

# THÈSE

présentée à

l'Université de Nice-Sophia Antipolis

École Doctorale : Sciences et Technologies de l'Information et de la  
Communication

pour obtenir le titre de

DOCTEUR

Spécialité : Traitement des Images

## Processus Ponctuels pour l'Extraction de Réseaux Linéïques dans les Images Satellitaires et Aériennes

par

Radu STOICA

Soutenue le 9 Février 2001 devant la commission composée de :

Président	M. Barlaud	UNSA
Rapporteurs	H. Le Men	IGN
	M.N.M. van Lieshout	CWI Amsterdam
	M. Schmitt	ENS Mines de Paris
Examineurs	M. Sigelle	ENST Paris
	X. Descombes	INRIA
	J. Zerubia	INRIA



# Table des matières

<b>Introduction</b>	<b>3</b>
<b>Chapitre 1. Etat de l'art</b>	<b>7</b>
1.1 Approches orientées contour et autres opérateurs . . . . .	9
1.1.1 Modélisation probabiliste des routes . . . . .	10
1.1.2 Banc de filtres. Décomposition multirésolution . . . . .	10
1.1.3 Lignes de crête. Vallées. . . . .	11
1.1.4 Réseaux de neurones . . . . .	11
1.1.5 Programmation dynamique . . . . .	12
1.1.6 Théorie de l'information . . . . .	13
1.1.7 Contours actifs . . . . .	14
1.2 Approches orientées région . . . . .	14
1.3 Apport de données externes . . . . .	16
1.4 Approche par processus objet . . . . .	18
<b>Chapitre 2. Fondements des Processus Ponctuels</b>	<b>21</b>
2.1 Rappel sur l'Intégration . . . . .	22
2.2 Processus Ponctuels . . . . .	23
2.2.1 Processus ponctuels marqués . . . . .	25
2.2.2 Processus ponctuels de Gibbs . . . . .	26
2.2.3 Processus ponctuels de Markov . . . . .	27
2.2.4 Stabilité des processus ponctuels . . . . .	28
2.2.5 Exemples et commentaires . . . . .	29
<b>Chapitre 3. Chaînes de Markov. Méthodes MCMC.</b>	<b>31</b>
3.1 Définition des Chaînes de Markov . . . . .	32
3.2 Propriétés des Chaînes de Markov . . . . .	33
3.2.1 Irréductibilité . . . . .	33

3.2.2	Atomes et ensembles petits . . . . .	34
3.2.3	Cycle et apériodicité . . . . .	35
3.2.4	Récurrence au sens de Harris . . . . .	35
3.2.5	Ergodicité . . . . .	36
3.2.6	Convergence géométrique . . . . .	37
3.3	Méthodes du type Monte Carlo par Chaînes de Markov . . . . .	39
3.3.1	Combinaison de mises-à-jour . . . . .	39
3.3.2	Metropolis-Hastings . . . . .	41
3.3.3	Metropolis-Hastings une variable à la fois . . . . .	42
3.3.4	Echantillonneur Gibbs . . . . .	42
3.3.5	Metropolis-Hastings-Green . . . . .	43
3.4	Simulation des processus ponctuels . . . . .	44
<b>Chapitre 4. "Candy" modèle</b>		<b>47</b>
4.1	Modèle probabiliste d'un réseau linéique . . . . .	49
4.1.1	"Candy" modèle : processus ponctuel de Gibbs . . . . .	54
4.1.2	Stabilité locale du "Candy" modèle . . . . .	56
4.2	Simulation du "Candy" model . . . . .	58
4.2.1	Naissance et mort d'un segment libre . . . . .	59
4.2.2	Naissance et mort d'un segment avec une connexion singulière . . . . .	60
4.2.3	Naissance et mort d'un segment avec une connexion simple . . . . .	62
4.2.4	Naissance et mort d'un segment avec une connexion double . . . . .	63
4.3	Choix des paramètres . . . . .	64
4.4	Réalisations . . . . .	66
<b>Chapitre 5. Etude de la dynamique du "Candy" modèle</b>		<b>69</b>
5.1	Irreductibilité. Récurrence. Ergodicité. . . . .	71
5.2	Convergence. Vitesse de mélangeance . . . . .	71
5.3	Taux d'acceptation. Choix des mouvements. . . . .	80
<b>Chapitre 6. Extraction de réseaux routiers en utilisant le "Candy" modèle</b>		<b>85</b>
6.1	Position du problème . . . . .	87
6.2	Energie d'attache aux données . . . . .	88
6.2.1	Définition . . . . .	88
6.2.2	Stabilité locale du "Candy" modèle complet . . . . .	96
6.3	Résultats . . . . .	96
6.4	Schéma de descente de température . . . . .	106

---

6.5	Validation des résultats . . . . .	108
6.5.1	Généralités . . . . .	108
6.5.2	Fôrets galleries . . . . .	108
<b>Chapitre 7.</b>	<b>Estimation des paramètres</b>	<b>113</b>
7.1	Fonction de vraisemblance . . . . .	115
7.2	Modification du "Candy" modèle . . . . .	117
7.3	Log-vraisemblance du "Candy" modèle . . . . .	119
7.4	Algorithme d'estimation . . . . .	121
7.5	Résultats . . . . .	122
<b>Conclusion</b>		<b>129</b>
<b>Références bibliographiques</b>		<b>141</b>
<b>Voila les gens !</b>		<b>143</b>



# Liste des figures

4.1	Région de répulsion autour du segment $\overline{QT}$ . . . . .	51
4.2	Interaction de type répulsion entre les segments . . . . .	52
4.3	Région d'attraction autour d'un segment . . . . .	52
4.4	Interaction de type attraction entre les segments . . . . .	53
4.5	Segments et leur différentes orientations . . . . .	54
4.6	Etude des paramètres $\omega_o$ et $\omega_r$ : a) configuration $S_a$ , b) configuration $S_b$ . .	65
4.7	Etude des paramètres $\omega_c$ et $\omega_f$ : a) configuration $S_a$ , b) configuration $S_b$ . .	66
4.8	Réalisations du "Candy" modèle sur une image de $341 \times 300$ pixels . . . . .	67
5.1	Etude de la dynamique en fonction du paramètre $\omega_c$ . . . . .	73
5.2	Etude de la dynamique en fonction du paramètre $\omega_f$ . . . . .	74
5.3	Etude de la dynamique en fonction du paramètre $\omega_o$ . . . . .	75
5.4	Etude de la dynamique en fonction du paramètre $\omega_r$ . . . . .	76
5.5	Etude de la dynamique en fonction du paramètre $\omega_t$ . . . . .	77
5.6	Etude de la dynamique en fonction des probabilités de choisir les mouvements	81
6.1	Segment dans une image . . . . .	89
6.2	Hypothèse $H_3$ : trois régions différentes . . . . .	90
6.3	Hypothèse $H_2$ : deux régions différentes . . . . .	92
6.4	Hypothèse $H_1$ : une région homogène . . . . .	93
6.5	Résultat de l'attache aux données sur une image SPOT . . . . .	95
6.6	Image SPOT T1, utilisée comme échantillon d'entraînement . . . . .	97
6.7	Détection des routes sur l'image T1 avec comme température initiale $H = 100.0$	99
6.8	Détection des routes sur l'image T1 avec comme température initiale $H = 50.0$	100
6.9	Extraction des routes : a) images SPOT T5 ; b) réseau routier détecté . . .	101
6.10	Extraction des routes : a) images SPOT T3 ; b) réseau routier détecté . . .	102
6.11	Extraction des routes sur une image aérienne : a) image initiale ; b) réseau routier détecté . . . . .	103

---

6.12	Image SPOT de la ville de Mexico . . . . .	104
6.13	Réseau routier détecté . . . . .	105
6.14	Extraction des routes sur une image radar ERS : a) image initiale ; b) réseau routier détecté . . . . .	107
6.15	Extraction des forets galleries : a) image SPOT ; b) vérité de terrain . . . . .	110
6.16	Résultat de l'extraction : a) $\omega_c = 2.5$ ; b) $\omega_c = 5.0$ ; c) $\omega_c = 7.5$ . . . . .	111
7.1	Réalisations du "Candy" modèle modifié . . . . .	120
7.2	Réalisation du modèle à estimer . . . . .	123
7.3	Estimation des paramètres à partir des moyennes des statistiques suffisantes	124
7.4	Estimation des paramètres à partir des moyennes des statistiques suffisantes	124
7.5	Estimation des paramètres à partir des moyennes des statistiques suffisantes	125
7.6	Estimation des paramètres à partir des statistiques suffisantes . . . . .	126
7.7	Estimation des paramètres à partir des statistiques suffisantes . . . . .	126
7.8	Estimation des paramètres à partir des statistiques suffisantes . . . . .	127



# Ars poetica

---

A partir de la deuxième moitié du XXème siècle, la puissance de calcul des ordinateurs ainsi que le marché des images aériennes et plus tard satellitaires n'ont cessé de croître. Le travail que nous avons mené s'inscrit comme un outil indispensable dans la société de la communication et de la nouvelle technologie, adressant un éventail d'activités économiques multiples ...

Voilà les premières lignes d'une introduction type qui est censée motiver le démarrage d'une thèse en traitement d'image. Ce sont d'excellentes raisons.

En ce qui me concerne, j'ai commencé cette thèse parce que j'ai voulu prolonger ma carte de séjour temporaire en tant qu'étudiant et également par curiosité.

Cette curiosité m'a permis de vivre une aventure pendant laquelle tout a été possible : rêver, faire des programmes qui "rament" trop pour être rentables, chercher, douter, croire toujours pouvoir tout recommencer le lendemain, et surtout rencontrer des gens qui dépensent leurs idées comme si c'était de l'argent.

En fait, je n'ai rien fait d'autre que jouer. D'accord, une partie lente. Chers lecteurs prenez du plaisir et asseyez-vous à la table de jeux. Un Mikado, pour commencer ? Jouons le jeu, alors ! Ah, j'étais en train d'oublier : je vous ai apporté des bonbons ...

---



# Introduction

En observation de la Terre, un des problèmes importants consiste en l'extraction d'information à des fins cartographiques à partir de capteurs optiques. Les principales sources d'information sont issues de l'imagerie aérienne et de l'imagerie spatiale (images SPOT, panchromatique et multispectrale, images Landsat, etc.).

L'ensemble de ces informations fournit un large éventail en terme de réponse spectrale pour lequel nous devons trouver des solutions pertinentes et robustes pour l'extraction d'objets cartographiques tels que : les réseaux routiers et hydrographiques, les villes, les bâtiments.

D'un point de vue applicatif, l'objectif général est de pouvoir fournir un outil d'aide à la décision le plus automatisé possible mais prenant en compte l'opérateur humain à la fois comme source d'information et comme structure validante.

Dans l'imagerie satellitaire et aérienne, l'augmentation de la résolution des capteurs optiques impose la position du problème à partir des objets géométriques présents dans l'image et non plus à partir des pixels.

Un outil théorique particulièrement adéquat pour l'analyse à partir d'objets géométriques, est le cadre offert par la théorie des processus ponctuels marqués. Nous n'allons plus aborder l'image comme une collection de valeurs numériques, mais plutôt comme un ensemble d'objets. Le problème est alors de localiser ces objets dans l'image et de les mesurer.

Le but de nos travaux est d'utiliser cette nouvelle approche pour trouver une réponse satisfaisante à un des problèmes, déjà classique mais toujours d'actualité, en traitement d'images : la détection des réseaux linéiques (par exemple : réseau routier, réseau hydrographique, etc.) dans les images satellitaires et aériennes.

Le premier chapitre passe en revue les méthodes utilisées en extraction d'objets et de réseaux linéïques, en particulier de réseaux routiers. Plusieurs techniques sont présentées brièvement : de la programmation dynamique jusqu'au contours actifs en passant par la théorie de l'information, et les réseaux de neurones.

Tous les travaux sur l'extraction des réseaux linéïques s'appuient sur des observations faites surtout par la communauté des cartographes travaillant sur les images satellitaires et aériennes. Les réseaux linéïques se présentent comme des bandes homogènes de différentes longueurs, épaisseurs et orientations, en fonction de leur caractéristiques : route nationale, autoroute, rivière, chemin de fer, etc.

Ces observations sont à la base de l'hypothèse générique de notre travail. Nous approchons les réseaux linéïques dans une image satellitaire ou aérienne, par un réseau des "segments de droite" connectés, de différentes longueurs, épaisseurs et orientations. L'emplacement du réseau dans l'image est aléatoire. Pour synthétiser, notre hypothèse suppose que les réseaux linéïques dans l'image sont la réalisation d'un processus ponctuel marqué.

C'est pour cette raison que la dernière partie du premier chapitre est dédiée aux premières méthodes, utilisées en traitement d'images, s'appuyant sur la théorie des processus ponctuels.

Nous envisageons donc de construire un processus ponctuel marqué qui soit capable de simuler et d'extraire des réseaux linéïques dans les images. Les chapitres deux et trois sont une introduction aux processus ponctuels et aux techniques de simulation par des méthodes de type Monte Carlo par chaînes de Markov. Dans ces deux chapitres, nous donnons des propriétés de stabilité qui vont imposer certaines contraintes sur la construction du modèle et de l'algorithme de simulation. En revanche, ces propriétés garantissent la convergence de la méthode proposée.

Le modèle probabiliste que nous construisons comporte deux composantes. Le premier terme, l'attache aux données, est relatif à l'emplacement du réseau dans l'image. Le deuxième terme gère les interactions entre les segments. Ce terme régularise la solution, jouant le rôle du terme a priori dans une approche bayésienne.

Le quatrième chapitre de ce mémoire contient la description du terme d'interaction

---

entre les segments qui forment le réseau. Nous appelons ce modèle "Candy" modèle. Les contraintes prises en compte par le "Candy" modèle sont : la dimension, la connectivité et la courbure moyenne du réseau. En outre, nous définissons des régions de répulsion et d'attraction autour d'un segment qui vont conférer la propriété markovienne au modèle proposé. C'est à dire qu'un segment ne va interagir qu'avec des segments qui se trouvent dans son voisinage.

La deuxième partie de ce chapitre est consacrée à la construction d'un algorithme de simulation adapté au "Candy" modèle. La dynamique qui est à la base de cet algorithme est une dynamique de type Monte Carlo par chaînes de Markov à sauts réversibles. Nous utilisons ce type de dynamique car elle nous donne la possibilité de simuler des modèles qui sont définis dans un espace d'état continu de dimension variable.

Le chapitre 5 est une étude des propriétés de la dynamique de simulation du "Candy" modèle. Contrôler la convergence d'un algorithme de type Monte Carlo à sauts réversibles est encore un problème ouvert. En s'appuyant sur la bibliographie, les propriétés du "Candy" modèle et des tests graphiques, nous essayons de caractériser la chaîne de Markov construite, et de tester sa convergence et sa vitesse de mélangeance. Cet étude ne sert pas du tout à montrer la supériorité de cette dynamique. Nous essayons de mieux comprendre son fonctionnement afin qu'elle soit utilisable pour la détection des réseaux linéiques dans les images.

Le terme d'attache aux données qui sert à trouver l'emplacement du réseau dans l'image est décrit dans le sixième chapitre. Ce terme s'appuie sur des tests d'hypothèse. Nous considérons qu'une route a des statistiques localement homogènes. Ces statistiques sont différentes du fond de la route. Pour construire les tests d'hypothèse nous utilisons une hypothèse gaussienne. En fin de chapitre, nous montrons des résultats sur l'extraction des réseaux routiers sur plusieurs types d'images : SPOT, ERS, aériennes.

Estimer les paramètres d'un modèle utilisé en traitement d'image est un tout premier pas vers des méthodes complètement non-supervisées. C'est le sujet abordé par le dernier chapitre de ce manuscrit. Nous traitons ici le cas du modèle d'interaction entre les segments. Pour estimer les paramètres du modèle, nous allons construire un algorithme qui utilise le maximum de vraisemblance via des méthodes de type Monte Carlo. Des résultats, ainsi que leur analyse, sont présentés à la fin du chapitre 7.

Les conclusions et les perspectives dégagées par ce travail sont présentées dans la toute dernière partie du manuscrit.

# Chapitre 1

## Etat de l'art

---

La forme des objets en vision par ordinateur est représentée par deux approches : la détection de contours ou la détection de régions [Cocquerez et Philipp (eds.), 1995].

Les routes sont des objets délicats. Elles ont été abordées très souvent dans le cadre de la détection de contour. Avec l'augmentation de la résolution des capteurs satellitaires et aériennes, les routes apparaissent plutôt comme des objets en forme de rubans homogènes, dont l'épaisseur varie selon le caractère de la route : autoroute, route nationale, sentier, etc ...

C'est pourquoi, dans ce chapitre, nous allons étudier des méthodes d'extraction d'objets orientés région ou bien contour, tout en mettant l'accent sur l'extraction de réseau linéique.

La première partie de ce chapitre est consacrée à l'extraction des objets dans l'image par approches pixéliques. Un paragraphe est dédié aux méthodes qui essayent d'améliorer les méthodes pixéliques par un apport de données externes : cartes géographiques, bases des données, etc. Les méthodes orientées objet, notamment celles qui utilisent comme formalisme la théorie des processus ponctuels, sont présentées dans la dernière partie du chapitre.

---





## 1.1 Approches orientées contour et autres opérateurs

La détection de réseaux linéiques a été souvent abordée comme un "sous-problème" de la détection des contours. C'est pourquoi nous avons étudié, dans un premier temps les travaux sur la détection de contours et sur le chaînage de contours incomplets.

La détection de contours suppose qu'un contour peut être détecté par de grandes valeurs du gradient de l'intensité de l'image. Les méthodes actuelles prennent aussi en compte l'orientation et la connectivité du contour. Il est vrai que la présence du bruit fait que ces méthodes sont extrêmement sensibles, et que le schéma de connectivité influence beaucoup la justesse de la forme déterminée.

Le plus souvent, l'extraction du réseau linéique est faite par une approche en deux étapes. Premièrement, un opérateur bas niveau (filtre de Canny-Deriche, transformée de Hough, opérateur morphologique, etc. ...) est utilisé pour extraire les composantes principales du réseau. Les éléments du réseau ainsi détectés sont connectés dans une deuxième étape. Parmi les critères utilisés pour effectuer le chaînage de contours ou de segments on peut mentionner : la longueur des segments, l'orientation, le contraste, la radiométrie moyenne etc.

Ces règles sont issues d'expériences psycho-visuelles, essayant d'expliquer le groupement perceptuel. Le groupement perceptuel est le processus qui, associant ou bien ignorant des informations locales dans l'image, conduit à la perception globale du contour des objets dans l'image.

[Mangin, 1994] a étudié la mise en pratique de ces principes. La complexité de ces hypothèses l'a mené à faire un compromis entre la complexité des modèles et l'implantation pratique. Malgré la manière originale d'aborder le problème, ses résultats sont assez peu convaincants. Néanmoins, son travail ouvre des directions intéressantes dans ce domaine.

[Urago, 1996], après une détection bas niveau, induit un champ markovien défini d'une manière explicite sur les pixels du contour. Finalement, les contours sont complétés en minimisant une fonction d'énergie. Dans cette approche, on n'estime pas directement les paramètres, mais on utilise la méthode des boîtes qualitatives, pour en faire une approximation. En outre, il existe la possibilité de rater des contours ou des routes si dans l'étape de bas niveau il y a des éléments que l'on ne détecte pas du tout.

### 1.1.1 Modélisation probabiliste des routes

La détection de réseaux routiers est une des applications immédiates de la détection des éléments linéiques dans une image. Cette modélisation géométrique des routes est appliquée notamment aux réseaux épais.

Les hypothèses génériques du modèle de la route sont :

- la radiométrie à l'intérieur de la route est assez homogène,
- l'épaisseur de la route varie lentement,
- la courbure de la route est régulière et localement rectiligne.

Les auteurs [Graffigne et Herlin, 1989] ajoutent comme information supplémentaire au modèle le gradient de l'image dans la région de la route. La détection de la route est faite en combinant une méthode déterministe et une méthode probabiliste. Après un détecteur de type Canny-Deriche et un prolongement déterministe (selon la direction du gradient), on effectue un prolongement probabiliste. Pour effectuer ce prolongement, le modèle de la route est mis sous la forme d'un processus de Gibbs. Une fonction d'énergie est construite. Elle contient des termes relatifs à la radiométrie de la route, à son épaisseur, et à la valeur du gradient. La route est ainsi approximée par l'estimateur du maximum de vraisemblance.

### 1.1.2 Banc de filtres. Décomposition multirésolution

Le filtrage reste toujours une solution attirante, surtout grâce au faible coût de calcul de cette opération, qui permet sur les processeurs actuels des traitements en temps réel.

Les vaisseaux sanguins dans l'imagerie médicale constituent des réseaux linéiques également [Poli et Vali, 1997]. Les auteurs proposent d'utiliser un banc de filtres gaussiens pour l'amélioration et la détection des vaisseaux de sanguins en imagerie médicale. Les filtres sont construits afin que l'on puisse différencier les diverses dimensions et orientations des vaisseaux. Un test de validité entre les différentes orientations détectées est également proposé.

Il est difficile de s'empêcher de faire un rapprochement avec l'analyse en ondelettes décrite par [Mallat, 1989].

### 1.1.3 Lignes de crête. Vallées.

Dans une image, les réseaux linéiques se présentent très souvent, comme des minima ou des maxima locaux de l'intensité le long d'une direction. Le formalisme des lignes de crêtes et de vallées présenté ci-dessous, [Eberly et al., 1994], garantit aussi leur invariance par rotation et translation.

Pour déterminer les lignes de crête et les vallées dans une image, on peut, par exemple, interpoler localement la fonction d'intensité de l'image [Haralick, 1983], et détecter les crêtes et les vallées analytiquement. Les résultats finaux sont obtenus après avoir subi des opérations morphologiques.

[Antoine, 2000] met bout-à-bout plusieurs techniques pour construire un détecteur de routes. Dans une première étape un détecteur de crête multi-échelle permet d'extraire l'axe central des routes de diverses épaisseurs. Le résultat est complété avec une transformée de Hough. Finalement, un processus markovien monodimensionnel est mis en place pour régulariser la solution. Les résultats montrent très peu de fausses alarmes, en ce qui concerne l'emplacement du réseau routier. Par rapport à la topologie, les résultats peuvent être améliorés par des opérateurs morphologiques, ou en introduisant des contraintes géométriques dans le modèle.

### 1.1.4 Réseaux de neurones

Des réseaux de neurones [Bogess, 1994] ont été entraînés pour détecter les routes sur des images LANDSAT. Pour cela, il utilise comme entrée du réseau le pixel à tester et ses 24 voisins, dans les 7 bandes spectrales de l'image. Malgré le fait que l'auteur annonce que cette technique donne peu de fausses détections, de nombreux pixels dans les régions problématiques ne sont pas détectés. L'auteur envisage de fournir une étude complémentaire relative à la connexion de segments après le résultat du réseau neuronal, et aussi de développer une approche mixte : réseau de neurones et méthodes probabilistes.

[Wolfer et al., 1994] présentent une approche très semblable à la précédente, en faisant de plus, une comparaison entre les diverses architectures de réseaux de neurones.

Malgré l'originalité de la démarche, les résultats dans les deux cas sont peu convaincants, même si les auteurs clament que leurs performances sont comparables aux méthodes statistiques. Cependant, les réseaux de neurones restent une technique à explorer, qui peut

s'avérer extrêmement efficace une fois le réseau entraîné.

### 1.1.5 Programmation dynamique

Le problème de suivi de routes, dans le cadre de la programmation dynamique, revient à l'optimisation d'une fonction d'énergie ou de cout, optimisation qui, placée dans ce contexte, est équivalente à la recherche d'un chemin optimal dans un graphe.

En adoptant le formalisme utilisé dans [Merlet et Zerubia, 1997], la fonction à minimiser s'écrit comme :

$$U = \sum_{s=S_0}^{S_1} \phi(s, q) \quad (1.1)$$

avec  $s$  étant un état (un pixel), et  $q$  une décision à prendre.  $\phi(s, q)$  est le coût de prendre la décision  $q$  en étant dans l'état  $s$ . Le successeur d'un état est  $\tau(s, q)$ . Les états initiaux et finaux sont  $S_0$  et (respectivement)  $S_1$ .

Il est possible de montrer que :

$$\forall s \neq s_0, U(s) \leftarrow \min_{q \in Q} \{ \phi(s, q) + U(\tau(s, q)) \} \quad (1.2)$$

avec les conditions initiales :

$$\begin{aligned} U(s_0) &= \min_{q \in Q} \{ \phi(S_0, q) \} \\ \forall s \neq S_0, U(s) &\leftarrow +\infty \end{aligned} \quad (1.3)$$

Tout le problème consiste dans la définition des fonctions de potentiel, qui doivent intégrer au mieux les caractéristiques d'une route, afin que les états trouvés comme solution correspondent aux pixels de la route dans l'image.

[Fischler et al., 1981] détectent les structures linéiques en deux étapes. Premièrement, ils appliquent deux opérateurs pour détecter des lignes : un opérateur qui donne très peu de fausses alarmes (DRO - Duda Road Operator), puis ils utilisent un deuxième opérateur tenant compte de la différence d'intensité entre les pixels de la route et les pixels du fond. Un coût inversement proportionnel à la probabilité de trouver un pixel sur une route est associé à chaque point de la région d'intérêt. La solution finale est obtenue par programmation dynamique en trouvant le chemin de coût minimal entre le point de départ et d'arrivée. Ces

points sont spécifiés au début de l'algorithme par l'utilisateur. Cette solution s'est révélée très fiable et reste comme référence dans ce domaine, mais cela nécessite des points d'amorce.

