon  $p_{\mathbf{d}}(\mathbf{x}|\theta_k)$ . Cette méthode est a priori plus simple à mettre en oeuvre, mais le choix d'une séquence optimale  $\{\epsilon_k\}$  reste un problème délicat [72].

Une alternative à ces deux stratégies est d'utiliser la méthode du gradient à pas optimal [37, 57, 154, 179, T7]. Dans ce cas, les paramètres suivent le schéma itératif suivant :

$$\theta_{k+1} = \theta_k + \rho(\theta_k) \nabla l_n(\theta_k),$$

le pas de descente  $\rho(\theta_k)$  étant choisi de façon optimale

$$l_n(\theta_k + \rho(\theta_k) \nabla l_n(\theta_k)) = \max_{\rho \in \mathbb{R}} l_n(\theta_k + \rho \nabla l_n(\theta_k)).$$

A chaque itération,  $\rho(\cdot)$  "dirige" les paramètres vers les valeurs qui maximisent la fonction de vraisemblance.

Lorsque  $k$  tend vers l'infini, et moyennant des conditions de régularité, toutes ces méthodes approchent l'estimateur du maximum de vraisemblance. Ces trois méthodes sont sensibles aux conditions initiales. Pour les modèles de familles exponentielles, la fonction de vraisemblance étant convexe, ce choix a peu importance. Cependant, une condition initiale trop éloignée de la solution peut induire des instabilités numériques. Un rééchantillonnage périodique du modèle permet de lever cette difficulté, mais il a un prix en termes de temps de calcul.

Supposons maintenant qu'une valeur  $\hat{\theta}_n$  a été obtenue en maximisant une approximation de la fonction de vraisemblance  $l_n(\theta)$ . Cette fonction a été construite à partir de  $n$  échantillons  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$  tirés selon  $p_{\mathbf{d}}(\mathbf{x}|\theta_0)$ . La différence entre  $\hat{\theta}_n$  et  $\hat{\theta}$ , estimateur vrai mais inconnu du maximum de vraisemblance, est caractérisée par un théorème de la limite centrale [71, 72].

Ainsi, quand  $n \rightarrow \infty$ , la variable aléatoire  $\sqrt{n}(\hat{\theta}_n - \hat{\theta})$  converge en loi vers une variable normale de moyenne nulle et de variance  $I(\hat{\theta})^{-1} \Gamma I(\hat{\theta})^{-1}$  :

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}) \rightarrow \mathcal{N}(0, I(\hat{\theta})^{-1} \Gamma I(\hat{\theta})^{-1}).$$

La matrice

$$I(\hat{\theta}) = \text{Var}_{\hat{\theta}}[t(\mathbb{X})] = -\nabla^2 l(\hat{\theta})$$

représente l'information de Fisher de  $\hat{\theta}$ , alors que  $\Gamma$  est la matrice de covariance asymptotique du gradient de Monte Carlo normalisé,  $\sqrt{n} \nabla l_n(\hat{\theta})$ .

En pratique, ces deux quantités sont très faciles à estimer. Clairement,  $I(\hat{\theta})$  peut être approximée par  $-\nabla^2 l_n(\hat{\theta}_n)$ . En ce qui concerne  $\Gamma$ , nous allons considérer l'estimateur

$$\frac{C_n}{\left\{ \frac{1}{n} \sum_{i=1}^n \exp \left[ t(\mathbb{X}_i)^t (\hat{\theta}_n - \psi_0) \right] \right\}^2}$$

où  $C_n$  la matrice de covariance empirique de  $(t(\mathbf{x}) - t(\mathbb{X})) \exp \left[ t(\mathbb{X})^t (\hat{\theta}_n - \psi_0) \right]$ , calculée à partir des échantillons  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$ .

Sous les hypothèses concernant les asymptotiques du maximum de vraisemblance, la matrice  $-\nabla^2 l_n(\hat{\theta}_n)$  peut être considérée comme l'inverse de la variance asymptotique de l'estimateur du maximum de vraisemblance [72, 143]. Ainsi, en prenant la racine carrée des éléments diagonaux de l'inverse de cette matrice, on obtient une estimation de l'erreur  $\hat{\theta} - \theta_0$ , qui représente la différence entre l'estimateur de maximum de vraisemblance et les vraies valeurs des paramètres du modèle.

### 6.2.3 Algorithme EM : données incomplètes

L'observation d'une structure dans sa totalité n'est pas toujours possible. Dans ce cas, une solution pour estimer les paramètres est l'utilisation d'algorithmes de type espérance-maximisation. L'idée d'un tel algorithme est de "compléter" les données en les simulant dans les régions où les observations ne sont pas disponibles. Ce type d'algorithme permet ainsi d'estimer les paramètres d'interaction du processus considéré dans sa totalité. Il donne ainsi la possibilité de réaliser leur inférence statistique à la fois dans les régions observables et les régions cachées [53, 67, 71, 199, 198, T11, T12].

La Figure 6.2 montre le schéma d'observation : seule la configuration de la région  $A$  peut être observée, alors qu'il existe probablement des structures "cachées" ou "manquantes" dans la région  $B$ . Comme la configuration des objets n'est observée que partiellement, les données sont considérées comme incomplètes. La représentation des régions  $A$  et  $B$  en deux carrés imbriqués n'est pas importante. La seule chose requise sur ces ensembles est qu'ils soient  $\nu$ -mesurables. Ceci équivaut à dire que des structures bien définies peuvent y coexister.

Soit  $p(\mathbf{x}, \mathbf{y}|\theta)$  la loi jointe de tous les objets, que l'on exprime sous la forme

$$p(\mathbf{x}, \mathbf{y}|\theta) = \frac{\exp\langle t(\mathbf{x}, \mathbf{y}), \theta \rangle}{Z(\theta)}$$

où  $\mathbf{x}$  désigne la configuration des objets observés (de  $A$ ) et  $\mathbf{y}$  celle des objets cachés (de  $B$ ). La constante de normalisation s'écrit ici

$$Z(\theta) = \int \exp\langle t(\mathbf{x}, \mathbf{y}), \theta \rangle \mu_A(d\mathbf{x}) \mu_B(d\mathbf{y})$$

où  $\mu_A$  et  $\mu_B$  désignent les restrictions à  $A$  et  $B$  de la mesure de référence poissonnienne  $\mu$  sur  $A \cup B$ .

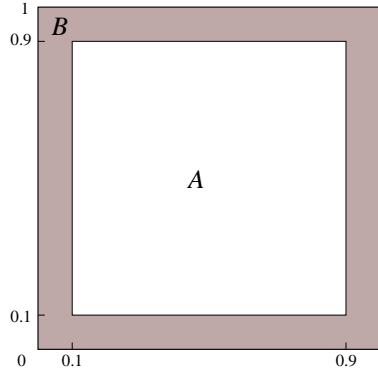


FIGURE 6.2 – Schéma d’observation dans le cas de données incomplètes : régions observée  $A$  et cachée  $B$

Dans ce contexte, la vraisemblance des paramètres est la loi marginale

$$L(\theta) = p(\mathbf{x}|\theta) = \int p(\mathbf{x}, \mathbf{y}|\theta) \mu_B(d\mathbf{y}). \quad (6.9)$$

Elle peut aussi s’écrire

$$L(\theta) = \frac{1}{Z(\theta)} \int \exp\langle t(\mathbf{x}, \mathbf{y}), \theta \rangle \mu_B(d\mathbf{y}) = \frac{Z(\theta|\mathbf{x})}{Z(\theta)}, \quad (6.10)$$

ce qui met en jeu la constante de normalisation  $Z(\theta|\mathbf{x})$  de la loi conditionnelle  $p(\mathbf{y}|\mathbf{x}, \theta)$ .

Considérons le rapport de vraisemblance par rapport à un paramètre connu  $\psi$ , puis prenons en le logarithme. Nous obtenons

$$l(\theta) = \log \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\psi)} = \log \frac{Z(\theta|\mathbf{x})}{Z(\psi|\mathbf{x})} - \log \frac{Z(\theta)}{Z(\psi)}.$$

Comme précédemment, nous exprimons les composantes de la logvraisemblance en termes d’espérance mathématiques (6.4). Ainsi

$$\frac{Z(\theta)}{Z(\psi)} = \mathbb{E} [\exp\langle t(\mathbb{X}, \mathbb{Y}), \theta - \psi \rangle] \quad (6.11)$$

et

$$\frac{Z(\theta|\mathbf{y})}{Z(\psi|\mathbf{y})} = \mathbb{E} [\exp\langle t(\mathbb{X}, \mathbb{Y}), \theta - \psi \rangle | \mathbb{X} = \mathbf{x}]. \quad (6.12)$$

Les expressions (6.11) et (6.12) ne sont pas calculables analytiquement. Mais elles peuvent être approximées par des techniques Monte Carlo, ce

qui conduit à l'expression suivante :

$$l_n(\theta) = \log \left( \frac{1}{n} \sum_{i=1}^n \exp \langle t(\mathbf{x}, \mathbb{Y}_i^*), \theta - \psi \rangle \right) - \log \left( \frac{1}{n} \sum_{i=1}^n \exp \langle t(\mathbb{X}_i, \mathbb{Y}_i), \theta - \psi \rangle \right), \quad (6.13)$$

$(\mathbb{X}_i, \mathbb{Y}_i), i = 1, 2, \dots$  désignant des échantillons de la loi jointe et  $(\mathbb{Y}_i^*)$  des échantillons de la loi conditionnelle aux objets observés  $\mathbf{x}$ .

Les résultats de convergence obtenus en approximant la vraisemblance des données complètes restent encore valables pour des données incomplètes [71]. En utilisant l'échantillonnage pondéré, le gradient de l'approximation de la logvraisemblance (6.13) s'écrit

$$\nabla l_n(\theta) = \mathbb{E} [n, \theta, \psi] t(\mathbb{X}, \mathbb{Y}) | \mathbb{X} = \mathbf{x} - \mathbb{E} [n, \theta, \psi] t(\mathbb{X}, \mathbb{Y}),$$

les espérances empiriques portant sur des échantillons simulés selon la loi de paramètre  $\psi$ .

Tous les ingrédients sont maintenant réunis pour la mise en oeuvre d'une méthode itérative d'optimisation locale. Pour les familles exponentielles, des résultats théoriques assurent que la vraisemblance augmente à chaque itération et qu'à l'infini la convergence s'établit vers l'optimum local le plus proche de la configuration initiale ([53], Thm. 1, pp. 101.). Des auteurs [53] démontrent également des résultats de type théorème centrale limite, qui devraient permettre d'évaluer la distance entre l'optimum local obtenu et l'optimum global inconnu. Cependant, il est très difficile d'affirmer quoi que ce soit sur la convexité de la vraisemblance (6.10).

# Chapitre 7

## Contributions

### 7.1 Convergence du recuit simulé pour des processus ponctuels marqués

Soit  $p(\mathbf{x}|\theta)$  la densité de probabilité

$$p(\mathbf{x}|\theta) = \frac{\exp[-U(\mathbf{x}|\theta)]}{Z(\theta)} \quad (7.1)$$

qui caractérise la forme cachée dans les données. Nous estimons cette structure en prenant le maximum de (7.1). Cette maximisation est effectuée par recuit simulé.

Le recuit simulé nécessite deux ingrédients : une méthode de simulation de la loi  $p(\mathbf{x}|\theta)$  et un schéma de décroissance de la température [69, 81, 86, 103, 189, T17]. Un premier résultat de convergence a été obtenu dans le cas des modèles définis sur des espaces discrets de configurations à nombre fini d'états [69]. Dans ce cas, le recuit simulé repose sur un échantillonneur de Gibbs et nécessite un schéma de refroidissement logarithmique. Sur le même modèle, [86] est parvenu à un résultat similaire lorsque le recuit simulé repose sur un algorithme de Metropolis-Hastings.

Le cas des processus ponctuels marqués est plus délicat, car l'espace de configuration est infini. La convergence d'un recuit simulé sur des espaces de configurations généraux a été étudiée par [81]. Le cas des processus ponctuels avec comme dynamique de simulation des processus spatiaux de naissance et de mort est traité par [189]. Ici nous présentons la convergence d'un recuit simulé pour des processus ponctuels, construit avec des dynamiques de Metropolis-Hastings adaptées [T17, T23].

**Lemme 3** *Soit  $\Omega_*$  l'ensemble de toutes les configurations maximisant  $p(\mathbf{x}|\theta)$  donnée par (7.1) et supposons  $\mu(\Omega_*) > 0$ . Soient également une suite po-*



sitive  $T_n$  telle que  $\lim_{n \rightarrow \infty} T_n = 0$  et les densités données par  $p_n(\mathbf{x}|\theta) = [p(\mathbf{x}|\theta)]^{\frac{1}{T_n}}$ . Alors, les mesures de probabilité  $\pi_n = \int p_n d\mu$  convergent en variation totale vers la distribution uniforme sur  $\Omega_*$ . En plus, les distributions  $\pi_n$  vérifient

$$\sum_{n=1}^{\infty} \|\pi_n - \pi_{n+1}\| < \infty.$$

**Preuve:** La preuve est donnée par ([189] Lemma 6 pp. 292-293, [T17] Thm. 2 pp. 1870).  $\square$

La chaîne de Markov qui simule les  $p_n$  n'est pas stationnaire. Nous avons besoin d'un outil supplémentaire pour prouver la convergence des algorithmes qui simulent des telles chaînes [69, 189, T17].

**Définition 37** Soit  $P(\cdot, \cdot)$  un noyau de transition sur  $(\Omega, \mathcal{F})$ . Le coefficient de contraction de Dobrushin est

$$c(P) = \sup_{\mathbf{x}, \mathbf{y} \in \Omega} \|P(\mathbf{x}, \cdot) - P(\mathbf{y}, \cdot)\|.$$

Dans ce qui suit, nous désignons par  $P_n^\delta$ , l'application itérée  $\delta$  fois du noyau de transition  $P_n$ . Le noyau  $P_n$  est défini par les mises à jour d'un algorithme Metropolis-Hastings avec des naissances uniformes et adaptées (5.1) et des morts uniformes (4.17). Ce noyau admet  $\pi_n$  comme loi invariante.

**Théorème 21** Soit  $(X_n)_{n \geq 0}$  un processus de Markov non homogène sur  $(\Omega, \mathcal{F})$  défini par les noyaux de transition  $P_n^\delta(\mathbf{x}, F)$  et considérons les conditions suivantes

$$\sum_{n=1}^{\infty} \|\pi_n - \pi_{n+1}\| < \infty \tag{7.2}$$

et

$$\lim_{n \rightarrow \infty} c(P_{n_0, n}) = 0 \tag{7.3}$$

respectées pour tous les  $n_0 \geq 0$  et les probabilités de transition  $P_{n_0, n}(\mathbf{x}, F) = \mathbb{P}(X_n \in F | X_{n_0} = \mathbf{x})$ . Alors, la limite  $\lim_{n \rightarrow \infty} \pi_n = \pi_\infty$  existe et la convergence  $\lim_{n \rightarrow \infty} \pi_0 P_{0, n} = \pi_\infty$  a lieu en variation totale, pour toute loi initiale  $\pi_0$ .

**Preuve:** Par construction,  $\pi_n$  est la loi invariante de  $P_n^\delta$ . La condition (7.2) résulte du Lemme 3. Si la condition (7.3) est aussi vérifiée, alors les énoncés ([189] Thm. 2 pp. 285-286, [T17] Thm. 3 pp. 1871) achèvent la démonstration.

□

La condition de Dobrushin (7.3) peut s'obtenir au moyen du résultat suivant, dont la preuve est d'intérêt pédagogique, car elle peut aussi servir à déterminer un schéma de refroidissement.

**Lemme 4** *Soit  $P_n^\delta(\mathbf{x}, F)$  le noyau de transition défini précédemment. Si le pas  $\delta$  est un entier fini, tel que  $n(\mathbf{x}) \leq \delta$  pour tout  $\mathbf{x} \in \Omega$ , alors le coefficient de Dobrushin vérifie la condition (7.3).*

**Preuve:** En utilisant les propriétés du coefficient de Dobrushin et d'après ([189] Lemma 3 pp. 287), il suffit de vérifier que

$$c(P_n) \leq 1 - \frac{1}{n}.$$

Nous utilisons les équations de Kolmogorov-Chapman pour calculer la probabilité  $P_n^\delta(\mathbf{x}, \Xi_0)$  de transiter d'une configuration quelconque  $\mathbf{x} \in \Omega$  à la configuration vide  $\Xi_0$  en  $\delta$  pas. Nous obtenons

$$P_n^\delta(\mathbf{x}, \Xi_0) \geq P_n^{n(\mathbf{x})}(\mathbf{x}, \Xi_0) P_n^{\delta-n(\mathbf{x})}(\emptyset, \Xi_0) \geq \left( \frac{p_b}{\nu(W)\Lambda^{1/T_n}} \right)^\delta = \Delta_n^\delta, \quad (7.4)$$

le terme  $\Delta_n$  étant calculé à partir d'une borne de probabilité d'acceptation d'une mort (4.14).

En développant cette inégalité (7.4) nous avons

$$P_n^\delta(\mathbf{x}, A) \geq (\Delta_n)^\delta \mathbf{1}\{A = \Xi_0\} \quad (7.5)$$

ainsi que

$$P_n^\delta(\mathbf{x}, A) \leq 1 - (\Delta_n)^\delta \mathbf{1}\{A \subseteq \Omega \setminus \Xi_0\}. \quad (7.6)$$

Nous regroupons les équations (7.5) et (7.6) pour obtenir

$$|P_n^\delta(\mathbf{x}, A) - P_n^\delta(\mathbf{y}, A)| \leq 1 - (\Delta_n)^\delta [\mathbf{1}\{A = \Xi_0\} + \mathbf{1}\{A \subseteq \Omega \setminus \Xi_0\}].$$

Si  $A = \Xi_0$  ou bien si  $A \subseteq \Omega \setminus \Xi_0$ , l'inégalité (7.1) devient

$$|P_n^\delta(\mathbf{x}, A) - P_n^\delta(\mathbf{y}, A)| \leq 1 - (\Delta_n)^\delta. \quad (7.7)$$

Si maintenant  $A = \Xi_0 \cup A_1$  avec  $A_1 \subseteq \Omega \setminus \Xi_0$ , il vient :

$$\begin{aligned} P_n^\delta(\mathbf{x}, A) &= \\ &= P_n^\delta(\mathbf{x}, \Xi_0) + P_n^\delta(\mathbf{x}, A_1) \\ &= P_n^\delta(\mathbf{x}, \Xi_0) + 1 - P_n^\delta(\mathbf{x}, \Xi_0) - P_n^\delta(\mathbf{x}, \Omega \setminus \Xi_0 \setminus A_1) \\ &= 1 - P_n^\delta(\mathbf{x}, B) \end{aligned}$$

où  $B = (\Omega \setminus \Xi_0 \setminus A_1) \subseteq (\Omega \setminus \Xi_0)$ . Il en résulte

$$|P_n^\delta(\mathbf{x}, A) - P_n^\delta(\mathbf{y}, A)| = |P_n^\delta(\mathbf{y}, B) - P_n^\delta(\mathbf{x}, B)| \leq 1 - (\omega_n)^\delta. \quad (7.8)$$

Les équations (7.7) et (7.8) donnent pour le coefficient de Dobrushin

$$c(P_n) \leq 1 - (\omega_n)^\delta.$$

Un schéma de refroidissement est ainsi trouvé si

$$(\Delta_n)^\delta = \left( \frac{pb}{\nu(W)\Lambda^{1/T_n}} \right)^\delta \geq \frac{1}{n}$$

ce qui donne, sous de faibles hypothèses

$$T_n \geq \frac{\delta \log \Lambda}{\log(n) + \delta \log \frac{pb}{\nu(W)}}, \quad (7.9)$$

quand  $n \rightarrow \infty$ . □

En prenant toutes les précautions nécessaires à une démarche scientifique, il est maintenant possible de répondre algorithmiquement à la question de la forme cachée dans les données.

**Algorithme 11**  $\mathbf{x} = \text{Recuit Simulé}(T_0, \delta, T)$

1. choisir une condition initiale  $\mathbf{x}_0$

2. pour  $i = 1$  jusqu'au  $T$  faire

$$\left\{ \begin{array}{l} \mathbf{x}_i = \text{miseAJour}(\mathbf{x}_{i-1}, T_{i-1}, \delta) \\ T_i = T_0 / [\log(i) + 1] \end{array} \right\}$$

3. sortir  $\mathbf{x}_T$ .

Cet Algorithme 11 a pour entrée une température initiale et le nombre maximal d'objets d'une configuration. En théorie, ces paramètres peuvent être calculés à partir du modèle de processus ponctuel utilisé. Cette tâche n'est pas toujours facile. Elle peut même amener à des valeurs qui engendrent un temps de calcul très important. En pratique, un compromis qui respecte le principe du recuit simulé dans ses lignes générales est considéré comme acceptable, pourvu, bien sûr, que les données soient de suffisamment bonne qualité. Les mises à jour se font par une dynamique Metropolis-Hastings adaptée, décrite par l'Algorithme 2 avec les propositions de naissance et de mort comme vu antérieurement.

Il faut aussi mentionner que l'ingrédient le plus important de cet algorithme, peu visible dans cette formulation algorithmique, la densité de probabilité  $p_d(\mathbf{x}|\theta)$  du modèle de forme que l'on souhaite détecter.

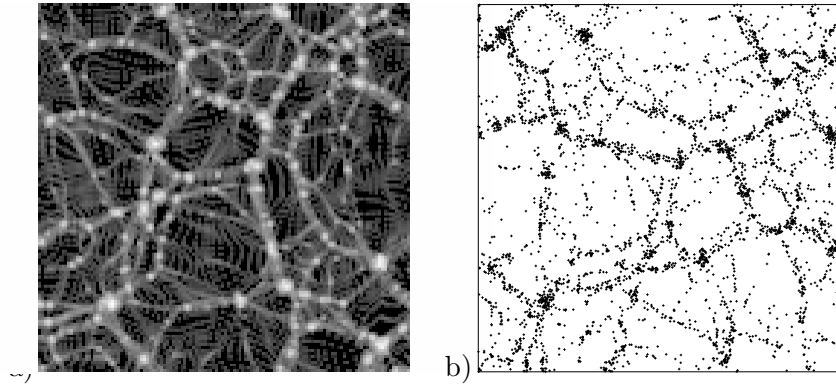


FIGURE 7.1 – a) Une carte de densité de matière noire ; b) La réalisation d'un champ de galaxies associées. L'unité de distance utilisée est le  $h^{-1}\text{Mpc}$  (megaparsec). Un parsec vaut 3,2616 années-lumière.

### 7.1.1 Détection de filaments galactiques : premiers pas

Nous présentons ici les premiers résultats de détection de filaments galactiques. Tous les détails mathématiques concernant l'implémentation, l'analyse des résultats, ainsi qu'une comparaison avec d'autres méthodes figurent dans [T20]. Nous ne montrons que l'application de la méthode proposée.

La procédure pour obtenir les données est décrite dans [116]. Un champ aléatoire décrivant la répartition spatiale de matière noire évolue dans le temps et fait apparaître des structures. La Figure 7.1a montre la densité de matière noire obtenue dans un tel champ. Les régions de faible densité sont appelées des vides. Les structures filamentaires sont des régions de forte densité qui délimitent des "cellules" vides. Le champ de galaxies est obtenu par un processus de Cox : conditionnellement à la densité de matière noire, les galaxies sont distribuées majoritairement dans les régions filamentaires, très peu d'entre elles se trouvant dans les régions vides.

Pour détecter les filaments, nous avons utilisé le modèle Candy, qui a déjà fait ses preuves en analyse d'image [T1, T13, T15]. Notre travail repose sur l'hypothèse que les filaments peuvent être approximés par de fins rectangles qui se connectent et s'alignent pour former un réseau. Par commodité, nous les appelons segments. Leur répartition spatiale est régie par la densité de probabilité d'un processus ponctuel marqué. Un estimateur du réseau s'obtient en maximisant cette densité de probabilité par recuit simulé.