[Serendero, 1989] utilise dans un premier temps pour l'extraction des réseaux linéïques un détecteur de lignes bas-niveau : le "chapeau-haut-de-forme" issu de la morphologie mathématique. Ensuite, des algorithmes de chaînage et de prolongement des éléments antérieurement détectés sont mis en place, ils sont fondés sur la programmation dynamique.

[Merlet et Zerubia, 1996] développent l'algorithme décrit dans [Fischler et al., 1981], en se plaçant implicitement sous l'hypothèse de la markovianité. En suivant cette logique, la fonction de coût est construite en fonction de cliques dans le voisinage d'un point. Une autre nouveauté consiste en l'ajout de la courbure locale. Cette contrainte aide au suivi dans la direction globale de la route.

[Barzohar et Cooper, 1996] construisent une procédure entièrement automatique pour la détection de routes épaisses. Leur modèle développe les idées présentées dans [Graffigne et Herlin, 1989]. Ils n'utilisent pas la courbure ni le gradient. Leur approche vise plutôt à trouver un modèle de type AR pour prendre en compte les variations entre les pixels de la route et ceux du fond. En utilisant ce modèle de route, et sous l'hypothèse des champs de Gibbs, les auteurs calculent le Maximum A Posteriori (MAP) de la route. Le MAP est calculé d'abord dans des petites imagettes, le résultat final est obtenu en utilisant les meilleures imagettes candidates. Il faut noter les bons résultats obtenus sur des images de routes de synthèse. L'inconvénient de cette méthode est que le résultat final est affecté par le choix des petites imagettes supposées contenir une route. Il est possible de connecter des imagettes où il n'y pas de route, ou bien d'arrêter l'algorithme avant que la détection ne soit complète.

### 1.1.6 Théorie de l'information

[Jedynak, 1995], [Geman et Jedynak, 1996] proposent une approche originale pour le suivi de routes épaisses. À partir d'un point d'amorce et d'une direction donnée, l'algorithme proposé effectue des tests, notamment sur la radiométrie des pixels dans la direction de la route et dans la direction orthogonale à la route. Les pixels sélectionnés comme pixels appartenant à une route seront les pixels qui minimisent une fonction d'entropie. L'originalité de ces travaux consiste dans l'approche par la théorie de l'information, via un jeu d'enfants, le jeu des vingt questions.

### 1.1.7 Contours actifs

Les contours actifs sont des contours qui, à partir d'un état initial s'adaptent aux contours souhaités, en suivant la minimisation d'une fonction d'énergie ou l'évolution d'une Equation aux Dérivées Partielles (EDP).

L'étude de ces techniques peut s'avérer utile dans le suivi de routes, ou bien pour l'amélioration ou le re-positionnement des structures linéiques après une détection "bas-niveau" faite au préalable.

Pour en savoir plus sur les contours actifs et les EDP, on peut se référer aux travaux de synthèse [Deriche et Faugeras, 1995].

[Fua et Leclerc, 1990] ont utilisé ce type de modèle pour la détection des routes et des bâtiments. Les auteurs prouvent que les courbes optimisées i.e. les solutions pour lesquelles la courbure est faible ou l'amplitude du gradient est constante, sont de bonnes approximations des contours. De plus, ils donnent une méthode adaptative pour déterminer les paramètres du modèle de l'énergie. Les courbes sont modélisées comme des rubans dont les bords sont des courbes parallèles. Un terme illustrant cette contrainte est ajouté à l'énergie d'attraction.

[Neuenschwander et al., 1997] améliorent les contours actifs classiques, en introduisant des points d'amorce. Ensuite, ces points sont rapprochés des contours selon l'intensité de l'image et son gradient ; les forces (l'énergie) d'attraction sont propagées de proche en proche selon le fonctionnement d'une fermeture éclair "ziplock".

[Dhéréte et Desachy, 1999] utilisent une batterie d'opérateurs linéiques sur plusieurs bandes spectrales de SPOT. Ensuite, une mise en correspondance entre une base de données géographiques et le réseau ainsi détecté est effectuée en utilisant la programmation dynamique. Le résultat final est obtenu en mettant en œuvre des contours actifs pour régulariser la solution.

## 1.2 Approches orientées région

Un des inconvénients majeurs des contours actifs est que l'optimisation de la fonction d'énergie est faite selon les principes de l'analyse variationnelle et de l'approximation par

différences finies ou par éléments finis. La solution trouvée est généralement une solution locale, ce qui nécessite une très bonne initialisation pour pouvoir converger vers le minimum global.

Une solution pour pallier cet inconvénient est d'utiliser des techniques de modélisation stochastiques, ce qui permettra l'utilisation d'algorithmes d'optimisation globale de type recuit simulé. Un autre avantage de la modélisation stochastique est la liberté d'ajouter d'une façon naturelle de nouvelles contraintes et, bien évidemment, la possibilité d'estimer les paramètres des modèles considérés.

[Ronfard, 1994] fait la transition entre les contours actifs variationnels et les modèles probabilistes en utilisant une approche par région. L'approche région suppose qu'un objet est localement homogène et teste si un pixel appartient ou non à l'intérieur de l'objet en mesurant la similitude entre son intensité et ses voisins.

En plaçant à des intervalles réguliers un ensemble de cellules de calcul sur le bord du contour de l'objet, on détermine des contours possibles dans une direction donnée et on se déplace peu à peu vers ces contours. Ces petits déplacements donnent le déplacement global du contour. Chaque déplacement est contrôlé par la distance de Ward, distance qui prend en compte le contour et les statistiques des régions à l'intérieur et à l'extérieur du contour.

[Figueiredo et al., 1997], considérant un contour fermé qui divise l'image en une région intérieure au contour et une région extérieure à celui-ci, proposent une estimation jointe des paramètres de la courbe du contour et des paramètres du modèle de l'image observée dans les deux régions. L'estimateur MAP des paramètres de ce modèle s'écrit comme :

$$(\hat{\theta}_{(K)}, \hat{\phi}) = \operatorname{argmax}\{p(\theta_{(K)}, \phi|I)\} = \operatorname{argmax}\{p(I|\theta_{(K)}, \phi)p(\theta_{(K)}, \phi)\} \quad (1.4)$$

$I$  est l'image observée,  $\theta_{(K)}$  sont les paramètres du contour (descripteurs de Fourier ou B-splines) et  $\phi$  sont les paramètres de l'image. Les probabilités a priori  $p(\theta_{(K)}, \phi)$  sont construites à partir du modèle de paramétrisation choisi et son ordre  $K$ . En supposant que les pixels sont des variables aléatoires indépendantes et qu'à l'intérieur (respectivement l'extérieur) de l'objet les pixels suivent la même loi  $\phi_{in}$  (respectivement  $\phi_{ext}$ ), on peut écrire la probabilité :

$$p(I|\theta_{(K)}, \phi) = \left( \prod_{(i,j) \in Int} p(I(i,j)|\phi_{in}) \right) \left( \prod_{(i,j) \in Ext} p(I(i,j)|\phi_{ext}) \right) \quad (1.5)$$

Ce modèle est applicable dans des situations où le gradient ne caractérise pas une région. Le modèle utilise tous les points dans l'image et pas seulement le voisinage d'un objet à segmenter. L'ordre du modèle joue un rôle très important. Le MAP est calculé de deux façons : la première avec un ordre  $K$  fixé, et la seconde calculant aussi l'ordre optimal. Le principe utilisé est le "Minimum Description Length" (MDL) qui affirme que le meilleur modèle est celui qui permet d'avoir la plus courte description des données observées et du modèle lui-même. Pour ce faire, un ordre  $K_{max}$  est choisi et les paramètres du modèle sont calculés et stockés. L'ordre est décrémenté au fur et à mesure. A la fin, la solution optimale est choisie.

[Germain et Réfrégier, 1996], [Réfrégier et al., 1997 (a)], [Réfrégier et al., 1997 (b)] supposent que l'image est divisée en deux régions : objet et fond. Le niveau de gris d'un pixel, noté  $i$  dans l'image, est caractérisé par :

$$s_i = a_i \omega_i + b_i (1 - \omega_i) \quad (1.6)$$

où  $\omega$  est le masque de l'objet,  $\omega_i = 1$  si  $i \in Int_{obj}$  et  $\omega_i = 0$  si  $i \in Ext_{obj}$ .  $a_i$  est le niveau de gris du pixel à l'intérieur de l'objet, et  $b_i$  est le niveau de gris du pixel à l'extérieur de l'objet. Si  $H_\omega$  représente l'hypothèse d'attribuer une certaine forme à l'objet, le choix de la forme optimale (MAP) de l'objet est donnée par le maximum de :

$$P(H_\omega, \mu^a, \mu^b | s) = P(s | (H_\omega, \mu^a, \mu^b)) \frac{P(H_\omega, \mu^a, \mu^b)}{P(s)} \quad (1.7)$$

avec  $\mu^a$  et  $\mu^b$  les statistiques des pixels respectivement à l'intérieur et à l'extérieur de l'objet. Sous certaines conditions d'indépendance et en fonction des modèles a priori choisis, le maximum de cette expression est trouvé en utilisant un algorithme stochastique itératif.

[Storvik, 1994] place les contours actifs dans le cadre bayésien, intégrant comme variable le nombre de points définissant le contour à déterminer. En englobant dans une fonction d'énergie des critères concernant à la fois la frontière de l'objet à délimiter et des informations concernant sa région, on trouve un contour optimal en minimisant la fonction d'énergie par des techniques de recuit simulé, utilisant une dynamique de type Metropolis-Hastings.

### 1.3 Apport de données externes

Une catégorie professionnelle très intéressée par l'utilisation des programmes de détection de routes est la communauté des spécialistes de Systèmes d'Information Géographique



(SIG).

Pour une bonne compréhension de la physique du phénomène et pour une aide à la modélisation, il convient de souligner les principales difficultés qui doivent être surmontées lors de la détection de routes [Bordes, 1997] :

- les conditions de prises de vue des images sont variables : en fonction des conditions météo, de l'heure de la prise de vue, la radiométrie et le contraste dans l'image varient d'une manière importante ;
- les caractéristiques de la route sont extrêmement variables : les routes "interagissent" avec beaucoup d'objets : nuages, arbres, voitures, ombre des objets à proximité de la route. Ces interactions produisent des variations de la radiométrie de la route et des interruptions du réseau routier ;
- l'aspect des routes est très lié à leur contexte d'apparition : en Europe, les routes sont beaucoup plus sinueuses qu'aux Etats Unis. Il en est de même entre les routes de montagne et celles de plaines. De plus la texture des champs, des bâtiments, des forêts peut se confondre avec la texture des routes ;
- le bruit dans l'image ;
- les fausses alarmes générées par la présence d'objets qui ont une radiométrie similaire à celle de la route.

Pour pallier cet inconvénient, les ingénieurs spécialistes de l'information géographique travaillent en combinant à une image satellitaire d'autres types d'information : plusieurs images (différentes largeurs de bande spectrale, des images prises à des intervalles de temps différents), base de données cartographiques [Bordes, 1997], cartes numérisées [Wu, 1987], [Roux, 1992], [Guerin, 1996], etc.

Pourtant, il se trouve que, dans certaines situations, le cumul d'information ne simplifie pas le problème. Les images, par exemple, peuvent avoir des résolutions différentes, ou bien subir des transformations préalables comme la translation ou la rotation [Wu, 1987]. Les cartes, malgré leur richesse en information, amènent des difficultés importantes [Guerin, 1996] : différentes résolutions, le langage symbolique des cartes, les normes comme l'écart

minimal entre objets, qui fait que la taille, la forme, la position des objets varient par rapport à leur référence sur le terrain.

[Rellier et al., 2000] proposent une méthode de mise en correspondance locale des cartes avec des images SPOT utilisant un champ markovien sur un graphe. La carte est utilisée pour initialiser le graphe. Ensuite, le graphe est modifié en minimisant une fonction d'énergie. L'énergie comporte des termes d'attache aux données relatifs à la carte et à l'image, et un terme imposant des contraintes topologiques pour bien conditionner le problème. La méthode donne de très bons résultats, mais dans le cas où une nouvelle route est construite, elle ne peut pas être utilisée comme technique de mise-à-jour des cartes.

## 1.4 Approche par processus objet

Jusqu'à présent, les images ont été analysées au niveau pixélique. C'est à dire qu'indépendamment de l'approche mathématique choisie (variationnelle ou bien stochastique) les mesures effectuées dans l'image se font toujours au niveau du pixel. Malheureusement ces méthodes sont très sensibles au bruit et aux minimas locaux.

[Tupin, 1997] se concentre précisément sur la détection des routes en imagerie radar RSO. L'auteur construit un champ markovien sur le graphe des segments détectés dans l'étape bas-niveau. Le modèle markovien donne les probabilités que deux segments soient reliés. Les routes sont complétées en minimisant, par recuit simulé, la fonction d'énergie induite par ce modèle. L'inconvénient de cette méthode est qu'on ne peut pas relier des segments qui n'ont pas été détectés dans l'étape de traitement bas-niveau.

Une nouvelle direction pour essayer d'éliminer cet inconvénient est d'aborder l'image non plus comme une collection de pixels, mais plutôt comme un ensemble d'objets. Le problème se pose alors de la façon suivante : déterminer le nombre d'objets et leurs caractéristiques. Ce problème se formule de façon rigoureuse dans le cadre des processus ponctuels marqués.

Soit un processus ponctuel dont la configuration  $x = \{x_1, \dots, x_n\}$  a la densité :

$$p(x) \propto \prod_{x_i \sim x_j} \alpha(x_i, x_j) \quad (1.8)$$

avec  $\sim$  une relation de voisinage et  $\alpha(\cdot, \cdot)$  la fonction d'interaction. Bien entendu l'interac-

tion peut être d'un ordre plus élevé que deux, par exemple  $\alpha(x_i, x_j, x_k, \dots)$ .

Ce type de processus peut être simulé en utilisant des algorithmes de *naissance et mort*, ou bien des méthodes Monte Carlo par chaînes de Markov à sauts réversibles. Les mouvements qui constituent cette dynamique sont les suivants : ajouter un objet à la configuration, enlever un objet à la configuration, modifier un objet à l'intérieur d'une configuration.

Le problème de détection des objets dans l'image est donc d'estimer la meilleure configuration  $x = \{x_1, \dots, x_n\}$  dans un espace continu à partir des données  $y$  représentées par une image discrète. En définissant un terme d'attache aux données  $f(y|x)$ , et un a priori  $p(x)$ , nous obtenons la distribution a posteriori :

$$p(x|y) \propto f(y|x)p(x) \quad (1.9)$$

Le maximum de l'équation (1.9) (*Maximum A Posteriori*) est la solution de notre problème.

A. Baddeley et M.N.M van Lieshout sont parmi les premiers à utiliser les processus ponctuels pour l'extraction des cellules dans les images microscopiques [Baddeley et van Lieshout, 1993]. Une cellule peut être approximée par un cercle de rayon aléatoire. Les coordonnées du centre du segment désignent un processus ponctuel non marqué. La marque est donnée par le rayon du cercle. Les auteurs détectent les objets dans l'images en maximisant une fonction de vraisemblance. La fonction de vraisemblance comporte deux termes : l'attache aux données et l'a priori. Pour le terme d'attache aux données, plusieurs modèles sont utilisés : l'image est en niveaux de gris et le bruit est additif gaussien ou exponentiel. Le cas où l'image est binaire et le bruit est de type poivre et sel est également analysé. En ce qui concerne le terme a priori, deux processus ponctuels sont utilisés : le processus de Strauss et le processus d'interaction d'aires. La solution est trouvée en utilisant des techniques déterministes (gradient de la plus grande pente, ICM) et stochastiques (recuit simulé).

L'algorithme de type recuit simulé s'appuie sur une dynamique utilisant des techniques de simulation des processus de naissance et mort. Une preuve pour un schéma de descente de température qui assure la convergence vers l'optimum global est donnée dans [Lieshout, 1994]. L'originalité de ce travail réside dans le fait que l'on est assuré de converger vers l'optimum global, même quand la dimension du sous-espace de configuration est variable ; donc quand le nombre d'objets dans l'image est une inconnu.

Une façon plus élaborée d'extraire les cellules en imagerie microscopique est de supposer qu'un objet a un contour de forme variable [Rue et Husby, 1998]. Par exemple, un contour peut être approximé par un polygone avec un nombre de points variable sous certaines contraintes d'orientation entre les côtés et de convexité imposées comme a priori. Pour prendre en compte le nombre d'objets, les auteurs [Rue et Syverseen, 1998], [Rue et Hurn, 1999] introduisent un processus ponctuel qui impose également que les différentes cellules ne se superposent pas.

Considerons le cas où l'image est constituée d'un seul objet entouré par un fond. Le contour est toujours supposé être un polygone déformable, avec un nombre de points variable. Alors, la détection du contour de l'objet peut se faire en imposant un processus ponctuel sur le nombre de points qui forment le polygone qui approxime le contour de l'objet [Pievatolo et Green, 1998]. Cette idée a été utilisée aussi par [Hansen et al., 2000] dans le cas d'une aide au diagnostic et d'un suivi automatique en imagerie médicale.

## Chapitre 2

# Fondements des Processus Ponctuels

---

Ce chapitre commence par le rappel des notions fondamentales sur l'intégration. Ensuite, les processus ponctuels et les propriétés qui les caractérisent sont introduites. En fin de chapitre, nous donnons quelques exemples de processus ponctuels.

Nous conseillons les références suivantes : [Dacunha et al., 1982], [Lusternik et Sobolev, 1989] pour les notions de base sur l'intégration et [Stoyan et al., 1987], [Ripley, 1977], [Ripley et Kelly, 1977], [Lieshout et al. (eds.),1999] pour la partie concernant les processus ponctuels.

---

## 2.1 Rappel sur l'Intégration

Soit  $\Omega$  un ensemble quelconque (*espace des réalisations*) et soit  $\mathcal{A} \in \mathcal{P}(\Omega)$ , où  $\mathcal{P}(\Omega)$  est une famille de parties de  $\Omega$  contenant  $\Omega$  (*ensemble des événements*). Si  $A$  et son complémentaire  $\bar{A}$  vérifient les propriétés suivantes :

$$\begin{aligned} A \in \mathcal{A} &\Rightarrow \bar{A} \in \mathcal{A} \\ A_j \in \mathcal{A}, \forall j \in \mathbb{N} &\Rightarrow \bigcup_{j=0}^{+\infty} A_j \in \mathcal{A} \\ \emptyset &\in \mathcal{A} \end{aligned} \quad (2.1)$$

on dit que  $\mathcal{A}$  est une *tribu ou  $\sigma$ -algèbre*. Une tribu très utilisée est la famille  $\mathcal{B}^d$  des boréliens de  $\mathbb{R}^d$ . Dans le cas monodimensionnel, cette  $\sigma$ -algèbre est générée par les intervalles  $(-\infty, x]$ ,  $x \in \mathbb{R}$ , les ensembles de Borel. Nous pouvons étendre cette définition à la dimension  $d$ .

Si  $\Omega$  est muni d'une  $\sigma$ -algèbre,  $(\Omega, \mathcal{A})$  est un *espace mesurable*. Intuitivement, les éléments de  $\mathcal{A}$  sont les parties de  $\Omega$  dont on peut mesurer la surface.

La fonction  $f : \Omega \rightarrow \mathbb{R}$  est dite  *$\mathcal{A}$ -mesurable* si, pour tout ensemble de Borel  $B \in \mathcal{B}^1$ , l'image réciproque de  $B$  par cette fonction,  $f^{-1}(B) = \{x \in \Omega : f(x) \in B\}$ , est un élément de la  $\sigma$ -algèbre  $\mathcal{A}$  associée à  $\Omega$ . La fonction indicatrice d'un ensemble mesurable est une fonction mesurable. Les fonctions coordonnées d'un point dans l'espace  $\mathbb{R}^d$  sont des fonctions Borel-mesurables, ainsi que toute fonction continue.

Une fonction  $\mu : \mathcal{A} \rightarrow \mathbb{R}$  est une *mesure* si et seulement si :

$$\begin{aligned} \forall A \in \mathcal{A} &\Rightarrow \mu(A) \geq 0 \\ A_j \in \mathcal{A}, \forall j \in \mathbb{N}, A_i \cap A_j = \emptyset &\Rightarrow \mu\left(\bigcup_{j=0}^{+\infty} A_j\right) = \sum_{j=0}^{\infty} \mu(A_j). \end{aligned} \quad (2.2)$$

Une mesure  $\mu$  sur l'espace mesurable  $(\Omega, \mathcal{A})$  définit le triplet  $(\Omega, \mathcal{A}, \mu)$ , dit *espace mesuré*. Une mesure  $\mu$  sur  $(\Omega, \mathcal{A})$  est  *$\sigma$ -finie*, s'il existe une réunion dénombrable  $\Omega = \cup_j A_j$  avec  $A_j \in \mathcal{A}$  telle que  $\mu(A_j) < \infty$ .

Les mesures les plus utilisées sont la mesure de Dirac  $\delta_\omega$  pour  $\omega \in \Omega$  :

$$\delta_\omega(A) = \begin{cases} 1, & \omega \in A \\ 0, & \text{sinon} \end{cases} \quad (2.3)$$

et la mesure de Lebesgue sur  $(\mathbb{R}^d, \mathcal{B}^d)$  :

$$\nu_d(Q) = (b_1 - a_1) \dots (b_d - a_d) \quad (2.4)$$

si  $Q = [a_1, b_1] \times \dots \times [a_d, b_d]$ .

La différence entre ces deux mesures est que la mesure de Dirac est concentrée sur des points isolés alors que la mesure de Lebesgue correspond à une mesure de volume et donne zéro pour les points isolés. Les mesures  $\mu$  sur  $(\mathbb{R}^d, \mathcal{B}^d)$  qui sont concentrées sur un nombre dénombrable des points  $x_1, x_2, \dots$  sont appelées *mesures atomiques*. Elles sont appelées *mesures de décompte (counting measures)* si, de plus, elles donnent pour chaque point la masse zéro ou un. Une *mesure diffuse* est une mesure qui donne zéro pour chaque point isolé, comme la mesure de Lebesgue.

Une *probabilité*  $P$  sur un espace mesurable  $(\Omega, \mathcal{A})$  est une mesure sur  $(\Omega, \mathcal{A})$  telle que  $P(\Omega) = 1$ . L'espace mesuré  $(\Omega, \mathcal{A}, P)$  sera appelé *espace de probabilités*.

Soit  $\mu$  une mesure et  $g$  une fonction positive mesurable définies respectivement sur  $\mathcal{A}$  et  $\Omega$ . Alors l'application :

$$\nu(A) = \int_A g d\mu, \quad \forall A \in \mathcal{A} \quad (2.5)$$

définit une nouvelle mesure sur  $(\Omega, \mathcal{A})$ .

Une mesure  $\nu$  est *absolument continue* par rapport à  $\mu$  si, pour  $A \in \mathcal{A}$  tel que  $\mu(A) = 0$ , on a  $\nu(A) = 0$ .

**THEOREME DE RADON-NIKODYM.** *Si  $\nu, \mu$  sont des mesures  $\sigma$ -finies de l'espace mesurable  $(\Omega, \mathcal{A})$ , alors il existe une fonction unique  $\mathcal{A}$ -mesurable, telle que :*

$$\nu(A) = \int_A g d\mu \quad (2.6)$$

*si et seulement si  $\nu$  est absolument continue par rapport à  $\mu$ .*

La fonction  $g$  est nommée *la dérivée de Radon-Nikodym* :  $g = d\nu/d\mu$ .

## 2.2 Processus Ponctuels

Soit l'espace mesuré  $(K, \mathcal{B}, \nu)$ , avec  $K$  un compact de  $\mathbb{R}^d$ ,  $\mathcal{B}$  la famille des boreliens et  $\nu$  la mesure de Lebesgue. Soit également l'espace mesurable  $(\Omega, \mathcal{F}, \mu)$  avec  $\Omega$  l'ensemble de configurations des points tel que :

$$\begin{aligned}\Omega &= \cup_{n=0}^{\infty} \phi_n, \\ \phi_n &= \{x_1 \dots x_n\} \subset K\end{aligned}\tag{2.7}$$

$\mathcal{F}$  est une  $\sigma$ -algèbre associée. Pour  $F \in \mathcal{F}$ , nous définissons :

$$\mu(F) = \exp(-\nu(K)) \left[ \mathbf{1}(\emptyset \in F) + \sum \frac{1}{n!} \int \int \dots \mathbf{1}(\{x_1, \dots, x_n\} \in F) \nu(dx_1) \dots \nu(dx_n) \right].\tag{2.8}$$

qui est la mesure du processus de Poisson d'intensité  $\nu$ .  $\mathbf{1}(\cdot)$  est la fonction indicatrice.

L'ensemble  $\Omega$  est la famille de tous les sous-ensembles  $\phi \subset K$  telle que l'ensemble  $\phi$  soit simple ( $x_i \neq x_j$  si  $i \neq j$ ), avec  $\mathcal{F}$  est la plus petite tribu rendant les applications  $\phi : \mathcal{B} \rightarrow \mathcal{F}$  mesurables.

Les processus ponctuels peuvent être abordés comme des ensembles aléatoires de points discrets ou comme des mesures aléatoires comptant les points dans des régions spatiales. Par conséquent, nous pouvons écrire  $x \in \Phi$  ce qui veut dire que  $x$  appartient à la séquence aléatoire  $\Phi$ , ou bien  $\Phi(B) = n$ , qui veut dire que l'ensemble  $B$  contient  $n$  points.

Formellement, un processus ponctuel  $\Phi$  dans  $K$  est une application mesurable d'un espace de probabilité dans l'espace mesuré  $(\Omega, \mathcal{F})$ . Intuitivement, c'est un choix aléatoire de  $\phi$  dans  $\Omega$ . Ceci génère une distribution  $P$  sur  $(\Omega, \mathcal{F})$ .

La distribution  $P$  d'un processus ponctuel  $\Phi$  est déterminée par les probabilités :

$$P(F) = P(\Phi \in F) = P(\{\omega \in K : \Phi(\omega) \in F\})\tag{2.9}$$

pour  $F \in \mathcal{F}$ .

Le processus ponctuel le plus simple est le processus dont les points n'interagissent pas entre eux, et dont les points sont uniformément distribués dans le domaine. Il est connu également sous le nom de processus ponctuel de Poisson. Sa distribution est donnée par :

$$P(\phi(B) = n) = \frac{\lambda^n \nu^n(B)}{n!} \exp(-\lambda \nu(B))\tag{2.10}$$

Le plus souvent les points d'un processus ponctuel interagissent entre eux. Dans ce cas la distribution du processus ponctuel s'écrit comme une mesure continue par rapport au processus de Poisson :

$$P(\phi \in F) = \int_F f(\phi) \mu(d\phi)\tag{2.11}$$



avec  $f(\phi)$  la densité du processus ponctuel.

Un processus ponctuel est *stationnaire* si ces statistiques ne changent pas lorsque les points du processus subissent une translation, i.e. si les distributions des processus  $\Phi = x_n$  et  $\Phi_x = x_n + x$  sont les mêmes :

$$P(\Phi \in F) = P(\Phi_x \in F) \quad (2.12)$$

pour tout  $F$  appartenant à  $\mathcal{F}$ . Si nous avons  $F_x = \{\phi_x \in \Omega : \phi \in F\}$ , la condition de stationnarité s'écrit :

$$P(F) = P(F_x) \quad (2.13)$$

Un processus ponctuel est *isotrope* si ces statistiques ne changent pas lorsque les points du processus ponctuel subissent une rotation autour de l'origine :

$$P(F) = P(rF) \quad (2.14)$$

avec  $rF = \{r\phi \in \Omega : \phi \in F\}$ . La stationnarité et l'isotropie impliquent l'invariance par rapport au mouvement.

La définition de l'intensité  $\Lambda$ , de  $\phi$  est :

$$\Lambda(B) = E\Phi(B) = \int \phi(B)P(d\phi). \quad (2.15)$$

Si le processus est stationnaire, l'invariance par translation implique :

$$\Lambda(B) = \lambda\nu(B) \quad (2.16)$$

$\lambda$  étant le nombre de points par unité de volume.

Les propriétés (2.12)-(2.16), elles ne sont valables au sens strict que dans tout l'espace euclidien.

### 2.2.1 Processus ponctuels marqués

Un processus ponctuel est dit *marqué* si l'on ajoute une marque (ou étiquette) à chaque point du processus. Un *processus ponctuel marqué* sur  $K$  est une séquence aléatoire  $\Psi = \{[x_n; m(x_n)]\}$  dans laquelle les points  $x_n$  constituent un processus ponctuel non-marqué dans  $K$  et les  $m(x_n)$  sont les marques correspondant respectivement à  $x_n$ . Soit  $M$  l'espace

des marques, et  $\mathcal{M}$  la  $\sigma$ -algèbre de Borel associée.

Exemple :  $x$  est le centre d'un segment,  $m(x)$  est le vecteur contenant l'orientation et la longueur du segment.

Il est possible d'interpréter un processus ponctuel marqué, comme un processus ponctuel habituel défini sur l'espace  $K \times M$ . Alors, pour les ensembles de Borel  $B \subset \mathcal{B}$  et  $L \subset \mathcal{M}$ , le nombre des points situés en  $B$  avec des étiquettes dans  $L$  est  $\Psi(B \times L)$ .

### 2.2.2 Processus ponctuels de Gibbs

Les processus de Gibbs viennent de la physique statistique. Ces processus sont décrits par la distribution de Gibbs qui caractérise l'état d'équilibre des particules à l'intérieur de systèmes physiques fermés.

Soit un processus ponctuel  $\Phi$  de distribution  $P$  avec exactement  $n$  points à l'intérieur de l'ensemble borné  $A$  et aucun des points du processus à l'extérieur de cet ensemble. La densité de probabilité est une fonction  $f : K^n \rightarrow [0, \infty)$  telle que :

$$P(\Phi \in F) = \int_{\phi \in F} f(\phi) \mu(d\phi) \quad (2.17)$$

avec  $F \in \mathcal{F}$ . L'expression de la densité de probabilité d'un processus de Gibbs s'écrit sous la forme :

$$f(\phi) = \exp(-U(\phi))/Z = \exp(-U(x_1, \dots, x_n))/Z \quad (2.18)$$

où  $Z$  est la fonction de partition et  $U : K^n \rightarrow \mathbb{R}^+$  est la fonction d'énergie. Souvent la fonction  $U$  est utilisée sous la forme d'une somme de potentiels d'interaction :

$$U(\phi) = \sum_{\psi \subset \phi} V(\psi) \quad (2.19)$$

avec  $V$  égal à zéro pour des sous-ensembles  $\psi$  de  $\phi$  de grande taille.

Le potentiel d'interaction le plus utilisé, est le *potentiel de paires*. Il est défini tel que  $V(\psi) = 0$  si le cardinal de  $\psi$  est strictement supérieur à 2. L'énergie d'un tel processus ponctuel peut être également mise sous la forme :

$$U(x_1, \dots, x_n) = \sum_{i=1}^n -\log \beta + \sum_{1 \leq i < j \leq n} \theta(x_i, x_j) \quad (2.20)$$

le paramètre  $\beta$  contrôle la densité des points et la fonction  $\theta : K^n \rightarrow \mathbb{R}$  définit les interactions par paire de points pour un processus ponctuel.

Soit  $\Phi$ , le processus de Gibbs avec un potentiel de paires et avec un nombre *aléatoire* de points dans  $B$ . Une généralisation du modèle de Gibbs est donnée par :

$$f(\phi) = c\beta^{\phi(B)} \exp\left(- \sum_{1 < i < j < \phi(B)} \theta(x_i, x_j)\right) \quad (2.21)$$

$c$  et  $\beta$  étant des constantes positives.  $c$  est la constante de normalisation.

### 2.2.3 Processus ponctuels de Markov

Soit  $\diamond$  une relation symétrique sur  $K$ . Deux points  $\zeta, \eta \in K$  qui vérifient la relation  $\diamond$  seront dit points *voisins*. Le *voisinage* d'un ensemble  $B \subset K$  est donné par :

$$V(B) = \{\zeta \in K : \zeta \diamond \eta, \eta \in B, \zeta \notin B\} \quad (2.22)$$

Un ensemble  $x \in \Phi_B = \{x_1 \dots x_n\}$  est une *clique* si tous les points de  $x$  sont voisins.

Une fonction  $f : \Omega \rightarrow [0, \infty)$  est une *fonction de Markov* par rapport à la relation  $\diamond$ , si pour tout  $x \in \Omega$  :

- (a)  $f(x) > 0 \Rightarrow f(y) > 0, \forall y \subset x$  (hérédité) ;
- (b) si  $f(x) > 0$  alors, pour  $\zeta \in K$  :

$$\frac{f(x \cup \zeta)}{f(x)} \quad (2.23)$$

ne dépend que de  $\zeta$  et de  $V(\zeta) \cap x$ .

Un *processus ponctuel de Markov* sur  $K$  est un processus ponctuel fini, ayant comme densité une fonction de Markov  $f$ .

Une fonction  $g : \Omega \rightarrow [0, \infty)$  est appelée *fonction d'interaction* si  $g(x) \neq 1$  implique que  $x$  est une clique.

Le résultat suivant fait l'analogie avec les champs de Markov, et le lien entre les processus de Gibbs et les processus markoviens. Ce théorème est l'extension du théorème de Hammersley-Clifford [Ripley et Kelly, 1977].

THEOREME. Une fonction  $f : \Omega \rightarrow [0, \infty)$  est une fonction de Markov si et seulement si il existe une fonction d'interaction  $g$  telle que :

$$f(x) = \prod_{y \subset x} g(y) \quad (2.24)$$

pour tous les  $x \in \Omega$ .

### 2.2.4 Stabilité des processus ponctuels

Soit  $\Phi$  un processus ponctuel dans  $K$ . Soit  $f$  sa densité de probabilité telle que  $f : \Omega \rightarrow [0, \infty)$ .

Pour définir la stabilité des processus ponctuels, plusieurs conditions sur l'énergie d'un système de particules et sur les potentiels d'interaction sont données dans [Ruelle, 1970]. Geyer [Geyer, 1999] donne la définition de deux conditions de stabilité des processus ponctuels.

DEFINITION 1 : Un processus ponctuel, avec une densité non-normalisée par rapport à la mesure  $\mu$ , est localement stable s'il existe  $M \in \mathbb{R}$  tel que :

$$f(\phi \cup \zeta) \leq M f(\phi), \forall \phi = \{x_1 \dots x_n\} \in \Omega, \zeta \in K \quad (2.25)$$

DEFINITION 2 : Un processus ponctuel, avec une densité non-normalisée par rapport à la mesure  $\mu$ , est stable au sens de Ruelle s'il existe  $M \in [1, \infty)$  tel que :

$$f(\phi) \leq M^n, \forall \phi = \{x_1 \dots x_n\} \in \Omega \quad (2.26)$$

Du point de vue physique, la stabilité locale impose que l'apport en énergie d'un nouveau point dans l'image soit limité. Ceci est une propriété plus forte que la stabilité au sens de Ruelle. Cette deuxième condition impose que l'énergie totale d'un système avec un nombre de points fixés soit borné.

Ces conditions sont importantes dans la simulations des chaînes de Markov ayant comme distribution d'équilibre la densité de probabilité d'un processus ponctuel. La stabilité des processus ponctuels est utilisée dans l'étude de l'irréductibilité, de la récurrence et de l'ergodicité de la chaîne que l'on souhaite simuler.

### 2.2.5 Exemples et commentaires

Les exemples et les considérations qui suivent sont utiles pour comprendre la relation entre les processus ponctuels, les processus de Gibbs et les processus ponctuels de Markov.

Soit  $B$  un ensemble borné de Borel dans  $K \subset \mathbb{R}^2$ , pourvu de l'algèbre  $\mathcal{B}^2$  et de la mesure de Lebesgue  $\nu_2$ .

La densité de probabilité d'un processus ponctuel fini, dont les points interagissent, (*interaction par paires*), est :

$$f(x) = \alpha\beta^n \prod_{i < j} g(x_i \diamond x_j) \quad (2.27)$$

avec  $x = \{x_1 \dots x_n\} \in B$ ,  $n$  le nombre de points,  $\alpha$  et  $\beta$  des constantes positives et  $g(\cdot)$  une fonction réelle positive. Pour que (2.27) soit intégrable, des conditions doivent être imposées sur  $g$ .  $\alpha$  est la constante de normalisation.

Soit  $d = \|x_i - x_j\|$  la norme euclidienne. Le *processus de Strauss* est défini par :

$$g(d) = \begin{cases} \gamma, & d < r \\ 1, & d \geq r \end{cases} \quad (2.28)$$

ou  $0 \leq \gamma \leq 1$  et  $r > 0$ . Soit  $s(x)$  le nombre de paires de points dont la distance entre les points est plus petite que  $r$ . La densité de probabilité devient :

$$f(x) = \alpha\beta^n \gamma^{s(x)} \quad (2.29)$$

Si  $\gamma = 1$ , nous obtenons un processus de Poisson sur  $B$  avec l'intensité  $\beta\nu_2(\cdot)$ .

Si  $\gamma < 1$  le processus va manifester une "répulsion" entre les points. Le modèle avec  $\gamma > 1$  a été étudié pour encourager "l'attraction" entre les points, mais dans ce cas, la densité de probabilité n'est pas intégrable [Baddeley et van Lieshout, 1993].

Si  $\gamma = 0$  alors :

$$f(x) = \begin{cases} \alpha\beta^n, & s(x) = 0 \\ 0, & s(x) > 0 \end{cases} \quad (2.30)$$

ce processus est connu sur le nom de *hard core process*. C'est l'équivalent d'un processus de Poisson d'intensité  $\beta\lambda(\cdot)$ , dont les points se trouvent à la distance minimale  $r$ .

Le logarithme de la densité de probabilité d'une certaine configuration  $x$  est une somme de pénalités imposées par la présence de chaque point  $x_i \in x$  et de l'interaction de chaque paire de points  $(x_i, x_j) \in x$ . Cette somme peut être interprétée comme une *énergie de potentiel*. Ce type d'interaction est très utilisé en physique statistique. Une probabilité sur  $\Omega$  est une "grande canonique" et  $\log g(\cdot)$  est le *potentiel d'interaction*.

La *relation de voisinage* définie auparavant, peut être abordée dans ce contexte comme suit : deux points  $\zeta, \eta \in K$  sont voisins,  $\zeta \diamond \eta$ , si  $\|\zeta - \eta\| < r$ . Alors la densité de probabilité devient :

$$f(x) = \alpha \beta^n \prod_{x_i \diamond x_j} g(\|x_i - x_j\|) \quad (2.31)$$

Il faut remarquer que si  $\zeta \in K, \zeta \notin x$  avec  $f(x) > 0$ , alors le rapport :

$$\frac{f(x \cup \zeta)}{f(x)} = \beta \prod_{x_i \diamond \zeta} g(\|x_i - \zeta\|) \quad (2.32)$$

ne dépend que de  $\zeta$  et des voisins de  $\zeta$  dans  $x$ . Cette propriété montre que toutes les interactions sont *locales*. En physique statistique,  $\log f(x \cup \zeta) - \log f(x)$  représente l'énergie nécessaire pour ajouter un nouveau point  $\zeta$  à la configuration  $x$ . En langage probabiliste, le rapport  $\frac{f(x \cup \zeta)}{f(x)}$  porte le nom d'intensité conditionnelle de Papangelou.

## Chapitre 3

# Chaînes de Markov. Méthodes MCMC.

---

La première partie de ce chapitre est dédiée à la présentation de notions fondamentales sur les chaînes de Markov. Après leur définition, nous présentons des propriétés qui garantissent la convergence des chaînes de Markov vers leur distribution d'équilibre : réversibilité, apériodicité, irréductibilité, récurrence, afin que le comportement de simulation de chaînes de Markov soit robuste aux conditions initiales, et que l'on puisse faire de l'inférence statistique. C'est pour cela que nous introduisons l'ergodicité d'une chaîne de Markov et le cadre d'application du Théorème de la Limite Centrale.

La deuxième partie passe en revue des techniques usuelles de simulation de chaînes de Markov, les dynamiques de type Metropolis-Hastings. Nous terminons cette partie avec la présentation d'une méthode plus récente de simulation : la dynamique à sauts réversibles de type Metropolis-Hastings-Green. Le dernier paragraphe de ce chapitre est dédié à la simulation des processus ponctuels.

Nous n'avons pas donné de preuve pour les résultats présentés dans ce chapitre. Le but de ce chapitre est de passer en revue les règles principales relatives à la construction d'une solution, qui s'appuie sur les chaînes de Markov. Pour plus de détails nous recommandons les références que nous avons suivi [Meyn et Tweedie, 1993], [Tierney, 1994], [Green, 1995], [Green, 1996], [Robert, 1996], [Robert et Casella, 1999], [Geyer et Møller, 1994], [Geyer, 1999], [Lieshout, 2000].

---

### 3.1 Définition des Chaînes de Markov

Soit l'espace de probabilités  $(S, \mathcal{F}, \mu)$ .

Une suite  $\{X_n\}$  de variables aléatoires forment une *chaîne de Markov* si :

$$\mathbf{P}(X_{n+1}/X_0, \dots, X_n) = \mathbf{P}(X_{n+1}/X_n). \quad (3.1)$$

La chaîne est dite *homogène* si :

$$\mathbf{P}(X_{t_1}, \dots, X_{t_{n+1}}/X_{t_0}) = \mathbf{P}(X_{t_1-t_0}, \dots, X_{t_{n+1}-t_0}/X_0) \quad \forall n \geq 1 \quad (3.2)$$

Un *noyau de transition* sur  $(S, \mathcal{F})$  est une application  $P : S \times \mathcal{F} \rightarrow [0, 1]$  telle que :

- (i) la fonction  $P(\cdot, A)$  est mesurable, pour tout  $A \in \mathcal{F}$  ;
- (ii) la fonction  $P(x, \cdot)$  est une mesure de probabilité sur  $(S, \mathcal{F})$ , pour tout  $x \in S$  fixé.

Une chaîne de Markov homogène dans le temps de distribution *invariante*  $\pi$ , est une séquence de variables aléatoires  $\{X_n; n \geq 0\}$  telle que le noyau de transition :

$$P(X_n, A) = P(X_{n+1} \in A/X_0, \dots, X_n) \quad (3.3)$$

satisfasse :

$$\pi(A) = \int_S \pi(dx) P(x, A) \quad (3.4)$$

pour tous les ensembles mesurables  $A$ .

Un noyau de transition  $P(x, A)$  est *réversible* si l'intégrale :

$$\int \int 1_A(x) 1_B(y) \pi(dx) P(x, dy) \quad (3.5)$$

est symétrique si l'on interchange  $A$  et  $B$ . Si l'on développe les intégrales de (3.5) :

$$\int_B \pi(dx) P(x, A) = \int_A \pi(dx) P(x, B) \quad (3.6)$$

nous obtenons que la probabilité de partir de l'état  $B$  et d'arriver dans l'état  $A$  est égale à la probabilité de partir de l'état  $A$  et d'arriver dans l'état  $B$ .

Si le noyau est stochastique,  $P(x, S) = 1$  et en considérant  $B = S$ , nous avons en intégrant (3.6) :

$$\int_S \pi(dx) P(x, A) = \int_A \pi(dx) = \pi(A) \quad (3.7)$$



ce qui démontre l'invariance de la distribution  $\pi$ . La réversibilité du noyau implique donc l'invariance de la distribution  $\pi$ .

La distribution de  $X_0$  est la distribution initiale de la chaîne. La probabilité conditionnelle de  $X_n$  sachant  $X_0$  s'écrit :

$$P\{X_n \in A/X_0\} = P^n(X_0, A) \quad (3.8)$$

avec  $P^n$  la  $n$ -ième application du noyau  $P$ .

La distribution invariante (stationnaire) est une *distribution d'équilibre* si et seulement si :

$$\lim_{n \rightarrow \infty} P^n(x, A) = \pi(A) \quad (3.9)$$

pour tous les ensembles mesurables  $A$  et pour tout  $x$ .

## 3.2 Propriétés des Chaînes de Markov

Dans ce paragraphe nous allons présenter les notions *d'irréductibilité*, *de récurrence* et *d'ergodicité* [Meyn et Tweedie, 1993], [Robert, 1996]. Ces propriétés garantissent la convergence des chaînes de Markov vers leur distribution d'équilibre.

Soit  $\tau_A$ , le temps d'arrêt en  $A$  :

$$\tau_A = \text{Inf}\{n \geq 1 : X_n \in A\} \quad (3.10)$$

avec  $\tau_A = \infty$  si  $X_n \notin A$ .

Nous notons :

$$L(x, A) = P_x(\tau_A < \infty) \quad (3.11)$$

la probabilité de retour en  $A$  en un nombre fini d'étapes.  $P_x(\cdot)$  est la probabilité conditionnelle de la chaîne en partant de  $x$ .

### 3.2.1 Irréductibilité

Une première mesure de robustesse d'une chaîne de Markov par rapport à ses conditions initiales est l'irréductibilité. Cette propriété est une des conditions nécessaires pour garantir

la convergence des algorithmes de simulation des chaînes de Markov.

Une chaîne de Markov discrète est *irréductible* si tous les états communiquent entre eux :

$$L(x, y) > 0, \forall x, y \in S \quad (3.12)$$

Dans le cas continu,  $L(x, y) = 0$  presque partout. Pour définir proprement cette notion, nous introduisons une mesure auxiliaire  $\varphi$  sur  $\mathcal{F}$ .

Une chaîne est  $\varphi$ -*irréductible* si la mesure  $\varphi$  possède la propriété suivante pour tout  $x \in S$  :

$$\varphi(A) > 0 \Rightarrow L(x, A) > 0 \quad (3.13)$$

La  $\varphi$ -irréductibilité nous assure que pour des ensembles "suffisamment grands" mesurés par  $\varphi$ , il est possible de les atteindre à partir de n'importe quel point initial en un temps fini.

Nous appelons la chaîne irréductible qui a une distribution invariante, chaîne *positive*.

### 3.2.2 Atomes et ensembles petits

Une chaîne de Markov possède *un atome*  $\alpha \in \mathcal{F}$ , s'il existe une mesure  $\nu$  telle que :

$$P(x, A) = \nu(A) \quad \forall x \in \alpha, \forall A \in \mathcal{F} \quad (3.14)$$

Cette notion implique que le noyau de transition est constant sur un ensemble de mesure non-nulle. C'est une contrainte trop importante pour le cas continu.

Un ensemble  $C$  est *petit* s'il existe  $m \in \mathbb{N}^*$  et une mesure  $\nu_m(A) > 0$  tels que :

$$P^m(x, A) \geq \nu_m(A) \quad \forall x \in C \quad \forall A \in \mathcal{F} \quad (3.15)$$

Le résultat suivant montre la relation entre l'irréductibilité et les ensembles petits.

**THEOREME :** *Soit  $(X_n)$   $\varphi$ -irréductible. Pour tout ensemble  $A \in \mathcal{F}$  tel que  $\varphi(A) > 0$ , il existe  $m \in \mathbb{N}^*$  et un ensemble petit  $C \subset A$  tels que  $\nu_m(C) > 0$ . De plus,  $S$  se décompose en une partition dénombrable d'ensembles petits.*

Le noyau de transition des chaînes discrètes a nécessairement un caractère atomique. En conclusion, nous avons présenté des notions qui nous permettent d'exercer dans le cas continu un contrôle aussi "précis" que sur les chaînes discrètes.

### 3.2.3 Cycle et apériodicité

Les ensembles  $A_1, \dots, A_m$  de  $\mathcal{F}$  forment un  $m$ -cycle si :

$$\begin{aligned} P(x, A_2) &= 1, & x \in A_1 \\ &\vdots \\ P(x, A_m) &= 1, & x \in A_{m-1} \\ P(x, A_1) &= 1, & x \in A_m \end{aligned} \tag{3.16}$$

et si :

$$\varphi([\cap_{i=1}^m A_i]^c) = 0 \tag{3.17}$$

Une chaîne est *apériodique* si elle est irréductible et si elle possède un cycle avec  $m = 1$ .

Si une chaîne a une distribution invariante  $\pi$  et, de plus si elle est irréductible et apériodique, alors  $\pi$  est unique et elle est également la distribution d'équilibre de la chaîne [Tierney, 1994].

### 3.2.4 Récurrence au sens de Harris

La condition d'irréductibilité nous assure que la chaîne de Markov peut atteindre tout point de l'espace d'intérêt. Mais cette propriété ne garantit pas un nombre de passages suffisamment grand dans le voisinage de ce point. Par exemple, dans le cas d'une optimisation globale, la convergence de la méthode n'est garantie que si l'on assure un balayage systématique de l'espace des états.

Soient :

$$\eta_A = \sum_{t=1}^{\infty} 1_A(x_t) \tag{3.18}$$

et :

$$U(x, A) = E_x[\eta_A] \tag{3.19}$$

le nombre de passages de  $(X_n)$  en  $A$  et le nombre moyen de passages en  $A$ .

Nous appelons  $A$  un *ensemble récurrent* si pour tout  $x \in S$  :

$$U(x, A) = \infty \tag{3.20}$$

et nous l'appelons *ensemble uniformément transient* s'il existe  $M < \infty$ , tel que :

$$U(x, A) \leq M \tag{3.21}$$

Un *ensemble transient* est une réunion dénombrable d'ensembles uniformément transients.

DEFINITION : Une chaîne  $(X_n)$  est récurrente s'il existe une mesure  $\varphi$  tel que  $(X_n)$  soit  $\varphi$ -irréductible et si  $U(x, A) = \infty$  pour tout  $x \in S$  et pour tout  $A \in \mathcal{F}$  tel que  $\varphi(A) > 0$ . Elle est transiente si elle est  $\varphi$ -irréductible et si  $S$  est transient.

La stabilité de la chaîne  $(X_n)$  est renforcée si le nombre de passages est infini pour presque toute trajectoire. Dans ce cas, la chaîne est dite *récurrente au sens de Harris*.

DEFINITION : Un ensemble  $A$  est récurrent au sens de Harris si

$$Q(x, A) = P_x(\eta_A = \infty) = 1, \quad \forall x \in A \quad (3.22)$$

La chaîne est récurrente au sens de Harris s'il existe une mesure  $\varphi$  telle que  $(X_n)$  soit  $\varphi$ -irréductible et si tout ensemble  $A$  tel que  $\varphi(A) > 0$  est récurrent au sens de Harris

Si une chaîne positive est récurrente, elle est dite *chaîne positive*. Les chaînes récurrentes au sens de Harris sont dites *positives au sens de Harris*.