Le recuit simulé a été implémenté selon une dynamique Metropolis-Hastings adaptée [T7, T20]. Au moment de ces travaux, nous n'étions pas encore en

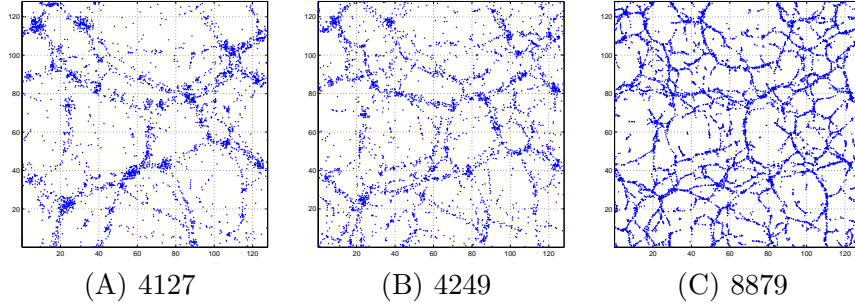


FIGURE 7.2 – Trois catalogues de galaxies bidimensionnelles obtenus par simulation contenant et le nombre de galaxies correspondant.

possession du résultat de convergence pour le recuit. Nous savions qu’un schéma logarithmique de refroidissement avait été établi dans des systèmes discrets, et nous l’avons d’abord testé sur notre méthode. Cependant, comme les données étaient jugées de très bonne qualité, et comme nous ne souhaitions que vérifier le principe d’application de la méthode, les résultats ont par la suite été obtenus par un recuit selon un schéma de refroidissement polynomial

$$T_{k+1} = cT_k,$$

$c$  étant une constante positive très proche de 1. L’algorithme a tourné pendant  $10^7$  itérations, la température étant refroidie toutes les  $10^3$  itérations.

Les données sur lesquels notre test a porté sont montrées Figure 7.2. Chaque jeu de données ou catalogue contient un nombre de galaxies différent formant un réseau ayant une topologie à part. Les paramètres du modèle Candy doivent être réglés avec soin pour obtenir une bonne représentation des filaments dans les données. Les marques ou les paramètres des segments ont été choisis afin que le modèle puisse détecter des filaments dans chaque jeu de données. Pour les catalogues (A) et (B), la longueur des segments a été choisie entre  $l_{\min} = 3$  et  $l_{\max} = 5$ , pour une épaisseur variant entre  $w_{\min} = 1$  et  $w_{\max} = 2$ . Pour le catalogue (C), des segments plus petits ont été utilisés  $l_{\min} = 2$ ,  $l_{\max} = 3$ ,  $w_{\min} = 0.95$ ,  $w_{\max} = 1.05$ . Pour les trois jeux de données, la région de connexité autour des segments a été défini par des cercles de même rayon  $r_c = 0.5$ . La courbure définissant l’orientation relative de deux segments adjacents a été fixé à  $\tau = 0.5$  radians pour les catalogues (A) et (B), et à  $0.75$  radians pour le catalogue (C).

Concernant les paramètres d’attache aux données et d’interaction, respectivement décrits par (3.8) et (3.6), plusieurs valeurs ont été testées. Le tableau 7.1 donne les valeurs avec lesquelles nos résultats sont été obtenus. Par ce choix, nous avons favorisé les segments connectés aux deux extrémités,

pénalisé fortement les segments libres, alors que les segments connectés à une seule extrémité sont pénalisés d'une manière moins forte. L'alignement relatif de deux segments a été également encouragé. Tout segment dont l'attache aux données ne passe pas les tests locaux n'est pas interdit mais une très faible probabilité lui est assignée. Par ce mécanisme, le réseau peut être prolongé dans des régions où les données sont de qualité moindre, mais où la configuration générale du réseau pourrait le suggérer.

TABLE 7.1 – Les paramètres d'interaction

Paramètres	Catalogue		
	A	B	C
$\log \gamma_2$	10	10	10
$\log \gamma_0$	-8	-7	-7
$\log \gamma_1$	-2	-2	-1
$\log \gamma_o$	-3	-3	-3
$u_{\max}$	-25	-25	-25

La Figure 7.3 montre les résultats obtenus. Selon l'analyse faite dans [T20], le meilleur résultat a été obtenu pour tous les catalogues avec le premier jeu de paramètres. Tous les jeux de paramètres détectent correctement les filaments que l'on peut trouver après inspection visuelle de données. Les agrégats de galaxies ne détruisent pas la construction de filaments. Les difficultés apparaissent dans les régions à plus faible concentration de galaxies. Cependant, la construction du réseau apparaît toujours comme naturelle du point de vue des cosmologues, qui ont également comparé notre démarche avec les techniques en cours à l'époque.

Ce travail a été le point de départ de ma coopération avec les amis cosmologues. Il soulève de nombreuses questions, dont on va parler dans la suite. Il a été également la première source de motivation pour la construction du modèle Bisous.

## 7.2 Maximum de vraisemblance pour le modèle Candy

L'application présentée au paragraphe précédent montre l'influence des paramètres d'interaction sur le résultat final. Dans cette partie, nous regardons le problème suivant : une structure est formée par des objets en interaction que nous supposons gouvernés par un processus ponctuel marqué. Nous sou-

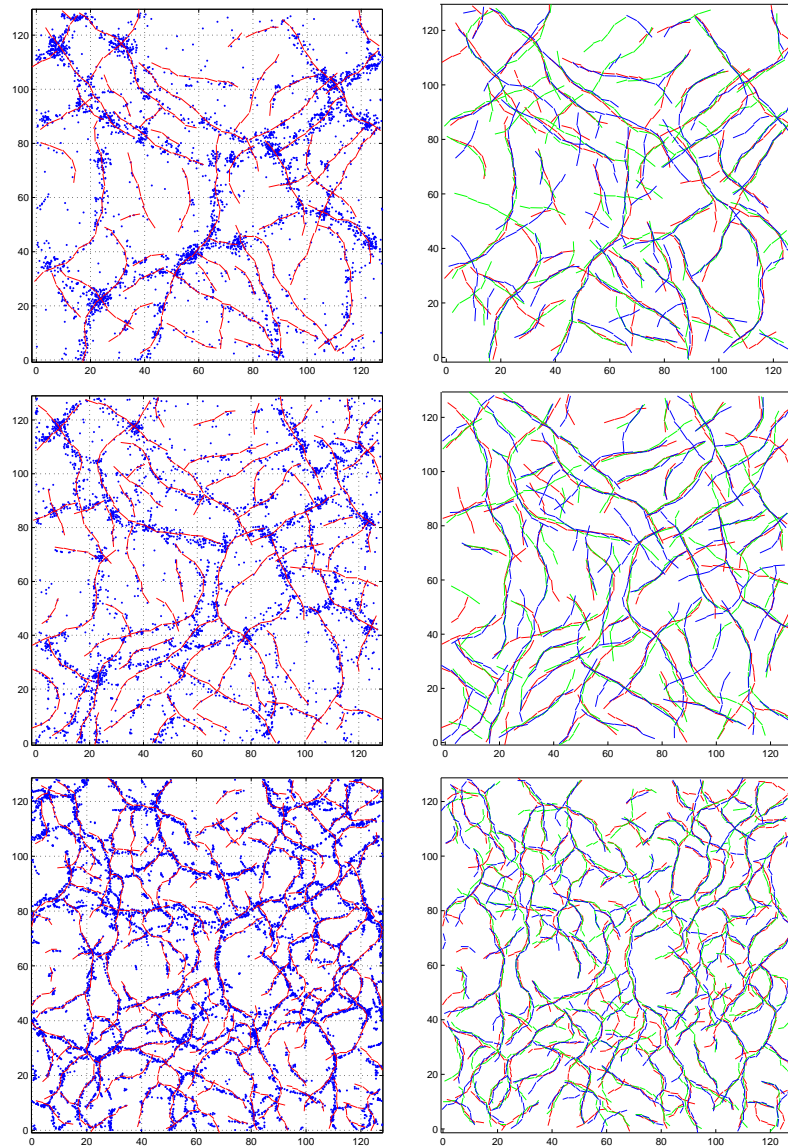


FIGURE 7.3 – Résultats obtenus pour les trois catalogues : A (en haut), B (au milieu) et C (en bas). A gauche, le "meilleur" résultat est superposé aux données. A droite on a superposé tous les résultats obtenus avec les jeux de paramètres du Tableau 7.1. Le premier jeu de paramètres est représenté en rouge, le second en bleu et le troisième en vert.

haitons estimer les paramètres d'interaction entre objets.

Nous présentons ici les résultats obtenus sur l'estimation des paramètres d'interaction du modèle Candy [T3, T7].

Le domaine d'observation est une partie compacte  $W \subset \mathbb{R}^2$  de volume strictement positif  $0 < \nu(W) < \infty$ . Un segment est un objet de centre  $w$  dans  $W$  et de caractéristiques  $m = (l, \theta)$ , longueur et orientation, dans l'espace de probabilité  $M = [l_{\min}, l_{\max}] \times [0, \pi)$ . La distribution des marques  $\nu_M$  est la loi uniforme sur  $M$ . Soit  $\mathbf{x} = \{(w_1, m_1), \dots, (w_n, m_n)\}$  une configuration finie de segments. Le modèle Candy est défini par la densité de probabilité par rapport à la mesure de référence poissonnienne (2.4)

$$p(\mathbf{x}) = p(\emptyset) \prod_{i=1}^n \exp \left[ \frac{l_i - l_{\max}}{l_{\max}} \right] \times \gamma_0^{n_0(\mathbf{x})} \gamma_1^{n_1(\mathbf{x})} \gamma_2^{n_2(\mathbf{x})} \gamma_r^{n_r(\mathbf{x})} \gamma_o^{n_o(\mathbf{x})} \quad (7.10)$$

où  $p(\emptyset)$  est la constante de normalisation, et où  $\gamma_2, \gamma_1, \gamma_0 > 0$  et  $\gamma_r, \gamma_o \in (0, 1)$  sont les paramètres d'interaction du modèle. Les statistiques exhaustives du modèle  $n_2(\mathbf{x}), n_1(\mathbf{x}), n_0(\mathbf{x}), n_r(\mathbf{x}), n_o(\mathbf{x})$  sont respectivement le nombre de segments connectés aux deux extrémités, le nombre de segments connectés à une seule extrémité, le nombre de segments non connectés, le nombre de paires de segments en répulsion et le nombre de paires de segments non alignés. Tous les détails concernant la définition des interactions et les propriétés du modèle Candy se trouvent dans [T7].

Pour ce problème d'estimation, on suppose les paramètres  $l_{\min}, l_{\max}, r_a$  et  $\tau$  connus. Cette hypothèse peut être débattue, car l'estimation des paramètres de portée des interactions est un problème encore ouvert. Dans ce contexte cependant, il n'est pas déraisonnable d'imaginer ces paramètres estimés par des techniques de statistique exploratoire ou bien d'analyse d'image [90, 141, 180].

Les modèles Candy et Bisous, comme la plupart des modèles avec lesquels nous avons travaillé, font partie de la famille exponentielle (6.2). La dynamique de simulation de ces modèles est construite à partir d'une chaîne de Markov irréductible, Harris récurrente et ergodique géométrique. Par conséquent, le cadre d'estimation de paramètres par maximum de vraisemblance peut être appliqué.

Ecrivons la densité de probabilité (7.10) sous la forme

$$p(\mathbf{x}|\theta) = \frac{\exp\langle t(\mathbf{x}), \theta \rangle h(\mathbf{x})}{Z(\theta)}$$



où  $t(\mathbf{x})$  est le vecteur de statistiques canoniques exhaustives

$$t(\mathbf{x}) = (n_2(\mathbf{x}), n_1(\mathbf{x}), n_0(\mathbf{x}), n_r(\mathbf{x}), n_o(\mathbf{x})),$$

$\theta$  le vecteur de paramètres

$$\theta = (\log \gamma_2, \log \gamma_1, \log \gamma_0, \log \gamma_r, \log \gamma_o)^T$$

et  $h(\mathbf{x}) = \prod_{i=1}^n \exp \left[ \frac{l_i - l_{\max}}{l_{\max}} \right]$ .

La procédure de maximisation nécessite les valeurs observées de  $t(\mathbf{x})$  et une valeur initiale de  $\theta_0$ . La logvraisemblance étant convexe, l'optimisation est en principe indépendante des conditions initiales. De plus, la simulation du modèle se fait à l'aide d'une chaîne de Markov qui possède de bonnes propriétés de convergence. En conséquence, les approximations Monte Carlo approchent très bien les valeurs théoriques et un théorème de la limite centrale permet d'évaluer les erreurs d'estimation. Cependant, les évaluations numériques par échantillonnage pondéré de la logvraisemblance et du gradient ne sont pas numériquement stables. En effet, ces calculs passent par l'évaluation d'expressions du type

$$\exp\langle t(X), \theta_k - \theta_{k+1} \rangle.$$

Si la distance entre la nouvelle et l'ancienne valeur des paramètres est trop grande, alors les valeurs numériques de (7.2) peuvent exploser.

La solution à ce problème est de rééchantillonner le modèle. En itérant cette procédure, on peut arriver suffisamment près de l'estimateur du maximum de vraisemblance et alors le calculer par une méthode directe ou de type Newton-Raphson. La méthode itérative que nous présentons est une méthode de gradient à pas optimal [57, T3, 179, T7].

**Algorithme 12** *Méthode itérative avec gradient à pas optimal pour l'approximation du MLE*

1. Initialiser  $\theta_1$  et  $k = 1$ .
2. Générer  $m$  échantillons de la loi  $p(\mathbf{x}|\theta_k)$  et calculer  $\nabla l_m(\theta_k)$ .
3. Pour chaque composante des paramètres  $i = \{1, \dots, 5\}$  et chaque composante du gradient  $\Delta_i$ , calculer les intervalles  $I_k^i = [\theta_k^i - \alpha \Delta_i, \theta_k^i + \alpha \Delta_i]$ ,  $\alpha > 0$  étant un paramètre scalaire de précision.
4. Maximiser la logvraisemblance dans chaque intervalle  $I_k^i$  par la méthode de la section dorée pour obtenir une nouvelle valeur  $\theta_{k+1}$ .
5. Si  $\|\theta_{k+1} - \omega_k\| > \tau_1$ , alors poser  $k = k + 1$  et aller au pas 2;  $\tau_1$  est une valeur seuil préfixée.

Valeurs initiales	Méthode du gradient	Monte Carlo MLE
$\log \gamma_0^i = -9.5$	$\widehat{\log \gamma_0^0} = -8.37$	$\widehat{\log \gamma_0^n} = -8.32$
$\log \gamma_1^i = -4.0$	$\widehat{\log \gamma_s^0} = -2.74$	$\widehat{\log \gamma_1^n} = -2.73$
$\log \gamma_2^i = 1.5$	$\widehat{\log \gamma_d^0} = 2.46$	$\widehat{\log \gamma_2^n} = 2.47$
$\log \gamma_o^i = -3.5$	$\widehat{\log \gamma_o^0} = -2.13$	$\widehat{\log \gamma_o^n} = -2.17$
$\log \gamma_r^i = -3.5$	$\widehat{\log \gamma_r^0} = -2.42$	$\widehat{\log \gamma_r^n} = -2.42$

TABLE 7.2 – Estimation des paramètres pour les données de la Figure 7.4.

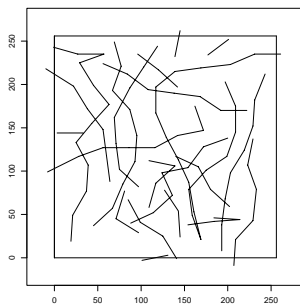
Ecart-type asymptotique	Ecart-type Monte Carlo
0.51	0.002
0.23	0.003
0.17	0.001
0.30	0.002
0.31	0.005

TABLE 7.3 – Erreurs d'estimation.

6. Si  $\|\nabla l_m(\theta_{k+1}) - \nabla l_m(\theta_k)\| > \tau_2$ , alors poser  $k = k + 1$  et aller au pas 3, sinon arrêter l'algorithme ;  $\tau_2$  est une valeur seuil préfixée.

La Figure 7.4 montre une réalisation du modèle Candy, ses paramètres d'interaction et les valeurs observées des statistiques exhaustives. L'Algorithme 12 a été implémenté en prenant comme données les statistiques observées de la Figure 7.4 et comme paramètres initiaux  $\theta_1$  ceux du Tableau 7.2 (première colonne). L'algorithme a tourné pendant 1000 itérations et il a utilisé  $\alpha = 10^{-3}$ ,  $\tau_1 = 3.0$  et  $\tau_2 = 10^{-6}$ . Le résultat obtenu est le vecteur de paramètres  $\theta_0$  qui figure dans la colonne du milieu du Tableau 7.2. La log-vraisemblance  $l_n(\theta)$  a été calculée à partir des  $n = 2 \times 10^4$  simulations du modèle Candy selon la loi  $p(\mathbf{x}|\theta_0)$ . Des profils de la logvraisemblance sont montrés Figure 7.5. Le maximum de  $l_n(\theta)$  est affiché en troisième colonne du Tableau 7.2.

L'écart-type asymptotique de l'estimateur du maximum de vraisemblance par rapport aux paramètres du modèle (inconnus en pratique) et l'erreur standard Monte-Carlo de l'approximation du maximum de vraisemblance par rapport au vrai maximum de vraisemblance sont donnés au Tableau 7.3.



Paramètres du modèle
$\log \gamma_0 = -8.5$
$\log \gamma_1 = -3$
$\log \gamma_2 = 2.5$
$\log \gamma_o = -2.5$
$\log \gamma_r = -2.5$

Statistiques exhaustives
$n_0 = 4$
$n_1 = 34$
$n_2 = 63$
$n_o = 12$
$n_r = 9$

FIGURE 7.4 – Réalisation du modèle Candy (en haut), ses paramètres d'interaction (tableau au milieu) et les valeurs observées des statistiques exhaustives (tableau du bas).

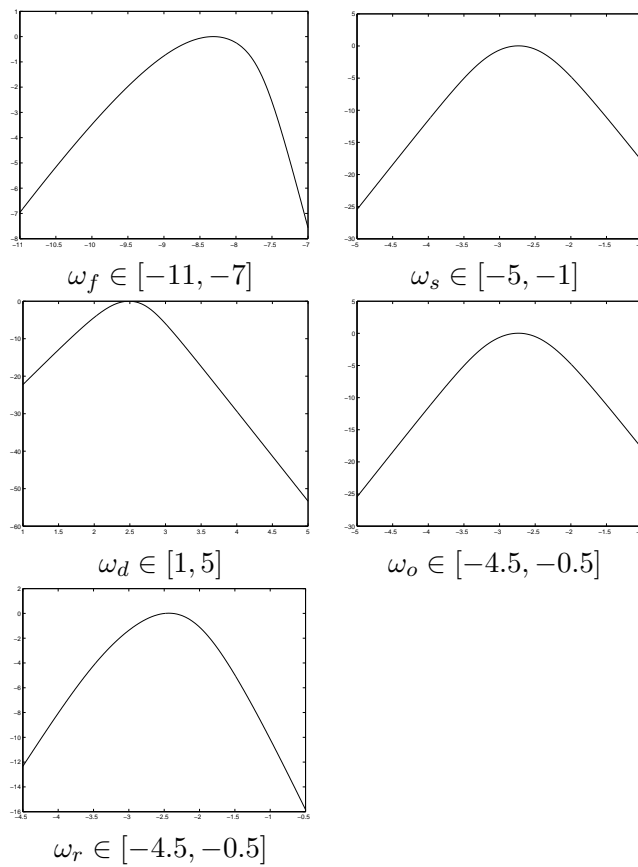


FIGURE 7.5 – Approximation Monte Carlo de la logvraisemblance en fonction des composantes individuelles, les autres composantes de  $\hat{\gamma}^0$  étant fixées.

### 7.3 Approche bayésienne pour la détection des structures

La détection de structures selon l'approche précédente demande la connaissance a priori des paramètres. Nous ne pouvons pas estimer ces paramètres par maximum de vraisemblance (par exemple), car cela requiert la connaissance de la structure que l'on cherche à détecter. La solution que nous avons choisie a été de fixer les paramètres au moyen de quelques critères sur la forme recherchée, puis de valider les résultats obtenus.

Pour détecter les filaments de données simulées, les paramètres ont été fixés par deux critères. D'abord la topologie du réseau : nous souhaitons obtenir des filaments avec beaucoup de segments connectés en leurs deux extrémités, suffisamment de segments connectés en une seule extrémité et très peu de segments isolés. La courbure du réseau a été également contrôlée pour obtenir des filaments "lisses" et se croisant préférentiellement à angle droit. Ensuite, ces paramètres de structure ont été ajustés par rapport aux paramètres d'attache aux données. Cela évite de détecter des filaments là où il n'y a pas, et permet de compenser la faible qualité de données par une éventuelle prolongation du réseau. Plusieurs essais ont permis de choisir des paramètres jugés visuellement satisfaisants.

Cette approche, suffisante au moment où ces recherches se déroulaient, a permis de révéler un certain nombre de nouveaux problèmes. L'un de ces problèmes est que l'inspection visuelle des résultats de détection n'est bien souvent pas possible. Il en est ainsi des catalogues de galaxies tridimensionnelles et des données d'épidémiologie animale concernant la répartition des élevages laitiers en France.

Il est alors commode de recourir à une approche bayésienne [T16, T21, T22, T23, T11, T12]. Cela revient à considérer la loi conjointe de la forme et des paramètres

$$p(\mathbf{x}, \theta) = p(\mathbf{x}|\theta)p(\theta), \quad (7.11)$$

$p(\theta)$  étant la loi a priori des paramètres et  $p(\mathbf{x}|\theta)$  la loi de la forme donnée par la densité de probabilité d'un processus ponctuel marqué.

La solution s'obtient en maximisant (7.11). Elle fournit simultanément la forme et les paramètres qui s'écrivent

$$\begin{aligned} (\hat{\mathbf{x}}, \hat{\theta}) &= \arg \max_{(\mathbf{x}, \theta) \in \Omega \times \Theta} p(\mathbf{x}|\theta)p(\theta) \\ &= \arg \max_{(\mathbf{x}, \theta) \in \Omega \times \Theta} \frac{\exp [U_d(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta) + U_p(\theta)]}{Z(\theta)Z_p}, \end{aligned} \quad (7.12)$$

en posant  $p(\theta) = \exp[-U_p(\theta)]/Z_p$ .

Cette manière de procéder n'élimine pas le problème d'estimation des paramètres, car nous devons choisir la loi a priori. Néanmoins, elle permet d'attribuer des plages de variation plutôt que des valeurs fixes aux paramètres de la forme, en laissant la machinerie algorithmique faire son travail.

Selon la forme de  $p(\theta)$  la solution (7.13) aura un caractère aléatoire plus ou moins prononcé. La structure détectée contient une partie "moyenne" qui est due à la robustesse des données, au modèle et à l'approche en général, ainsi qu'une partie "variance" qui est due aux fluctuations induites par ces mêmes facteurs. Nous souhaitons renforcer la partie "moyenne" et réduire la partie "variance". Dans ce qui suit, nous présentons une approche qui permet la visualisation de la "moyenne" de la forme.

Soit  $\mathcal{T}$  un compact de  $W$ . La forme aléatoire  $X$  intersecte la région  $\mathcal{T}$  avec la probabilité

$$\mathbb{P}(\mathcal{T} \cap \mathbb{X} \neq \emptyset) = \mathbb{E}[\mathbf{1}\{\mathcal{T} \cap \mathbb{X} \neq \emptyset\}]. \quad (7.13)$$

Cette probabilité de contact peut être approximée par des méthodes Monte Carlo

$$\hat{\mathbb{P}}(\mathcal{T} \cap \mathbb{X} \neq \emptyset) = \frac{1}{U} \sum_{u=1}^U \mathbf{1}\{\mathcal{T} \cap X_u \neq \emptyset\}, \quad (7.14)$$

où  $X_1, X_2, \dots, X_U$  sont des réalisations de formes aléatoires générées selon  $p(\mathbf{x}, \theta)$ .