### 3.2.5 Ergodicité

Les chaînes simulées par les méthodes de Monte Carlo sont construites avec cette propriété de positivité. Nous introduisons la propriété *d'ergodicité* qui nous garantit la convergence de la chaîne vers sa distribution d'équilibre indépendamment des conditions initiales.

Soit la norme de la variation totale :

$$\| \mu_1 - \mu_2 \|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)| \quad (3.23)$$

DEFINITION : Pour une chaîne  $(X_n)$  positive au sens de Harris, un atome  $\alpha$  est ergodique si

$$\lim_{n \rightarrow \infty} |P^n(\alpha, \alpha) - \pi(\alpha)| = 0 \quad (3.24)$$

Pour une chaîne positive récurrente et apériodique, tout atome est ergodique.

THEOREME : Si  $(X_n)$  est positive au sens de Harris et apériodique,

$$\lim_{n \rightarrow \infty} \left\| \int P^n(x, \cdot) \mu(dx) - \pi \right\|_{TV} = 0 \quad (3.25)$$

pour toute distribution initiale  $\mu$ .

L'égalité :

$$\| \mu \|_{TV} = \sup_{|h| \leq 1} \left| \int h(x) \mu(dx) \right| \quad (3.26)$$

implique la convergence :

$$\lim_{n \rightarrow \infty} |E_\mu(h(X_n)) - E^\pi h(x)| = 0 \quad (3.27)$$

pour toute fonction bornée  $h$ .

### 3.2.6 Convergence géométrique

Une mesure de la qualité des algorithmes de type Monte Carlo par chaînes de Markov est l'étude de la vitesse de convergence de  $P^n$  vers  $\pi$ . Une telle évaluation permet la construction de critères d'arrêt pour ces algorithmes et dans le cas de l'existence de vitesses suffisamment grandes, elle permet l'application du Théorème de la Limite Centrale.

La généralisation de la norme de la variation totale en la norme  $\| \cdot \|_h$  est donnée par :

$$\| \mu \|_h = \sup_{\|g\| \leq h} \left| \int g(x) \mu(dx) \right| \quad (3.28)$$

DEFINITION : Une chaîne  $(X_n)$  est dite géométriquement  $h$ -ergodique, avec  $h \geq 1$  sur  $S$ , si  $(X_n)$  est Harris positive de loi stationnaire  $\pi$ , vérifie  $E^\pi[h] < \infty$  et s'il existe  $r_h > 1$  tel que

$$\sum_{n=1}^{\infty} r_h^n \|P^n(x, \cdot) - \pi\|_h < \infty \quad (3.29)$$

pour tout  $x \in S$ . Le cas  $h = 1$  correspond à l'ergodicité géométrique.

La  $h$ -ergodicité géométrique impose que la convergence de  $P^n$  vers  $\pi$  se fasse au moins avec une vitesse géométrique car l'équation (3.29) implique :

$$\| P^n(x, \cdot) - \pi \|_h \leq M r_h^{-n} \quad (3.30)$$

avec

$$M = \sum_{n=1}^{\infty} r_h^n \| P^n(x, \cdot) - \pi \|_h \quad (3.31)$$

DEFINITION : La chaîne  $(X_n)$  est uniformément ergodique si

$$\lim_{n \rightarrow \infty} \sup_{x \in S} \| P^n(x, \cdot) - \pi \| = 0 \quad (3.32)$$

La notion d'*ergodicité uniforme* impose un taux de convergence uniforme sur l'espace entier. Elle complète l'ergodicité géométrique et elle sert de base au Théorème de la Limite Centrale.

Soit  $g(x)$  une fonction  $\pi$ -intégrable,

$$s = E_{\pi}g(X) = \int g(x)\pi(dx) \quad (3.33)$$

avec  $\pi$  la distribution d'équilibre d'une chaîne Harris récurrente.

Soient également :

$$\hat{s}_n = \frac{1}{n} \sum_{i=1}^n g(X_i) \quad (3.34)$$

les sommes partielles calculées à partir d'une réalisation tronquée quelconque  $X_1, \dots, X_n$  de la chaîne.

Si la fonction  $g$  satisfait la condition de Lyapunov :

$$\int \|g(x)\|^{2+\epsilon} \pi(dx) < \infty \quad (3.35)$$

avec  $\epsilon > 0$ , alors le Théorème de la Limite Centrale implique :

$$\sqrt{n}(s_n - s) \rightarrow \mathcal{N}(0, \gamma^2) \quad (3.36)$$

où :

$$\gamma^2 = \text{Var}_{\pi}(g(X)) + 2 \sum_{k=1}^{\infty} \text{Cov}_{\pi}(g(X_t)g(X_{t+k})) \quad (3.37)$$

Le Théorème de la Limite Centrale permet de contrôler la convergence des algorithmes de Monte Carlo par chaînes de Markov au sens où, lorsque  $\gamma^2 > 0$ , nous pouvons vérifier si les moyennes  $s_n(g)$  convergent vers la quantité  $E_{\pi}[g]$ . La difficulté majeure de ce résultat

est l'estimation de la quantité  $\gamma^2$ . L'importance de ce théorème est que l'on sait, qu'à partir d'un large nombre d'échantillons nous pouvons calculer des approximations des quantités d'intérêt, mais aucune indication n'est donnée concernant le nombre d'échantillons nécessaire pour un calcul fiable.

### 3.3 Méthodes du type Monte Carlo par Chaînes de Markov

Soit  $(S, \mathcal{F}, \mu)$  un espace mesuré et  $h$  une fonction mesurable réelle positive sur  $S$ . Si :

$$0 < c = \int h(x)\mu(dx) < \infty \quad (3.38)$$

soit :

$$f(x) = \frac{1}{c}h(x) \quad (3.39)$$

nous disons que  $h(x)$  est la densité de probabilité non-normalisée et que  $f(x)$  est la densité de probabilité normalisée.

Il se trouve que, dans beaucoup de situations, le calcul de la constante de normalisation  $c$  ne peut pas se faire. Dans ce cas, les problèmes ne peuvent pas être résolus du point de vue analytique.

Les méthodes Monte Carlo par Chaînes de Markov dites MCMC représentent une solution à cette difficulté : par la construction d'un noyau stochastique, il est possible de simuler une chaîne de Markov dont la distribution d'équilibre converge en loi vers  $h$ .

[Geyer, 1999] considère qu'il est plus naturel de parler de "mise-à-jour" de l'état du système, que de l'échantillonnage de la loi  $h$ . Nous allons adopter ce point de vue dans la suite de ce manuscrit.

#### 3.3.1 Combinaison de mises-à-jour

Il y a deux grandes méthodes pour combiner les mécanismes de mise-à-jour : composition et mélange. La composition correspond à la multiplication des noyaux de transition, et le mélange à leur combinaison linéaire.

La propriété fondamentale d'une méthode MCMC est qu'elle possède une distribution stationnaire.

Si un noyau de transition  $P_1$  est suivi au pas d'après par le noyau  $P_2$ , alors la *composition* des deux noyaux s'écrit :

$$(P_1P_2)(x, A) = \int P_1(x, dy)P(y, A) \quad (3.40)$$

Il est facile de prouver que si  $\pi$  est stationnaire pour  $P_i$  avec  $i = 1 \dots d$ , alors  $\pi$  est stationnaire pour le noyau composé  $P_1P_2 \dots P_d$ .

Nous pouvons combiner les mises-à-jour d'une manière statistique. Le *mélange* des noyaux  $P_i$  se fait en assignant à chacun d'entre eux une probabilité  $p_i$ . Alors, le noyau de transition sera :  $P = \sum p_i P_i$ . Comme dans le cas précédent, la mesure  $\pi$  est invariante pour le noyau  $P$ .

Une autre façon de combiner les mises-à-jour est le *mélange dépendant de l'état (state-dependent mixing)* proposé par [Green, 1995].

Supposons que l'on dispose d'un nombre fini ou infini de noyaux de transition  $P_i(x, A)$ ,  $i \in I$ , et de probabilités de mélange  $p_i(x)$ . Le noyau de transition sera alors :

$$P(x, A) = \sum_{i \in I} p_i(x) P_i(x, A) \quad (3.41)$$

Alors, pour faire un mouvement lorsque l'on se trouve dans l'état  $x$ , on choisit le noyau  $P_i$  avec la probabilité  $p_i(x)$  et on simule l'état suivant avec la probabilité  $P_i(x, \cdot)$ .

Il se trouve qu'aucune des propriétés des  $P_i$  n'est transmise au noyau  $P$ , c'est à dire que  $\pi$  ne l'est pas obligatoirement stationnaire pour le noyau  $P$ .

Soit les noyaux  $K_i(x, A) = p_i(x)P_i(x, A)$ . Alors :  $P = \sum_{i \in I} K_i(x, A)$ . Si les  $K_i$  sont réversibles alors  $P$  est aussi réversible, et si  $P$  est stochastique, alors  $\pi$  est une distribution stationnaire de  $P$ .

Nous avons :

$$p_i(x) = K_i(x, S) \quad (3.42)$$

et :

$$P_i(x, A) = \frac{K_i(x, A)}{K_i(x, S)} \quad (3.43)$$



nous pouvons donc considérer la chaîne définie à partir des  $K_i$ . Comme certains mouvements ne vont pas pouvoir être faits à partir de l'état  $x$ , il se peut que  $K_i(x, S) = 0$  pour certains  $i$ . Alors :

$$\sum_i K_i(x, S) \leq 1, \forall x \in S \quad (3.44)$$

Soit  $d$  la probabilité de ne pas modifier l'état  $x$  :

$$d(x) = 1 - \sum_i K_i(x, \Omega), x \in \Omega \quad (3.45)$$

et le noyau :

$$K'(x, A) = d(x)\mathbb{I}(x, A) \quad (3.46)$$

avec  $\mathbb{I}$  le noyau identité :

$$\mathbb{I}(x, A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \quad (3.47)$$

L'intégrale :

$$\int \int 1_A(x)1_B(y)\pi(dx)K'(x, dy) \quad (3.48)$$

est symétrique si l'on change  $A$  en  $B$ . Cela implique que le noyau  $K'(x, \cdot)$  est réversible par rapport à la mesure  $\pi$ . Si nous ajoutons le noyau  $K'$  à l'ensemble des noyaux  $K_i$ , la somme devient stochastique. Le noyau de transition  $K' + \sum_{i \in I} K_i$  est réversible par rapport à  $\pi$  si les  $K_i$  sont réversibles par rapport à  $\pi$ .

Alors, utilisant le noyau  $K' + \sum_{i \in I} K_i$ , nous avons le mécanisme de mise à jour suivant :

1. Choisir  $i$  avec la probabilité  $p_i$ . Avec la probabilité  $1 - \sum_{i \in I} p_i$  n'effectuer aucun mouvement et sauter l'étape suivante.
2. Simuler un nouvel état à partir de  $x$  avec la distribution  $P_i(x, \cdot)$ .

### 3.3.2 Metropolis-Hastings

Une mise à jour de type Metropolis-Hastings simule une distribution non-normalisée  $h$ , par rapport à la mesure  $\mu$ . Cette technique utilise une *distribution auxiliaire (loi instrumentale)*  $q(x \rightarrow \cdot)$ .

La distribution auxiliaire doit être mesurable par rapport à  $\mu$ ,  $q(x \rightarrow dy)$  peut être évaluée pour tous les  $(x, y)$  et pour tout  $x$ , il est possible de simuler  $q(x \rightarrow \cdot)$  par rapport

à  $\mu$ .

ALGORITHME :

1. *Simuler*  $y \sim q(x \rightarrow \cdot)$
2. *Calculer le rapport :*

$$R = \frac{h(y)q(y \rightarrow dx)}{h(x)q(x \rightarrow dy)} \quad (3.49)$$

3. *Accepter le nouvel état  $y$  avec la probabilité  $\min(1, R)$ .*

Si  $q(x \rightarrow dy) = q(y \rightarrow dx)$ , alors  $R = \frac{h(y)}{h(x)}$ . C'est la mise à jour de type Metropolis. Elle a l'avantage d'éliminer le calcul de la densité auxiliaire.

### 3.3.3 Metropolis-Hastings une variable à la fois

Supposons que  $x = (x_1, \dots, x_d)$  et que  $\mu = \mu_1 \times \dots \times \mu_d$ . Il existe une variante de l'algorithme Metropolis-Hastings qui ne change qu'une seule variable à la fois. La densité auxiliaire est de la forme  $q_i(x_i \rightarrow \cdot)$ .

ALGORITHME :

1. *Simuler*  $y \sim q_i(x \rightarrow \cdot)$ . Où  $y = (x_1, \dots, x_{i-1}, x_y, x_{i+1}, \dots, x_d)$
2. *Calculer le rapport :*

$$R = \frac{h(y)q_y(x_y \rightarrow dx_i)}{h(x)q_i(x_i \rightarrow dx_y)} \quad (3.50)$$

3. *Accepter le nouvel état  $y$  avec la probabilité  $\min(1, R)$ .*

### 3.3.4 Echantillonneur Gibbs

L'échantillonneur de Gibbs est une variante de Metropolis-Hastings une variable à la fois. La densité non-normalisé  $h$ , est aussi la densité auxiliaire  $q_i(x_i \rightarrow \cdot)$ . Si  $c$  est la constante de normalisation, le taux d'acceptation devient :

$$R = \frac{h(y)q_y(x_y \rightarrow dx_i)}{h(x)q_i(x_i \rightarrow dx_y)} = \frac{h(y)h(x)/c}{h(x)h(y)/c} = 1 \quad (3.51)$$

La différence entre l'échantillonneur de Metropolis et celui de Gibbs est que le mécanisme de mise à jour de type Gibbs accepte tous les nouveaux états, mais les états sont générés selon la densité vers laquelle nous souhaitons converger. L'échantillonneur de Gibbs peut être utilisé si la densité de probabilité que l'on veut simuler peut être calculée localement.

Cette contrainte ne s'applique pas dans le cas de Metropolis car les nouveaux états sont générés selon une densité auxiliaire uniforme sur l'ensemble des états. Le taux d'acceptation de ces nouvelles configurations peut être inférieur à 1. Cela veut dire que certaines configurations proposées ne sont pas acceptées. En revanche, le calcul de la densité auxiliaire est plus rapide que pour l'échantillonneur de Gibbs.

### 3.3.5 Metropolis-Hastings-Green

La dynamique de type Metropolis-Hastings-Green [Green, 1995] est aussi connue sous le nom d'algorithme de type Monte Carlo à sauts réversibles (*Reversible Jump Monte Carlo Markov Chains - RJMCMC*). Cette formulation nous permet de gérer des problèmes où la dimension du sous-espace de configurations est à déterminer et où le choix de mouvements se fait en fonction de l'état.

Le principe de cette dynamique est similaire à celui de la dynamique de Metropolis-Hastings. Supposons que la chaîne de Markov que nous souhaitons simuler se trouve dans l'état  $x$ . Nous choisissons d'une façon aléatoire un mouvement de type  $i$  qui déplace  $x$  dans  $y$  avec la probabilité  $q_i(x \rightarrow dy)$ .

Le nouvel état  $y$ , est accepté avec la probabilité :

$$\alpha_i(x \rightarrow y) = \min\{1, R\} \quad (3.52)$$

où :

$$R = \frac{f(y)q_i(y \rightarrow dx)}{f(x)q_i(x \rightarrow dy)} \quad (3.53)$$

Quand la dimension de l'espace de configurations est variable, une mise en correspondance des dimensions de  $x$  et  $y$  est nécessaire.

Pour ce faire, un vecteur de variables aléatoires continues  $u$  est généré indépendamment de  $x$ . La nouvelle configuration  $y$  est obtenue en utilisant une fonction déterministe inversible

$y'(x, u)$ . Le taux d'acceptation est modifié par le Jacobien de la transformation :

$$R = \frac{f(y)}{f(x)q(u^{(i)})} \left| \frac{\partial y'}{\partial(x, u)} \right| \frac{p(\text{choisir de proposer le saut inverse})}{p(\text{choisir de proposer un saut avant})} \quad (3.54)$$

avec  $q(u^{(i)})$  la densité du vecteur  $u$ .

Si la dimension des configurations  $x$  et  $y$  est la même, nous sommes dans le cas de la dynamique de Metropolis-Hastings.

Dans une application pratique avec plusieurs sous-espaces de configuration, nous devons pouvoir les parcourir en totalité.

Cette opération va être faite en construisant des mouvements entre les différents sous-espaces. Ces mouvements vont être choisis aléatoirement. Les différents mouvements doivent être construits de telle façon que la chaîne de Markov que l'on souhaite simuler soit apériodique, irréductible et réversible.

Ces conditions nous assurent de la possibilité de parcourir tout l'espace des états et de la convergence de la distribution de la chaîne vers sa distribution d'équilibre.

L'algorithme devient alors :

ALGORITHME :

1. Choisir avec la probabilité  $p_i(x)$ , un mouvement de transition,  $i$ . Avec la probabilité  $1 - \sum_{i \in I} p_i(x)$  ne pas faire ce mouvement, et rester dans l'état  $x$  ;
2. Simuler  $y \sim q_i(x, \cdot)$  ;
3. Calculer le rapport de Green  $R$  selon l'équation (3.54) :
4. Accepter le nouvel état  $y$ , avec la probabilité  $\min(1, R)$ .

### 3.4 Simulation des processus ponctuels

Nous allons commencer ce paragraphe par la présentation de la méthode la plus ancienne pour la simulation des processus ponctuels, les *algorithmes de type naissance et mort* [Lieshout, 2000].

Ces algorithmes n'autorisent que deux types de mouvements : ajouter un point à la configuration (*naissance*) et enlever un point à la configuration (*mort*). Pour modifier un point de la configuration, nous pouvons toujours nous servir des mouvements de naissance et mort.

Le processus simulé reste dans l'état  $X_n = x$ , pendant le temps  $T_n$ . Le temps  $T_n$ , est une variable aléatoire avec une distribution exponentielle.

Si nous souhaitons passer dans l'état  $y = x \cup \zeta$ , nous avons le taux de naissance global :

$$B(x) = \int_{\mathcal{K}} b(x, \zeta) d\mu(\zeta) \quad (3.55)$$

La forme la plus utilisée pour le taux de naissance local est donnée par l'intensité de Papangelou :

$$b(x, \zeta) = \frac{f(x \cup \zeta)}{f(x)} \quad (3.56)$$

Si nous souhaitons enlever un point de la configuration, c'est à dire passer dans l'état  $y = x \setminus \zeta$ , avec  $\zeta \in x$ , le taux de mort global est donné par :

$$D(x) = \sum_{\zeta \in x} d(x, \zeta) \quad (3.57)$$

Le plus souvent le taux de mort local est choisi constant i.e.  $d(x, \zeta) = 1$ .

ALGORITHME :

1.  $T_n$  est distribué selon une loi exponentielle de moyenne  $1/(B(x) + D(x))$ .  $T_n$  est choisi indépendamment des autres temps de séjour.

2. Le mouvement de type mort est proposé avec la probabilité  $D(x)/(B(x) + D(x))$ . Un point quelconque  $\zeta$  de la configuration est enlevé avec la probabilité :

$$\frac{d(x, \zeta)}{D(x)} \quad (3.58)$$

3. Le mouvement de type naissance est proposé avec la probabilité  $B(x)/(B(x) + D(x))$ . Un nouveau point  $\zeta$  est ajouté à la configuration initiale. Le point est généré par rapport à la densité de probabilité :

$$\frac{b(x, \zeta)}{B(x)} \quad (3.59)$$

*et la mesure  $\mu$ .*

La simulation des processus ponctuels peut donc se faire à l'aide de deux grandes familles de méthodes : les algorithmes de type naissance-mort, et les techniques de simulation de type Monte Carlo par chaînes de Markov que nous avons vu précédemment.

Les deux familles de méthodes utilisent trois types de mouvements : naissance, mort ou déplacement d'un point à l'intérieur de la configuration.

Les premiers types de mouvements (naissance et mort) impliquent la modification de la dimension de sous-espace de configuration. Les dynamiques de Monte Carlo à sauts réversibles RJMCMC [Green, 1995] sont particulièrement bien adaptées à ce problème.

Les deux techniques de simulation, processus naissance-mort et RJMCMC ont été comparées [Geyer et Møller, 1994], [Geyer, 1999], [Møller, 1999], [Imberty et Descombes, 2000]. Les algorithmes de type naissance-mort convergent en un nombre plus faible d'itérations. L'inconvénient de ces algorithmes est que les probabilités pour ajouter ou bien enlever un point dans la configuration se calculent d'une façon globale [Ripley, 1979]. Ce qui est très coûteux en temps de calcul.

Les méthodes RJMCMC nécessitent plusieurs itérations car beaucoup de propositions de nouvelle configuration sont refusées. Par contre, le calcul du taux d'acceptation se fait localement, ce qui fait que les algorithmes de type Monte Carlo à sauts réversibles sont plus rapides et beaucoup plus simples à mettre en œuvre.

## Chapitre 4

# "Candy" modèle

---

Nous présentons dans ce chapitre un modèle probabiliste pour les réseaux linéïques [Descombes et al., 1999b], [Stoica et al., 2000c]. Nous considérons qu'un reseau linéïque est composé de plusieurs segments. Ces segments sont caractérisés par leur position, orientation, épaisseur et longueur. Ces grandeurs sont des variables aléatoires indépendantes. Un réseau linéïque sera donc la réalisation d'un processus ponctuel marqué.

Dans le premier paragraphe nous allons présenter le modèle d'un segment aléatoire, que nous avons appelé "Candy" modèle. En s'appuyant sur ce nouveau modèle, nous allons ensuite modéliser un réseau linéïque.

---





## 4.1 Modèle probabiliste d'un réseau linéique

Soit le réseau linéique  $S$  composé par les segments  $s_i$  :

$$\begin{aligned} S &= \{s_i, \quad i = 1, \dots, n\} \\ n &\in \mathbb{N} \end{aligned} \tag{4.1}$$

Un réseau linéique est la réalisation d'un processus ponctuel marqué dont la densité de probabilité non-normalisée est :

$$f(S) \propto \beta^{n(S)} \prod_{s_i \in S} g(s_i) \prod_{s_i \diamond s_j, i < j} h(s_i, s_j) \tag{4.2}$$

avec  $\beta$  la densité de segments.  $s_i \diamond s_j$  signifie que les deux segments interagissent. Les fonctions  $g$  et  $h$  "pénalisent" la dimension, l'état et les interactions des segments formant le réseau.

Les paramètres qui caractérisent un segment  $s_i$  sont des variables aléatoires continues :

$$\begin{aligned} (x_i, y_i) &\in K \subset \mathbb{R}^2 \quad \nu(K) < \infty \\ w_i &\sim \mathcal{U}([w_{min}, w_{max}]) \\ l_i &\sim \mathcal{U}([l_{min}, l_{max}]) \\ \theta_i &\sim \mathcal{U}([0, 2\pi]) \end{aligned} \tag{4.3}$$

avec  $\mathcal{U}$  la loi uniforme,  $(x_i, y_i)$  les coordonnées du centre du segment,  $w_n, l_n, \theta_n$  sont respectivement l'épaisseur, la longueur et l'orientation du segment. Le vecteur de paramètres  $m_i = [w_i \quad l_i \quad \theta_i]$  est la marque du processus ponctuel induit par  $(x_n, y_n)$ .  $\nu$  est la mesure de Lebesgue.

Un réseau est composé par des segments connectés. Un segment a deux extrémités qui peuvent être connectées. Leur coordonnées sont  $t_i = T(t_{x_i}, t_{y_i})$  et  $q_i = Q(q_{x_i}, q_{y_i})$  (cf. Figure 4.1).

Si une des extrémités du segment  $s_i$  se trouve dans un petit voisinage  $\epsilon$  d'une extrémité quelconque du segment  $s_j$ , telle que  $\nu(\epsilon) > 0$ , nous considérons les deux segments comme *connectés*.

Un segment a des états. Il peut être libre, c'est à dire qu'il n'a aucune connexion ( $s = s^\emptyset$ ). Si le segment est connecté, il peut avoir une seule extrémité connectée ( $s = s^{t||q}$ ) ou bien les deux extrémités connectées ( $s = s^{tq}$ ). Le nombre de connexions d'un segment n'est pas

limité.

La fonction  $g$  pénalise la dimension et la connectivité du réseau. Son expression est donnée par :

$$g(s_i) = g_1(s_i) \times g_2(s_i) \quad (4.4)$$

Le premier terme :

$$g_1(s_i) = \exp\left(\frac{l_{s(i)} - l_{max}}{l_{max}}\right) \quad (4.5)$$

pénalise les segments qui ne sont pas très longs.

Le second terme :

$$g_2(s_i) = \begin{cases} g_{21}, & \text{si } s_i = s_i^\emptyset \\ g_{22}, & \text{si } s_i = s_i^{t||q} \\ g_{23} = 1, & \text{si } s_i = s_i^{tq} \end{cases} \quad (4.6)$$

prend en considération l'état du segment. Les segments libres sont fortement pénalisés. Les segments qui ont une seule extrémité connectée sont moins pénalisés. La fonction  $g$  ne sanctionne pas les segments avec les deux extrémités connectées. En général, nous avons :  $g_{21} \ll g_{22} < 1$ .

Nous souhaitons que les segments ne se superposent pas et qu'ils soient alignés. C'est la fonction  $h$  qui contrôle ces phénomènes. Nous considérons deux types d'interaction entre deux segments : la répulsion et l'attraction.

Pour définir l'interaction de répulsion, nous devons définir une région de répulsion autour du segment comme dans la Figure 4.1,  $R_r(s) = C_r[O = (x, y), r_r = l/2]$  où  $C_r[O, r]$  est le disque de centre  $O$  et rayon  $r$ .

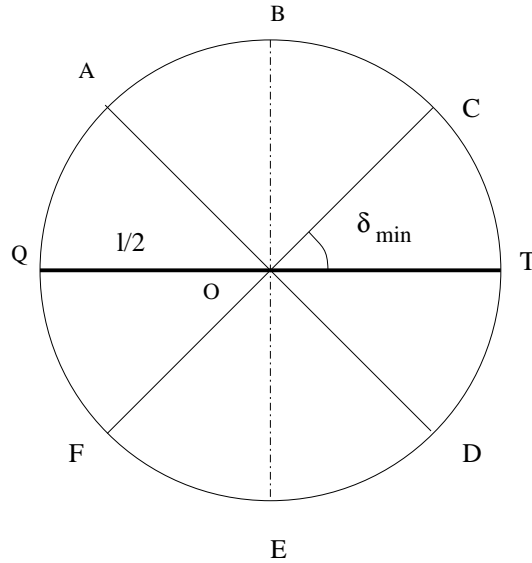
Les segments  $s_i$  et  $s_j$  présentent une interaction de type *répulsion* si :

$$\begin{cases} (x_i, y_i) \in R_r(s_j) \quad \text{ou} \quad (x_j, y_j) \in R_r(s_i) \\ \Xi(s_i, s_j) \quad \text{est vérifiée} \end{cases} \quad (4.7)$$

où  $\Xi(s_i, s_j)$  est une condition qui autorise le croisement entre les segments. Pour la définir, considérons  $\delta$  comme suit :

Nous définissons  $\delta$  :

$$\delta = \begin{cases} \frac{\|\theta_j - \theta_i\| - \pi/2}{\pi}, & \text{si } \|\theta_j - \theta_i\| \leq \pi \\ \frac{\|\theta_j - \theta_i\| - 3\pi/2}{\pi}, & \text{si } \pi < \|\theta_j - \theta_i\| \leq 2\pi \end{cases} \quad (4.8)$$

FIG. 4.1 – Région de répulsion autour du segment  $\overline{QT}$ 

$\theta_i, \theta_j$  étant les orientations des segments. La condition  $\Xi(s_i, s_j)$  est vérifiée si et seulement si :

$$\delta < \delta_{\min} \quad (4.9)$$

avec  $\delta_{\min}$  l'angle minimal entre les segments croisés.

Nous montrons un exemple d'interaction de répulsion sur la Figure 4.2. Les segments  $\overline{TQ}$  et  $\overline{T_2Q_2}$  présentent une interaction de type répulsion. Les segments  $\overline{TQ}$  et  $\overline{T_1Q_1}$  satisfont  $\delta \geq \delta_{\min}$ , ce qui signifie que leur position relative est considérée comme un croisement dans le réseau, et non pas comme une répulsion. Nous avons la même situation pour les segments  $\overline{T_1Q_1}$  et  $\overline{T_2Q_2}$  qui ne sont pas en interaction de répulsion, mais dans une position de croisement.

Par conséquent, si deux segments interagissent d'une façon répulsive, la fonction  $h$  est :

$$h(s_i, s_j) = h_r \leq 1 \quad (4.10)$$

Deux segments peuvent s'attirer réciproquement. Un segment possède donc une région d'attraction  $R_a$  comme décrit dans la Figure 4.3. Cette région est définie par deux cercle centrés sur chacune des extrémités du segment :  $R_a(s) = C_{at}[t(s), r_a = l/4] \cup C_{aq}[q(s), r_a = l/4]$ .

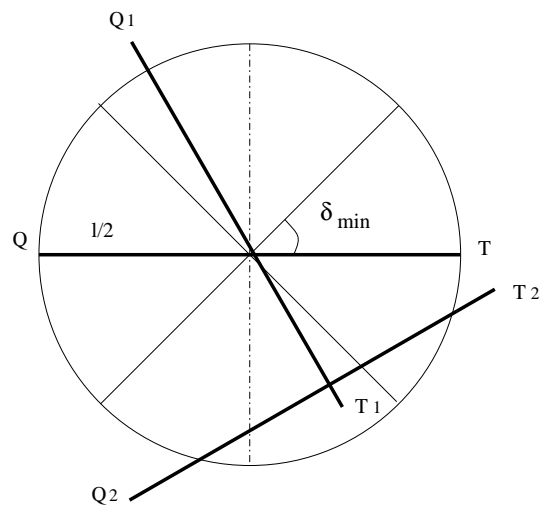


FIG. 4.2 – Interaction de type répulsion entre les segments

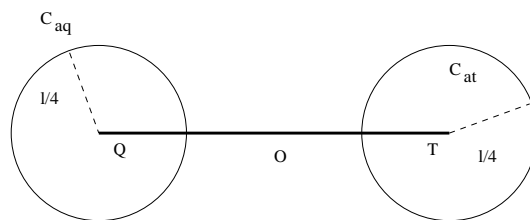


FIG. 4.3 – Région d'attraction autour d'un segment

Deux segments présentent une interaction de type *attraction* si toutes les conditions suivantes sont accomplies :

$$\begin{cases} (x_i, y_i) \notin R_r(s_j) \\ (x_j, y_j) \notin R_r(s_i) \\ A(s_i, s_j) \end{cases} \quad (4.11)$$

avec la condition  $A(s_i, s_j)$  qui est satisfaite, si une des propositions suivantes est vérifiée :

$$\begin{aligned} t_i \in R_a(s_j) \quad \text{et} \quad q_i \notin R_a(s_j) \\ q_i \in R_a(s_j) \quad \text{et} \quad t_i \notin R_a(s_j) \\ t_j \in R_a(s_i) \quad \text{et} \quad q_j \notin R_a(s_i) \\ q_j \in R_a(s_i) \quad \text{et} \quad t_j \notin R_a(s_i) \end{aligned} \quad (4.12)$$

Nous montrons un exemple sur la Figure 4.4. Les segments  $\overline{TQ}$  et  $\overline{T_1Q_1}$  se trouvent en interaction de type attraction. Les segments  $\overline{T_1Q_1}$  et  $\overline{T_2Q_2}$  sont également en interaction de type attraction, alors que les segments  $\overline{TQ}$  et  $\overline{T_2Q_2}$  se trouvent en interaction de type répulsion.

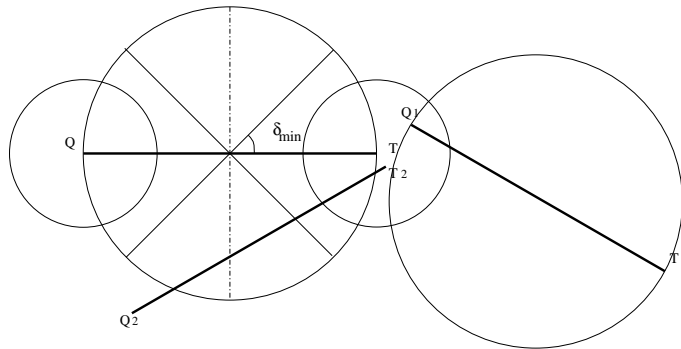


FIG. 4.4 – Interaction de type attraction entre les segments

Nous voulons que les segments soient bien alignés. Pour ce faire, nous devons vérifier si les segments ont la même orientation ou pas. C'est pour quoi nous définissons la courbure :

$$\tau = \begin{cases} \tau_1 = \frac{\|\theta_j - \theta_i\|}{\pi}, & \theta_j \neq \theta_i \\ \tau_2 = \frac{\theta_c}{\pi}, & \theta_j = \theta_i \end{cases} \quad (4.13)$$

où  $\theta_i, \theta_j$  sont les orientations des segments. Si les deux segments ont la même orientation, nous devons définir  $\theta_c$ , l'angle entre un des segments et la droite formée par les centres des deux segments (cf. Figure 4.5).

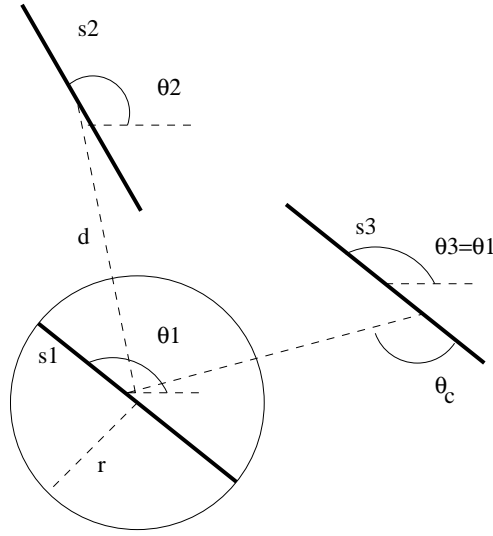


FIG. 4.5 – Segments et leur différentes orientations

Si deux segments se trouvent dans une interaction de type attraction, la fonction  $h$  pénalise les segments qui ne sont pas alignés :

$$h(s_i, s_j) = h_o \leq 1 \quad \text{si} \quad \tau > \tau_{\max} \quad (4.14)$$

Le système de voisinage est symétrique. Cette construction nous garantit que la chaîne de Markov qui simule le modèle aura de bonnes propriétés comme l'apériodicité, l'irréductibilité et la réversibilité.

Le processus ponctuel que nous venons de définir porte le nom de "*Candy*" modèle.

#### 4.1.1 "Candy" modèle : processus ponctuel de Gibbs

Nous pouvons placer le processus ponctuel "Candy" dans le cadre des processus ponctuels de Gibbs. La densité de probabilité d'un processus ponctuel de Gibbs est à une constante près :

$$f(S) \propto \exp(-U_I(S)) \quad (4.15)$$

avec  $U_I(S)$  l'énergie totale du système.

Le "Candy" modèle se met sous la forme d'un processus ponctuel de Gibbs, en écrivant

l'énergie de la manière suivante :

$$U_I(S) = -\log\left(\prod_{s_i \in S} g(s_i) \prod_{s_i \diamond s_j, i < j} h(s_i, s_j)\right) \quad (4.16)$$

d'où

$$U_I(S) = -\left(\sum_{s_i \in S} \log g(s_i) + \sum_{s_i \diamond s_j, i < j} \log h(s_i, s_j)\right) \quad (4.17)$$

Prenons une forme simplifiée du "Candy" modèle avec  $g_1(s) = 1$ , ce qui est équivalent à ne plus pénaliser la longueur des segments. Alors, l'énergie de Gibbs d'une réalisation du "Candy" modèle peut s'écrire :

$$U_I(S) = \sum_{i=1}^5 N_i \omega_i \quad (4.18)$$

Nous avons dans la configuration  $S$  :

$$\begin{aligned} N_1 &= N_c \\ N_2 &= N_f \\ N_3 &= N_o \\ N_4 &= N_r \\ N_5 &= N_t \end{aligned} \quad (4.19)$$

où  $N_c$  est le nombre de segments avec une seule extrémité connectée,  $N_f$  est le nombre de segments libres,  $N_o$  le nombre de paires de segments qui ne sont pas bien alignés,  $N_r$  le nombre de paires de segments qui se trouvent en interaction de répulsion et  $N_t$  le nombre total de segments.

Les interactions correspondantes sont données par :

$$\begin{aligned} \omega_1 &= \omega_c = -\log g_{22} \\ \omega_2 &= \omega_f = -\log g_{21} \\ \omega_3 &= \omega_o = -\log h_o \\ \omega_4 &= \omega_r = -\log h_r \\ \omega_5 &= \omega_t = -\log \beta \end{aligned} \quad (4.20)$$

La densité du "Candy" modèle peut s'écrire alors :

$$f(S) \propto \exp\left(-\sum_{i=1}^5 N_i \omega_i\right) = \exp(-\langle \vec{N}, \vec{\omega} \rangle) \quad (4.21)$$

où  $\langle \cdot, \cdot \rangle$  symbolise le produit scalaire.

Cette écriture est utile non seulement du point de vue de l'implantation informatique pour simulation du "Candy" modèle, mais elle permet également le contrôle des statistiques suffisantes dans le modèle. Nous allons montrer dans les chapitres suivants que l'estimation des différents paramètres d'interaction peut être formalisée à partir de l'Equation (4.21).

#### 4.1.2 Stabilité locale du "Candy" modèle

Nous avons rappelé que la stabilité locale des processus ponctuels est une propriété très importante utilisée pour la simulation du processus. Elle est une condition nécessaire pour que la chaîne, qui a comme distribution d'équilibre la densité d'un processus ponctuel localement stable, soit irréductible, Harris récurrente et ergodique.

**PROPOSITION :** *Le processus ponctuel décrit par la densité de probabilité du "Candy" modèle a la propriété de stabilité locale.*

**PREUVE :** Soit  $f(S)$  la densité du "Candy" modèle écrite sous la forme :

$$f(S) \propto \beta^{n(S)} \prod_{\zeta \in S} \exp\left[\frac{l_\zeta - l_{max}}{l_{max}}\right] \times \gamma_1^{n_f(S)} \gamma_2^{n_s(S)} \gamma_3^{n_r(S)} \gamma_4^{n_o(S)} \quad (4.22)$$

avec  $n(S)$  le nombre total de segments dans la configuration  $S$ ,  $n_f(S)$  le nombre de segments libres,  $n_s(S)$  le nombre de segments avec une seule extrémité connectée,  $n_r(S)$  est le nombre de paires de segments en interaction de type répulsion, et  $n_o(S)$  est le nombre de paires de segments non-alignés. Nous avons aussi  $\gamma_1 = g_{21}$ ,  $\gamma_2 = g_{22}$ ,  $\gamma_3 = h_r$ ,  $\gamma_4 = h_o \in (0, 1)$

Soit également  $\lambda(\zeta/S)$  l'intensité conditionnelle de Papangelou :

$$\lambda(\zeta/S) = \frac{f(S \cup \zeta)}{f(S)} \quad (4.23)$$

Si nous ajoutons un segment libre à la configurations  $S$ , alors :

$$\lambda(\zeta/S) < \beta \gamma_1 \quad (4.24)$$

Si nous ajoutons un segment avec un seule extrémité connectée, alors ce segment peut être connecté à :

(i) un segment libre

$$\lambda(\zeta/S) < \beta \frac{1}{\gamma_1} \gamma_2^2 \quad (4.25)$$



(ii) un segment avec une seule extrémité connectée :

$$\lambda(\zeta/S) < \beta \quad (4.26)$$

(iii) un segment avec deux extrémités connectées :

$$\lambda(\zeta/S) < \beta\gamma_2 \quad (4.27)$$

Si un segment avec deux extrémités connectées est ajouté, alors ce segment peut connecter :

(iv) deux segments libres :

$$\lambda(\zeta/S) < \beta \frac{1}{\gamma_1^2} \gamma_2^2 \quad (4.28)$$

(v) un segment libre et un segment avec une seule extrémité connectée :

$$\lambda(\zeta/S) < \beta \frac{1}{\gamma_1} \quad (4.29)$$

(vi) un segment libre et un segment avec deux extrémités connectées :

$$\lambda(\zeta/S) < \beta \frac{1}{\gamma_1} \gamma_2 \quad (4.30)$$

(vii) deux segments avec une seule extrémité connectée :

$$\lambda(\zeta/S) < \beta \frac{1}{\gamma_2} \quad (4.31)$$

(viii) deux segments avec deux extrémités connectées :

$$\lambda(\zeta/S) < \beta \quad (4.32)$$

Soit  $M = \beta \frac{1}{(\gamma_1 \gamma_2)^{12}}$ , car les relations considérées au paravant peuvent être modifiées par le nombre maximal de segments que l'on peut connecter ou déconnecter par hasard.

Nous obtenons alors, la propriété de stabilité locale :

$$\lambda(\zeta/S) < M \quad (4.33)$$

## 4.2 Simulation du "Candy" model

Nous définissons pour chaque segment le nombre de connexions à chaque extrémité  $c_t$  et  $c_q$ .

Pour un segment *libre*, on a  $c_t = c_q = 0$ . Un segment a une *connexion singulière* si une des conditions suivantes est respectée :

$$\begin{cases} c_t > 1 & \text{et} & c_q = 0 \\ c_q > 1 & \text{et} & c_t = 0 \end{cases} \quad (4.34)$$

Un segment a une *connexion simple* si une des conditions est accomplie :

$$\begin{cases} c_t = 1 & \text{et} & c_q = 0 \\ c_q = 1 & \text{et} & c_t = 0 \end{cases} \quad (4.35)$$

Si, pour un segment, nous avons  $c_t \geq 1$  et  $c_q \geq 1$ , alors il a une *connexion double*.

Le réseau linéique  $S$  contient  $n_t$  segments. Nous avons  $n_d$  segments libres et  $n_c$  segments connectés :

$$n_t = n_c + n_d \quad (4.36)$$

Les segments connectés se divisent en segments avec les deux extrémités connectées  $n_{c2}$  et des segments avec une seule extrémité connectée  $n_{c1}$  :

$$n_c = n_{c2} + n_{c1} \quad (4.37)$$

A leur tour, les segments avec une seule extrémité connectée se divisent en segments avec une connexion singulière  $n_{sg}$  et en segments avec une connexion simple  $n_{sl}$  :

$$n_{c1} = n_{sg} + n_{sl} \quad (4.38)$$

Pour simuler le "Candy" modèle, nous allons construire les mouvements d'une dynamique de type RJMCMC. Nous avons trois types de mouvements : ajouter ou enlever un segment à la configuration initiale, et modifier un segment à l'intérieur d'une configuration. C'est la combinaison de tous ces mouvements qui assure la réversibilité, l'irréductibilité, l'apériodicité, la récurrence et l'ergodicité de la chaîne de Markov qui a comme distribution d'équilibre la distribution de "Candy" modèle.

Voici la liste des mouvements que nous utilisons :

- naissance et mort d'un segment libre
- naissance et mort d'un segment avec une connexion simple
- naissance et mort d'un segment avec une connexion singulière
- naissance et mort d'un segment avec une connexion double
- modification de l'orientation d'un segment avec une seule extrémité connectée
- modification de la longueur d'un segment avec une seule extrémité connectée
- modification de la position d'un segment avec une connexion simple
- modification de la position d'un segment avec une connexion singulière.

Les mouvements de type *modification* ne présentent aucune difficulté en ce qui concerne le calcul du taux d'acceptation, car la dimension du sous-espace de configuration reste inchangée. Ils sont équivalents à une dynamique de type Metropolis-Hastings classique.

Les mouvements de type *naissance et mort* agissent sur la dimension du sous-espace de configuration. Dans le paragraphe qui suit nous allons expliquer comment nous calculons le taux d'acceptation pour ces mouvements, selon les principes de la dynamique de type RJMCMC.

Pour optimiser le temps de calcul, nous considérons que si l'on ajoute un segment libre à la configuration, il ne sera pas connecté par hasard avec d'autres segments. Il en est de même pour la naissance des segments avec une seule extrémité connectée, c'est à dire le segment il ne sera pas connecté par hasard aux deux extrémités.

#### 4.2.1 Naissance et mort d'un segment libre

Le mouvement de type naissance permet d'ajouter un segment libre  $\zeta$  à la configuration courante  $S$ . Le nouvel état est  $S' = S \cup \zeta$ . Le mouvement de type mort est le mouvement inverse : enlever un segment libre à la configuration.

Soit le vecteur aléatoire  $u$  choisi indépendamment de la configuration initiale  $S$  :

$$u = [(x_u, y_u), l_u, \theta_u, w_u] \quad (4.39)$$

avec la densité de probabilité :

$$q(u) = \frac{1}{\nu(K)} \times \frac{1}{l_{max} - l_{min}} \times \frac{1}{w_{max} - w_{min}} \times \frac{1}{2\pi} \quad (4.40)$$

Les paramètres du vecteur  $u$  représentent précisément les paramètres du nouveau segment dans la configuration :

$$\zeta = \zeta(u) = \begin{bmatrix} x_\zeta = x_u \\ y_\zeta = y_u \\ l_\zeta = l_u \\ \theta_\zeta = \theta_u \\ w_\zeta = w_u \end{bmatrix} \quad (4.41)$$

Par conséquent, la transformation inversible est la fonction identité :

$$s'(S, u) = (s'_1 = s_1, \dots, s'_n = s_n, \zeta = u) \quad (4.42)$$

Le Jacobien de la transformation est :

$$\left| \frac{\partial s'}{\partial(S, u)} \right| = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{vmatrix} = 1 \quad (4.43)$$

Soit  $P_n$  la probabilité de choisir la naissance d'un segment libre. Soit  $P_m$  la probabilité de choisir la mort d'un segment libre. Alors la probabilité d'enlever un segment libre donné de la nouvelle configuration est  $P_m \times \frac{1}{n_d+1}$ .

Le mouvement de naissance est accepté avec la probabilité :

$$\alpha = \min\{1, R\} \quad (4.44)$$

avec le taux d'acceptation donné par :

$$R = \frac{f(S \cup \zeta)}{f(S)} \times \frac{P_m}{P_n} \times \frac{2\pi\nu(K)(l_{\max} - l_{\min})(w_{\max} - w_{\min})}{n_d + 1} \quad (4.45)$$

Le taux d'acceptation du mouvement inverse s'obtient par l'inversion de l'Eq. (4.45) :

$$\tilde{R} = \frac{f(S \setminus \zeta)}{f(S)} \times \frac{P_n}{P_m} \times \frac{n_d}{2\pi\nu(K)(l_{\max} - l_{\min})(w_{\max} - w_{\min})} \quad (4.46)$$

#### 4.2.2 Naissance et mort d'un segment avec une connexion singulière

Ce mouvement ajoute un segment avec une connexion singulière à la configuration initiale. Un segment de longueur, épaisseur et orientation aléatoires est connecté à l'extrémité

non-libre d'un segment déjà connecté.

Soit le vecteur aléatoire  $u$  :

$$u = [(x_u, y_u), l_u, \theta_u, w_u] \quad (4.47)$$

avec  $(x_u, y_u)$  uniformément choisi dans un ensemble mesurable  $\epsilon$  de mesure  $\nu(\epsilon) > 0$ .  $l_u, \theta_u, w_u$  sont aussi uniformément choisis dans l'espace des paramètres du segment.

La densité de probabilité du vecteur  $u$  est :

$$q(u) = \frac{1}{\nu(\epsilon)} \times \frac{1}{l_{max} - l_{min}} \times \frac{1}{w_{max} - w_{min}} \times \frac{1}{2\pi} \quad (4.48)$$

Le nouvel élément dans la configuration sera le segment caractérisé par :

$$\zeta = \zeta(u) = \begin{bmatrix} x_\zeta = x_u + x_{sg} \\ y_\zeta = y_u + y_{sg} \\ l_\zeta = l_u \\ \theta_\zeta = \theta_u \\ w_\zeta = w_u \end{bmatrix} \quad (4.49)$$

avec  $(x_{sg}, y_{sg})$  les coordonnées de l'extrémité d'un segment pour lequel la connexion singulière est possible.

Le nouveau segment est donc construit en utilisant une nouvelle représentation : les coordonnées d'une extrémité, la longueur, l'orientation et l'épaisseur. Cette nouvelle représentation permet de choisir pour le vecteur  $u$  une densité facilement calculable.

La transformation inversible  $s'(S, u)$  :

$$s'(S, u) = (s'_1 = s_1, \dots, s'_n = s_n, \zeta(u)) \quad (4.50)$$

a comme Jacobien :

$$\left| \frac{\partial s'}{\partial(S, u)} \right| = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 1 \end{vmatrix} = 1 \quad (4.51)$$

Soient  $P_{ng}$  et  $P_{mg}$  les probabilités respectives de choisir le mouvement de faire une naissance d'un segment avec une connexion singulière et de choisir le mouvement d'enlever un segment avec une connexion singulière.

Nous avons donc  $P_{ng} \times \frac{1}{2n_{c2} + n_{c1}}$  et  $P_{mg} \times \frac{1}{n'_{sg}}$ , pour les probabilités de faire les mouvements de naissance et de mort d'un segment avec une connexion singulière.  $n'_{sg}$  est le nombre de segments avec une connexion singulière de la configuration  $S'$ .

Si on ajoute à la configuration un segment avec une connexion singulière, le taux d'acceptation est :

$$R = \frac{f(S \cup \zeta)}{f(S)} \times \frac{P_{mg}}{P_{ng}} \times \frac{(2n_{c2} + n_{c1}) \times 2\pi\nu(\epsilon)(l_{max} - l_{min})(w_{max} - w_{min})}{n'_{sg}} \quad (4.52)$$

si l'on inverse le rapport (4.52), nous obtenons le taux d'acceptation du mouvement inverse :

$$\tilde{R} = \frac{f(S \setminus \zeta)}{f(S)} \times \frac{P_{ng}}{P_{mg}} \times \frac{n_{sg}}{(2n'_{c2} + n'_{c1}) \times 2\pi\nu(\epsilon)(l_{max} - l_{min})(w_{max} - w_{min})} \quad (4.53)$$

$n'_{c1}$  est le nombre de segments avec une seule extrémité connectée de la configuration  $S'$ , et  $n'_{c2}$  est le nombre de segments avec les deux extrémités connectées de la configuration  $S'$ .

### 4.2.3 Naissance et mort d'un segment avec une connexion simple

Nous allons donner directement les taux d'acceptation pour ce mouvement car son principe est le même que celui du mouvement présenté dans le paragraphe précédent.

Soient  $P_{nl}$  et  $P_{ml}$  les probabilités respectives de choisir le mouvement de créer un segment avec une connexion simple et de choisir le mouvement d'enlever un segment donné avec une connexion simple.