Considérons maintenant une grille finie qui partitionne le domaine  $W$  en cellules de même volume. Pour chaque cellule, nous pouvons calculer (7.13) en utilisant (7.14) et ne considérer que les cellules qui ont la probabilité de contact supérieure à une valeur critique donnée. Par cette opération de seuillage, nous obtenons une carte des probabilités de contact ou des régions qui sont le plus souvent "visitées" par le modèle de structure. Cette carte constitue en effet une moyenne visuelle de la forme cachée dans les données.

### 7.3.1 Détection d'agrégats en épidémiologie animale : obtention d'une solution

Nous avons utilisé cette méthodologie pour détecter des agrégats dans des données d'épidémiologie animale [T16]. Ces données sont présentées dans la partie 3.2. Dans une fenêtre  $W$ , nous observons une population de points qui sont les positions des fermes sur le territoire français et leur score de susceptibilité de maladie. Un agrégat est vu comme une région dans laquelle un ensemble de fermes géographiquement proches, ont tendance à avoir un score

important. L'ensemble des agrégats est approximé par une configuration de disques de rayon aléatoire que l'on modélise par la densité de probabilité d'un processus ponctuel marqué  $p(\mathbf{x}|\theta)$ . Ce processus spécifie le positionnement des disques dans le champ de données ainsi que leur position relative au moyen d'un terme d'énergie d'attache aux données (3.10) et d'un terme d'énergie d'interaction (3.11).

La fenêtre d'observation est le rectangle minimum qui contient la position de toutes les fermes, c'est à dire  $W = [0, 317] \times [0, 318]$ . Le rayon des disques est choisi dans l'intervalle  $[1, 10]$ . Dans les deux cas, l'unité de longueur représente  $3km$  dans le monde réel.

L'énergie d'attache aux données (3.10) est construite en effectuant un test de Student sur le score des maladies des fermes couvertes par un disque donné. Ce test vérifie si la moyenne des scores de ces fermes est plus grande qu'une valeur seuil préfixée  $d_0 = 3.75$ . Un disque doit également couvrir un nombre minimum de fermes  $n_0 = 4$ . En général, les disques de grand rayon ont moins de chance d'être acceptés que ceux de petit rayon. Ceci s'explique par la nature de la maladie (la mammite bovine) qui fait que les disques recouvrant un grand nombre de fermes ont beaucoup plus de chance d'être rejetés par le test de Student. Les grands disques sont plus souvent acceptés s'ils se situent dans une région déjà couverte de petits disques. En même temps, la trop grande présence de petits disques augmente le coût calculatoire de la méthode. L'utilisation simultanée de grands et petits disques diminue les effets de lissage. Il y a donc un compromis à trouver. A défaut d'avoir une connaissance a priori de ce paramètre, nous lui avons attribué une loi uniforme.

Le vecteur  $\theta$  ne contient que des paramètres de l'énergie d'interaction (3.11). Nous avons  $\theta = (\log \gamma_a, \log \gamma_o)$  défini sur l'espace des paramètres  $\Theta = [0, 0.5] \times [-0.1, 0]$ . Le paramètre du modèle d'interaction par aire est fixé afin que les grands disques isolés soient fortement pénalisés. Le paramètre du modèle de Strauss n'est pas trop fort, permettant ainsi aux disques de se superposer. Cependant, si le nombre de superpositions de disques est important, tout nouveau disque sera fortement pénalisé. Là encore, nous n'avons pas de connaissance a priori sur le poids de chacune de ces composantes. Par conséquent, nous leur attribuons à chacun une loi a priori uniforme  $p(\theta)$ .

Le recuit simulé a été lancé pour maximiser  $p(\mathbf{x}, \theta)$  et ainsi détecter les agrégats d'un jeu de données représentant un ensemble de fermes en 1996. La Figure 7.6a montre le résultat obtenu. La configuration des disques fait apparaître un agrégat central au milieu de la France et quelques agrégats plus petits autour.

La position des fermes est donnée par le centre de la commune à laquelle les fermes appartiennent. Un point sur la carte peut donc représenter la position de plusieurs fermes. Par conséquent, l'interprétation visuelle du résultat n'est pas possible. Cet argument ainsi que la nature de la solution proposée justifient le calcul de probabilités de contact pour vérifier la "fiabilité" des agrégats. Pour cela, la fenêtre d'observation a été divisée en cellules de même surface  $9\text{km}^2$ . Pour chaque cellule la probabilité (7.14) a été calculée. Nous obtenons ainsi une carte des visites du modèle qui est représentée en Figure 7.6b.

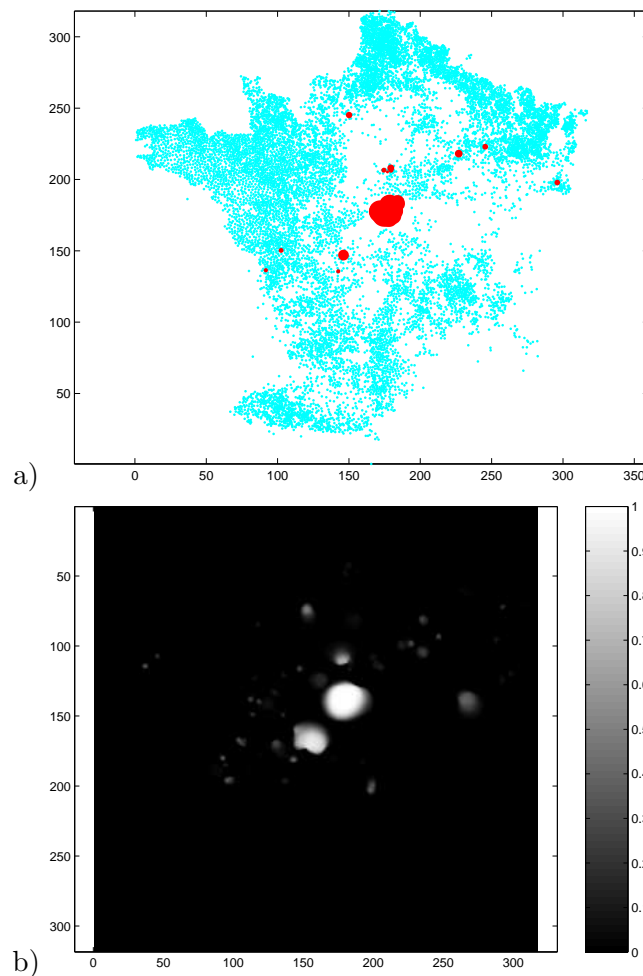


FIGURE 7.6 – a) Configuration de disques (zones rouges) indiquant la présence d'agrégats dans les données de l'année 1996 : les fermes sont réparties sur presque tout le territoire de la France (région bleue); b) Les régions les plus visitées par le modèle indiquant la "force" de chaque agrégat.



Le résultat de la détection et la carte des visites du modèle sont conjointement interprétés. Les petits agrégats sont peu importants car leur probabilité de contact varie entre 0.3 et 0.6. Une exception notable est l'agrégat au point (150, 150) qui a une probabilité de contact supérieure à 0.9. Dans cette optique, il résulte que l'agrégat le plus important est celui du point (175, 175).

Les cartes des visites du modèle réduisent donc le "bruit" de notre méthode et permettent de se prononcer plus facilement sur la qualité de la détection. Cependant cette technique n'est pas une méthode de validation à proprement parler car nous sommes toujours sous l'hypothèse du modèle et tous les calculs sont locaux. A ce stade, cette technique n'est juste qu'un outil efficace de visualisation.

### 7.3.2 Détection de filaments galactiques : obtention d'une solution

La coopération avec les amis cosmologues a continué [T21, T22, 184, T25]. Nous montrons ci-dessous les premiers résultats obtenus sur des données réelles, issues de catalogues de galaxies observées [T21].

Notre méthode a été testée sur des échantillons du catalogue 2dFGRS. Un tel échantillon est montré Figure 7.7a. Ce catalogue contient 7588 galaxies situées dans un volume rectangulaire  $W$  de dimensions  $88.6 \times 169.1 \times 20.7$ . L'unité de longueur est le mégaparsec  $h^{-1}\text{Mpc}$ . Un parsec  $1pc$  est la distance égale approximativement à 3.261 années lumière, alors que  $h \approx 0.71$  est une constante sans dimension appelée paramètre de Hubble. Un point de la Figure 7.7a représente la position d'une galaxie. A titre indicatif, le diamètre d'une galaxie est en moyenne 30 kpc.

Les filaments galactiques sont approximés par des cylindres. Ces cylindres se positionnent dans le champ de galaxies dans les régions où les galaxies ont tendance à s'aligner. De plus, les cylindres peuvent se connecter en formant des filaments. Ces propriétés sont intégrées dans la densité de probabilité d'un processus ponctuel marqué  $p(\mathbf{x}|\theta)$ . Le positionnement des cylindres dans le champ de galaxies est le terme d'attache aux données ; il est décrit par (3.8). Les relations entre cylindres forment le terme d'interaction ; il est décrit par (3.6).

Les dimensions des cylindres sont  $r = 0.5$  et  $h = 6.0$ . Le rayon d'attraction permettant la connexion des cylindres est  $r_a = 0.5$  et les paramètres de courbure définissant l'alignement et le croisement des cylindres sont  $\tau_{\parallel} = \tau_{\perp} = 0.15$ . Ceci impose un angle d'au plus  $\approx 30^\circ$  entre deux cy-

lindres connectés et d'au moins  $\approx 80^\circ$  pour deux cylindres qui se croisent.

Le paramètre d'attache aux données est  $u_{\max} = -25$ . Concernant les paramètres de connexité du terme d'interaction, leur domaine de définition respectif est  $\log \gamma_0 \in [-12.5, -7.5]$ ,  $\log \gamma_1 \in [-5, 0]$  et  $\log \gamma_2 \in [0, 5]$ . Si les cylindres sont très proches et pas en position orthogonale, alors on interdit cette configuration en posant  $\gamma_h = 0$ . Ce choix de paramètre favorise la connexion entre cylindres, en encourageant la formation de filaments plutôt longs dans les régions favorables du champ de galaxies, tout en interdisant l'agrégation de cylindres. Cependant, nous n'avons pas de connaissance précise au sujet du poids respectif de chacun de ces paramètres. C'est pourquoi une loi uniforme est choisie pour  $p(\theta)$ .

Pour maximiser la loi conjointe  $p(\mathbf{x}, \theta)$ , nous avons utilisé le recuit simulé. L'algorithme a tourné pendant 250000 itérations. Des échantillons ont été prélevés et la température a été modifiée toutes les 250 itérations. La température a été refroidie à partir de  $T_0 = 1$  selon un schéma logarithmique. La configuration finale des cylindres est montrée Figure 7.7b. Les cylindres détectés apparaissent bien dans les régions où les galaxies semblent les plus alignées.

Comme au cas précédent, la solution fournie par le recuit simulé n'est pas unique et la convergence n'est atteinte qu'à l'infini. Ici non plus, le diagnostic visuel ne peut pas être effectué. Par conséquent, le calcul des probabilités de contact (7.13) est utilisé pour attribuer une forme "moyenne" aux filaments galactiques. Le domaine  $W$  a été divisé en petites cellules cubiques de même volume  $1 h^{-3} \text{Mpc}^3$ . La carte des probabilités de contact est montrée Figure 7.7c. Le réseau de filaments apparaît clairement. Sur la carte des visites du modèle, un filament est une composante connexe faite de cellules voisines de probabilité de contact  $\geq 95\%$ . Ici une cellule a 26 voisins au maximum.

Beaucoup de filaments détectés ont une structure simpliste. Cependant, de nombreux filaments présentent une structure morphologique complexe. Ceci est un aspect nouveau qui obligera peut être les cosmologistes à changer la définition initiale des filaments, à savoir un simple lien ou pont entre agrégats de galaxies. Les Figures 7.8 et 7.9 montrent deux structures de ce type.

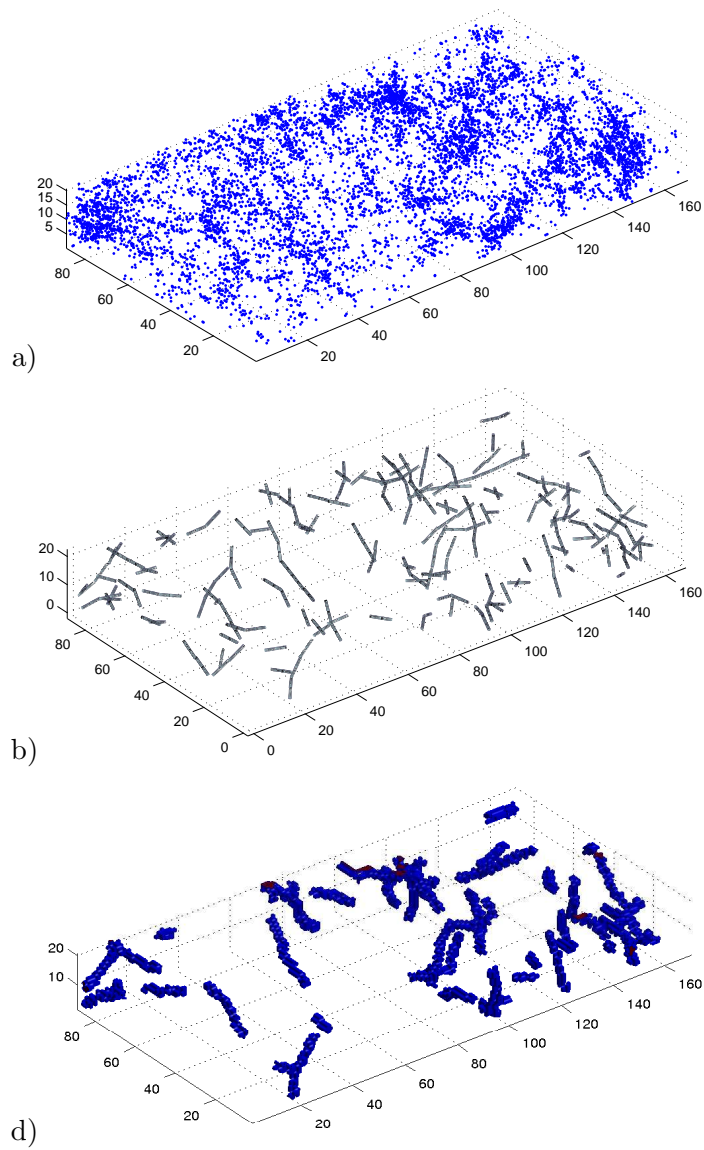


FIGURE 7.7 – a) Données : le catalogue N250 ; b) Configuration de cylindres obtenues en utilisant l’algorithme du recuit simulé ; c) Carte des probabilités de contact seuillée à 95%.

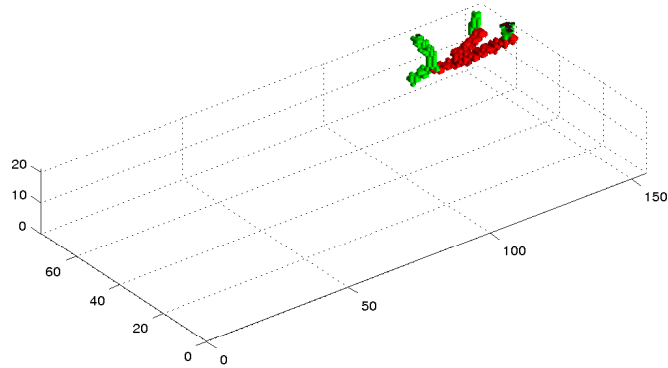


FIGURE 7.8 – Filament de morphologie simple détecté dans le catalogue *NGP250*. La carte des probabilités de contact seuillée à 50% (couleur vert) et à 95% (couleur rouge).

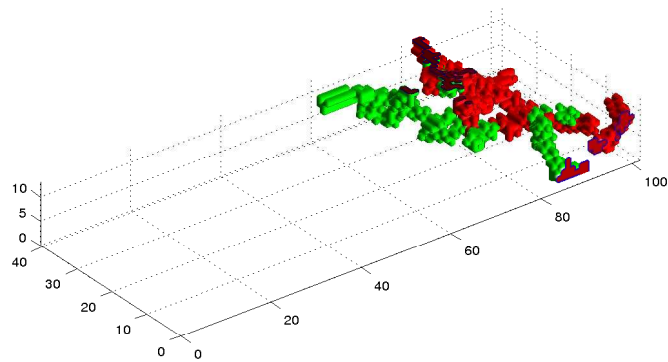


FIGURE 7.9 – Filament de morphologie complexe détecté dans le catalogue *NGP150*. La carte des probabilités de contact seuillée à 50% (couleur vert) et à 95% (couleur rouge).

## 7.4 Evaluation des résultats. Caractérisation morpho-statistique des structures.

L'approche bayésienne ne résout pas le problème de l'estimation des paramètres. Les cartes de probabilités de contact ne sont pas une méthode de validation statistique. Ceci est un problème ouvert de notre méthodologie. Il faut toutefois s'interroger sur la stabilité de la méthode vis-à-vis des paramètres du modèle. Le cadre mathématique de notre méthodologie permet d'évaluer les résultats obtenus, à partir des statistiques exhaustives du modèle. En effet, ces statistiques caractérisent entièrement la structure dans tout le domaine d'observation. Nous montrons ci-dessous deux exemples de détection d'agrégats en épidémiologie et de filaments galactiques.

### 7.4.1 Détection d'agrégats en épidémiologie animale : vérification et analyse des résultats

Nous allons vérifier que la structure détectée est totalement tributaire des données, et n'est pas un artefact aléatoire quelconque du modèle ou de la méthodologie proposée.

Deux expériences ont été mises en oeuvre sur chaque jeu de données des années 1996 à 2000. La première consiste à lancer notre algorithme de détection précédent en posant  $T_0 = 1$  et en laissant la température inchangée tout le long de la dynamique. Cela revient tout simplement à simuler  $p(\mathbf{x}, \theta)$ . Nous avons besoin de simuler à température fixée pour avoir des dynamiques de simulation stationnaires. Ceci est une condition nécessaire à l'inférence statistique. Les paramètres du modèle sont les mêmes pour tous les jeux de données. Les échantillons ainsi obtenus ont ensuite servi à calculer les moyennes empiriques des statistiques suffisantes du modèle.

Les résultats obtenus sont montrés au Tableau 7.4. Les statistiques exhaustives du modèle sont des descripteurs de la structure d'agrégats. Elles permettent de quantifier la présence d'agrégats dans les données. Le nombre total de disques composant la structure et sa surface sont liés à la dimension et à la géométrie des agrégats. Le nombre de paires de disques superposés est une caractéristique plus subtile, car il est possible de l'interpréter comme la "concentration" ou "l'intensité" d'un agrégat dans le champ de données. La présence d'agrégats dans les données est indiquée par l'importance des valeurs numériques de ces trois caractéristiques. Par exemple, la présence d'agrégats en 1996 est montrée par le nombre moyen de 74.10 disques, une surface moyenne de l'union des disques de 312.46, et une "concentration" des agrégats indiquée par un nombre moyen de 555.08 paires de disques superposés. Ce sont les plus grandes valeurs du Tableau 7.4. A la lumière de

cette interprétation, les agrégats les plus importants sont détectés en 1996, puis en 1997 et 1998. Ce sont dans les années 1999 et 2000 que la présence d'agrégats est la plus faible.

Statistiques exhaustives	Jeux de données				
	1996	1997	1998	1999	2000
$\bar{n}(\mathbf{x})$	74.10	50.15	33.35	20.02	24.75
$\bar{\nu}(Z(\mathbf{x}))$	312.46	196.66	167.13	89.53	129.59
$\bar{n}_o(\mathbf{x})$	555.08	247.73	131.48	16.25	25.12

TABLE 7.4 – Moyenne des statistiques exhaustives pour chaque jeu de données :  $\bar{n}(\mathbf{x})$  est le nombre moyen de disques qui forment l'agrégat,  $\bar{\nu}(Z(\mathbf{x}))$  est la surface moyenne de la structure d'agrégats et  $\bar{n}_o(\mathbf{x})$  est le nombre moyen de paires de disques superposés.

La deuxième expérience est une procédure en deux étapes qui a également été appliquée à chaque jeu de données. Nous avons commencé par permuter les scores cellulaires de chaque ferme, laissant leur position inchangée. Nous avons ainsi généré 100 champs de données permutes. De cette manière, la structure d'agrégats des données est détruite. Une détection d'agrégats a ensuite été lancée sur chaque champ généré. Pour chacun de ces champs la moyenne des statistiques exhaustives a été calculée. Finalement, la valeur maximale des 100 moyennes a été retenue.

Les résultats figurent au Tableau 7.5. Comme espéré, les données permutes ne contiennent presque pas d'agrégats. Toutes les valeurs des statistiques exhaustives sont plus petites que celles des données originelles. Ce test différencie les deux situations à l'aide d'une  $p$ -valeur qui est plus petite que 1%. De cette façon, nous sommes assurés que la détection d'agrégats est due à la structure des données, et non pas au modèle.

Statistiques suffisantes	Champs de données permutes				
	1996	1997	1998	1999	2000
$\max \bar{n}(\mathbf{x})$	2.36	3.85	3.09	4.14	2.06
$\max \bar{\nu}(Z(\mathbf{x}))$	13.83	17.63	19.48	25.96	9.12
$\max \bar{n}_o(\mathbf{x})$	2.62	7.12	6.10	9.65	1.02

TABLE 7.5 – Maximum des moyennes des statistiques exhaustives calculées sur 100 champs de données permutes correspondant au cinq jeux de données.

### 7.4.2 Détection de filaments galactiques : vérification et analyse des résultats

L'utilisation de statistiques exhaustives et de tests de permutation s'étend naturellement à la détection de filaments galactiques [T21, T22]. Nous présentons ici une comparaison entre le catalogue des observations  $2dF$  et 22 catalogues simulés, appelés aussi catalogues MOCK. Les catalogues simulés ont été construits à partir de modèles physiques. Une comparaison entre les observations et les simulations serait utile pour actualiser les modèles en tenant compte des observations les plus récentes.

Pour tous les jeux de données, nous avons utilisé les mêmes paramètres que pour le processus de cylindres connectés, la loi a priori et la dynamique de simulation. Tous les détails se trouvent dans [T22]. Pour chaque catalogue nous avons enregistré la moyenne empirique des statistiques exhaustives. Les résultats sont présentés au Tableau 7.6.

Comme pour les agrégats, la simulation a été effectuée à température fixée. Cela permet une comparaison quantitative des observations et des données simulées via la loi des statistiques exhaustives. La Figure 7.10 montre les boîtes à moustache des statistiques des catalogues simulés et réels. Les boîtes des catalogues MOCK sont indexées par leur indice (de 1 à 22). L'intérêt immédiat des boîtes à moustaches est qu'elles permettent une comparaison visuelle simultanée de toutes les données. Cela se fait par rapport au centre, l'étendue et la portée des distributions considérées.

Les lois de la statistique  $n_0$  sont comparées sur le graphique en haut de la Figure 7.10. Cette statistique représente le nombre de cylindres isolés (0-connectés). Ainsi, un nombre important de cylindres isolés indique un réseau plutôt fragmenté. Nous observons que seul le catalogue MOCK 3 présente un réseau moins fragmenté que les données  $2dF$ . Un nombre important de catalogues simulés a un réseau qui est beaucoup plus fragmenté que les données : la médiane de ces catalogues est clairement plus élevée que celle des données réelles. Il y a toutefois des catalogues de médiane similaire à celle des données réelles, et il est de même pour la forme générale de la loi de la statistique. Ces catalogues sont 1, 10, 11, 15 et 18. Pour tester ces similarités, nous avons effectué un test de Kolmogorov-Smirnov. Les  $p$ -valeurs des catalogues 1 et 18 ont été respectivement de 0.96 et 0.13. Pour les catalogues 10, 11 et 15 les  $p$ -valeurs obtenues ont été toutes inférieures à 0.002. Nous concluons que parmi les catalogues simulés, il n'y en a que deux qui sont similaires aux données du point de vue de la loi de la statistique  $n_0$ . La majorité des catalogues MOCK présentent des réseaux beaucoup plus fragmentés que celui des données réelles.

Jeux de données	Statistiques <b>exhaustives</b>		
	$\bar{n}_2$	$\bar{n}_0$	$\bar{n}_1$
2dF	1.94	5.30	11.66
MOCK 1	2.53	5.62	13.16
MOCK 2	0.48	6.20	7.52
MOCK 3	1.29	4.65	6.88
MOCK 4	1.55	9.33	15.45
MOCK 5	1.45	10.63	9.24
MOCK 6	0.38	6.21	8.96
MOCK 7	1.36	9.08	8.12
MOCK 8	0.18	6.91	4.27
MOCK 9	2.07	6.09	9.76
MOCK 10	1.62	4.40	11.91
MOCK 11	1.28	4.65	10.14
MOCK 12	2.65	7.97	11.25
MOCK 13	0.73	6.48	7.08
MOCK 14	0.36	7.30	16.44
MOCK 15	0.98	4.36	8.47
MOCK 16	2.75	11.04	22.88
MOCK 17	0.30	5.96	7.67
MOCK 18	2.15	5.11	10.44
MOCK 19	1.59	8.02	10.99
MOCK 20	1.27	8.79	10.50
MOCK 21	2.77	10.57	11.06
MOCK 22	1.79	8.10	17.26

TABLE 7.6 – Moyenne des statistiques exhaustives des observations et des simulations :  $\bar{n}_2$  moyenne du nombre de cylindres 2-connectés,  $\bar{n}_1$  moyenne du nombre de cylindres 1-connectés et  $\bar{n}_0$  moyenne du nombre de cylindres 0-connectés.



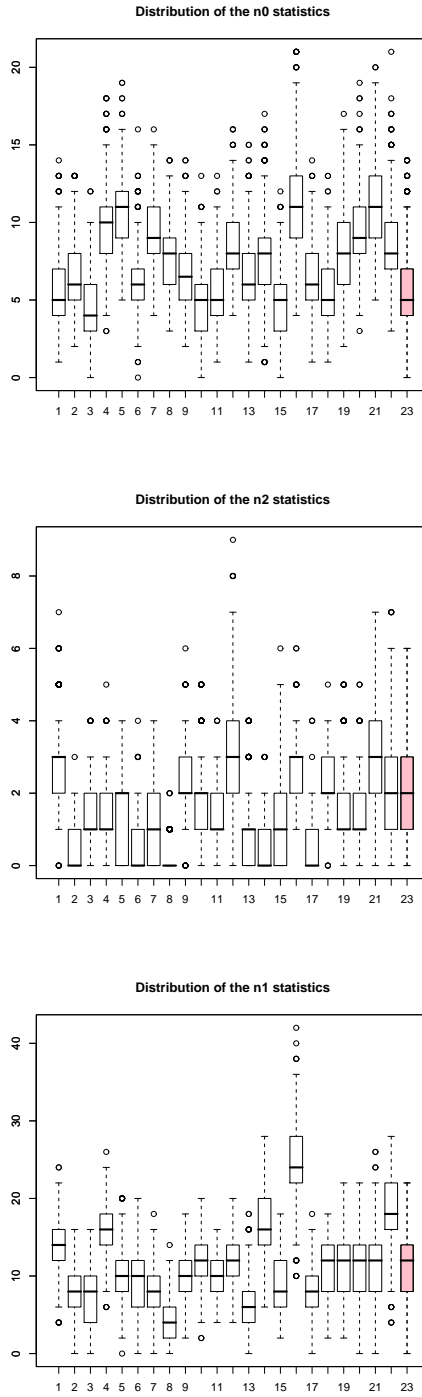


FIGURE 7.10 – Comparaison des distributions des statistiques exhaustives des données réelles (en rose) et simulées (en blanc).

Le panneau central de la Figure 7.10 compare les lois de la statistique  $n_2$ . Cette statistique donne le nombre de cylindres connectés aux deux extrémités (2-connectés). Une configuration présentant un nombre considérable de tels cylindres est un réseau composé de longs filaments. Nous observons que les catalogues 12 et 21 ont des réseaux contenant des filaments plus longs que les données. La loi du catalogue 22 est très similaire à celles des données. La  $p$ -valeur du test de Kolmogorov qui lui est associé est de 0.09. Pour décider vis-à-vis des catalogues 1, 9, 16 et 18, un test de Kolmogorov-Smirnov unilatéral a été effectué. L'hypothèse alternative a été que la loi du  $n_2$  des réelles présente plus souvent de grandes valeurs que celles des catalogues MOCK. Toutes les  $p$ -valeurs ont été très petites, ce qui permet d'affirmer que ces quatre catalogues ont des filaments plus courts que ceux dans les données réelles. Nous concluons qu'à quelques exceptions près, les filaments du catalogue  $2dF$  sont plus longs que ceux des catalogues simulés.

En bas de la Figure 7.10 se trouve la comparaison des lois de la statistique  $n_1$ . Cette statistique représente le nombre de cylindres connectés à une seule extrémité (1-connectés). Une configuration de valeur importante est un réseau de grande densité de filaments. En regardant ces trois statistiques ensemble, nous pouvons avoir une idée générale de la topologie du réseau. Par exemple, si les valeurs de  $n_1$  et  $n_2$  sont toutes deux fortes, cela indique un réseau qui a ses filaments semblables à des spaghettis ou à un arbre avec de longues branches. Si maintenant les valeurs de  $n_1$  et  $n_0$  sont simultanément fortes, cela indique un réseau qui a ses filaments semblables à des macaronis ou à un bosquet de petites branches. Par conséquent, les catalogues 1, 4, 14, 16 et 22 possèdent un réseau qui est beaucoup plus dense que celui du catalogue  $2dF$ . Les boîtes à moustache des catalogues 18, 19, 20 et 21 sont presque identiques à celle des données réelles, ce qui indique l'identité des lois correspondantes. Ceci a été confirmé par un test Kolmogorov-Smirnov. Les catalogues 10 et 12 ont la même médiane que les données, tandis que la loi est plus concentrée et symétrique. Le test de Kolmogorov-Smirnov a montré que ces distributions diffèrent de celle de la statistique  $n_1$  des données. Nous concluons que du point de vue de la statistique  $n_1$ , quatre catalogues simulés ont la même loi que les données, cinq autres ont clairement un réseau beaucoup plus dense, tandis que les autres catalogues ont des réseaux moins denses ou bien contiennent moins d'extrémités 1-connectés.

En conclusion, une seule statistique ne permet pas de discriminer les données simulées et observées. Cependant, si les trois statistiques sont prises ensemble, une différence évidente apparaît entre les simulations et les observations. D'un point de vue topologique, les réseaux des catalogues simulés sont plus fragmentés et contiennent des filaments plus courts que celui des

Réalizations binomiales	Statistiques exhaustives		
	max $\bar{n}_2$	max $\bar{n}_0$	max $\bar{n}_1$
MOCK 1	0	0.02	0
MOCK 2	0	0.015	0
MOCK 3	0	0.01	0
MOCK 5	0	0.015	0
MOCK 6	0	0.03	0
MOCK 7	0	0.02	0
MOCK 8	0	0.015	0

TABLE 7.7 – Maximum de la moyenne des statistiques exhaustives des champs binomiaux générés pour quelques catalogues MOCK.

données. Quant à la densité de filaments des catalogues simulés, elle se situe dans un intervalle qui inclut celle des données réelles. Les limites de cet intervalle sont dues à la variabilité morphologique des catalogues simulés.

Pour vérifier que le réseau de filaments existe vraiment, nous avons effectué un test de permutation. Pour cela, pour chaque jeu de données on a laissé fixe le nombre de galaxies mais leur position a été générée uniformément dans  $W$ . Nous avons donc un processus binomial. Pour chaque catalogue nous avons généré 100 réalisations d'un processus binomial. Pour chaque réalisation, une détection de filaments a été effectuée, en utilisant les mêmes paramètres pour le modèle et pour la dynamique de simulation que ceux des données d'origine. Les moyennes des statistiques exhaustives ont été calculées pour chaque champ, puis leur maximum sur les 100 réalisations a été pris. Les résultats sont montrés au Tableau 7.7.

Nous observons que la méthode ne détecte pas de filaments dans les champs binomiaux. Cette méthode détecte des filaments que dans les jeux de données où il se trouve quelque chose. La détection est due à la présence d'une structure dans les données et non pas à un effet aléatoire quelconque dans la méthode. Comme dans le cas des agrégats en épidémiologie animale, cette vérification ne fait que certifier que des filaments existent bien dans les données. Ce test ne certifie pas la qualité de la détection.

## 7.5 Construction d'une forme moyenne par un estimateur d'ensembles de niveaux

Les cartes de visites sont un outil qui a fait ses preuves du point de vue pratique. Cependant, à ce stade, les estimateurs des probabilités de contact ne sont qu'un outil de calcul local. En utilisant cet outil, nous souhaitons

en savoir plus sur la moyenne de la forme détectée. Pour cela, nous devons pouvoir caractériser la région occupée par les cellules de probabilité de contact supérieure à une valeur donnée. Ce travail a été publié [T5], et nous présentons ci-dessous ses idées principales.

Les processus ponctuel marqués font partie d'une famille plus grande d'objets mathématiques que l'on rencontre en géométrie stochastique. Ces objets sont les ensembles aléatoires. Il y a dans la littérature plusieurs façons de définir l'espérance d'un ensemble aléatoire, comme par exemple l'espérance d'Aumann ou la moyenne de Vorob'ev [10, 134, 181]. L'espérance d'Aumann est avantageuse pour moyenner des ensembles aléatoires convexes. Ceci est dû au fait que cette espérance est souvent convexe, même dans le cas où l'ensemble aléatoire est ni convexe, ni même connexe. L'espérance de Vorob'ev s'applique aux ensembles aléatoires compacts. Elle n'est toutefois pas recommandée pour des ensembles aléatoires de mesure nulle [10].

Nous avons opté pour l'espérance de Vorob'ev pour deux raisons. La première est que nos cartes de visites sont formulées comme des ensembles de niveau, et que l'espérance de Vorob'ev est définie à partir du même formalisme. La deuxième raison est que les ensembles aléatoires de nos applications ne sont pas convexes et pas de mesure nulle.

### 7.5.1 Ensembles de niveaux : un outil pour moyenner des ensembles aléatoires

Soit  $(\mathcal{S}, \mathcal{A}, \mathbb{P})$  un espace de probabilité et  $(W = [0, 1]^d, \mathcal{B}_W, \nu)$  un espace mesuré où  $\mathcal{B}_W$  est la tribu borélienne sur  $W$  et  $\nu$  la mesure de Lebesgue. Nous considérons également  $\mathcal{C}$  la classe des parties compactes de  $W$ .

Un ensemble aléatoire compact  $\Gamma$  dans  $W$  est une application mesurable de  $\mathcal{S}$  dans  $\mathcal{C}$ , au sens où

$$\forall C \in \mathcal{C}, \quad \{\omega : \Gamma(\omega) \cap C \neq \emptyset\} \in \mathcal{A}.$$

La fonction de recouvrement ou la probabilité de contact ou la fraction volumique est donnée par

$$p(w) = \mathbb{P}(w \in \Gamma).$$

Pour  $\alpha \in [0, 1]$ , l'ensemble de niveau correspondant est

$$Q_\alpha = \{w \in W : p(w) > \alpha\}.$$

L'ensemble  $Q_\alpha$  est déterministe, et parfois noté simplement  $\{p > \alpha\}$ .

**Définition 38** L'espérance de Vorob'ev est l'ensemble borélien  $\mathbb{E}_V \Gamma$  vérifiant

$$\nu(\mathbb{E}_V \Gamma) = \mathbb{E}[\nu(\Gamma)]$$

ainsi que

$$\{p > \alpha^*\} \subset \mathbb{E}_V \Gamma \subset \{p \geq \alpha^*\},$$

où

$$\alpha^* = \inf\{\alpha \in [0, 1] : \nu(Q_\alpha) \leq \mathbb{E}[\nu(\Gamma)]\}.$$

L'espérance de Vorob'ev est un ensemble de niveau  $\alpha^*$  de volume égal au volume moyen de l'ensemble aléatoire  $\Gamma$ . La Figure 7.11 montre l'évolution du volume  $F(\alpha) = \nu(Q_\alpha)$  de l'ensemble  $Q_\alpha$  en fonction du niveau  $\alpha$ . Il s'agit d'une fonction càdlàg avec éventuellement des plateaux. Les plateaux de  $p(w)$  impliquent des discontinuités de  $\nu(Q_\alpha)$ , alors que les plateaux de  $\nu(Q_\alpha)$  impliquent des discontinuités de  $p(w)$ .

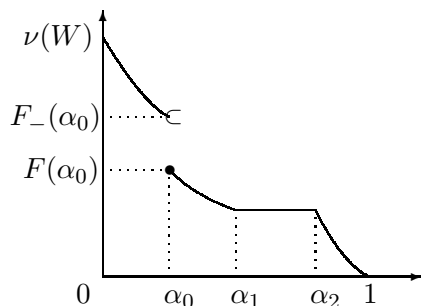


FIGURE 7.11 – Comportement de la fonction  $F(\alpha) = \nu(Q_\alpha)$ .

L'unicité de  $\mathbb{E}_V(\Gamma)$  est assurée si  $F$  est continue en  $\alpha^*$ . Dans ce cas, nous pouvons choisir

$$\mathbb{E}_V(\Gamma) = \{p \geq \alpha^*\}$$

qui est un ensemble compact puisque  $p(w)$  est semi-continue supérieure [135]. Il est possible de montrer que  $\mathbb{E}_V$  minimise

$$B \rightarrow \mathbb{E}[\nu(B \Delta \Gamma)]$$

sous la contrainte  $\nu(B) = \mathbb{E}[\nu(\Gamma)]$ ,  $\Delta$  étant la différence symétrique ([135], Thm. 2.3, pp. 177).

Cependant, l'utilisation directe de l'espérance de Vorob'ev soulève quelques difficultés. Premièrement, la probabilité de contact n'est analytiquement calculable que sur un nombre limité de modèles. Deuxièmement, les ensembles de niveau doivent être calculés en tout point de  $W$ . Pour surmonter ces inconvénients et construire des estimateurs, des approximations de type Monte Carlo et des discrétisations doivent être envisagées.

## 7.5.2 Estimation des ensembles de niveaux

Soient  $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ ,  $n$  réalisations i.i.d de  $\Gamma$ . L'approximation de  $p(w)$  s'écrit naturellement

$$p_n(w) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{w \in \Gamma_i\}}.$$

Alors, l'estimateur *plug-in* d'un ensemble de niveau  $\alpha$  est défini par

$$Q_{n,\alpha} = \{p_n > \alpha\}. \quad (7.15)$$

La convergence d'un tel estimateur a été déjà traitée dans la littérature [41, 42, 130, 131, 133, 158, 188]. Pour l'étude de cette convergence plusieurs distances peuvent être envisagées. Nous nous intéressons plus particulièrement à la pseudo-distance définie par

$$d(A, B) = \nu(A \Delta B)$$

pour tout  $A, B \in \mathcal{B}_W$ . La convergence  $L^1$  a été prouvée sous des hypothèses très faibles sur  $p(w)$ ; elle n'a pas besoin d'être continue [42]. Des résultats de convergence ont également été obtenus avec la distance de Hausdorff, moyennant des conditions de régularité et de continuité supplémentaires [130, 131, 42]. Le cas où  $p(w)$  est une densité qui peut être estimée par une méthode à noyau a été également traité dans la littérature. Des vitesses de convergence sont calculées par [133, 158, 188] et la normalité asymptotique est établie [120]. Une estimation adaptative en tenant compte de la régularité de  $p(w)$  peut être également effectuée [176].

Dans ce qui suit, nous généralisons la démarche proposée par [42]

Soit  $r > 0$  le pas définissant la grille  $W_r = W \cap r\mathbb{Z}^d$ . Nous cherchons à approximer  $Q_{n,\alpha}$  par une quantité  $Q_{n,\alpha}^r$ . Cette quantité doit être proche de  $Q_{n,\alpha}$  quand  $r$  tend vers 0 et elle doit être facile à calculer.

Pour tout borélien  $B$  de  $W$  et tout  $r \in 2^{-\mathbb{N}}$ , son approximation discrète  $B^r$  est définie par

$$B^r = \bigsqcup_{w \in B \cap r\mathbb{Z}^d} [w, w + r)^d.$$

Soit  $N_r(\partial B)$ , le nombre de cellules de  $W_r$  nécessaires au recouvrement de la frontière de  $B$

$$N_r(\partial B) = \text{Card}\{w \in r\mathbb{Z}^d : [w, w + r)^d \cap \partial B \neq \emptyset\},$$

et soit  $\overline{\dim}_{\text{box}}(\partial B)$ , la dimension "upper box"

$$\overline{\dim}_{\text{box}}(\partial B) = \limsup_{r \rightarrow 0} \frac{\log N_r(\partial B)}{-\log r}.$$

A partir de ces définitions, la proposition suivante est obtenue.

**Proposition 17** *Si  $\overline{\dim_{\text{box}}}(\partial B) < d$ , alors pour tout  $\varepsilon > 0$ , il existe  $r_\varepsilon$  tel que*

$$0 < r < r_\varepsilon \implies d(B^r, B) \leq r^{d - \overline{\dim_{\text{box}}}(\partial B) - \varepsilon}.$$

Le résultat suivant prolonge celui obtenu par ([42], Thm. 3, pp. 14). En ajoutant la discrétisation, les estimateurs *plug-in* des ensembles de niveau peuvent être utilisés en pratique.

**Proposition 18** *Supposons qu'il existe  $\kappa > 0$  tel que  $\overline{\dim_{\text{box}}}(\partial \Gamma) \leq d - \kappa$  presque sûrement. Alors, pour tout  $\alpha$  tel que  $\nu(\{p = \alpha\}) = 0$ , nous avons alors*

(i) *avec la probabilité 1*

$$\lim_{\substack{r \rightarrow 0 \\ n \rightarrow \infty}} \nu(Q_{n,\alpha}^r \Delta Q_\alpha) = 0,$$

(ii) *pour tout  $\varepsilon > 0$ ,*

$$\mathbb{E}[\nu(Q_{n,\alpha}^r \Delta Q_\alpha)] \leq r^\kappa + 2 \exp^{-2n\varepsilon^2} + F(\alpha - \varepsilon) - F(\alpha + \varepsilon).$$

Le niveau  $\alpha$  de l'estimateur  $Q_{n,\alpha}^r$  est une valeur fixée. Le niveau dans l'espérance de Vorob'ev est une quantité qui doit être estimée. Pour cela, nous procédons dans deux étapes. Premièrement, nous utilisons une approximation  $\alpha_n^*$  dans l'expression de l'espérance de Vorob'ev. Cette quantité est connue dans la littérature sous le nom d'espérance de Kovyazin [104]. Deuxièmement, cette quantité est discrétisée et proposée comme estimateur de l'espérance de Vorob'ev.

**Définition 39** *L'espérance de Kovyazin est le borélien  $K_n$  vérifiant*

$$\nu(K_n) = \frac{1}{n} \sum_{i=1}^n \nu(\Gamma_i)$$

*et tel que*

$$\{p_n > \alpha_n^*\} \subset K_n \subset \{p_n \geq \alpha_n^*\},$$

*où*

$$\alpha_n^* = \inf\{\alpha \in [0, 1] : \nu(\{p_n > \alpha\}) \leq \nu(K_n)\}. \quad (7.16)$$

**Théorème 22** *Supposons  $\nu(\{p = \alpha^*\}) = 0$ . Alors nous avons :*

$$\lim_{n \rightarrow \infty} \nu(K_n \Delta \mathbb{E}_V \Gamma) = 0.$$

*presque sûrement.*

La discrétisation de l'espérance de Kovyazin se fait en considérant le borélien  $K_{n,r}$  tel que

$$\{p_n > \alpha_{n,r}^*\}^r \subset K_{n,r} \subset \{p_n \geq \alpha_{n,r}^*\}^r,$$

où

$$\alpha_{n,r}^* = \inf\{\alpha \in [0, 1] : \nu(\{p_n > \alpha\}^r) \leq \nu(K_n)\}.$$

L'ensemble  $K_{n,r}$  est l'estimateur opératoire que nous proposons. Sa construction suit la procédure suivante. Premièrement, nous approximations le volume moyen de l'ensemble aléatoire  $\Gamma$  par  $\nu(K_n)$ . Cette approximation détermine le nombre de cellules de la grille à sélectionner. Ces cellules sont ensuite réunies pour former l'ensemble  $K_{n,r}$ , en fonction de l'estimation de leur probabilité de recouvrement  $p_n$ .

Le résultat suivant montre que l'estimateur ainsi construit est consistant, et qu'il améliore les estimateurs existants, même si le calcul exact de  $\nu(K_n)$  n'est pas facile pour des ensembles aléatoires de structure complexe.

**Théorème 23** *Supposons que pour  $\kappa > 0$ , nous avons  $\overline{\dim_{\text{box}}}(\partial\Gamma) \leq d - \kappa$  presque sûrement, et que  $\nu(\{p = \alpha^*\}) = 0$  et  $\nu(\{p = \beta^*\}) = 0$  avec  $\alpha^* = \inf\{\alpha \in [0, 1] : \nu(Q_\alpha) \leq \mathbb{E}[\nu(\Gamma)]\}$  et  $\beta^* = \sup\{\alpha \in [0, 1] : \nu(\{p > \alpha\}) \geq \mathbb{E}[\nu(\Gamma)]\}$ . Alors, presque sûrement :*

$$\lim_{\substack{r \rightarrow 0 \\ n \rightarrow \infty}} \nu(K_{n,r} \Delta \mathbb{E}_V \Gamma) = 0.$$

La preuve de ce résultat s'appuie sur les deux lemmes suivants.

**Lemme 5** *Supposons que  $\mathbb{P}(\overline{\dim_{\text{box}}}(\partial\Gamma) \leq d - \kappa) = 1$  pour  $\kappa > 0$ . Alors, avec la probabilité un*

$$\alpha^* \leq \liminf_{\substack{n \rightarrow \infty \\ r \rightarrow 0}} \alpha_{n,r}^* \leq \limsup_{\substack{n \rightarrow \infty \\ r \rightarrow 0}} \alpha_{n,r}^* \leq \beta^*.$$

*Nous avons également*

$$\alpha^* \leq \liminf_{n \rightarrow \infty} \alpha_n^* \leq \limsup_{n \rightarrow \infty} \alpha_n^* \leq \beta^*.$$

**Lemme 6** *Supposons que  $\mathbb{P}(\overline{\dim_{\text{box}}}(\partial X) \leq d - \kappa) = 1$  pour  $\kappa > 0$ . Alors, avec la probabilité un*

$$\limsup_{\substack{n \rightarrow \infty \\ r \rightarrow 0}} (K_{n,r}, K_n) \leq 2 \left[ \lim_{\substack{\alpha \rightarrow \alpha^* \\ \alpha < \alpha^*}} F(\alpha) - F(\beta^*) \right].$$

**Preuve:** La preuve est donnée par ([T5], Thm. 2.5, pp. 50). Nous écrivons

$$\nu(K_{n,r} \Delta \mathbb{E}_V \Gamma) \leq \nu(K_{n,r} \Delta K_n) + \nu(K_n \Delta \mathbb{E}_V \Gamma)$$

et nous utilisons le Théorème 22 et les lemmes 5 et 6 pour conclure.  $\square$



### 7.5.3 Application de l'espérance de Vorob'ev pour les champs de lacets browniens

Dans cette application, nous considérons  $n$  ensembles compacts i.i.d  $\Gamma_i = U_i + L_i$  de  $\mathbb{R}^2$ , les  $U_i$  étant des points uniformes de  $[0, 0.25] \times [0, 0.25]$  et les  $L_i$  des compacts délimités par un lacet brownien. Ces ensembles sont appelés aussi des lacets browniens remplis [66]. Un lacet brownien  $L$  est un chemin dans  $\mathbb{R}^2$  qui part de 0 et qui est conditionné pour retourner en 0. Un lacet brownien rempli est le complémentaire de l'unique composante non bornée de  $\mathbb{R}^2 \setminus L$ , comme le montre la Figure 7.12a.