Alors  $P_{nl} \times \frac{1}{2n_d + n_{c1}}$  et  $P_{ml} \times \frac{1}{n'_{sl}}$  sont les probabilités de faire les mouvements de naissance et de mort d'un segment avec une connexion simple.  $n'_{sl}$  est le nombre de segments avec une connexion simple de la configuration  $S'$ .

Si on ajoute à la configuration un segment avec une connexion simple, le taux d'acceptation est :

$$R = \frac{f(S \cup \zeta)}{f(S)} \times \frac{P_{ml}}{P_{nl}} \times \frac{(2n_d + n_{c1}) \times 2\pi\nu(\epsilon)(l_{max} - l_{min})(w_{max} - w_{min})}{n'_{sl}} \quad (4.54)$$

si l'on inverse le rapport (4.54), nous obtenons le taux d'acceptation du mouvement inverse :

$$\tilde{R} = \frac{f(S \setminus \zeta)}{f(S)} \times \frac{P_{nl}}{P_{ml}} \times \frac{n_{sl}}{(2n'_d + n'_{c1}) \times 2\pi\nu(\epsilon)(l_{max} - l_{min})(w_{max} - w_{min})} \quad (4.55)$$

$n'_d$  est le nombre de segments libres de la configuration  $S'$ , et  $n'_{c1}$  est le nombre de segments avec une seule extrémité connectée de la configuration  $S'$ ,

#### 4.2.4 Naissance et mort d'un segment avec une connexion double

Ce mouvement ajoute ou enlève à la configuration un segment avec une connexion double. Deux extrémités de deux segments quelconques sont choisies. La distance entre les deux extrémités doit être comprise entre  $l_{min}$  et  $l_{max}$ . Le nouveau segment sera donc connecté aux deux extrémités ainsi choisies.

Soit le vecteur aléatoire  $u$  :

$$u = [(x_{1u}, y_{1u}), (x_{2u}, y_{2u}), w_u] \quad (4.56)$$

avec les coordonnées des points choisies uniformément dans un ensemble mesurable  $\epsilon$  de mesure  $\nu(\epsilon) > 0$ .

La densité de probabilité du vecteur  $u$  est :

$$q(u) = \frac{1}{\nu(\epsilon)^2} \times \frac{1}{w_{max} - w_{min}} \quad (4.57)$$

Le nouveau segment sera donc :

$$\zeta = \zeta(u) = \begin{bmatrix} t_x(\zeta) = x_{1u} + x_{sl}(i) \\ t_y(\zeta) = y_{1u} + y_{sl}(i) \\ q_x(\zeta) = x_{2u} + x_{sl}(j) \\ q_y(\zeta) = y_{2u} + y_{sl}(j) \\ w_\zeta = w_u \end{bmatrix} \quad (4.58)$$

avec  $(x_{sl}(i), y_{sl}(i))$  et  $(x_{sl}(j), y_{sl}(j))$  les coordonnées des extrémités où le nouveau segment sera connecté.

La transformation inversible  $s'(S, u)$  est :

$$s'(S, u) = (s'_1 = s_1, \dots, s'_n = s_n, \zeta(u)) \quad (4.59)$$

et son Jacobien :

$$\left| \frac{\partial s'}{\partial(S, u)} \right| = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 1 \end{vmatrix} = 1 \quad (4.60)$$

Soient  $P_{nd}$  et  $P_{md}$  les probabilités respectives de choisir le mouvement naissance et de choisir le mouvement mort d'un segment avec une connexion double. Les probabilités pour effectuer une naissance ou une mort d'un segment avec une connexion double sont  $P_{nd} \times \frac{1}{n_{pc}}$  et  $P_{md} \times \frac{1}{n_{c2}}$ .  $n_{pc}$  est le nombre de couples d'extrémités dont la distance prend des valeurs entre  $l_{min}$  et  $l_{max}$ .

Le taux d'acceptation pour le mouvement de type naissance est :

$$R = \frac{f(S \cup \zeta)}{f(S)} \times \frac{P_{md}}{P_{nd}} \times \frac{\nu(\epsilon)^2 \times (w_{max} - w_{min}) \times n_{pc}}{n'_{c2}} \quad (4.61)$$

où  $n'_{c2}$  étant le nombre de segments avec les deux extrémités connectées de la configuration  $S'$ .

En inversant le rapport 4.61, nous obtenons, le taux d'acceptation pour le mouvement inverse :

$$\tilde{R} = \frac{f(S \setminus \zeta)}{f(S)} \times \frac{P_{nd}}{P_{md}} \times \frac{n_{c2}}{\nu(\epsilon)^2 \times (w_{max} - w_{min}) \times n'_{pc}} \quad (4.62)$$

$n_{pc}^{(i)}$  étant le nombre de couples d'extrémités dont la distance prend des valeurs entre  $l_{min}$  et  $l_{max}$ , de la configuration  $S'$ .

### 4.3 Choix des paramètres

Pour fixer les valeurs des paramètres du "Candy" modèle, nous avons en recours à la méthode des boîtes qualitatives [Azencott, 1992], [Urago, 1996].

Nous avons construit des configurations de segments. En imposant que l'énergie d'une configuration soit préférable à une autre, nous déduisons des inégalités permettant d'encadrer les valeurs des paramètres.



Soit la Figure (4.6a) Le segment  $s_3$  se trouve en interaction de type répulsion avec le segment  $s_2$ . Dans la Figure (4.6b) le segment  $s_3$  se trouve en position de mauvais alignement avec le segment  $s_2$ .

La différence d'énergie entre les deux configurations est :

$$\Delta U = U(S_b) - U(S_a) = \omega_o - \omega_r \quad (4.63)$$

Donc si  $\Delta U < 0$ , cela veut dire que nous pénalisons plus fortement les segments qui se superposent que ceux qui sont mal alignés.

Nous allons calculer la différence d'énergie des configurations  $S_a$  et  $S_b$  représentées sur la Figure (4.7) :

$$\Delta U = U(S_b) - U(S_a) = \omega_c + \omega_r + \omega_o - \omega_f \quad (4.64)$$

le segment  $s_4$  de la configuration  $S_a$  est libre, mais dans la configuration  $S_b$  il est en position de répulsion avec le segment  $s_3$  et de non-alignement avec le segment  $s_2$ . En imposant  $\Delta U < 0$ , nous forçons les segments à former un réseau connecté.



a)

b)

FIG. 4.6 – Etude des paramètres  $\omega_o$  et  $\omega_r$  : a) configuration  $S_a$ , b) configuration  $S_b$

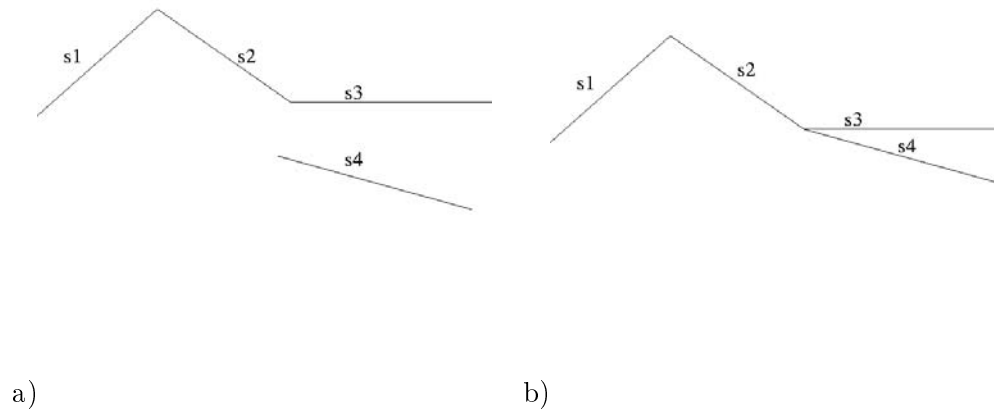
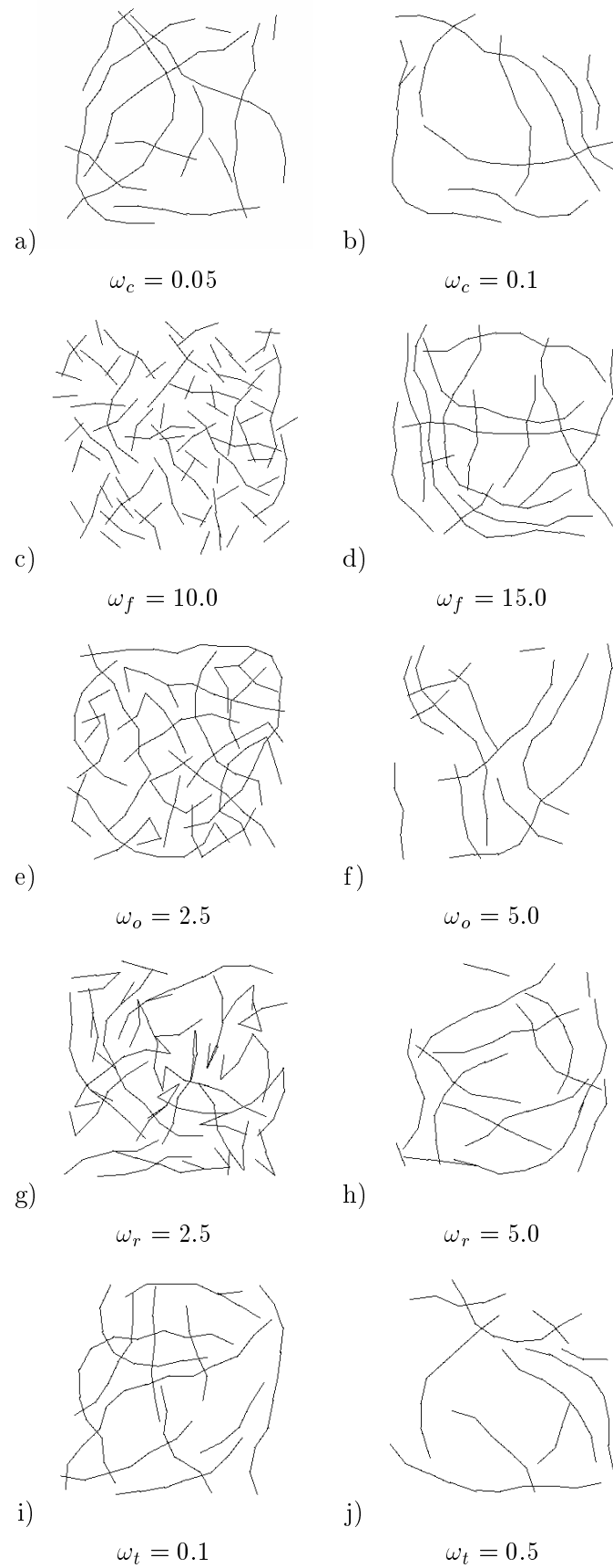


FIG. 4.7 – Etude des paramètres  $\omega_c$  et  $\omega_f$  : a) configuration  $S_a$ , b) configuration  $S_b$

#### 4.4 Réalisations

Dans la Figure (4.8), nous montrons quelques réalisations du "Candy" modèle. Nous pouvons observer que selon les paramètres choisis nous pouvons obtenir des structures linéiques avec des topologies différentes.

Pour des raisons informatiques, pendant la simulation du modèle nous avons introduit une bande de "sécurité" autour des bords de l'image : aucun segment n'est admis dans cette région. Cette décision peut donc introduire un *effet d'île* : le réseau a plus une tendance de retourner dans l'espace image que d'y sortir.

FIG. 4.8 – Réalisations du "Candy" modèle sur une image de  $341 \times 300$  pixels



## Chapitre 5

# Etude de la dynamique du "Candy" modèle

---

Nous avons construit un processus ponctuel de Gibbs capable de simuler et d'extraire les réseaux linéïques dans les images. L'estimateur du réseau est donné en minimisant l'énergie de ce processus. Les algorithmes de simulation et d'optimisation utilisent une dynamique de type Monte Carlo par chaînes de Markov à sauts réversibles.

Ce chapitre est une analyse des propriétés de la chaîne de Markov simulée. Les arguments énoncés ci-après n'ont pas un caractère de preuve. En s'appuyant sur des expérimentations et la littérature existante, nous essayons de mettre en évidence les avantages et les inconvénients d'une telle approche.

---



## 5.1 Irreductibilité. Récurrence. Ergodicité.

[Geyer et Møller, 1994] ont construit un algorithme de type Metropolis-Hastings pour simuler des processus ponctuels. L'algorithme s'appuie sur uniquement trois mouvements : naissance et mort d'un point, déplacement d'un point.

Leur algorithme est un cas particulier de la dynamique MCMC à sauts réversibles [Green, 1995]. Ces algorithmes génèrent une chaîne de Markov réversible avec comme densité d'équilibre la densité de probabilité du processus ponctuel que l'on souhaite simuler. Leur technique produit une chaîne Harris récurrente, et si la probabilité d'atteindre la configuration vide existe, la chaîne est aussi apériodique et Harris ergodique.

[Geyer, 1999] prouve que si l'intensité conditionnelle de Papangelou d'un processus ponctuel est bornée, alors la chaîne simulée par la technique de Metropolis-Hastings est irréductible, Harris récurrente et ergodique.

Notre dynamique utilise directement la dynamique RJCMC et pas un cas particulier. Notre choix se fonde sur le fait que le modèle est plus complexe que ceux présentés dans les références citées. Nous avons également plus de mouvements.

Cependant, les algorithmes sont tout à fait équivalents. Pour les mêmes mouvements, les taux d'acceptation ont une forme identique. Les mouvements que nous avons de plus par rapport à [Geyer et Møller, 1994], suivent toujours le même principe : naissance et mort de points dans un ensemble mesurable. Le choix des mouvements se fait de la même façon. La seule différence consiste dans la manière de calculer le taux d'acceptation.

Comme le "Candy" modèle est localement stable, nous affirmons en nous appuyant sur les arguments antérieurement mentionnés que la dynamique proposée simule une chaîne de Markov : réversible, apériodique, irréductible, récurrente et ergodique.

## 5.2 Convergence. Vitesse de mélangeance

Les bonnes propriétés énoncées dans le paragraphe précédent montrent que la chaîne de Markov que l'on simule converge, mais elles ne donnent aucune indication sur la vitesse de convergence des algorithmes. C'est à dire au bout de combien d'itérations nous pouvons être sûrs que l'échantillon que l'on prélève représente bien la distribution d'équilibre.

[Møller,1999] calcule des vitesses de convergence pour des processus ponctuels assez simples. La dynamique qui est à la base de ces algorithmes de simulation utilise des processus de naissance et mort.

Des méthodes d'étude de la convergence des dynamiques de type RJMCMC sont proposées dans [Brooks et Giudici, 1998]. Le critère de convergence développé par les auteurs, est utilisable dans le cas où il est difficile d'identifier les statistiques suffisantes du modèle. De plus, pour mettre en œuvre cette méthode il faut simuler plusieurs chaînes en parallèle.

La construction du modèle et la dynamique que nous avons mise en œuvre nous permettent de tenir compte de l'information fournie par les statistiques suffisantes du modèle, et donc d'utiliser pour l'étude de la convergence, les méthodes graphiques [Robert, 1996].

Ces techniques empiriques ne constituent pas de preuve de convergence. Elles sont plutôt un moyen rapide et simple de détection des disfonctionnements de la chaîne. L'avantage de ces méthodes consiste aussi à ne simuler qu'une seule chaîne.

Dans une première étape, nous avons suivi l'évolution des moyennes cumulées :

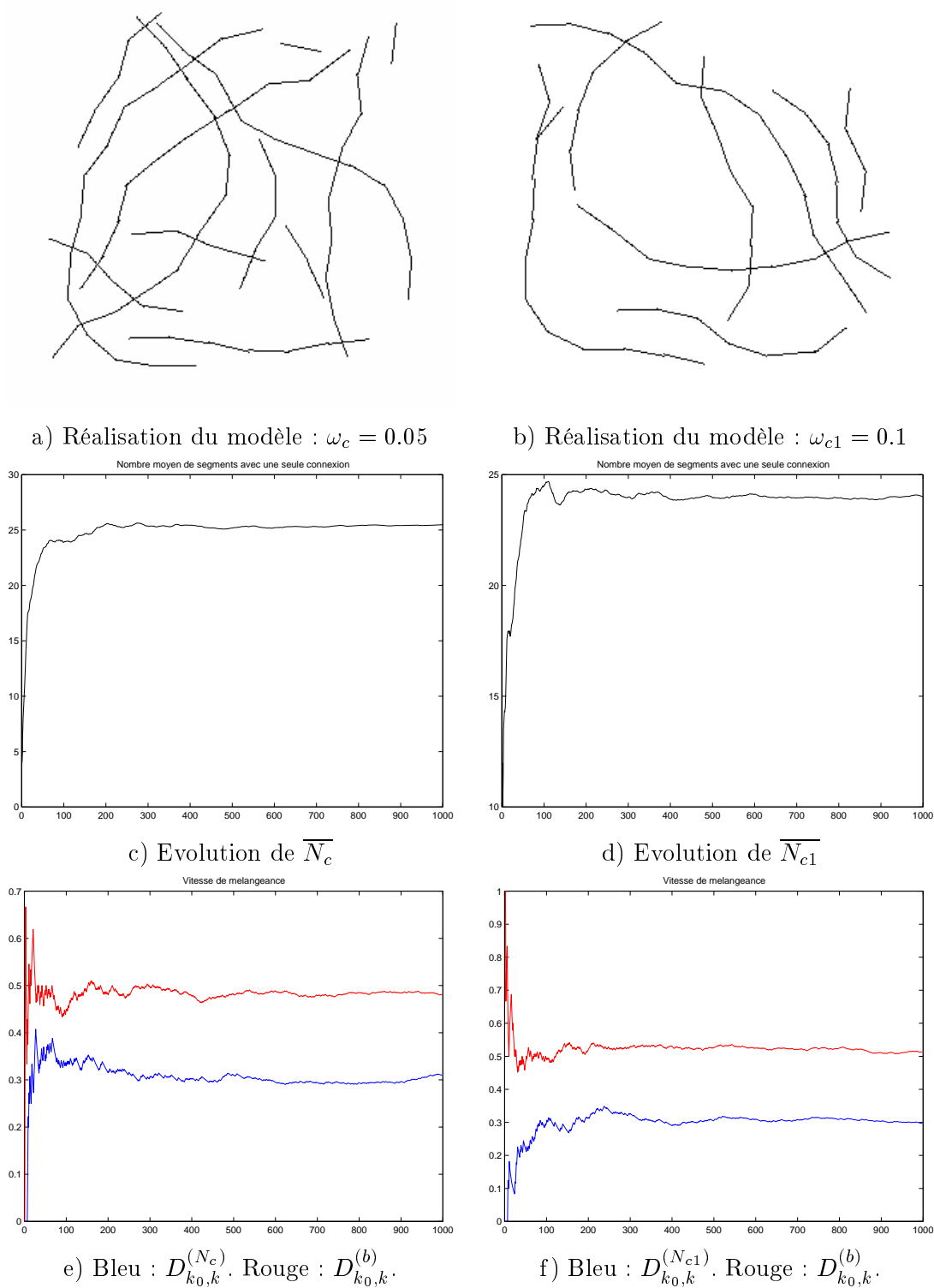
$$\overline{N}_i = \frac{1}{K} \sum_{k=1}^K N_i^{(k)} \quad (5.1)$$

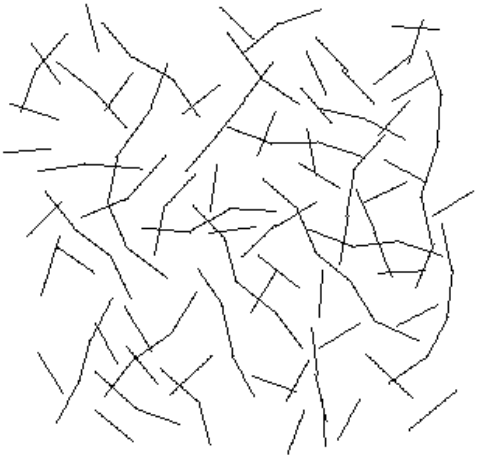
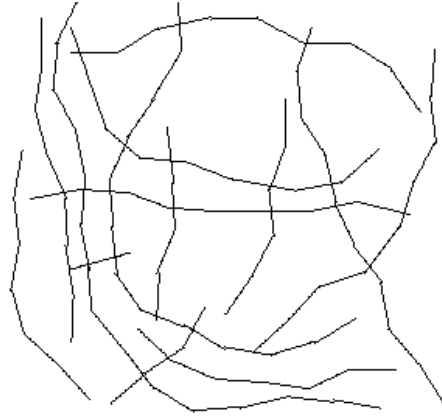
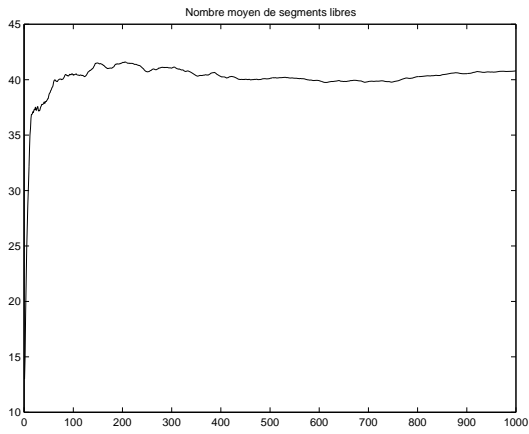
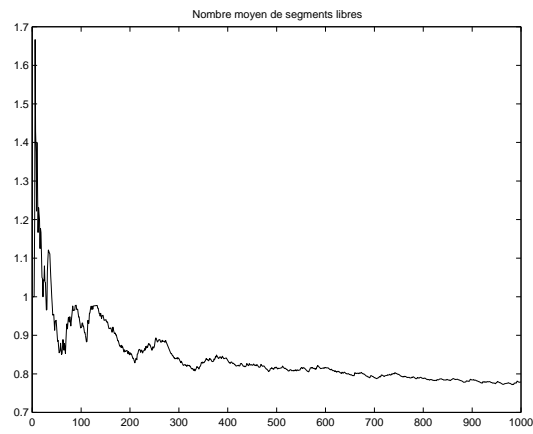
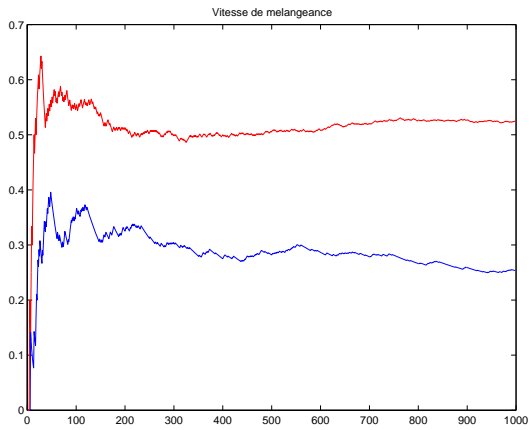
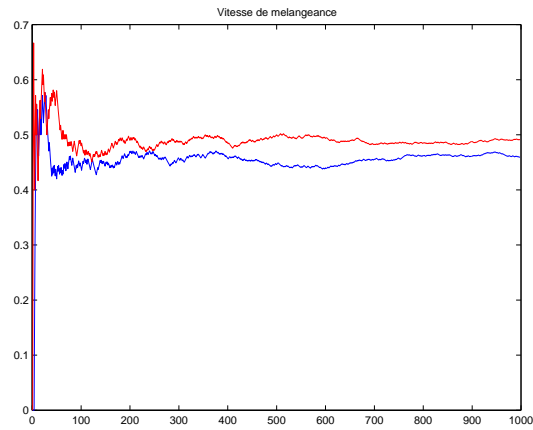
Dans les Figures 5.1, 5.2, 5.3, 5.4 et 5.5, nous avons montré à côté des réalisations l'évolution des différentes moyennes des statistiques suffisantes du "Candy" modèle.