Par cet exemple, nous souhaitons montrer un ensemble aléatoire pour lequel le Théorème 23 s'applique sans que la frontière des ensembles  $\Gamma_i$  soit lisse. La dimension de Hausdorff d'une frontière brownienne est  $4/3$  [109].

Nous avons simulé 1000 lacets browniens remplis et avons calculé la fonction empirique de recouvrement  $p_n(w)$  pour tous les points  $w$  d'une grille sur  $[0, 1]^2$  à pas de 0.01. La Figure 7.12b montre les lignes de niveau. Le calcul de  $p_n(w)$  permet d'obtenir l'estimateur  $K_{n,r}$  de l'espérance de Vorob'ev (la région colorée de la Figure 7.12b). Pour ce faire, nous avons calculé la surface moyenne des lacets browniens remplis puis  $\alpha^*$  en utilisant (7.16).

A l'exception de l'intensité du processus ponctuel de germes, ce modèle a des similarités avec le modèle beaucoup plus élaboré de la soupe de lacets browniens [110]. Les applications en écologie fournissent une motivation supplémentaire à considérer ce modèle [156]. L'idée est d'étudier l'occupation d'un territoire par des animaux qui cherchent de la nourriture. Dans [48, 156] leur déplacement est modélisé par un mouvement brownien ou un brownien fractionnaire, et l'enveloppe convexe du chemin est étudiée.

### 7.5.4 Application de l'espérance de Vorob'ev aux filaments galactiques

Nous appliquons le calcul de l'espérance de Vorob'ev à l'étude de filaments galactiques. Les données sur lesquelles nous travaillons sont les positions des galaxies du catalogue *NGP150* (cf. Fig. 7.13). Ce catalogue contient 2499 galaxies situées dans le domaine  $W = [0, 53.1] \times [0, 101.5] \times [0, 12.4]$ . L'unité de longueur est le mégaparsec  $h^{-1}Mpc$ .

La probabilité de recouvrement s'écrit

$$p(w) = \int_{\Omega} \int_{\Theta} \mathbf{1}\{w \in \mathbf{x}\} p(\mathbf{x}|\theta) p(\theta) d\mu(\mathbf{x}) d\theta, \quad (7.17)$$

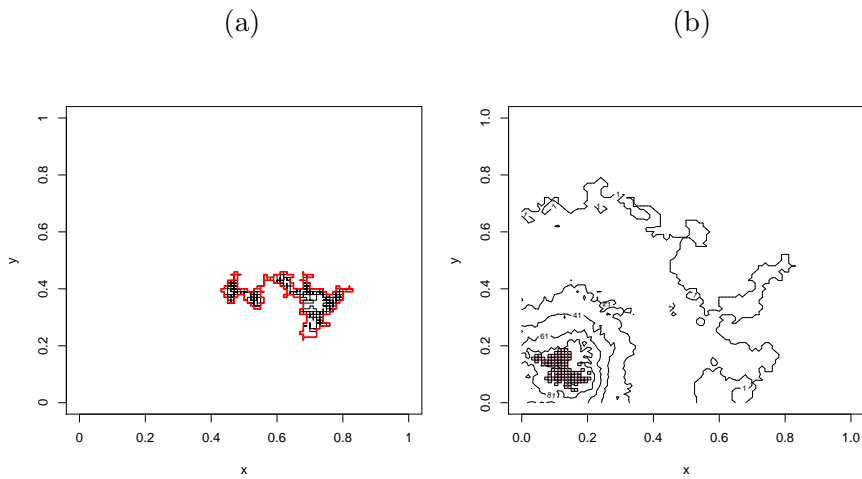


FIGURE 7.12 – a) Lacet brownien (noir) et la frontière du lacet brownien rempli (rouge) ; b) Espérance de Vorob'ev (région coloriée) et les ensembles de niveau pour  $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ . Les valeurs numériques obtenues ont été  $\alpha^* = 0.096$  et  $\nu(K_n) = 8.64 \cdot 10^{-3}$ .

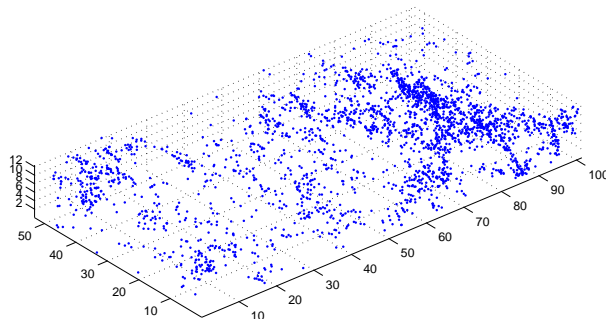


FIGURE 7.13 – Le catalogue des galaxies NGP150.

et représente la probabilité que le point  $w \in W$  soit rencontré par un réseau de cylindres qui se connectent et s'alignent. La configuration de cylindres est gouvernée par la loi bivariable  $p(\mathbf{x}, \theta) = p(\mathbf{x}|\theta)p(\theta)$ . Tous les détails concernant cette loi ont déjà été précisés. Nous rappelons que la loi conditionnelle  $p(\mathbf{x}|\theta)$  est la densité de probabilité d'un processus ponctuel marqué gérant le positionnement et les interactions des cylindres dans le champ de galaxies. La loi a priori  $p(\theta)$  prend en compte les incertitudes concernant le poids des différents paramètres du modèle ; elle est uniforme sur le domaine de définition des paramètres considérés.

La mesure de Lebesgue des ensembles de niveau induits par (7.17) est donnée par la fonction

$$F(\alpha) = \int_W \mathbf{1}\{p(w) > \alpha\} \nu(dw).$$

Pour approximer la formule (7.17) par la méthode Monte Carlo, il est possible de générer des échantillons indépendants  $X_1, \dots, X_n$  d'un réseau de cylindres, en utilisant  $n$  algorithmes Metropolis-Hastings adaptés qui soient initialisés chacun avec une condition initiale différente. Pour économiser le temps de calcul, l'échantillonnage a été effectué sur une seule simulation. Après une période transitoire (burn-in), les échantillons ont été prélevés de façon suffisamment espacée pour qu'ils puissent être considérés comme non corrélés.

L'expression analytique de la fonction de recouvrement  $p(w)$  n'est pas disponible. Aussi est il difficile de vérifier la continuité de  $F(\alpha)$ . Par conséquent, nous supposons que les conditions du Théorème 23 sont respectées. Nous supposons aussi que le modèle de réseau filamentaire est le vrai modèle. En effet, sous l'hypothèse que le réseau galactique est construit à partir de cylindres connectés, nous considérons que le cadre bayésien réduit le risque d'ajuster un mauvais modèle aux données. La validation du modèle pour ce problème de reconnaissance des structures reste encore un problème ouvert [15].

La dynamique de simulation décrite plus haut a été utilisée à température fixe pour simuler une suite de  $n = 1000$  réalisations de  $\Gamma$ . L'approximation Monte Carlo de  $F(\alpha)$  est montrée Figure 7.14. Après inspection visuelle, nous n'avons pu observer de discontinuité. La grille de maille  $r$  induit dans le domaine  $W$  des cellules de même volume  $1h^{-1}Mpc$ . En utilisant les échantillons produits, nous obtenons les valeurs empiriques  $\nu(K_n) = 1158$  et  $\alpha_n^* = 0.36$ .

L'estimateur de l'espérance de Vorob'ev est montré Figure 7.15. Cet outil qui fournit la forme moyenne du réseau filamentaire complète notre

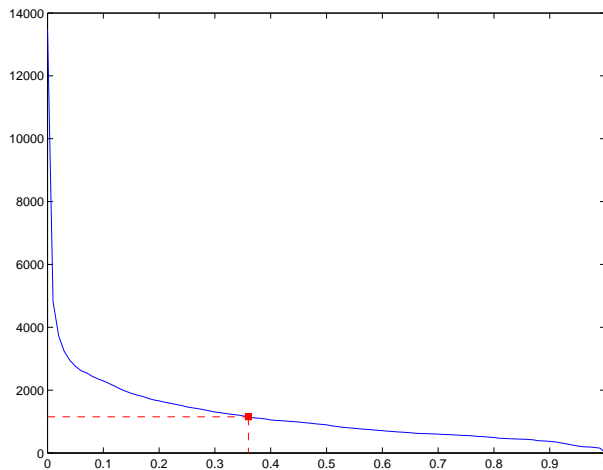


FIGURE 7.14 – Approximation Monte-Carlo de la fonction  $F(\alpha)$  pour le catalogue *NGP150*.

méthodologie de détection de formes. C'est son intérêt principal. La définition et la construction des moments d'ordre supérieur d'une structure reste toutefois encore un problème ouvert.

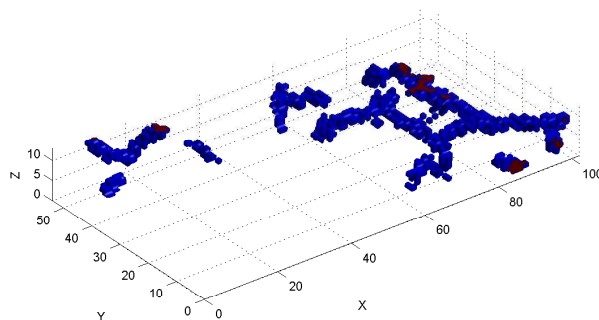


FIGURE 7.15 – L'estimateur de l'espérance de Vorob'ev pour le catalogue de galaxies *NGP150*.

## 7.6 Ensembles de niveau, statistiques d'ordre et lois à queue régulière pour l'étude des perturbations planétaires sur les comètes

Les comètes qui proviennent du nuage d'Oort et qui traversent notre Système Solaire, peuvent être perturbées par la force de gravitation exercée par

les grandes planètes. Ces planètes sont Jupiter, Saturne, Uranus et Neptune. L'effet de ces perturbations sur la dynamique d'une comète peut être négligeable mais peut aller jusqu'à l'éjection de la comète dans l'Univers.

Nous considérons comme perturbation planétaire la variation  $\Delta z = z_f - z_i$  d'un paramètre de la trajectoire d'une comète. Ce paramètre est l'inverse du demi grand axe de l'ellipse sur laquelle la comète se déplace. Cette variation a lieu entre l'entrée et la sortie de la comète dans de notre système solaire, puisque c'est dans cet intervalle que la comète peut être perturbée par une grande planète. Ces perturbations dépendent aussi de deux autres paramètres de la trajectoire d'une comète : la distance périhélique et le cosinus l'angle de l'inclinaison, notés respectivement  $q$  et  $\cos i$ .

Plusieurs milliers de perturbations ont été simulées en utilisant un intégrateur numérique, en formant un jeu de données spatialisées dans le plan  $(q, \cos i) \in W = [0, 32AU] \times [-1, 1]$ . La complexité et la masse du calcul d'intégration sont des faits mathématiquement reconnus. Il s'agit d'un cas particulier du problème des  $N$  corps : les corps considérés pour ce calcul sont le Soleil, la comète et les quatre grandes planètes. En supposant que ces perturbations sont la réalisations d'un champ aléatoire, nous souhaitons utiliser la modélisation stochastique pour remplacer l'intégration numérique par la simulation d'une loi de probabilité. Ceci est un projet très ambitieux. C'est pour cela que nous travaillons par étapes. La première étape est de fournir une description statistique des perturbations. Nous présentons ici les points clés de ce travail qui a été publié dans [T19].

### 7.6.1 Statistiques d'ordre

Les demi-grands axes des trajectoires des grandes planètes sont : pour Jupiter  $a_J = 5.2AU$ , Saturne  $a_S = 9.6AU$ , Uranus  $a_U = 19.2AU$  et Neptune  $a_N = 30.1AU$ . Les perturbations sont disposées uniformément dans le plan  $W$ . Cependant, les marques  $\Delta z$  doivent présenter des hétérogénéités : les perturbations dans les régions près des orbites des grandes planètes sont plus fortes que dans les autres régions de l'espace.

Pour mettre en évidence cette absence de stationnarité par analyse statistique, nous considérons une partition de  $W = \cup_{i=1}^n W_i$ . Les cellules  $W_i$  doivent être suffisamment petites pour que les marques à l'intérieur de chaque cellule puissent être considérées comme i.i.d., mais suffisamment grandes pour qu'elles contiennent un nombre suffisamment grand de perturbations pour que l'inférence statistique puisse être effectuée. Nous avons opté pour des cellules de volume égal à  $0.1AU \times 0.1$ , contenant environ 1500 perturbations.

Nous sommes intéressés par trois questions. Les deux premières questions portent sur les queues et la symétrie des lois de perturbation. La troisième question est la suivante : est-ce que les marques des perturbations ont une structure spatiale particulière dans le plan  $W$  considéré ?

Pour répondre à ces questions, les statistiques d'ordre ou les  $q$ -quantiles empiriques de chaque cellule ont été calculés. La plupart de ces valeurs indiquent que les perturbations sont réparties autour de l'origine, tandis qu'aucune structure particulière n'apparaît dans le plan des positions des marques des perturbations.

La situation est toutefois complètement différente quand on considère des quantiles extrêmes tels que 0.01, 0.05, 0.95, 0.99 : autour des demi-grands axes des grandes planètes ces quantiles prennent des valeurs importantes. Pour vérifier si ces valeurs sont dues à la présence de lois à queue lourde, l'indicateur  $z_q - n_q$  a été utilisé. Le premier terme est la quantile empirique d'ordre  $q$  ; le deuxième est la quantile d'ordre  $q$  de la loi normale pour laquelle les paramètres ont été respectivement estimés par  $z_{0.50}$  et  $0.5(z_{0.84} - z_{0.16})$ . Aux quantiles proches de 1, des valeurs positives de cet indicateur suggèrent un comportement de type loi à queue lourde des données. Bien évidemment, cet indicateur peut être utilisé pour des quantiles proches de 0.

La Figure 7.16 montre les valeurs de l'indicateur  $z_{0.99} - \hat{n}_{0.99}$ . Nous observons que les ensembles de niveau ont une forme de flèche autour des orbites des grandes planètes. La flèche pointe de droite à gauche, et cette forme est perdue quand le cosinus de l'angle d'inclinaison se rapproche de  $-1$ . Les niveaux des ensembles sont d'autant plus grands que la planète est grande. Ceci est en accord avec l'idée très raisonnable que plus la planète est grande plus elle peut perturber la trajectoire d'une comète.

Pour vérifier si l'apparition de cette structure est bien présente dans les données et n'est pas un artefact de la méthodologie<sup>1</sup>, nous avons effectué deux vérifications. La première consiste à appliquer le théorème central limite aux statistiques d'ordre pour effectuer un test qui rejette l'hypothèse de normalité [47, T19]. Nous avons pu ainsi observer que la normalité est essentiellement rejetée dans les régions autour des orbites des grandes planètes. La deuxième vérification consiste à calculer les valeurs moyennes des indicateurs  $z_q - n_q$  à l'aide d'une procédure de type bootstrap, ce qui nous a permis de confirmer les résultats obtenus sur les quantiles extrêmes [T19].

Il est tout à fait possible d'utiliser des indicateurs comme la différence de

---

1. Le seul paramètre de cette méthodologie est la dimension de la cellule  $W_i$ .

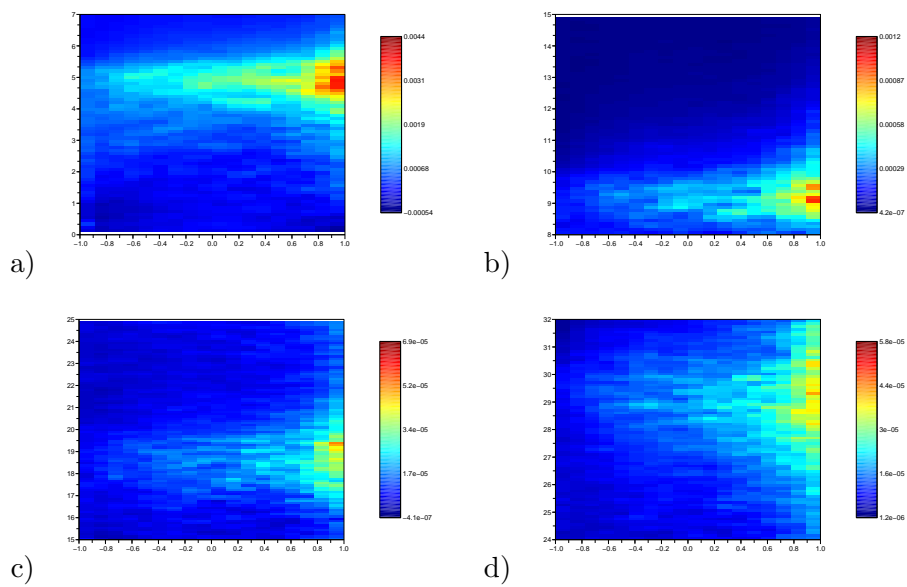


FIGURE 7.16 – Valeurs de l'indicateur basé sur la différence entre les quantiles empiriques et les quantiles de la loi normale  $z_{0.99} - \widehat{n}_{0.99}$  pour les perturbations planétaires. Les intervalles des valeurs pour l'indicateur autour des orbites de chaque grandes planètes sont : a) Jupiter ( $[-0.0005, 0.0044]$ ), b) Saturne ( $[4.2 \times 10^{-7}, 0.0012]$ ), c) Uranus ( $[-4.1 \times 10^{-7}, 6.9 \times 10^{-5}]$ ) et d) Neptune ( $[1.2 \times 10^{-6}, 5.8 \times 10^{-5}]$ ). Pour chaque courbe l'axe des  $y$  est la distance périhélique initiale et l'axe des  $x$  est le cosinus de l'angle d'inclinaison.

quantiles pour vérifier la symétrie de la loi. En considérant  $z_q - |z_{1-q}|$ , nous nous attendons à obtenir des valeurs proches de 0 quand la distribution est symétrique. Notre analyse indique que la distribution des perturbations peut être symétrique ou pas, indépendamment de la position des grandes planètes. La conclusion de cette analyse est que la symétrie et tout particulièrement les queues de distribution sont des éléments à prendre en compte dans un modèle probabiliste de perturbations.

### 7.6.2 Lois stables

Les lois stables sont une classe très riche de lois de probabilités qui permettent prendre en compte des caractéristiques comme les queues lourdes et la symétrie.

Une variable aléatoire  $Z$  est stable si pour tout  $a, b > 0$ , il existe  $c > 0$  et  $d \in \mathbb{R}$  telles que

$$aZ_1 + bZ_2 \stackrel{\mathcal{L}}{=} cZ + d,$$

où  $Z_1$  et  $Z_2$  sont des copies indépendantes de  $Z$  et " $\stackrel{\mathcal{L}}{=}$ " est l'égalité en loi.

Une loi stable est caractérisée par quatre paramètres  $\alpha \in (0, 2]$ ,  $\beta \in [-1, 1]$ ,  $\gamma \geq 0$  et  $\delta \in \mathbb{R}$ . On la note par  $\mathcal{S}_\alpha(\beta, \gamma, \delta)$ . Le rôle de chaque paramètre est le suivant :  $\alpha$  détermine la vitesse avec laquelle la loi converge vers 0,  $\beta$  contrôle la symétrie de la loi, alors que  $\gamma$  et  $\delta$  sont respectivement des paramètres d'échelle et de position. La Figure 7.17 montre l'influence de chaque paramètre sur la forme de la loi de probabilité.

La transformée linéaire d'une variable aléatoire stable est aussi une variable aléatoire stable. Si  $\alpha \in (0, 2)$ , alors  $\mathbb{E}|Z|^p < \infty$  pour tout  $0 < p < \alpha$  et  $\mathbb{E}|Z|^p = \infty$  pour tout  $p \geq \alpha$ . La loi est gaussienne si  $\alpha = 2$ . La variable stable avec  $\alpha < 2$  a une variance infinie, et ses queues de distribution sont asymptotiquement équivalentes à une loi de Pareto [177]. Plus précisément

$$\begin{cases} \lim_{z \rightarrow \infty} z^\alpha \mathbb{P}\{Z > z\} &= \frac{(1+\beta)}{2} \sigma, \\ \lim_{z \rightarrow \infty} z^\alpha \mathbb{P}\{Z < -z\} &= \frac{(1-\beta)}{2} \sigma. \end{cases}, \quad (7.18)$$

où  $\sigma = C_\alpha \gamma^\alpha$ ,  $C_\alpha = \frac{1-\alpha}{2\Gamma(2-\alpha)\cos(\pi\alpha/2)}$  si  $\alpha \neq 1$  et  $C_\alpha = \frac{2}{\pi}$  sinon. La loi est symétrique si  $\beta = 0$ . Quand  $\alpha < 1$ , le support de la loi  $\mathcal{S}_\alpha(\beta, \gamma, 0)$  est la demi-droite positive si  $\beta = 1$  et la demi-droite négative si  $\beta = -1$ . Si  $\alpha > 1$ , le moment du premier ordre existe et vaut le paramètre de translation  $\delta$ .

Une des difficultés techniques de l'étude des lois stables est que, sauf quelques cas remarquables (Gaussien, Cauchy, Lévy), les densités de probabilité n'ont



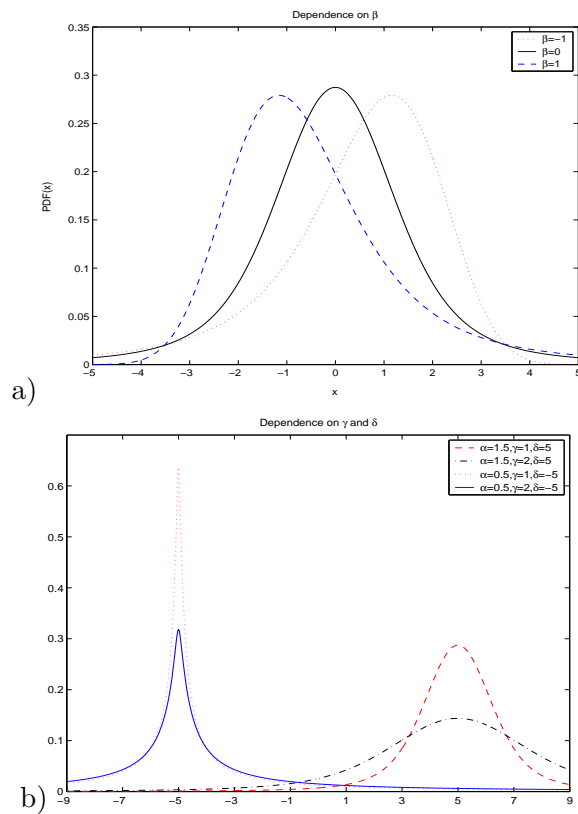


FIGURE 7.17 – Influence des paramètres sur la forme d’une distribution stable : a) le paramètre  $\beta$ , b) les paramètres  $\alpha$ ,  $\gamma$  et  $\delta$ .

pas d'expression analytique explicite. Ceci fait que l'estimation des paramètres des lois stables est toujours un problème ouvert. Plusieurs méthodes existent dans la littérature [114, T19]. Cependant la plupart des ces méthodes ne fonctionnent que lorsque  $\alpha \in (0, 2]$ . Pour notre problème nous ne savons pas si les données peuvent relever d'une loi stable ou pas. Dans ce cas, les lois à queue régulière peuvent être envisagées.

Une variable aléatoire a une loi à queue régulière de paramètre  $\alpha \geq 0$  s'il existe des nombres  $p, q \geq 0, p + q = 1$  et une fonction à variation lente<sup>2</sup>  $L$  tels que

$$\begin{cases} \lim_{z \rightarrow \infty} z^\alpha L(z) \mathbb{P}\{Z > z\} & = p, \\ \lim_{z \rightarrow \infty} z^\alpha L(z) \mathbb{P}\{Z < -z\} & = q. \end{cases} \quad (7.19)$$

Il est important d'observer que les conditions (7.18) peuvent être déduites de (7.19) en prenant  $L(z) = 1/\sigma$  et  $p = (1 + \beta)/2$ .

Sous l'hypothèse que l'échantillon a la propriété asymptotique (7.18), nous avons utilisé l'algorithme d'estimation des paramètres proposé par [50, 51, 114]. L'algorithme estime trois paramètres  $\hat{\alpha}, \hat{\beta}$  et  $\hat{\sigma}$ . Le paramètre  $\hat{\delta}$  est approximé par la moyenne empirique de l'échantillon quand  $\alpha > 1$ . Dans le cas où cet algorithme est utilisé pour des données de loi stable, le paramètre  $\alpha$  devrait avoir des valeurs dans l'intervalle  $(0, 2]$ . Le point fort de l'algorithme est qu'il peut être aussi utilisé dans le cas où les données ne présentent pas un caractère de loi stable. Dans ce cas, les données sont supposées provenir d'une loi à queue régulière. Le point faible de l'algorithme est que dans ce dernier cas nous n'avons pas d'information concernant la partie centrale de la loi. Néanmoins, cet algorithme permet déjà une caractérisation assez complète d'une large gamme de lois.

La Figure 7.18 montre l'application de l'algorithme pour estimer le paramètre  $\alpha$  sur des données de perturbations planétaires. Les ensembles de niveau inférieurs à 2 forment clairement une région autour de l'orbite de chaque grande planète. Effectivement, cette région est moins bien précisée que dans le cas des statistiques d'ordre. Ceci est sans doute dû au nombre de perturbations par cellule qui est assez limité par rapport aux exigences de l'algorithme. Cependant, les deux résultats sont cohérents et indiquent un comportement de type loi à queue lourde pour les perturbations autour des orbites des grandes planètes.

Plusieurs tests ont été effectués pour étayer cette conclusion [T19]. Premièrement, l'utilisation de l'équivalence asymptotique entre les lois de type Pareto et le théorème centrale limite des statistiques d'ordre a permis de valider l'estimation du coefficient  $\alpha$ . Deuxièmement, pour les régions où  $\alpha \leq 2$ , un

---

2. Une fonction à variation lente  $L(z)$  est telle que  $\lim_{z \rightarrow \infty} \frac{L(\lambda z)}{L(z)} = 1$  pour tout  $\lambda > 0$ .

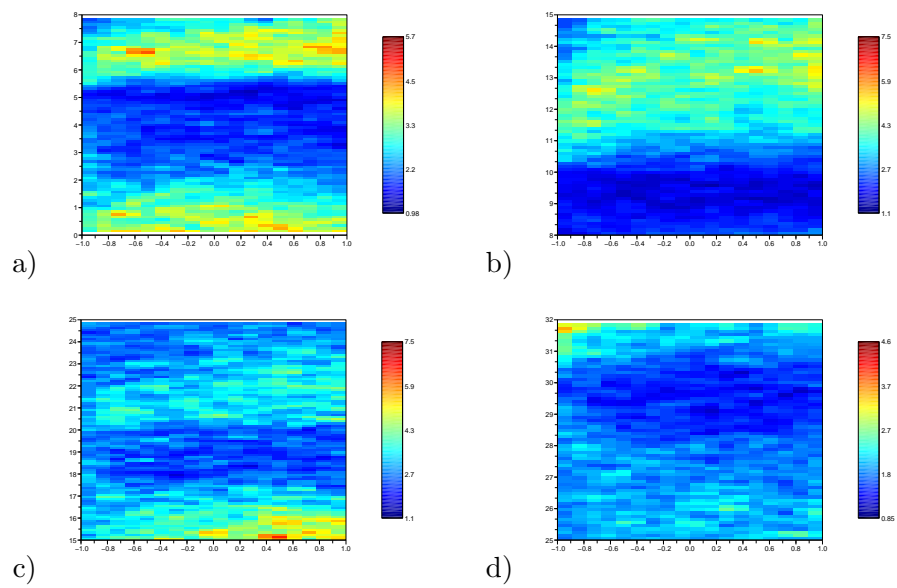


FIGURE 7.18 – Résultat de l'estimation du paramètre  $\alpha$  pour les perturbations autour des grandes planètes. Les intervalles correspondants sont : a) Jupiter  $[0.98, 5.7]$ , b) Saturne  $[1.1, 7.5]$ , c) Uranus  $([1.1, 7.5])$  et d) Neptune  $([0.85, 4.6])$ .

test du  $\chi^2$  a été fait pour valider le résultat de l'estimation de tous les paramètres. Finalement, pour remédier à l'inconvénient de l'algorithme, une autre loi a été proposée pour les régions où  $\alpha > 2$ , et un test du  $\chi^2$  a été construit pour valider les résultats.

L'expression de cette autre loi est

$$f(z) = \frac{C_{\kappa,\alpha}}{1 + |\kappa z - \omega|^{\alpha+1}}, \quad (7.20)$$

où avec  $C_{\kappa,\alpha}$  est une constante de normalisation,  $\kappa$  le paramètre d'échelle,  $\omega$  le paramètre de position et  $\alpha$  la vitesse de variation de la queue de distribution. L'estimation des paramètres de (7.20) se fait en plusieurs étapes. D'abord  $\alpha$  est estimé au moyen de l'algorithme précédent. Puis  $\omega$  est estimé comme moyenne empirique des données. La constante de normalisation  $C_{\kappa,\alpha}$  et le paramètre d'échelle sont finalement estimés par la méthode des moments.

Les  $p$ -valeurs du test autour de l'orbite du Jupiter sont montrées Figure 7.19. Nous observons que le test valide l'estimation des paramètres dans la plupart des cellules. Ceci renforce l'idée qu'autour des orbites des grandes planètes, les perturbations ont un comportement de type loi à queue lourde. En même temps, pour améliorer les résultats de ce test, nous pensons à encore une autre loi qui serait à même de prendre en compte la symétrie, ainsi qu'à une augmentation de données pour mieux estimer les paramètres.

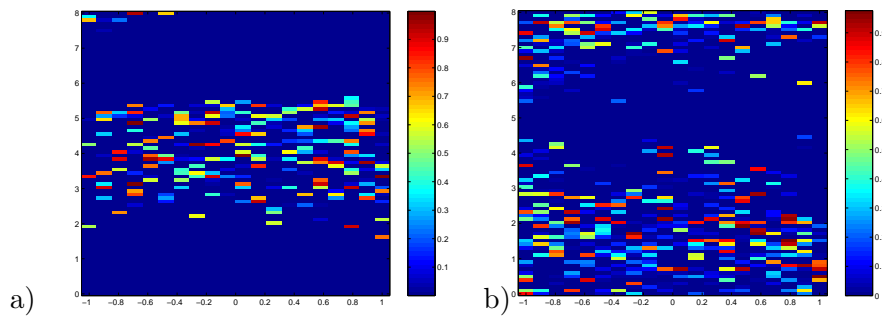


FIGURE 7.19 – Les  $p$ -valeurs du test  $\chi^2$  autour de l'orbite de Jupiter : a) région  $\alpha \leq 2$ , région  $\alpha > 2$ .

### 7.6.3 Théorie d'Öpik

Nos collaborateurs astronomes ont pu interpréter et valider nos résultats à la lumière de la théorie d'Öpik. Cette théorie étudie les rencontres proches

d'une comète avec une planète qui gravite autour du Soleil. Il s'agit d'un problème à trois corps, et dans ce cas des formules mathématiques sont disponibles pour calculer les perturbations planétaires. La Figure 7.20 montre des courbes de niveau des perturbations autour de Jupiter qui ont été calculées dans [T19]. Nous pouvons observer que ces lignes de niveau ont des caractéristiques similaires aux celles mises en évidence par l'analyse statistique.

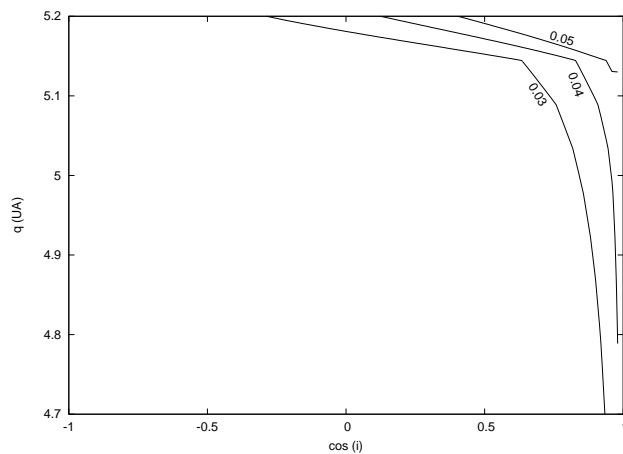


FIGURE 7.20 – Courbes de niveau des perturbations autour de Jupiter, calculées à l'aide de la théorie d'Öpik.

## Quatrième partie

# Conclusion et perspectives

Nos travaux de recherche ont abouti à la construction d'une méthodologie, d'une "machinerie algorithmique" pour détecter et caractériser la structure de données spatialisées.

Cette construction procède en trois étapes. D'abord, un modèle de forme ou de structure est proposé à partir des données observées. Puis, une dynamique de simulation est construite en adéquation avec le modèle. Enfin, des procédures statistiques sont mises au point pour inférer les caractéristiques de la structure cachée et les paramètres du modèle.

Chacune de ces étapes est attachée à un domaine particulier des probabilités et des statistiques. La modélisation repose sur des processus ponctuels, notamment marqués. La dynamique de simulation utilise les chaînes de Markov. L'inférence s'appuie sur l'analyse bayésienne, le recuit simulé, le maximum de vraisemblance, les tests.

La synthèse de ces trois étapes se fait au confluent de trois domaines : la géométrie aléatoire, les chaînes de Markov et les statistiques appliquées. Cette synthèse nous a permis d'aborder des applications concrètes en analyse d'image, en science de l'environnement et en astronomie, dans une relation d'interdépendance avec les mathématiques.

Cette dernière partie présente des conclusions et des perspectives qui découlent de notre travail. Une analyse générale de l'approche est faite en début du chapitre. Des questions et des thèmes de recherche sont ensuite présentés, en suivant la structure du document : modélisation de la forme, simulation de type Monte Carlo et inférence statistique. La toute dernière partie est dédiée à la présentation des deux directions de recherche, qui constituent les bases de mon nouveau projet de recherche.

## Chapitre 8

# Bilan et nouvelles directions de recherche

### 8.1 Analyse générale de l'approche

#### 8.1.1 Et les autres méthodes ?

D'une certaine manière, notre approche est équivalente à d'autres méthodes non supervisées<sup>1</sup> mises en place pour détecter des structures du même type. Dans les deux situations, la philosophie est la même : un calcul local qui dépend des données, des statistiques, un filtrage, suivi par une procédure de lissage ou de régularisation. Le premier calcul localise la solution, le deuxième suggère une certaine structure de solution.

Deux grandes familles de méthodes mathématiques suivent ce principe : les méthodes probabilistes et les méthodes déterministes. Les méthodes probabilistes, dont la notre, résument le modèle sous la forme d'une densité de probabilité. Les méthodes déterministes assemblent les éléments de la forme via un critère qui est ensuite injecté dans une procédure d'optimisation, par exemple les approches à base d'équations aux dérivées partielles [56]. Les deux écoles se disputent la suprématie, du point de vue de la précision de la solution, de la vitesse de calcul ou du naturel de l'approche ... Elles admettent toutefois que les critères qu'elles définissent pour modéliser les structures sont souvent très similaires.

Cette classification n'est pas très stricte. Il est également possible de mentionner une troisième famille de méthodes, plus proche de l'ingénierie, qui associe la caractérisation locale et la régularisation d'une manière très effi-

---

1. En reconnaissance des formes, une méthode non supervisée est complètement automatique, au sens où elle ne nécessite pas d'intervention humaine, une fois les paramètres du modèle connus.



cace du point de vue numérique. Ces méthodes sont rarement étayées par des résultats de convergence. Il n'en demeure pas moins qu'elles projettent de fédérer les avantages des deux courants majoritaires. C'est une voie à ne pas négliger.

Que choisir entre une approche probabiliste ou déterministe ? Nous pensons, qu'il faut garder de la souplesse face un problème pratique ; et surtout éviter de proclamer à l'avance, ou tout de suite après, la supériorité d'une méthode face à une autre.

Notre préférence va pour le cadre probabiliste, et cela pour deux raisons. D'un point de vue esthétique, nous pensons qu'il y a une certaine beauté à supposer que l'on ne contrôle pas tout, qu'il existe une part d'incertitude dans la solution proposée. D'un point de vue plus pratique, formuler une réponse sous forme de loi de probabilité donne implicitement la possibilité d'utiliser tout l'appareil intégrateur fourni par la théorie des probabilités et des statistiques.

La solution est construite en définissant ce que l'on cherche. Cependant, qu'est-ce exactement qu'une forme, une structure, ou bien, de façon plus concrète qu'un agrégat ou un filament galactique ? Ce sont des questions encore ouvertes. Cela tient à la richesse des données, à la multitude de méthodes existantes et à la spécificité de chaque discipline scientifique.

### **8.1.2 Pour quoi ça marche ?**

Une des raisons pour lesquelles notre "machinerie" fonctionne est qu'elle est construite en accord avec des principes classiques des sciences, à savoir l'approximation d'une structure complexe par des objets simples, l'analyse bayésienne et l'optimisation sous contrainte.

L'approximation d'une forme par des objets simples en interaction est notre hypothèse principale. Cette idée se rencontre dans beaucoup de domaines comme la physique, la biologie et bien sûr les mathématiques. Elle permet l'analyse et le découpage de la structure. Cette hypothèse est faite en anticipant le mécanisme résumant la forme, qui est dans notre cas une loi de probabilité. Les lois que nous construisons possèdent souvent la propriété de Markov. C'est par cette propriété que la complexité de la forme est brisée tout en préservant les capacités de synthèse du modèle. Dans nos exemples, cette complexité est décrite par des interactions symétriques entre couples d'objets. Cette approche est clairement avantageuse par rapport à d'autres techniques qui proposent des analyses très poussées, pas forcément dictées par un souci de synthèse ou bien d'intégration des résultats obtenus.

La deuxième hypothèse fondamentale utilisée est que toute l'information contenue dans les données à propos de la forme figure dans la fonction de vraisemblance. C'est le principe de vraisemblance de l'analyse bayésienne [162]. Le terme d'attache aux données peut être vu comme une vraisemblance, capable de situer les objets dans le champ de données. Le terme d'interaction entre objets peut être vu comme la loi a priori de la structure toute entière. Prises conjointement, les deux composantes forment la loi a posteriori, i.e. la loi de la forme cachée sachant les données. Pour être plus en accord avec certains points de vue, nous devrions préciser que nous travaillons plutôt sur des vraisemblances pénalisées. Mais comme nous utilisons des lois a priori pour les paramètres de nos modèles en de nombreuses situations, nous ajoutons une dimension bayésienne à notre travail.

La troisième idée générale est l'optimisation sous contrainte. Le terme à optimiser est l'énergie d'attache aux données. Ce terme agit comme un filtre ou un opérateur morphologique [175]. Cependant, les problèmes formulés à partir de ce seul type de filtre sont souvent mal posés. L'énergie d'interaction joue alors le rôle de contrainte qui régularise la solution. Ce problème de reconnaissance de formes peut être ainsi vu comme un problème inverse.

Le cadre offert par ces trois principes est très solide. Cependant, si le modèle est en complet désaccord avec les données et si la structure induite par le modèle a priori n'est pas ce que l'on souhaite trouver, alors les résultats seront difficilement acceptables. De même, si la dynamique de simulation n'a pas de bonnes propriétés de convergence, les résultats obtenus devront être considérés avec beaucoup de précaution.

En même temps, les mathématiques ne permettent pas de répondre à toutes les questions. Par exemple, quelles sont les véritables interactions qui définissent notre structure? Faut-il considérer le terme d'attache aux données comme une partie de la mesure de référence? Comment construire des dynamiques de simulation appropriées? Comment régler au mieux la méthode vis-à-vis des données? Pour l'instant, ce sont des problèmes ouverts.

## 8.2 Questions concernant la modélisation de forme

### 8.2.1 Extension à l'infini, percolation et transition de phase

Les processus ponctuels que nous avons construits comportent des interactions entre les objets qui composent la forme. La loi de ces processus dépend des observations par un terme d'attache aux données. Ce terme introduit de la non stationnarité dans le modèle. Les interactions ne portent que sur les

paires d'objets, et elles sont symétriques. Elles reposent sur la distance entre objets et leur marques. Le processus qui en résulte est localement stable et markovien au sens de Ripley-Kelly [161].

La stabilité locale garantit l'intégrabilité du modèle. C'est également une condition nécessaire à la convergence des dynamiques de Metropolis-Hastings. Le caractère markovien nous assure que les interactions sont de portée finie. Cela signifie qu'un objet n'interagit qu'avec les objets situés en deça d'un rayon fixe. Ce rayon dépend des paramètres du modèle. Le théorème de Hammersley-Clifford permet de caractériser ces processus par une mesure de Gibbs.

Les modèles sont définis dans une fenêtre d'observation finie  $W \subset \mathbb{R}^d$ . Nous souhaitons étudier l'existence de ces mesures quand le domaine d'observation s'étend vers  $\mathbb{R}^d$ . Notre motivation première est la cosmologie [116, 153]. En effet, la possibilité d'augmenter l'angle d'observation de notre univers justifie l'utilisation des modèles capables d'être étendus à l'infini.

Une présentation des résultats connus sur l'extension des processus ponctuels markoviens est donnée par [180, 141]. Pour des résultats plus récents nous nous référons aux travaux de [54, 55]. Il existe déjà dans la littérature des résultats concernant l'extension des processus ponctuels markoviens. Nous souhaitons synthétiser les résultats existants et les adapter à notre cadre, où les données et la superposition des interactions jouent un rôle clef.

De point de vue méthodologique, une telle extension permettrait d'aller vers une fusion de deux courants d'utilisation des processus ponctuels. Le premier courant suppose que l'observation disponible peut s'étendre à l'infini et conduit à des résultats asymptotiques. Le deuxième courant suppose que l'observation n'est disponible que de façon locale et met en jeu des procédures d'inférence à partir de simulations numériques.

L'existence d'une mesure de Gibbs ne garantit pas son unicité. Une transition de phase signifie existence mais non unicité d'une mesure de Gibbs. Il est bien connu qu'une transition de phase se manifeste surtout quand les interactions du modèle sont fortes, ce qui est le cas de nos applications. Pour chaque modèle, il serait donc intéressant de déterminer la gamme de valeurs des paramètres pour laquelle une transition de phase a lieu. Des résultats concernant l'existence d'une transition de phase existent pour quelques modèles, mais les valeurs critiques exactes des paramètres sont inconnues. Une discussion sur la transition de phase du modèle de Widom-Rowlinson est présentée dans [141]. Les références suivantes y sont également mentionnées [125, 85, 119]. Nous souhaitons poursuivre ces pistes dans le cadre de nos modèles et nos applications.

Il se pourrait que la présence d'un terme d'attache aux données réduise les effets non désirés d'une transition de phase. Néanmoins, la connaissance des valeurs critiques de paramètres, pourrait largement améliorer notre méthodologie. Nous pensons notamment à la construction de lois a priori judicieuses, quand un cadre bayésien est utilisé pour la détection de structures.

L'extension du domaine des observations demande l'extension des modèles. Il est par conséquent naturel de se demander si les structures détectées se prolongent à l'infini. Ce phénomène est connu sous le nom de percolation. Par exemple en cosmologie, la question se pose de l'existence des filaments galactiques infinis. La percolation continue a déjà été étudiée pour des schémas booléens [125, 180], ainsi que pour des modèles de type Quermass [39]. Nous souhaitons poursuivre ces voies, afin d'étudier le comportement de nos modèles d'un point de vue théorique, mais aussi par simulation, sous l'influence de paramètres critiques de percolation. Il est peut être possible d'établir la présence de la percolation dans un jeu de données par l'utilisation de procédures d'estimation des paramètres.

### **8.2.2 Processus de Cox, processus permanents et déterminants**

Une fois les données observées, il y a deux manières d'aborder la détection des structures. La première suppose que la forme est la cause des observations. Cette idée se retrouve dans la construction de processus de Cox ou de Neymann-Scott [141, 146]. La deuxième suppose que la forme est conséquence des données observées. Cette idée se retrouve dans la construction de processus ponctuels marqués ou de champs de Markov.

La construction de formes par des processus ponctuels marqués permet de caractériser la forme d'une manière naturelle, à l'aide des statistiques exhaustives du modèle. La caractérisation de formes par des processus de type Cox se fait au moyen des fonctions d'intensité ou de covariance ou encore des mesures moments. Les processus de Cox permettent aussi d'effectuer des simulations conditionnelles une fois la structure connue.

Nous souhaitons rapprocher ces deux manières d'aborder la forme. Dans ce but, un premier pas serait de construire des modèles pour lesquels le terme d'attache aux données serait une loi de probabilité que l'on sait simuler conditionnellement quand la forme est connue [T25].

Ces deux modélisations ont souvent leur loi accessible à une constante de normalisation près. Pour palier cet inconvénient, la modélisation par les

processus permanentaux et déterminentaux se profile comme un choix possible [108, 124]. Les processus permanentaux peuvent être vus comme une classe particulière de processus de Cox. Ces processus simulent des populations de points qui ont tendance à s'agréger. Les processus déterminentaux s'obtiennent en suivant une démarche similaire à celle des processus permanentaux. Ces processus simulent des populations de points qui ont tendance à se repousser. L'intérêt de ces processus est que leur densité de probabilité est analytiquement accessible.

En conséquence, les processus permanentaux et déterminentaux peuvent s'avérer intéressants pour des problèmes de détection de formes. De plus, ces modèles pourraient créer un lien supplémentaire entre les champs aléatoires et les structures induites par les processus ponctuels. Plusieurs questions se posent : comment introduire des marques qui définiraient les objets composant la forme ? Est-il possible de construire des fonctions d'intensité et de covariance qui traduisent des propriétés géométriques de la forme ? Serait-il possible de prendre en compte des attractions et des répulsions simultanément ? La densité de probabilité est obtenue par la troncature du spectre d'une fonction de covariance. Cette approximation doit pouvoir être contrôlée. Cependant, il est important de connaître ses effets si l'inférence est effectuée à partir de cette densité de probabilité. Des procédures d'estimation de paramètres ont été proposées pour les processus déterminentaux par [108]. Là aussi, il serait intéressant de savoir dans quelle mesure les procédures proposées diminuent l'effet mentionné par [13].

### 8.3 Questions concernant la simulation MCMC

La simulation de lois de processus ponctuels marqués par itérations markoviennes est toujours un problème ouvert. En effet, cette approche comporte deux grandes familles de méthodes, la première famille, classique, garantissant la convergence à l'infini, la deuxième famille des méthodes parfaites fournissant avec précision le temps de convergence. La difficulté vient de ce que les méthodes classiques ont besoin d'approximer le temps de convergence, alors que les méthodes parfaites sont fortement dépendantes des paramètres du modèle.

Si l'on doit simuler un processus ponctuel de faibles interactions, et si les propriétés du modèle le permettent, nous pouvons utiliser des méthodes de simulation exacte. Dans ce cas, les temps de convergence sont raisonnables. Cependant, les structures émanent souvent de modèles à interactions fortes. De plus, la procédure de maximisation par recuit simulé requiert un renforcement des interactions de la loi simulée. Dans ce cas, les méthodes classiques

de simulation sont plutôt utilisées.

Les perspectives sont ici d'améliorer ces différents aspects.

### 8.3.1 Méthodes classiques de simulation

Notre choix préféré pour simuler un processus ponctuel est la dynamique de Metropolis-Hastings. Ceci tient à ce qu'il a été possible de construire des algorithmes pour nos modèles qui leur sont bien adaptés (par l'utilisation de noyaux de transition spécifiques) et qui ont de bonnes propriétés de convergence.