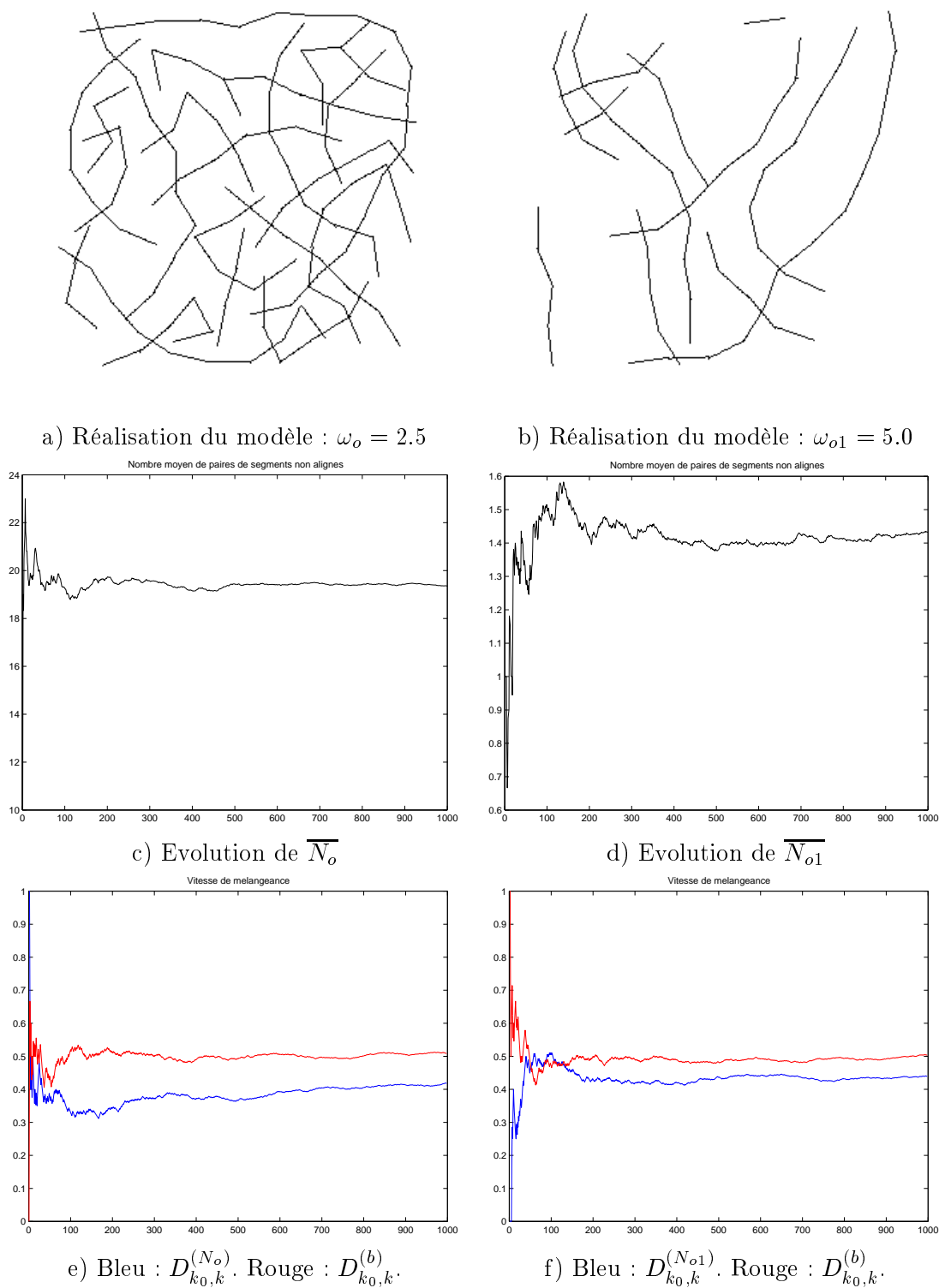
Nous avons laissé tourner la dynamique pendant un temps très long :  $10^6$  itérations. Nous rappelons que par itération nous entendons un seul mouvement, et qu'un mouvement traite un seul objet à la fois. C'est à dire que le nombre d'itérations est comparable à celui de la simulation d'un modèle de Ising sur une grille  $100 \times 100$  pendant 100 itérations, ce qui est plus que raisonnable pour les méthodes de traitement d'image utilisant les champs de Markov.

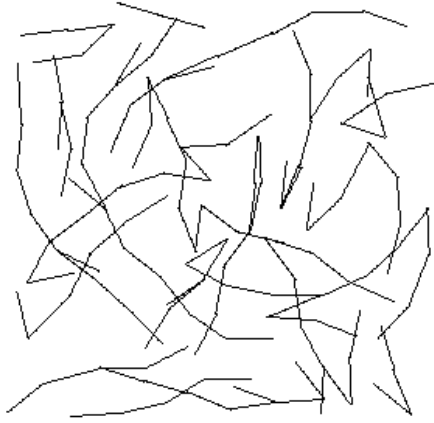
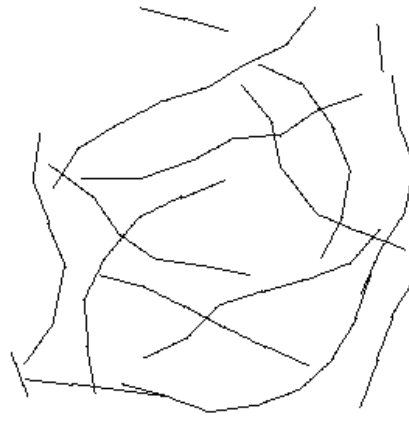
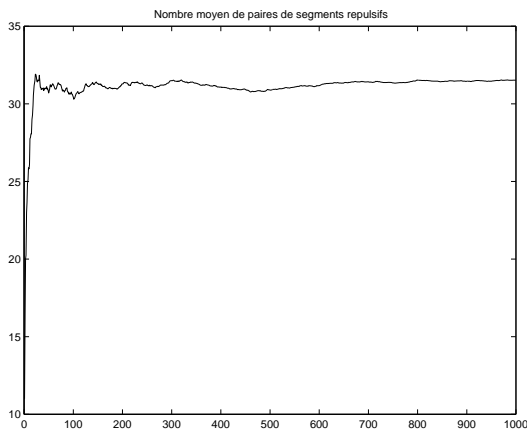
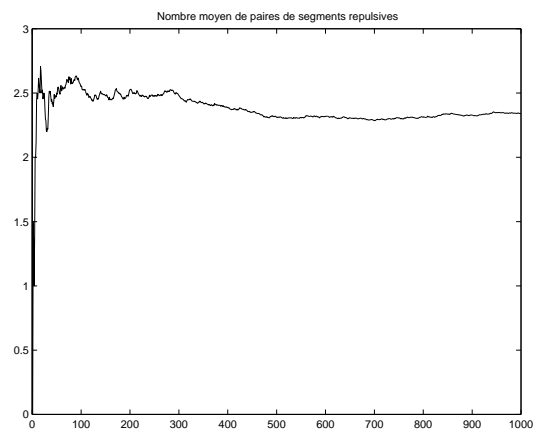
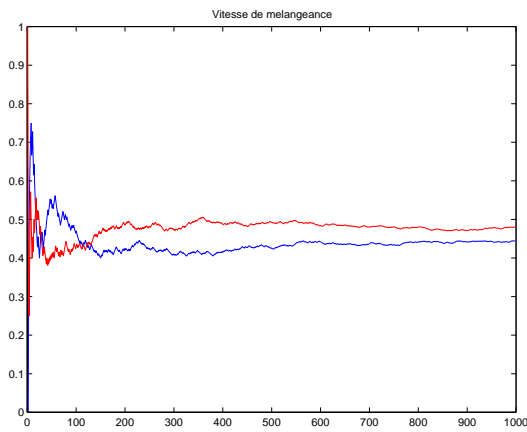
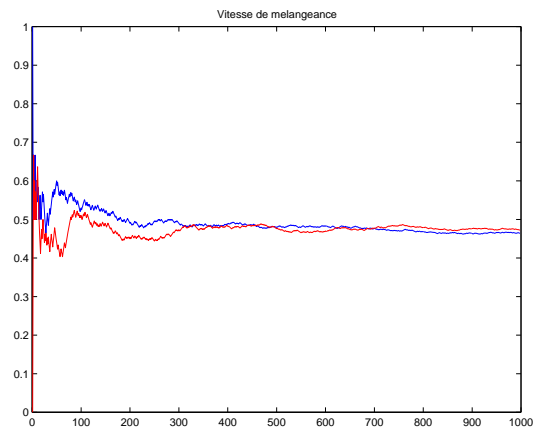
Les échantillons ont été prélevés toutes les  $10^3$  itérations. Nous remarquons que la stabilité des statistiques apparaît au bout de  $2.5 \times 10^5$  itérations, et que la valeur des statistiques suffisantes correspond aux choix des paramètres.

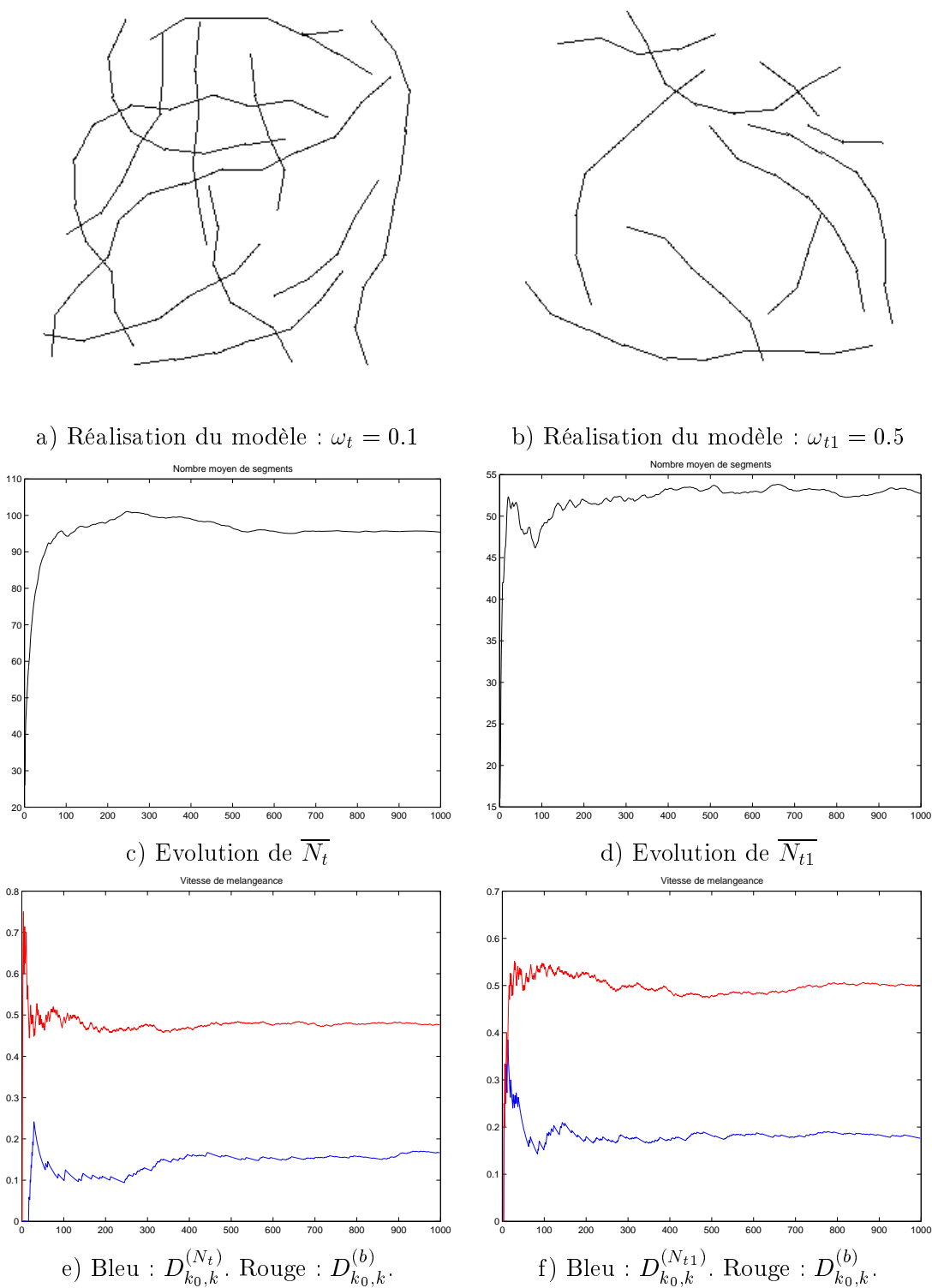


FIG. 5.1 – Etude de la dynamique en fonction du paramètre  $\omega_c$

a) Réalisation du modèle :  $\omega_f = 10.0$ b) Réalisation du modèle :  $\omega_{f1} = 15.0$ c) Evolution de  $\overline{N_f}$ d) Evolution de  $\overline{N_{f1}}$ e) Bleu :  $D_{k_0,k}^{(N_f)}$ . Rouge :  $D_{k_0,k}^{(b)}$ .f) Bleu :  $D_{k_0,k}^{(N_{f1})}$ . Rouge :  $D_{k_0,k}^{(b)}$ .FIG. 5.2 – Etude de la dynamique en fonction du paramètre  $\omega_f$

FIG. 5.3 – Etude de la dynamique en fonction du paramètre  $\omega_o$

a) Réalisation du modèle :  $\omega_r = 2.5$ b) Réalisation du modèle :  $\omega_r = 5.0$ c) Evolution de  $\overline{N_r}$ d) Evolution de  $\overline{N_{r1}}$ e) Bleu :  $D_{k_0,k}^{(N_r)}$ . Rouge :  $D_{k_0,k}^{(b)}$ .f) Bleu :  $D_{k_0,k}^{(N_{r1})}$ . Rouge :  $D_{k_0,k}^{(b)}$ .FIG. 5.4 – Etude de la dynamique en fonction du paramètre  $\omega_r$

FIG. 5.5 – Etude de la dynamique en fonction du paramètre  $\omega_t$

La stationarité des statistiques suffisantes du modèle est une condition nécessaire pour la convergence [Robert, 1996], mais cette stationarité peut correspondre à l'exploration d'un mode unique de la densité  $f(S)$ .

Nous sommes donc intéressés par la capacité de la chaîne à passer d'un mode à un autre suffisamment vite. Pour évaluer la vitesse de mélangeance de la chaîne, nous allons utiliser une méthode graphique qui s'appuie sur les sommes cumulées (CUSUM) [Yu et Mykland,1998], [Brooks, 1998], [Brooks et Roberts, 1998].

Soient  $S_1, \dots, S_k$ , les réalisations de la chaîne simulant le modèle, et :

$$\hat{\mu} = \frac{1}{k - k_0} \sum_{t=k_0+1}^k N(S_t) \quad (5.2)$$

la moyenne empirique d'une des statistiques suffisantes quelconque du modèle. La méthode d'évaluation CUSUM est unidimensionnelle donc, pour chaque statistique suffisante du modèle, il va falloir répéter l'opération.

Soit la somme partielle :

$$\hat{\sigma}_T = \sum_{t=k_0+1}^T [N(S_t) - \hat{\mu}] \quad (5.3)$$

avec  $T = k_0 + 1, \dots, n$ .

En traçant  $\{\hat{\sigma}_T\}$  pour  $T = k_0 + 1, \dots, n$ , nous pouvons étudier la vitesse de mélangeance et la corrélation entre les  $S_k$ . Si le graphe de  $\{\hat{\sigma}_T\}$  est bruité et concentré autour de 0, alors la chaîne est fortement mélangeante. Si, au contraire le graphe est régulier avec de longues excursions évitant 0, la chaîne est faiblement mélangeante.

En pratique, faire un diagnostic à partir d'une seule courbe est une méthode peu objective. Pour pallier cet inconvénient, les auteurs [Yu et Mykland,1998] proposent de faire une comparaison de  $\{\hat{\sigma}_T\}$  avec une somme cumulée test  $\{\hat{\sigma}_T^b\}$ .

La somme test est construite de la même façon que  $\{\hat{\sigma}_T\}$ , à partir de variables aléatoires indépendantes identiquement distribuées selon la loi normale  $\mathcal{N}(\hat{\mu}, \hat{v}^2)$  avec  $\hat{v}^2$  la variance empirique des statistiques suffisantes  $N(S)$ .

Pour donner une mesure quantitative de la vitesse de mélangeance, un indicateur de bruitage est construit à partir de la courbe  $\{\hat{\sigma}_T\}$  [Brooks, 1998]. Soit alors :

$$d_T = \begin{cases} 1 & \text{si } \sigma_T > \sigma_{T-1} \text{ et } \sigma_T > \sigma_{T+1} \\ & \text{ou } \sigma_T < \sigma_{T-1} \text{ et } \sigma_T < \sigma_{T+1} \\ 0 & \text{sinon} \end{cases} \quad (5.4)$$

avec  $T = k_0 + 1, \dots, k - 1$ . Alors :

$$D_{k_0,k} = \frac{1}{k - k_0} \sum_{T=k_0+1}^{k-1} d_T \quad (5.5)$$

prend des valeurs entre 0 et 1. Une courbe  $\hat{\sigma}_T$  très lisse vérifiera  $D_{k_0,k} \rightarrow 0$  ce qui signifie que la chaîne a une très faible vitesse de mélangeance. Au contraire, une courbe  $\hat{\sigma}_T$  très "bruitée" vérifiera  $D_{k_0,k} \rightarrow 1$ , indiquant une chaîne avec une forte vitesse de mélangeance.

Un calcul simple [Brooks, 1998] montre que  $\mathbb{E}(d_T) = 1/2$ , si les  $\sigma_T$  sont des variables aléatoires indépendantes symétriquement distribuées autour de leur moyenne. Nous allons donc tester la vitesse de mélangeance d'une chaîne simulée en faisant la comparaison entre les moyennes empiriques utilisant les courbes  $\{D_{k_0,k}\}$  et  $\{D_{k_0,k}^b\}$ . La courbe  $\{D_{k_0,k}^b\}$  utilise les sommes partielles test  $\hat{\sigma}_T^b$ , qui sont construites à partir des variables aléatoires indépendantes identiquement distribuées selon une loi gaussienne.

Les Figures 5.1, 5.2, 5.3, 5.4 et 5.5 montrent également, pour chaque paramètre, la comparaison entre la courbe  $\{D_{k_0,k}\}$  construite à partir des statistiques suffisantes, et  $\{D_{k_0,k}^b\}$ . Nous pouvons grâce à cette méthode, comparer la vitesse de mélangeance de deux chaînes différentes.

Nous remarquons le bon comportement de la chaîne pour les paramètres  $\omega_f$ ,  $\omega_o$  et  $\omega_r$ . Les vitesses de mélangeance relatives aux paramètres  $\omega_c$  et  $\omega_t$  sont beaucoup plus petites. Dans ce cas, nous devons faire un compromis : soit nous renforçons les paramètres ce qui va augmenter la vitesse de mélangeance tout en détruisant la structure de réseau linéique, soit nous gardons la structure linéique en tolérant une vitesse de mélangeance moins rapide.

Néanmoins, il faut observer que les valeurs vers lesquelles les sommes  $\{D_{k_0,k}\}$  pour les paramètres  $\omega_c$  et  $\omega_f$  convergent sont très proches de celles obtenues dans [Brooks, 1998], pour simuler un modèle d'Ising sur une grille de dimensions  $100 \times 100$  avec un échantillonneur de Gibbs. Cela montre donc que la dynamique que nous avons proposée est tout à fait comparable, de ce point de vue aux techniques utilisant les champs de Markov, employées

en traitement d'image.

### 5.3 Taux d'acceptation. Choix des mouvements.

Un inconvénient important des dynamiques de type MCMC à sauts réversibles est le faible taux d'acceptation. Cela veut dire que beaucoup de propositions de nouvelles configurations sont rejetées.

Nous rappelons le taux d'acceptation pour un mouvement de la dynamique type RJMCMC :

$$R = \frac{f(y)}{f(x)q(u^{(i)})} \left| \frac{\partial y'}{\partial(x, u)} \right| \frac{p(\text{choisir de proposer le saut inverse})}{p(\text{choisir de proposer un saut avant})} \quad (5.6)$$

Pour augmenter le taux d'acceptation, il est possible de jouer sur la densité du vecteur  $u$  [Giudici et Roberts, 1998]. En effet, le vecteur  $u$  est tiré de telle façon que les nouvelles configurations aient le plus de chance d'être acceptées.

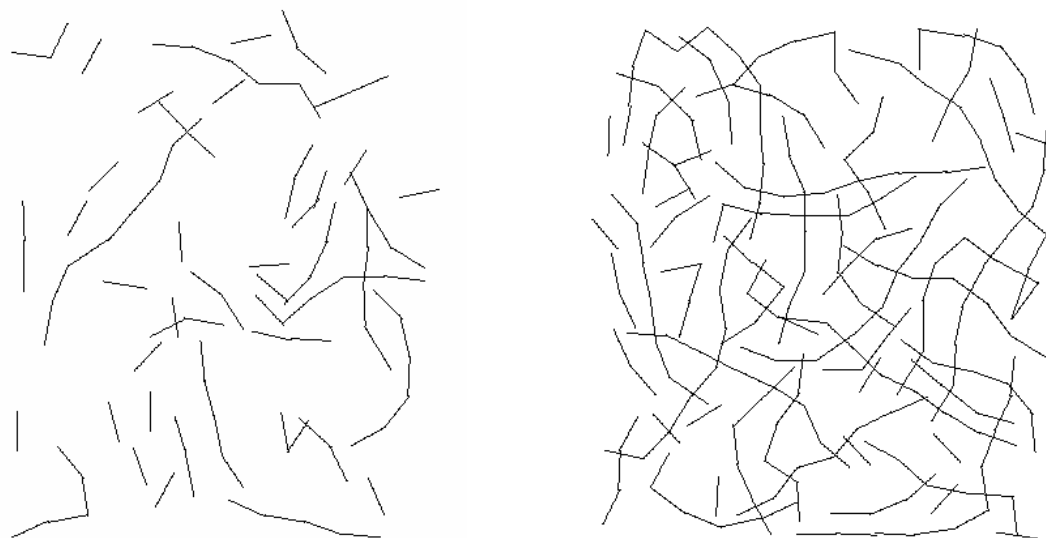
Une autre façon consiste à ne pas rejeter tout de suite une nouvelle configuration proposée [Green et Mira, 1999]. Si un mouvement n'est pas accepté, alors le mouvement ne va pas être rejeté tout de suite. Une autre proposition sera faite avec une autre densité  $q_1(u)$ . Cette nouvelle densité, donc implicitement le nouveau taux d'acceptation, dépendront de la proposition rejetée antérieurement.

Les mouvements pour la simulation du "Candy" modèle sont construits à partir d'un vecteur  $u$  qui possède une densité uniforme. L'intérêt d'une telle densité consiste en la simplicité du calcul du taux d'acceptation. Un autre aspect important est que ainsi nous pouvons proposer des configurations qui sont pénalisées par le modèle. Par exemple, la dynamique peut proposer d'ajouter à la configuration des segments qui sont pas correctement alignés ou bien qui se trouvent en interaction de type répulsion.

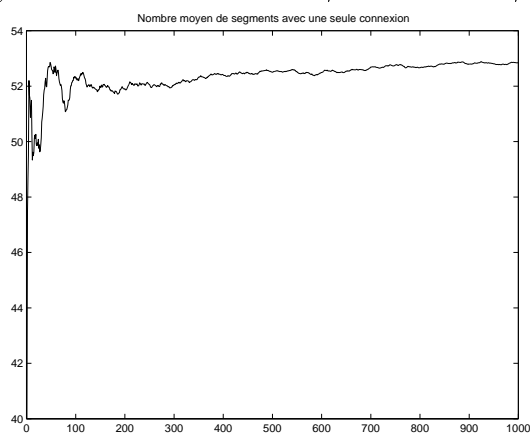
C'est-à-dire que si une route prend une direction qui est pénalisée par le "Candy" modèle, il est possible du point de vue probabiliste de proposer un prolongement du réseau dans cette direction, donc effectuer une détection plus complète du réseau.

En revanche, cela veut dire également que nous allons aussi proposer beaucoup de mouvements qui ne vont pas être acceptés, car le modèle pénalise fortement ce type de configuration.





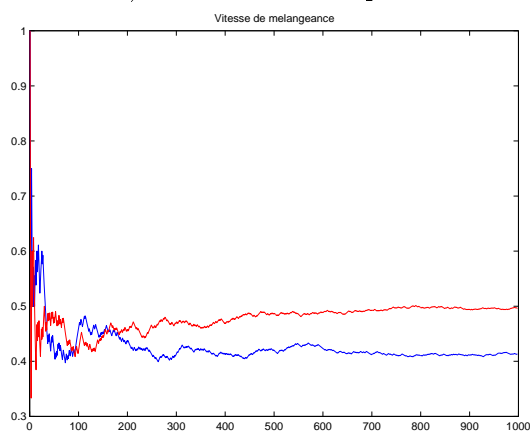
a) Réalisation avec  $P_{nl} = 4/17$  et  $P_{ml} = 1/17$     b) Réalisation avec  $P_{nl}^{(1)} = 12/27$  et  $P_{ml}^{(1)} = 3/27$



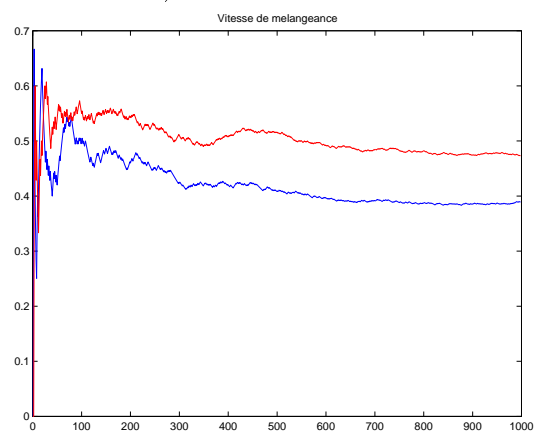
c) Evolution de  $\overline{N_c}$  pour



d) Evolution de  $\overline{N_{c1}}$



e) Bleu :  $D_{k_0,k}^{(N_c)}$ . Rouge :  $D_{k_0,k}^{(b)}$ .



f) Bleu :  $D_{k_0,k}^{(N_{c1})}$ . Rouge :  $D_{k_0,k}^{(b)}$ .

FIG. 5.6 – Etude de la dynamique en fonction des probabilités de choisir les mouvements

Le choix des mouvements proposant les nouveaux états du système est directement lié au taux d'acceptation de la dynamique. Théoriquement, il est possible de choisir les mouvements en fonction de l'état [Green, 1995]. Nous ne connaissons pas, jusqu'à présent, d'applications utilisant ce type de combinaison de noyaux de transition.

C'est pourquoi nous avons choisi de combiner les mouvements d'une manière purement stochastique. Cette technique garantit la convergence de la chaîne vers la distribution d'équilibre, car elle est équivalente à une combinaison linéaire de noyaux construits avec de bonnes propriétés [Geyer, 1999].