Pour simuler un processus ponctuel marqué avec interactions, le point clef est la construction du noyau de transition. Ce noyau met en jeu trois types de transitions possibles : ajouter, enlever et modifier un objet. Si ces transitions s'effectuent en utilisant des lois d'instrumentation suffisamment simples, il est possible de montrer que la chaîne simulée par ce noyau est apériodique, irréductible, récurrente et ergodique. Une telle chaîne a comme unique distribution d'équilibre la loi du processus ponctuel étudié. Les échantillons ainsi obtenus permettent l'utilisation du théorème centrale limite [73, 72, 190, 128, 141, 164]. Il est souvent possible de proposer un noyau de transition un peu plus élaboré, qui aide le modèle à former des structures [T7, T17]. Cependant, il n'est pas toujours facile de trouver les bonnes lois d'instrumentation. Ces lois doivent à la fois être adaptées au modèle et garantir toutes les propriétés de convergence de la chaîne simulée [T6, 194].

La construction de dynamiques de simulation qui produisent des échantillons utilisables pour l'inférence statistique est un problème qui se pose à chaque fois qu'un nouveau modèle est proposé.

Dans certaines situations, les dynamiques de Metropolis-Hastings adaptées peuvent être vues comme des dynamiques Monte Carlo à sauts réversibles (RJMCMC) [76, 79, T6]. Ces algorithmes ont un réel intérêt quand l'espace d'états n'est pas composé que d'états communiquants récurrents. Les algorithmes RJMCMC ont reçu une attention particulière en traitement d'image [60, 61]. Ceci est dû à la capacité de ces dynamiques à effectuer des sauts entre sous-espaces de configuration de dimensions différentes. Elle est parfois confondue avec la capacité du modèle proposé de détecter le nombre d'objets contenus dans l'image. Cette idée est en partie justifiée [179, T15]. Cependant, nous sommes assez d'accord aujourd'hui avec le point de vue exprimé par [72] qui soutient que l'on rencontre plutôt rarement des preuves des propriétés de convergence des dynamiques RJMCMC. Une exception notable est [33] qui établit de façon rigoureuse l'équivalence entre des dyna-

miques RJMCMC et les algorithmes de type naissance et mort.

Au sujet de la dynamique de simulation proposée pour simuler des modèles de pavages en  $T$  [T6], nous souhaitons enrichir le noyau de transition de l’algorithme en cours pour garantir la récurrence et l’ergodicité de la chaîne simulée. Dans un premier temps, nous pensons essayer l’idée proposée par C. Lantuéjoul d’enlever une composante connexe entière. Cela permettrait d’adopter directement le cadre des dynamiques Metropolis-Hastings et d’appliquer les techniques de convergence utilisées dans [73, 72].

Les dynamiques construites peuvent être utilisés dans le cadre d’un algorithme de Gibbs, permettant ainsi l’échantillonnage de la loi conjointe de la forme et des paramètres [T16, T21, T22, T25]. Dans notre cas, les paramètres sont spécifiés par une loi a priori, plaçant notre travail dans la perspective bayésienne [164, 162]. Nous souhaitons continuer dans cette direction en essayant d’intégrer à notre approche de nouvelles idées qui visent à améliorer les propriétés de l’algorithme de simulation [80].

Des stratégies particulières s’imposent quand le nombre d’objets de la configuration devient particulièrement important [60, 61]. Dans cette situation, la construction de noyaux de transition qui ajoutent et qui enlèvent plusieurs objets à la fois peut être envisagée. Il serait aussi peut être intéressant de s’inspirer des algorithmes de packing, même si leur construction s’appuie en général sur des principes différents de ceux des chaînes de Markov [26, 115, 182].

En général, le contrôle du comportement de ces dynamiques s’effectue en suivant l’évolution des séries chronologiques de statistiques exhaustives du modèle simulé. A ce propos, nous souhaitons enrichir notre arsenal de techniques en considérant comme point de départ les approches décrites dans [164].

### 8.3.2 Simulation exacte

Après la grande émulation de la fin des années 1990 et du début des années 2000, le nombre d’articles sur la simulation exacte est devenu moins important. Les explications données par certains sont la technicité du sujet, la difficulté d’y travailler avec des étudiants et le fait que le sujet soit à risque. Cependant, les perspectives que l’on peut déceler dans [107, 141, 164] sont bien réelles.

Du point de vue de la simulation des processus ponctuels marqués avec application à la détection des structures, nous sommes intéressés par des méthodes capables d’échantillonner des modèles à interactions fortes. Cela

signifie que nous avons besoin d’algorithmes flexibles pour simuler les modèles sur des gammes de paramètres les plus larges possibles.

Nous souhaitons poursuivre l’étude et l’adaptation théorique et pratique des algorithmes existants à notre problème. Un premier pas consisterait à regarder les performances de l’algorithme de Fill sur différents processus ponctuels marqués [64, 185, 138]. Une deuxième idée est d’étudier l’applicabilité aux processus ponctuels des idées de simulation parfaite dans le sens direct [201]. Finalement, une troisième piste serait de considérer les algorithmes de type ”simulated tempering” [58, 139]. Ces derniers algorithmes ont d’emblée un double avantage. D’abord, ils permettent de simuler des modèles à très basse température tout en réduisant les problèmes liés aux fortes corrélations entre échantillons et au faible mélange de la chaîne simulée. Ces algorithmes permettent aussi d’utiliser des algorithmes qui ont des noyaux de transition adaptés au modèle.

## 8.4 Questions concernant l’inférence statistique

La ”machinerie” que nous avons construite suppose les paramètres du modèle connus. Pour les paramètres d’interaction cette condition a été relâchée en les modélisant par une loi a priori. Mais dans les deux cas le problème de l’estimation des paramètres se pose.

### 8.4.1 Détection conjointe de la forme et des paramètres

Dans le cas de modèles à base de champs de Markov, les travaux de [106, 203, 204, 202] ont abordé simultanément les problèmes de la détection (ou restauration) de la structure et de l’estimation des paramètres. Le principe de ces méthodes est connu plus souvent sous le nom de gradient stochastique. Un lien entre les méthodes basées sur le gradient stochastique et les algorithmes de type EM a été établi par [106, 53]. Le gradient stochastique a été appliqué à l’estimation des paramètres de processus ponctuels par [144, 72].

Nous souhaitons étudier l’approche introduite par [106, 203] pour essayer de l’appliquer à notre cadre de détection de structures. Cela demandera de l’adapter au cadre des processus ponctuels. Un des points importants devrait être l’étude de la vitesse de convergence des dynamiques Monte Carlo mises en place pour simuler le modèle conditionnellement aux paramètres.



### 8.4.2 Simulation de la loi a posteriori

La simulation de la loi a posteriori  $p(\theta|\mathbf{x})$ , c'est-à-dire de la loi des paramètres conditionnellement à la forme observée est une question liée au problème précédent. En effet, nous sommes capables de simuler la loi de la forme conditionnellement aux paramètres du modèle,  $p(\mathbf{x}|\theta)$ . Si l'on est capable de simuler les deux lois conditionnelles, cela implique que l'on peut simuler la loi conjointe  $p(\mathbf{x}, \theta)$ , qui caractérise à la fois la forme et les paramètres.

La difficulté avec la simulation de la loi a posteriori est qu'elle met en jeu le calcul du rapport des constantes de normalisation. Ce calcul doit être fait à son tour par simulation. Une excellente idée qui élimine ce calcul est présentée dans [137]. Elle consiste à introduire des variables auxiliaires dans une dynamique de Metropolis-Hastings. L'incorporation dans le taux d'acceptation de la densité de probabilité de la variable auxiliaire simplifie son expression, au sens où le calcul des constantes de normalisation n'est plus nécessaire. De plus, le choix de variables auxiliaires est très large et très peu contraint par les hypothèses de travail.

Cette liberté de choix est à la fois le point fort et le point faible de la méthode. Du point de vue théorique, il est très facile de choisir une variable auxiliaire. De point de vue numérique, les choses sont un peu différentes, au sens où ce choix ne garantit pas forcément de bonnes propriétés de mélange pour la chaîne obtenue en incluant les variables auxiliaires. Concrètement, la chaîne peut rester bloquée dans le même état pendant de longs moments, chose qui nuit sérieusement à la qualité des échantillons obtenus. Les auteurs préviennent de la possibilité de cet effet non désiré sur certains modèles et proposent des heuristiques pour éviter le problème.

En s'inspirant de cette idée, nous avons proposé une méthode qui approxime la loi a posteriori d'une manière contrôlée [T18]. Les résultats numériques sont très prometteurs. Nous travaillons en ce moment pour trouver un encadrement théorique à la loi approximée [27, 28, 52]. Un tel résultat permettrait de transposer notre idée à d'autres techniques d'échantillonnage approximé [78].

### 8.4.3 Choix de modèle

Valider statistiquement des résultats demande soit la connaissance à priori de la forme qui se cache dans les données, soit être capable de simuler un champ de données connaissant la forme qui s'y trouve. En analyse d'image, la forme peut être considérée comme connue puisque on peut la voir. Cepen-

dant, si la structure de données est plus complexe, il devient plus difficile d'accéder à cette visualisation. Effectuer la démarche inverse, c'est-à-dire simuler des données en supposant connue la forme, peut se faire si l'on prend en compte la structure des données et le modèle de la forme. Ces considérations font que cette question reste pour le moment un problème ouvert.

L'échantillonnage de la loi a posteriori permet également effectuer des tests d'hypothèse en vue d'un choix de modèle.

Pour vérifier les résultats obtenus par notre méthode, nous avons proposé des tests de Monte Carlo basés sur les statistiques exhaustives du modèle. Sous l'hypothèse du modèle, ces tests ont montré qu'une structure était bien présente dans les données. Nous souhaitons poursuivre ce genre de tests en les enrichissant le plus possible pour pouvoir comparer des modèles.

L'intérêt pour ce problème a pris un essor particulier grâce à l'apparition de l'analyse par résidus des processus ponctuels [15, 142]. Les résidus sont définis comme la différence entre la somme d'une fonction aux points d'une configuration et l'intégrale de cette fonction sur tout le domaine. L'outil mathématique à la base de cette notion est la formule de Georgii-Nguyen-Zessin [70, 147]. Il existe également un lien fort entre cette méthode et l'estimation des paramètres par pseudovraisemblance. Les propriétés mathématiques des résidus ont été étudiées dans [12], alors que des théorèmes centraux limites ont été établis par [38, 9]. L'utilisation des résidus dans le contexte des processus ponctuels spatiotemporels est abordée dans [174, 205]

De point de vue applicatif, ce développement s'est concrétisé dans une série d'outils graphiques pour qualifier l'ajustement d'un modèle à un jeu de données. Cela facilite le choix d'un meilleur modèle [14, 8].

Notre perspective dans cette direction est d'adapter ces idées pour notre problème de reconnaissance de formes. Cela revient d'abord à détecter la structure, puis à estimer ses paramètres et finalement calculer des résidus pour valider le modèle. Il y a indéniablement un lien fort avec les questions liées à l'estimation des paramètres. Techniquement, cela peut demander d'incorporer le terme d'attache aux données dans la mesure de référence. Certains outils existants pourraient aussi être adaptés en vue de valider le terme d'attache aux données, en plus du terme d'interaction.

## 8.5 Projet : processus stochastiques et géométrie aléatoire

Considérons quelques situations :

- un réseau routier évolue en fonction de l'activité humaine et économique qu'il connecte - sur une durée de plusieurs années des routes apparaissent et disparaissent ;
- les fissures au sol se propagent et se ramifient en formant un réseau - cette structure évolue au cours du temps jusqu'à atteindre une certaine stabilité en fonction de facteurs environnementaux
- lors de la collision entre deux satellites, des débris spatiaux se forment - au tout début, ces débris se déplacent dans l'espace d'une manière plutôt agrégé - puis le mouvement se stabilise, les débris étant repartis sur certaines orbites.

A chaque instant, notre approche de détection et de caractérisation des formes s'applique à chacune de ces situations en répondant à la question "quelle est la forme qui se cache dans les données?". Cependant, la présence de la dimension temporelle apporte plus d'information et permet de poser des nouvelles questions, comme par exemple :

- comment se développent les réseaux routiers en fonction de la croissance des villes ou la diminution des forêts ?
- quelle est la probabilité qu'une fracture géologique atteigne une certaine location ?
- y a-t-il une corrélation entre l'emplacement d'un gisement de minerais et la position de fractures géologiques ?
- quelle est la carte de probabilité d'impact d'un nouveau satellite avec les débris existants ?

Les situations décrites précédemment sont des exemples de processus stochastiques spatiotemporels. A temps fixé, ces processus peuvent être décrits par les outils existants de la géométrie stochastique et des statistiques spatiales, en apportant ainsi des éléments de réponses aux questions formulées. Cependant, l'intégration de ces éléments nécessite un mécanisme mathématiquement robuste. L'ajout de la dimension temporelle impose naturellement d'adapter des outils en provenance de la théorie des processus stochastiques à la description de dynamiques d'objets géométriques aléatoires. Cette idée a été lancée par Yu. Davydov dans le groupe de travail "Géométrie Stochastique" de l'Université Lille 1. Et ceci constitue la direction de recherche majeure vers laquelle nous souhaitons nous orienter pour les années à venir.

Nous pensons démarrer sur cette nouvelle voie, en partant comme toujours d'applications et de questions concrètes. Pour répondre à ces questions, nous souhaitons construire des modèles, leur associer des dynamiques de simulation et bâtir des procédures d'inférence statistique. De cette manière, la re-

lation bénéfique d'interdépendance entre les mathématiques et ses domaines d'application se trouve renforcée.

### 8.5.1 Modélisation spatio-temporelle

Les bases théoriques des processus spatio-temporels sont décrites dans la littérature [45, 46]. Cependant, pour répondre aux questions pratiques il faut étudier des processus plus spécifiques. Nous avons actuellement deux familles de processus en vue : les processus ponctuels séquentiels et les processus de cristallisation.

Les processus ponctuels séquentiels sont des processus ponctuels pour lesquels compte l'ordre des points d'une configuration. Ils ont été introduits par M. N. M. van Lieshout. Ceci permet de s'affranchir des interactions symétriques et d'ajouter le temps comme caractéristique de la forme [194]. Les mesures de Campbell et les mesures moments de ces processus ont été calculées par [193] et une application des processus séquentiels en modélisation et analyse d'images vidéo est présentée dans [196]. Une autre application concerne les processus de type absorption stochastique. Par rapport aux processus spatiaux de naissance et de mort, l'absorption stochastique n'est qu'un processus de naissance. En plaçant l'absorption stochastique dans le cadre des processus séquentiels, on obtient un modèle sur lequel le maximum de vraisemblance peut être utilisé pour en estimer les paramètres [195].

Les processus de cristallisation sont des processus germination-croissance qui évoluent dans le temps. A partir d'une configuration de points initiaux auxquels on a attaché un temps de début de croissance, des formes géométriques appelées grains se développent dans le temps en occupant l'espace environnant. Un point et son grain sont appelés cristal, et le mécanisme de développement du grain est appelée cristallisation. La réalisation d'un tel processus peut être vue comme un ensemble aléatoire, ou champ aléatoire, ou encore pavage si l'on s'intéresse à l'union des frontières des cristaux. Pour une définition rigoureuse de ces processus, ainsi que pour des références, nous recommandons [49, 136]. Dans [49], la structure finale obtenue est un champ aléatoire qui assigne à chaque point de l'espace la valeur du premier temps à laquelle ce point est touché par un cristal. Sous des conditions assez générales portant sur la forme du cristal et sa vitesse de croissance, il est prouvé que ce champ est mélangeant. Ce résultat est appliqué pour estimer des paramètres du modèle.

Pour la suite, nous souhaitons étudier et appliquer ces processus à des problèmes concrets. Nous pensons que ces deux familles de processus offrent un cadre de modélisation commode pour introduire la composante tempo-

relle dans nos problèmes de détection et de caractérisation des structures. Nous sommes intéressés par la construction de statistiques spatio-temporelles qui permettent de décrire la forme observée. Dans le cas du processus de cristallisation, la considération de germes en interaction est une possible généralisation du modèle. Sur un plan théorique nous aimerions pouvoir établir des liens entre les processus séquentiels et les processus de cristallisation.

Au sujet de l'introduction du temps dans la modélisation d'une structure, il faut mentionner les excellents travaux effectués dans le groupe de recherche animé par E. B. Vedel-Jensen concernant les processus spatio-temporels de Lévy [89, 93, 94, 95]. Ces travaux sont plutôt tournés vers les processus de type Cox, alors que les dynamiques associées relèvent des processus de Lévy. Ils constituent un complément de notre approche et pourraient aider, inspirer voire accompagner d'une manière harmonieuse la voie que nous souhaitons aborder.

La difficulté principale de l'introduction du temps est que dans un espace spatio-temporel on ne peut pas forcément appliquer les mêmes principes de modélisation que sur l'espace des formes.

Une forme est vue comme une entité complexe construite à partir d'objets simples en interaction. Peut-on définir une structure dans le temps? Cette structure peut-elle être vue comme une entité composée d'objets simples? Si oui, quels sont ces objets simples? Ce sont des questions qui nous font tous un peu rêver. En même temps, il ne faut pas perdre de vue que des éléments de réponse à ces différents problèmes peuvent être apportés en considérant le temps comme une dimension supplémentaire de l'espace des formes.

### **8.5.2 Simulation de processus stochastiques des formes**

L'ajout d'une dimension temporelle à la modélisation des structures amène à des changements dans les méthodes de simulation, comme par exemple dans [172, 40]. D'autres méthodes, basées sur le filtrage particulière, devraient aussi être étudiées [34, 152].

Il est difficile de formuler des perspectives dans cette direction car la dimension temporelle pèse à la fois sur la définition du modèle et sur la complexité de la stratégie de simulation. Ce qui est clair, c'est que nous avons besoin de dynamiques de simulation capables de produire des échantillons de suffisamment bonne qualité pour pouvoir procéder à l'inférence statistique.

### 8.5.3 Statistiques pour les processus stochastiques des formes

L'utilisation des distributions de Palm et des capacités de Choquet comme outils de construction de statistiques descriptives de configurations de points est bien connue [190, 141, 180]. Nous souhaitons compléter notre approche de détection de structures avec ce genre de statistiques. Cela nous permettrait d'étudier des corrélations entre formes différentes, notamment les formes construites avec des objets de plusieurs types. La généralité mathématique de cette approche donnerait la possibilité de comparer des résultats provenant de différentes méthodes de détection.

Il semble que cette manière de construire des statistiques exploratoires peut être étendue à un espace de formes enrichi d'une dimension temporelle supplémentaire. Comme nous souhaitons ajouter une dimension temporelle aux modèles de formes que nous mettrons en place, cette même dimension temporelle devrait être ajoutée aux statistiques décrivant la forme. Une telle construction doit être réalisée avec prudence et ne doit pas être une extension "automatique" des outils mathématiques : une telle extension risquerait ne pas avoir de lien réel avec le phénomène étudié.

### 8.5.4 Outils probabilistes et d'inférence statistique pour caractériser le chaos

Le travail sur la dynamique des comètes se poursuit. En ce moment, nous appliquons la méthodologie proposée dans [T19] à des modèles plus simples sur lesquels on peut calculer un nombre important de grandeurs. Cela devrait nous permettre de simuler de nouvelles perturbations et de tester l'influence de ces perturbations sur la dynamique des comètes.

A long terme, nous souhaitons tester et consolider notre méthodologie pour qu'elle puisse appliquer à d'autres systèmes dynamiques perturbés.

# Bibliographie

- [1] T. Arak, P. Clifford, and D. Surgailis. Point-based polygonal models for random graphs. *Advances in Applied Probability*, 25(2) :348–372, 1993.
- [2] K. B. Athreya, H. Doss, and J. Sethuraman. On the convergence of the markov chain simulation method. *The Annals of Statistics*, 24 :69–100, 1996.
- [3] G. J. Babu and E. D. Feigelson. Spatial point processes in astronomy. *Journal of Statistical Planning and Inference*, 50 :311–326, 1996.
- [4] F. Baccelli and B. Blaszczyzyn. *Stochastic Geometry and Wireless Networks : Volume I Theory*. Now Publishers, 2010.
- [5] F. Baccelli and B. Blaszczyzyn. *Stochastic Geometry and Wireless Networks : Volume II Applications*. Now Publishers, 2010.
- [6] A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan (eds.). *Case studies in spatial point process modeling*. Lecture Notes in Statistics, Springer, 2006.
- [7] A. J. Baddeley. Spatial sampling and censoring. In O. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry, Likelihood and Computation*. CRC Press/Chapman and Hall, Boca Raton, 1999.
- [8] A. J. Baddeley. Validation of statistical models for spatial point patterns. In G. J. Babu and E. D. Feigelson, editors, *Statistical challenges in modern astronomy IV*, volume 371. Astronomical Society of Pacific, Conference Series, 2007.
- [9] A. J. Baddeley, J. F. Coeurjolly, E. Rubak, and R. P. Waagepetersen. A logistic regression estimating function for spatial gibbs point processes. *Research Report Centre for Stochastic Geometry and Advanced Bioimaging*, (2), 2013.
- [10] A. J. Baddeley and I. S. Molchanov. Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision*, 8 :79–92, 1998.

- [11] A. J. Baddeley and J. Møller. Nearest-neighbour markov point processes and random sets. *International Statistical Review*, 57(2) :89–121, 1989.
- [12] A. J. Baddeley, J. Møller, and A. G. Pakes. Properties of residual for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60 :627–649, 2008.
- [13] A. J. Baddeley and B. W. Silverman. A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, 40 :1089–1093, 1984.
- [14] A. J. Baddeley and R. Turner. **spatstat** : an **r** package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6) :1–42, 2005.
- [15] A. J. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society, Series B*, 67(5) :617–666, 2005.
- [16] A. J. Baddeley and T. R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3) :283–322, 2000.
- [17] A. J. Baddeley and M. N. M. van Lieshout. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47 :601–619, 1995.
- [18] A. A. Barker. Monte Carlo calculations of the radial distribution function for a proton electron plasma. *Australian Journal of Physics*, 18 :119–133, 1965.
- [19] O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout (eds.). *Stochastic Geometry Likelihood and Computation*. Chapman and Hall/CRC, Boca Raton, 1999.
- [20] M. Berman and T. R. Turner. Approximating point process likelihoods with glim. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41 :31–38, 1992.
- [21] K. K. Berthelsen and J. Møller. A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society*, 33 :351–367, 2002.
- [22] K. K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30 :549–564, 2003.
- [23] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2) :192–236, 1974.



- [24] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24(3) :179–195, 1975.
- [25] J. Besag, R. Milne, and S. Zachary. Point process limits of lattice processes. *Journal of Applied Probability*, 19(1) :210–216, 1982.
- [26] A. Bezrukov, D. Stoyan, and M. Bargiel. Spatial statistics for simulated packings of spheres. *Image Analysis & Stereology*, 20 :203–206, 2001.
- [27] P. Billingsley. *Probability and Measure*. John Wiley and Sons, 1995.
- [28] P. Billingsley. *Convergence of Probability Measures. Second Edition*. John Wiley and Sons, 1999.
- [29] J.-M. Billiot, J.-F. Coeurjolly, and R. Drouilhet. Maximum pseudo-likelihood estimator for exponential family models of marked gibbs point processes. *Electronic Journal of Statistics*, 2 :234–264, 2008.
- [30] A. Brix and J. Chadoeuf. Spatio-temporal modelling of weeds by shot-noise g cox processes. *Biometrical Journal*, 44(1) :83–99, 2001.
- [31] A. Brix and J. Møller. Space-time multi type log gaussian cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics*, 28(3) :471–488, 2001.
- [32] R. F. Brøndum and E Rubak. *Stochastic channel modelling : a Bayesian approach using reversible jump Markov chain Monte Carlo methods*. Master thesis - University of Aalborg, 2006.
- [33] O. Cappé, C. P. Robert, and T. Rydén. Reversible jump, birth-and-death and more general continuous time markov chain monte carlo samplers. *Journal of the Royal Statistical Society, Series B*, 65(3) :679–700, 2003.
- [34] F. Caron, P. del Moral, A. Doucet, and M. Pace. On the conditional distributions of spatial point processes. *Advances in Applied Probability*, 43(2) :301–307, 2011.
- [35] Y. C. Chin and A. J. Baddeley. On connected component markov point processes. *Advances in Applied Probability*, 31(2) :279–282, 1999.
- [36] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications. Third Edition*. John Wiley and Sons, 2013.
- [37] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, Paris, 1994.
- [38] J.-F. Coeurjolly and F. Lavancier. Residuals and goodness-of-fit tests for stationary marked gibbs point processes. *Journal of the Royal Statistical Society, Series B*, 75(2) :247–276, 2013.
- [39] D. Coupier and D. Dereudre. Continuum percolation for the Quermass model. *Preprint ArXiv*, 1201.2344v1 :-, 2012.

- [40] O. Cronie. *Different Aspects of Inference for Spatio-Temporal Point Processes*. PhD Thesis - Chalmers University of Technology and University of Gothenburg, 2010.
- [41] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25 :2300–2312, 1997.
- [42] A. Cuevas, W. Gonzalez-Manteiga, and A. Rodriguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48 :7–19, 2006.
- [43] D. J. Daley and G. Last. Descending chains, the lilypond model, and mutual-nearest-neighbour matching. *Advances in Applied Probability (SGSA)*, 37(3) :604–628, 2005.
- [44] D. J. Daley, H. Stoyan, and D. Stoyan. Fraction of a poisson germ model with maximally non-overlapping spherical grains. *Advances in Applied Probability (SGSA)*, 31(3) :610–624, 1999.
- [45] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume 1 : Elementary Theory and Methods, Second Edition*. Springer, 2003.
- [46] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume 2 : General Theory and Structure, Second Edition*. Springer, 2008.
- [47] H. A. David. *Order Statistics*. John Wiley and Sons, 1981.
- [48] Yu. Davydov. On convex hull of d-dimensional fractional Brownian motion. *Statistics and Probability Letters*, 82(1) :37–39, 2012.
- [49] Yu. Davydov and A. Illig. Ergodic properties of crystallization processes. *Journal of Mathematical Sciences*, 163(4) :375–381, 2009.
- [50] Yu. Davydov and V. Paulauskas. On the estimation of the parameters of multivariate stable distributions. *Acta Applicandae Mathematicae*, 58 :107–124, 1999.
- [51] Yu. Davydov and V. Paulauskas. Recent results on p-stable convex compact sets with applications. *Fields Institute Communications*, 44 :127–139, 2004.
- [52] Yu. Davydov and V. Rotar. On asymptotic proximity of distributions. *Journal of Theoretical Probability*, 22(1) :82–98, 2009.
- [53] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1) :94–128, 1999.
- [54] D. Dereudre. Gibbs delaunay tessellations with geometric hardcore conditions. *Journal of Statistical Physics*, 131(1) :127–152, 2008.
- [55] D. Dereudre, R. Drouilhet, and H.-O. Georgii. Existence of gibbsian point processes with geometry dependent interactions. *Probability Theory and Related Fields*, 153(3-4) :643–670, 2012.

- [56] R. Deriche and O. Faugeras. Les edp en traitement des images et vision par ordinateur. *Rapport de Recherche INRIA*, (2697), 1995.
- [57] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod. Estimation of markov random field prior parameters using markov chain monte carlo maximum likelihood. *IEEE Transactions on Image Processing*, 8 :954–963, 1999.
- [58] G. Döge, K. Mecke, J. Møller, and R.P. Waagepetersen. Grand canonical simulations of hard-disk systems by simulated tempering. *International Journal of Modern Physics C*, 15(1) :129–147, 2004.
- [59] I. L. Dryden, M. R. Scarr, and C. C. Taylor. Bayesian texture segmentation of weed and crop images using reversible jump markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 52(1) :31–50, 2003.
- [60] X. Descombes (ed.). *Applications de la Géométrie Stochastique à l'Analyse d'Images (Série Signal et Image IC2)*. Hermès Lavoisier, 2011.
- [61] X. Descombes (ed.). *Stochastic Geometry for Image Analysis (Digital Signal and Image Processing series)*. John Wiley and Sons, 2012.
- [62] R. J. Erskine. Mastitis control in dairy herds. In O.M. Radostis, editor, *Herd health : food animal production medecine*. W.B. Saunders Company, Philadelphia, 2001.
- [63] P. A. Fernandez and R. Fernandez. Perfect simulation for interacting point processes, loss networks and Ising models. *Stochastic Processes and their Applications*, 102 :63–88, 2002.
- [64] J. A. Fill. An interruptible algorithm for perfect sampling via Markov chains. *The Annals of Applied Probability*, 8(1) :131–162, 1998.
- [65] D. Flipo. *Chaînes de Markov. Agrégation de Mathématiques*. Cours photocopié, Université Lille 1, 2007.
- [66] C. Garban and J.A. Trujillo Ferreras. The expected area of the filled planar Brownian loop is  $\pi/5$ . *Communication in Mathematical Physics*, 264 :797–810, 2006.
- [67] A. E. Gelfand and B. P. Carlin. Maximum-likelihood estimation for constrained- or missing- data models. *Canadian Journal of Statistics*, 21 :303–311, 1993.
- [68] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp (eds.). *Handbook of Spatial Statistics*. Chapman and Hall/CRC, Boca Raton, 2010.
- [69] S. Gemand and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.

- [70] H.-O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics*, 48 :31–51, 1976.
- [71] C. J. Geyer. On the convergence of monte carlo maximum likelihood calculations. *Journal of Royal Statistical Society, Series B*, 54(1) :261–274, 1994.
- [72] C. J. Geyer. Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry, Likelihood and Computation*. CRC Press/Chapman and Hall, Boca Raton, 1999.
- [73] C. J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21(4) :359–373, 1994.
- [74] C. J. Geyer and E. A. Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of Royal Statistical Society, Series B*, 54(3) :657–699, 1992.
- [75] M. Goulard, A. Särkä, and P. Grabarnik. Parameter estimation for marked gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, 23(3) :365–379, 1996.
- [76] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4) :711–732, 1995.
- [77] P. J. Green and S. Richardson. Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, 97(460) :1055–1070, 2002.
- [78] A. Grelaud, C. P. Robert, J.-M. Marin, F. Rodolphe, and J.-F. Taly. ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, 4(2) :317–336, 2009.
- [79] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society, Series B*, 56(4) :549–603, 1994.
- [80] J. E. Griffin and S. G. Walker. On adaptive metropolis-hastings methods. *Statistics and Computing*, 23(1) :123–134, 2013.
- [81] H. Haario and E. Saksman. Simulated annealing process in general state space. *Advances in Applied Probability*, 23(4) :866–893, 1991.
- [82] O. Häggström and R. Meester. Nearest neighbor and hard sphere models in continuum percolation. *Random Structures & Algorithms*, 9(3) :295–315, 1996.
- [83] O. Häggström and K. Nelander. Exact sampling for from anti-monotone systems. *Statistica Neerlandica*, 52(3) :360–380, 1998.

- [84] O. Häggström and K. Nelander. On exact simulation of markov random fields using coupling from the past. *Scandinavian Journal of Statistics*, 26(3) :395–411, 1999.
- [85] O. Häggström, M. N. M. van Lieshout, and J. Møller. Characterization results and markov chain monte carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 5(4) :641–658, 1999.
- [86] B. Hajek. Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2) :311–329, 1988.
- [87] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.
- [88] D. Hearn and M. P. Baker. *Computer Graphics. Second Edition*. Prentice-Hall, 1994.
- [89] G. Hellmund, M. Prokešová, and E. B. Vedel Jensen. Lévy-based cox point processes. *Advances in Applied Probability (SGSA)*, 40(3) :603–629, 2008.
- [90] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, 2008.
- [91] J. Jensen and H. R. Kunsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46 :475–486, 1994.
- [92] J. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability*, 1(3) :445–461, 1991.
- [93] K. Yr. Jonsdottir, U. Hahn, and E. B. Vedel Jensen. Inhomogeneous spatial point processes, with a view to spatio-temporal modelling. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Spatial point process modelling and its applications*. Publicacions de la Universitat Jaume I, Castellon, Spain, 2004.
- [94] K. Yr. Jonsdottir and E. B. Vedel Jensen. Gaussian radial growth. *Image Analysis & Stereology*, 24(2) :117–126, 2005.
- [95] K. Yr. Jonsdottir, J. Schmiegel, and E. B. Vedel Jensen. Lévy-based growth models. *Bernoulli*, 14(1) :62–90, 2008.
- [96] J. Kahn. Existence of Gibbs measure for a model of T-tessellation. *arXiv :1012.2182v1*, pages –, 2010.
- [97] F. P. Kelly and B. D. Ripley. A note on strauss’s model for clustering. *Biometrika*, 63 :357–360, 1976.
- [98] W. S. Kendall and J. Møller. Perfect metropolis-hastings simulation of locally stable point processes. *Research Report R-99-2001, Department of Mathematical Sciences, Aalborg University*, 1999.

- [99] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability (SGSA)*, 32 :844–865, 2000.
- [100] W. S. Kendall and G. Montana. Small sets and markov transition densities. *Stochastic Processes and their Applications*, 99(2) :177–194, 2002.
- [101] J. F. C. Kingman. *Poisson processes*. Clarendon Press, Oxford, 1993.
- [102] J. F. C. Kingman. Poisson processes revisited. *Probability and Mathematical Statistics*, 26(1) :77–95, 2006.
- [103] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598) :671–680, 1983.
- [104] S. A. Kovyazin. On the limit behavior of a class of empirical means of a random set. *Theory of Probability and its Applications*, 30(4) :814–820, 1986.
- [105] C. Lacoste, X Descombes, and J. Zerubia. Point processes for unsupervised line network extraction in remote sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10) :1568–1579, 2005.
- [106] S. Lakshmanan and H. Derrin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8) :799–813, 1989.
- [107] C. Lantuéjoul. *Geostatistical Simulation Models and Algorithms*. Springer-Verlag, Berlin, 2002.
- [108] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Preprint ArXiv*, 1205.4v3 :-, 2013.
- [109] G. Lawler, O. Schramm, and W. Werner. The dimension of the planar Brownian frontier is  $4/3$ . *Mathematical Research Letters*, 8 :401–411, 2001.
- [110] G. Lawler and W. Werner. The brownian loop soup. *Probability Theory and Related Fields*, 128 :565–588, 2004.
- [111] A.B. Lawson and D.G.T. Denison. Spatial cluster modelling : an overview. In A.B. Lawson and D.G.T. Denison, editors, *Spatial cluster modelling*. Chapman and Hall/CRC, 2002.
- [112] A.B. Lawson and M. Kulldorff. A review of cluster detection methods. In A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J-F. Viel, and R. Bertollini, editors, *Disease mapping and risk assessment for public health*. John Wiley and Sons, 1999.
- [113] S. Z. Li. *Markov Random Field Modeling*. Springer-Verlag, London, 2009.
- [114] S. Liu. *Lois stables et processus ponctuels : liens et estimation des paramètres*. PhD thesis - Université Lille 1, 2010.

- [115] K. Lochmann, A. Anikeenko, A. Elsner, N. Medvedev, and D. Stoyan. Statistical verification of crystallization in hard sphere packings under densification. *The European Physical Journal B*, 53 :67–76, 2006.
- [116] V. J. Martinez and E. Saar. *Statistics of the galaxy distribution*. Chapman and Hall, 2002.
- [117] S. Mase. Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbsian processes. *The Annals of Applied Probability*, 5(3) :603–612, 1995.
- [118] S. Mase. Marked gibbs processes and asymptotic normality of maximum pseudo-likelihood estimators. *Mathematische Nachrichten*, 209(1) :151–169, 2000.
- [119] S. Mase, J. Møller, D. Stoyan, R. P. Waagepetersen, and G. Döge. Packing densities and simulated tempering for hard core Gibbs point processes. *Annals of the Institute of Statistical Mathematics*, 53(4) :661–680, 2001.
- [120] D. Mason and W. Polonik. Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability*, 19(3) :1108–1142, 2009.
- [121] J. Mateu and F. Montes. Pseudo-likelihood inference for Gibbs processes with exponential families through generalized linear models. *Statistical Inference for Stochastic Processes*, 4(2) :125–154, 2001.
- [122] Georges Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.
- [123] K. Matthes, J. Kerstan, and J. Mecke. *Infinitely Divisible Point Processes*. John Wiley and Sons, 1978.
- [124] P. McCullagh and J. Møller. The permanental process. *Advances in Applied Probability (SGSA)*, 38(4) :873–888, 2006.
- [125] R. Meester and R. Roy. *Continuum percolation*. Cambridge University Press, 1996.
- [126] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1) :101–121, 1996.
- [127] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6) :1087–1092, 1953.
- [128] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2009.
- [129] R. E. Miles and M. S. Mackisack. A large class of random tessellations with the classic Poisson polygon distributions. *Forma*, 17 :1–17, 2002.
- [130] I. Molchanov. Uniform laws of large numbers for empirical associated functionals of random closed sets. *Theory of Probability and its Applications*, 32 :556–560, 1987.

- [131] I. Molchanov. Empirical estimation of distribution quantiles of random closed sets. *Theory of Probability and its Applications*, 35 :594–600, 1990.
- [132] I. Molchanov. *Statistics of the Boolean Model for Practitioners and Mathematicians*. John Wiley and Sons, 1997.
- [133] I. Molchanov. A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*, 25 :235–242, 1998.
- [134] I. Molchanov. Random closed sets : results and problems. In O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic geometry, likelihood and computation*. CRC Press/Chapman and Hall, Boca Raton, 1999.
- [135] I. Molchanov. *Theory of random sets*. Springer, 2005.
- [136] J. Møller. Generation of johnson-mehl crystals and comparative analysis of models for random nucleation. *Advances in Applied Probability*, 27(2) :367–383, 1995.
- [137] J. Møller, A. N. Pettitt, R. W. Reeves, and K. K. Berthelsen. An efficient Markov chain Monte Carlo method for distributions with intractable normalizing constants. *Biometrika*, 93 :451–458, 2006.
- [138] J. Møller and K. Schladitz. Extensions of fill’s algorithm for perfect simulation. *Journal of the Royal Statistical Society, Series B*, 61(4) :955–969, 1999.
- [139] J. Møller and K. Schladitz. Perfect simulation for sample-based inference. *Research Report, Department of Mathematical Sciences, Aalborg University*, R-99-2011, 1999.
- [140] J. Møller and R. P. Waagepetersen. Markov connected component fields. *Advances in Applied Probability*, 30(1) :1–35, 1998.
- [141] J. Møller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, 2004.
- [142] J. Møller and R. P. Waagepetersen. Modern statistics for spatial point processes (with discussion). *Scandinavian Journal of Statistics*, 34(4) :643–711, 2007.
- [143] A. Monfort. *Cours de statistique mathématique*. Economica, 1997.
- [144] R. A. Moeed and A. J. Baddeley. Stochastic approximation of the MLE for a spatial point pattern. *Scandinavian Journal of Statistics*, 18 :39–50, 1991.
- [145] D. J. Murdoch and P. J. Green. Exact sampling from continuous state space. *Scandinavian Journal of Statistics*, 25(3) :483–502, 1998.
- [146] J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B*, 20(1) :1–43, 1958.



- [147] X. X. Nguyen and H. Zessin. Integral and differential characterizations of the Gibbs process. *Mathematische Nachrichten*, 88 :105–115, 1979.
- [148] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, 1984.
- [149] Y. Ogata and M. Tanemura. Estimation for interaction potentials of spatial point processes through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics*, 33 :315–338, 1981.
- [150] Y. Ogata and M. Tanemura. Likelihood analysis of spatial point patterns. *Journal of Royal Statistical Society, Series B*, 46 :496–518, 1984.
- [151] Y. Ogata and M. Tanemura. Likelihood estimation of soft-core interaction potentials for gibbsian point patterns. *Annals of the Institute of Statistical Mathematics*, 41 :583–600, 1989.
- [152] M. Pace. *Méthodes et modèles stochastiques pour le suivi multi-objets*. PhD thesis - Université de Bordeaux I, 2011.
- [153] B. Pandey and S. Bharadwaj. A two-dimensional analysis of percolation and filamentarity. *Monthly Notices of the Royal Astronomical Society*, 357 :1068–1076, 2005.
- [154] W. H. Press, B. P. Flannery, and S. A. Teukolsky. *Numerical Recipes in C : the art of scientific computing*. Cambridge Univeristy Press, 1988.
- [155] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9 :223–252, 1996.
- [156] J. Randon-Furling, S.N. Majumdar, and A. Comtet. Convex hull of N planar Brownian motions : exact results and an application to Ecology. *Physical Review Letters*, 103 :140602, 2009.
- [157] R.-D. Reiss. *A Course on Point Processes*. Springer-Verlag, New-York, 1993.
- [158] P. Rigollet and R. Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15 :1154–1178, 2009.
- [159] B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B*, 39 :172–212, 1977.
- [160] B. D. Ripley. *Statistical inference for spatial processes*. Cambridge University Press, 1988.
- [161] B. D. Ripley and F. P. Kelly. Markov point processes. *Journal of the London Mathematical Society*, 15 :188–192, 1977.
- [162] C. P. Robert. *The Bayesian choice. From decision-theoretic foundations to computational implementation - Second Edition*. Springer, 2007.

- [163] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, New-York, 1999.
- [164] C. P. Robert and G. Casella. *Monte Carlo statistical methods - Second Edition*. Springer, 2004.
- [165] G. O. Roberts and A. F. M. Smith. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications*, 49 :207–216, 1994.
- [166] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1) :95–110, 1996.
- [167] L. Roques and F. Hamel. Mathematical analysis of the optimal habitat configurations for species persistence. *Mathematical Biosciences*, 210 :34–59, 2007.
- [168] H. Rue and M. Hurn. Bayesian object identification. *Biometrika*, 3 :649–660, 1999.
- [169] H. Rue and O.K. Husby. Identification of partly destroyed objects using deformable templates. *Statistics and Computing*, 8 :221–228, 1998.
- [170] H. Rue and A.R. Syversveen. Bayesian object recognition with baddeley’s delta loss. *Advances in Applied Probability (SGSA)*, 30 :64–84, 1998.
- [171] D. Ruelle. *Statistical Mechanics : Rigorous Results*. Imperial College Press, World Scientific Publishing,, 1999.
- [172] A. Särkkä and E. Renshaw. The analysis of marked point patterns evolving through space and time. *Computational Statistics and Data Analysis*, 51 :1698–1718, 2006.
- [173] D. A. Saunders, R. J. Hobbs, and C. R. Margules. Biological consequences of ecosystem fragmentation : a review. *Conservation Biology*, 5 :18–32, 1991.
- [174] F. P. Schoenberg. Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464) :789–795, 2003.
- [175] J. Serra. *Image analysis and mathematical morphology*. Academic Press, 1982.
- [176] A. Singh, C. Scott, and R. Nowak. Adaptive Hausdorff estimation of density level sets. *Annals of Statistics*, 37 :2760–2782, 2009.
- [177] A.V. Skorokhod. Asymptotic formulas for stable distribution laws. In *Selected Translations in Mathematical Statistics and Probability 1*, pages 157–161, 1961.

- [178] A. G. Steenbeek, M. N. M. van Lieshout, R. S. Stoica, P. Gregori, and K. K. Berthelsen. Mppplib, a c++ library for marked point processes. *CWI*, 2003.
- [179] R. S. Stoica. *Processus Ponctuels pour l'Extraction de Réseaux Linéiques*. PhD thesis - Université de Nice Sophia-Antipolis, 2001.
- [180] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley and Sons, 1995.
- [181] D. Stoyan and I. S. Molchanov. Set-valued means of random particles. *Journal of Mathematical Imaging and Vision*, 7 :111–121, 1997.
- [182] D. Stoyan, A. Wagner, H. Hermann, and A. Elsner. Statistical characterization of the pore space of random systems of hard spheres. *Journal of Non-Crystalline Solids*, 357 :1508–1515, 2011.
- [183] D. J. Strauss. A model for clustering. *Biometrika*, 62 :467–475, 1975.
- [184] E. Tempel, R. S. Stoica, and E. Saar. Evidence for spin alignment of spiral and elliptical galaxies in filaments. *Monthly Notices of the Royal Astronomical Society*, 428 :1827–1836, 2013.
- [185] E. Thönnnes. Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability*, 31(1) :69–87, 1999.
- [186] E. Thönnnes. A primer on perfect simulation. In K. R. Mecke and D. Stoyan, editors, *Statistical physics and spatial statistics*. Springer - Verlag, Berlin, 2000.
- [187] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4) :1701–1728, 1994.
- [188] A. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25 :948–969, 1997.
- [189] M. N. M. van Lieshout. Stochastic annealing for nearest neighbour point processes with application to object recognition. *Advances in Applied Probability*, 26 :281–300, 1994.
- [190] M. N. M. van Lieshout. *Markov Point Processes and their Applications*. Imperial College Press, London, 2000.
- [191] M. N. M. van Lieshout. Propagation of spatial interaction under superposition. In G. B. M. Heuvelink and M. J. P. M. Lemmens, editors, *Accuracy 2000. Proceedings of the 4th International Symposium on Spatial Accuracy Assesment in Natural Ressources and Environmental Sciences*, pages 687–694, Delft, 2000. Delft University Press.
- [192] M. N. M. van Lieshout. *Private communication*. CWI Amsterdam, 2002.
- [193] M. N. M. van Lieshout. Campbell and moment measures for finite sequential spatial processes. In M. Hušková and M. Janžura, editors,

- Proceedings of the 6th International Conference on Stereology, Spatial Statistics and Stochastic Geometry*, page 215–224, Prague, 2006. Matfyzpress.
- [194] M. N. M. van Lieshout. Markovianity in space and time. In D. Denteener, F. den Hollander, and E. Verbitskiy, editors, *Dynamics & Stochastics : Festschrift in honor of M. S. Keane*, pages 154–168. IMS Lecture Notes - Monograph Series, 2006.
- [195] M. N. M. van Lieshout. Maximum likelihood estimation for random sequential adsorption. *Advances in Applied Probability (SGSA)*, 38(4) :889–898, 2006.
- [196] M. N. M. van Lieshout. Depth map calculation for a variable number of moving objects using Markov sequential object processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7) :1308–1312, 2008.
- [197] M. N. M. van Lieshout. Spatial point process theory. In A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, editors, *Handbook of Spatial Statistics*. Chapman and Hall/CRC, 2010.
- [198] M. N. M. van Lieshout and A. J. Baddeley. Extrapolating and interpolating spatial patterns. In A. B. Lawson and D. G. T. Denison, editors, *Spatial cluster modelling*. CRC Press/Chapman and Hall, Boca Raton, 2002.
- [199] M. N. M. van Lieshout and E. W. van Zwet. Exact sampling from conditional Boolean models with applications to maximum likelihood inference. *Advances in Applied Probability*, 33 :339–353, 2001.
- [200] B. Widom and J. S. Rowlinson. New model for the study of liquid-vapor phase transition. *The Journal of Chemical Physics*, 52 :1670–1684, 1970.
- [201] D. B. Wilson. How to couple from the past using a read-once source of randomness. *Random Structures & Algorithms*, 16(1) :85–113, 2000.
- [202] G. Winkler. *Image analysis, random fields and Markov chain Monte Carlo methods (second edition)*. Springer, 2003.
- [203] L. Younes. Estimation and annealing for gibbsian fields. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 24 :269–294, 1988.
- [204] L. Younes. Parametric inference for imperfectly observed Gibbsian fields. *Probability Theory and Related Fields*, 82 :625–645, 1989.
- [205] J. Zhuang. Second-order residual analysis of spatio-temporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society, Series B*, 6(4) :635–653, 2006.