Pour accélérer la convergence de la dynamique, nous avons donné une forte probabilité d'occurrence aux mouvements qui aident le modèle. Théoriquement, nous pouvons converger vers la distribution d'équilibre seulement avec des mouvements de type naissance et mort de segments libres. Pour rendre utilisable la dynamique, nous avons construit les mouvements de connexion.

Dans la Figure 5.6 nous étudions la dynamique selon le poids des mouvements proposant la naissance ou la mort d'un segment avec une connexion simple, c'est à dire les mouvements relatifs au prolongement du réseau.

Nous observons que les structures linéiques se forment uniquement quand les mouvements de connexion sont prédominantes (cf. Figure 5.6b). Assigner une probabilité très forte aux mouvements de naissance des segments connectés diminue le taux d'acceptation de ces mouvements. Mais cette diminution n'est pas significative dans le calcul de l'expression du taux d'acceptance, car elle est compensée par le poids du rapport des densités du "Candy" modèle.

La stationarité des moyennes cumulées n'a pas été atteinte dans la Figure 5.6c, donc la convergence d'une telle dynamique ne peut être atteinte que si une probabilité considérablement grande est fixée pour les mouvements aidant le modèle. Nous remarquons que les valeurs de vitesse de mélangeance sont très proches (cf. les Figures 5.6e et 5.6f), mais dans ce cas où une des réalisations n'a pas convergé, il est inutile de comparer les vitesses de mélangeance des chaînes. Cet exemple montre également les limites des méthodes graphiques : le diagnostic ne peut pas être fiable à partir d'une seule courbe, la dynamique n'est pas "regardée" dans sa totalité mais en fonction d'un seul paramètre, ou d'une seule statistique suffisante.

Etablir la convergence des méthodes de simulations de chaînes de Markov, donc déterminer des critères d'arrêt fiables reste un problème ouvert [Robert et Casella, 1999].

Une voie encore peu explorée est la simulation de plusieurs chaînes en parallèle. Intuitivement, nous pouvons ainsi réduire la variabilité, la dépendance aux conditions initiales et, par conséquent, obtenir un diagnostic plus fiable qu'à partir d'une seule simulation de la chaîne. En revanche, une grande attention doit être prêtée à ce type d'approche. Une connaissance a priori de la densité d'équilibre  $f$  est nécessaire. Sinon des conditions initiales trop concentrées autour d'un mode local de la distribution d'équilibre  $f$  ne vont pas apporter beaucoup plus qu'une seule simulation de la chaîne que l'on souhaite analyser.

Une autre direction très prometteuse est la simulation exacte. Dans un espace d'état fini, [Propp et Wilson, 1996] ont montré qu'il est possible d'obtenir un échantillon qui suit exactement la distribution d'équilibre de la chaîne. Le principe de la méthode est de remonter dans le *passé* de la chaîne utilisant une suite unique de variables aléatoires jusqu'à ce que les chemins ainsi obtenus aboutissent tous au même état au moment  $t = 0$ . Le nombre de chaînes est considérablement réduit si des *relations d'ordre* peuvent être construites.

Le cas des processus ponctuels est plus complexe car l'espace des configurations n'est pas fini. [Kendall et Thonnes, 1997] ont déjà construit des algorithmes de simulation exacte pour des processus ponctuels qui sont construits à partir de dynamiques de type naissance-mort. Pour les processus ponctuels localement stables, [Kendall et Møller, 1999] ont construit des relations d'ordre en s'appuyant sur l'intensité de Papangelou et l'inclusion. Des algorithmes de simulation exacte utilisant une dynamique de type Metropolis-Hastings ont pu être ainsi construits. La convergence d'autres processus ponctuels a été analysée utilisant la simulation exacte par [Häggström et al., 1999].

La stabilité locale de "Candy" modèle et les considérations mentionnés ci-dessous nous mènent à envisager dans un futur proche la possibilité de l'implantation de la simulation exacte. Cela peut s'avérer utile non seulement dans le cas de l'analyse de la convergence, mais aussi pour l'estimation des paramètres du modèle.



## Chapitre 6

# Extraction de réseaux routiers en utilisant le "Candy" modèle

---

Nous présentons dans ce chapitre une application du "Candy" modèle : l'extraction de réseaux routiers dans des images satellitaires ou aériennes [Stoica et al., 2000a], [Stoica et al., 2000b].

Nous supposons que les réseaux routiers dans une image satellitaire ou aérienne peuvent être approchés par un ensemble de segments connectés.

Nous construisons un processus ponctuel de Gibbs dont l'énergie comporte deux termes. Le premier terme, l'énergie d'attache aux données, détecte l'emplacement du réseau. Le deuxième terme, l'énergie d'interaction entre les segments composant le réseau, est donné par l'énergie de Gibbs du "Candy" modèle.

L'estimateur des réseaux routiers est trouvé en minimisant la fonction d'énergie totale du modèle. Pour ce faire, nous utilisons un algorithme de type recuit simulé qui s'appuie sur la dynamique RJMCMC construite pour la simulation du "Candy" modèle.

---



## 6.1 Position du problème

Soit  $D$  l'image observée. L'image est discrète sur une grille finie,  $T$ , telle que  $d_t$  représente le niveau de gris d'un pixel de l'image avec la coordonnée  $t \in T$ .

Soit  $S$  le réseau linéique que l'on souhaite détecter. Le réseau est constitué par des segments  $s_i$ .

$$\begin{aligned} S &= \{s_i, i = 1, \dots, n\} \\ n &\in N \end{aligned} \tag{6.1}$$

avec  $s_i = (p_i, m_i)$ , un segment dans la configuration. Les  $p_i = (x_i, y_i)$  sont les coordonnées des centres des segments. L'épaisseur, la longueur et l'orientation du segment sont données par le vecteur des paramètres  $m_i = (w_i, h_i, \theta_i)$ .

Nous définissons l'image d'un segment  $I(s)$ . L'image est formée à partir de la marque du segment. La *silhouette* d'un segment est l'intersection entre son image et la grille finie  $T$  :  $W(s) = I(s) \cap T$ .

Notre problème consiste donc à détecter le nombre de segments, leur emplacement dans l'image ainsi que leurs paramètres.

Cette approche orientée objet se différencie des autres travaux en traitement d'image utilisant des processus ponctuels [Baddeley et van Lieshout, 1993], [Lieshout, 1994], [Rue et Husby, 1998], [Rue et Syverseen, 1998], [Rue et Hurn, 1999] par le fait que nous ne nous plaçons pas dans le cadre bayésien.

Dans le premier chapitre, nous avons énoncé les hypothèses principales qui font de l'extraction des réseaux routiers une tâche difficile : variabilité de la radiométrie, interruptions dans le réseau, bruit dans l'image, similarité avec d'autres objets dans l'image. La complexité des réseaux routiers dans les images nous empêche de pouvoir écrire une vraisemblance pour ce type d'objet. Par conséquent, nous sommes obligés de travailler d'une façon complètement non supervisée.

Pour remédier à cette difficulté, nous nous plaçons dans le cadre des processus ponctuels de Gibbs. L'estimateur des réseaux linéiques dans les images n'est pas donné par le *Maximum A Posteriori* (MAP) mais par le minimum de l'énergie du modèle.

La distribution des segments (grande canonique) est alors :

$$f(S) \propto \exp(-U(S)) = \exp(-(U_D(S) + U_I(S))) \quad (6.2)$$

$U(S)$  est l'énergie du modèle.  $U_D(S)$  est l'énergie d'attache aux données qui est relative à l'emplacement du réseau dans l'image.  $U_I(S)$  est l'énergie interne du modèle. Cette énergie est l'énergie de Gibbs du "Candy" modèle définie dans le chapitre précédent.

L'estimateur du réseau est trouvé en minimisant la fonction d'énergie  $U(S)$  :

$$\hat{S} = \min_{S \in \mathcal{S}} \{U_D(S) + U_I(S)\} \quad (6.3)$$

avec  $\mathcal{S}$  l'espace des configurations.

Pour trouver l'optimum global de la fonction d'énergie nous mettons en place un algorithme de type recuit simulé qui utilise une dynamique de type RJMCMC construite pour la simulation du "Candy" modèle.

Cet algorithme simule d'une manière itérative la loi de probabilité :

$$f(S, H) \approx [f(S)]^{\frac{1}{H}} \quad (6.4)$$

tout en faisant décroître la température  $H$ . Quand  $H \rightarrow 0$ , le résultat de la simulation converge en loi vers l'optimum global  $\hat{S}$ .

## 6.2 Energie d'attache aux données

### 6.2.1 Définition

L'énergie d'attache aux données  $U_D(S)$  doit indiquer l'emplacement du réseau dans l'image.

Nous faisons l'hypothèse que les valeurs des pixels appartenant à la route sont homogènes dans la direction de la route, et qu'il existe aussi un contraste important entre la route et le fond.

Nous considérons par conséquent qu'un segment se trouve sur une route si les statistiques des pixels correspondant à sa silhouette sont différentes des statistiques des segments



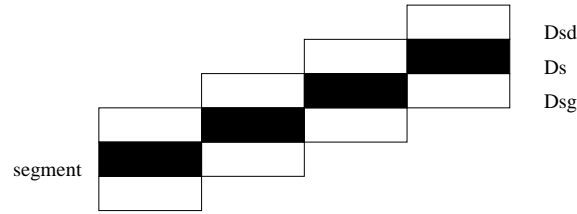


FIG. 6.1 – Segment dans une image

à "droite" et à "gauche" par rapport à l'orientation de la route (cf. Fig. 6.1).

L'énergie d'attache aux données doit aussi permettre d'éviter de placer des segments sur les contours des autres objets dans l'image ou bien au milieu d'une région homogène.

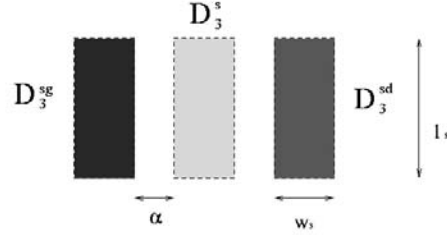
Pour pouvoir incorporer ces considérations dans une seule fonction, nous allons construire l'énergie d'attache aux données à partir de tests d'hypothèses [Germain et Réfrégier, 1996], [Réfrégier et al., 1997 (a)], [Réfrégier et al., 1997 (b)].

Nous définissons trois régions comme dans la Figure 6.2 :

- $D_s = W(s)$  est la région de l'image qui contient les pixels appartenant à la silhouette du segment que l'on souhaite analyser ;  $n_{D_s}$  est le nombre de pixels correspondant ;
- $D_{sd} = W(sd)$  est la région de l'image qui contient les pixels appartenant à la silhouette du segment qui se trouve à la droite du segment à analyser ;  $n_{D_{sd}}$  est le nombre de pixels correspondant ;
- $D_{sg} = W(sg)$  est la région de l'image qui contient les pixels appartenant à la silhouette du segment qui se trouve à la gauche du segment à analyser ;  $n_{D_{sg}}$  est le nombre de pixels correspondant.

Les masques ont donc les dimensions d'un segment. L'écart  $\alpha$  entre les segments est nécessaire pour éviter les erreurs dues à la discrétisation. Les pixels qui se trouvent entre les silhouettes des segments ne sont pas pris en compte.

Soit alors l'hypothèse  $H_3$  : nous avons trois régions différentes. Pour écrire la vraisem-

FIG. 6.2 – Hypothèse  $H_3$  : trois régions différentes

blance de cette hypothèse nous allons supposer que les pixels à l'intérieur de chaque région suivent une loi gaussienne, cette hypothèse étant largement utilisée en traitement d'image.

La fonction de vraisemblance pour l'hypothèse  $H_3$  est :

$$\begin{aligned}
 L(H_3) &= \prod_{i=1}^{n_{D_3^s}} \frac{1}{\sqrt{2\pi\sigma_{D_3^s}^2}} \exp\left(-\frac{(x_i^{D_3^s} - \mu_{D_3^s})^2}{2\sigma_{D_3^s}^2}\right) \times \\
 &\times \prod_{i=1}^{n_{D_3^{sd}}} \frac{1}{\sqrt{2\pi\sigma_{D_3^{sd}}^2}} \exp\left(-\frac{(x_i^{D_3^{sd}} - \mu_{D_3^{sd}})^2}{2\sigma_{D_3^{sd}}^2}\right) \times \\
 &\times \prod_{i=1}^{n_{D_3^{sg}}} \frac{1}{\sqrt{2\pi\sigma_{D_3^{sg}}^2}} \exp\left(-\frac{(x_i^{D_3^{sg}} - \mu_{D_3^{sg}})^2}{2\sigma_{D_3^{sg}}^2}\right)
 \end{aligned} \tag{6.5}$$

avec  $\mu_{D_3^s}, \sigma_{D_3^s}^2, \mu_{D_3^{sd}}, \sigma_{D_3^{sd}}^2, \mu_{D_3^{sg}}, \sigma_{D_3^{sg}}^2$ , les estimées de la moyenne et de la variance pour chaque région.

En prenant le logarithme de l'Eq. (6.5), il vient :

$$\begin{aligned}
\log L(H_3) &= -\frac{n_{D_3^s}}{2} \log(2\pi\sigma_{D_3^s}^2) - \sum_{i=1}^{n_{D_3^s}} \frac{(x_i^{D_3^s} - \mu_{D_3^s})^2}{2\sigma_{D_3^s}^2} \\
&- \frac{n_{D_3^{sg}}}{2} \log(2\pi\sigma_{D_3^{sg}}^2) - \sum_{i=1}^{n_{D_3^{sg}}} \frac{(x_i^{D_3^{sg}} - \mu_{D_3^{sg}})^2}{2\sigma_{D_3^{sg}}^2} \\
&- \frac{n_{D_3^{sd}}}{2} \log(2\pi\sigma_{D_3^{sd}}^2) - \sum_{i=1}^{n_{D_3^{sd}}} \frac{(x_i^{D_3^{sd}} - \mu_{D_3^{sd}})^2}{2\sigma_{D_3^{sd}}^2}
\end{aligned} \tag{6.6}$$

et d'où :

$$\begin{aligned}
\log L(H_3) &= -\frac{n_{D_3^s} + n_{D_3^{sg}} + n_{D_3^{sd}}}{2} \log(2\pi\sigma_{D_3^s}^2) \\
&- \frac{n_{D_3^{sg}}}{2} \log(2\pi\sigma_{D_3^{sg}}^2) - \frac{n_{D_3^{sd}}}{2} \log(2\pi\sigma_{D_3^{sd}}^2)
\end{aligned} \tag{6.7}$$

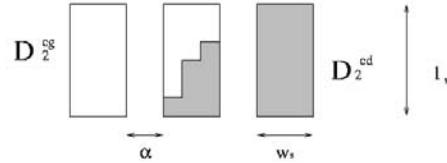
Définissons maintenant deux régions autour d'un segment comme dans la Figure 6.3 :

- $D_2^{cd}$  est la région de l'image qui contient les pixels à droite du plus grande axe de symétrie du segment à analyser ;  $n_{D_2^{cd}}$  est le nombre de pixels correspondant. Cette région est formée par les pixels de la région  $D_3^{sd}$  et les pixels de la région  $D_3^s$  qui sont plus proche du point de vue statistique de la région  $D_3^{sd}$  que  $D_3^{sg}$  ;
- $D_2^{cg}$  est la région de l'image qui contient les pixels à gauche du plus grande axe de symétrie du segment à analyser ;  $n_{D_2^{cg}}$  est le nombre de pixels correspondant. Cette région est formée par les pixels de la région  $D_3^{sg}$  et les pixels de la région  $D_3^s$  qui sont plus proche du point de vue statistique de la région  $D_3^{sg}$  que  $D_3^{sd}$ .

Les pixels qui se trouvent dans la région désignée par l'écart  $\alpha$  ne sont pas pris en compte.

Soit l'hypothèse  $H_2$  : nous avons un contour. En gardant l'hypothèse gaussienne, la fonction de vraisemblance s'écrit :

$$\begin{aligned}
L(H_2) &= \prod_{i=1}^{n_{D_2^{cd}}} \frac{1}{\sqrt{2\pi\sigma_{D_2^{cd}}^2}} \exp\left(-\frac{(x_i^{D_2^{cd}} - \mu_{D_2^{cd}})^2}{2\sigma_{D_2^{cd}}^2}\right) \times \\
&\times \prod_{i=1}^{n_{D_2^{cg}}} \frac{1}{\sqrt{2\pi\sigma_{D_2^{cg}}^2}} \exp\left(-\frac{(x_i^{D_2^{cg}} - \mu_{D_2^{cg}})^2}{2\sigma_{D_2^{cg}}^2}\right)
\end{aligned} \tag{6.8}$$


 FIG. 6.3 – Hypothèse  $H_2$  : deux régions différentes

où  $\mu_{D_2^{cd}}, \sigma_{D_2^{cd}}^2, \mu_{D_2^{cg}}, \sigma_{D_2^{cg}}^2$  sont les estimées de la moyenne et de la variance de chaque région.

La log-vraisemblance est donc :

$$\begin{aligned} \log L(H_2) &= \frac{n_{D_2^{cg}}}{2} \log(2\pi\sigma_{D_2^{cg}}^2) - \sum_{i=1}^{n_{D_2^{cg}}} \frac{(x_i^{D_2^{cg}} - \mu_{D_2^{cg}})^2}{2\sigma_{D_2^{cg}}^2} \\ &\quad - \frac{n_{D_2^{cd}}}{2} \log(2\pi\sigma_{D_2^{cd}}^2) - \sum_{i=1}^{n_{D_2^{cd}}} \frac{(x_i^{D_2^{cd}} - \mu_{D_2^{cd}})^2}{2\sigma_{D_2^{cd}}^2} \end{aligned} \quad (6.9)$$

ce qui donne :

$$\log L(H_2) = -\frac{n_{D_2^{cg}} + n_{D_2^{cd}}}{2} - \frac{n_{D_2^{cg}}}{2} \log(2\pi\sigma_{D_2^{cg}}^2) - \frac{n_{D_2^{cd}}}{2} \log(2\pi\sigma_{D_2^{cd}}^2) \quad (6.10)$$

Définissons enfin une région homogène autour d'un segment comme dans la Figure 6.4 :

•  $D_1^h$  est la région de l'image qui contient les pixels appartenant à la région homogène définie autour du segment à analyser ;  $n_{D_1^h}$  est le nombre de pixels correspondant. Cette région est donc formé par la réunion des pixels des trois silhouettes  $D_3^{sg}$ ,  $D_3^{sd}$  et  $D_3^s$ .

Comme dans les cas précédents les pixels compris dans l'espace créé par l'écart  $\alpha$  ne sont pas pris en compte.

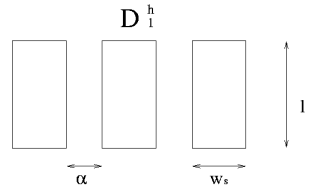


FIG. 6.4 – Hypothèse  $H_1$  : une région homogène

La fonction de vraisemblance pour l'hypothèse  $H_1$  : *nous avons une région homogène*, s'écrit sous le même contexte gaussien :

$$L(H_1) = \prod_{i=1}^{n_{D_1^h}} \frac{1}{\sqrt{2\pi\sigma_{D_1^h}^2}} \exp\left(-\frac{(x_i^{D_1^h} - \mu_{D_1^h})^2}{2\sigma_{D_1^h}^2}\right) \quad (6.11)$$

avec  $\mu_{D_1^h}, \sigma_{D_1^h}^2$  les estimées de la moyenne et de la variance des pixels appartenant à la région homogène.

La log-vraisemblance s'écrit :

$$\log L(H_1) = \frac{n_{D_1^h}}{2} \log(2\pi\sigma_{D_1^h}^2) - \sum_{i=1}^{n_{D_1^h}} \frac{(x_i^{D_1^h} - \mu_{D_1^h})^2}{2\sigma_{D_1^h}^2} \quad (6.12)$$

d'où :

$$\log L(H_1) = -\frac{n_{D_1^h}}{2} - \frac{n_{D_1^h}}{2} \log(2\pi\sigma_{D_1^h}^2) \quad (6.13)$$

Soit le test statistique :

$$\gamma(s) = -\min\{\log L(H_3) - \log L(H_1), \log L(H_3) - \log L(H_2)\} \quad (6.14)$$

Bien que la radiométrie des réseaux routiers puisse beaucoup varier selon les zones de l'image, il se trouve que les valeurs des pixels du réseau présentent souvent une forme de

vallée ou de crête par rapport au fond de l'image. Nous pouvons ainsi remarquer qu'en fonction du type d'image observée, nous avons des routes "blanches" ou "noires".

Pour renforcer le test statistique  $\gamma(s)$  d'attache aux données d'un Segment, nous ajoutons un terme d'énergie correspondant aux routes "blanches" :

$$U_{crete}(s) = -\min\{(\mu_{D_3^s} - \mu_{D_3^{sg}})(\mu_{D_3^s} - \mu_{D_3^{sd}})\} \quad (6.15)$$

ou bien un terme d'énergie correspondant aux routes "noires" :

$$U_{vallee}(s) = -\min\{(\mu_{D_3^{sg}} - \mu_{D_3^s})(\mu_{D_3^{sd}} - \mu_{D_3^s})\} \quad (6.16)$$

L'énergie d'attache aux données pour un segment est définie comme suit :

$$U_D(s_i) = \begin{cases} 0 & \text{si } \gamma(s_i) < m \\ \gamma(s_i) & \text{si } \gamma(s_i) \in [m, M] \\ \infty & \text{si } \gamma(s_i) > M \end{cases} \quad (6.17)$$

$m$  et  $M$  sont des constantes positives. Cette construction est nécessaire pour que le "Candy" modèle complété par l'attache aux données soit localement stable.

L'énergie d'attache aux données pour toute la configuration est :

$$U_D(S) = \sum_{i=1}^n U_D(s_i) \quad (6.18)$$

Pour illustrer le rôle de l'énergie d'attache aux données, nous avons fait l'expérience suivante : nous avons fixé la longueur et l'épaisseur des segments ainsi que le nombre d'orientations possibles. En chaque point de l'image, nous avons calculé l'énergie d'attache aux données pour toutes les orientations possibles. Le maximum de niveau de gris en un point de l'image correspond au minimum de l'énergie, par rapport à l'orientation, du segment centré dans ce point.

Dans la Figure 6.5a), nous avons un fragment d'une image satellitaire SPOT (résolution de 10m par pixel). Le résultat de l'attache aux données est montré dans la Figure 6.5b).

L'expression complète de l'énergie du système est alors :

$$U(S) = U_D(S) + U_I(S) = \sum_{s_i \in S} U_D(s_i) - \left( \log g(s_i) + \sum_{s_i \diamond s_j, i < j} \log h(s_i, s_j) \right) \quad (6.19)$$

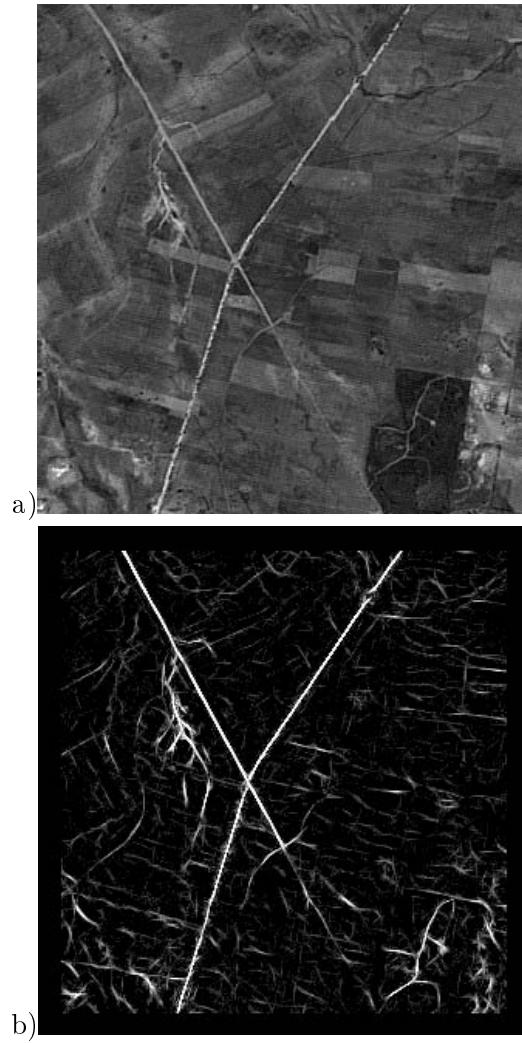


FIG. 6.5 – Résultat de l'attache aux données sur une image SPOT

Nous appelons ce modèle "Candy" modèle complet. Si nous mettons la densité du "Candy" modèle complet sous la forme de densités de processus ponctuels marqués avec interaction entre les points, nous obtenons :

$$f(S) \propto \beta^{n(S)} \prod_{i=1}^n d(s_i) g(s_i) \prod_{i < j, s_i \diamond s_j} h(s_i, s_j) \quad (6.20)$$

avec  $d(s_i) = \exp(-U_D(s_i))$  pour tout  $i = 1, \dots, n$ .

### 6.2.2 Stabilité locale du "Candy" modèle complet

PROPOSITION : *Le processus ponctuel décrit par la densité de probabilité de "Candy" modèle complet a la propriété de stabilité locale*

PREUVE : Soit l'intensité conditionnelle de Papangelou pour le "Candy" modèle complet que nous pouvons écrire en tenant compte de l'Eq. (4.33)

$$\lambda(\zeta/S) < d(\zeta)M \quad (6.21)$$

avec M la constante de stabilité locale de "Candy" modèle.

Le terme  $d(\zeta)$  est construit comme suit :

$$d(\zeta) = \exp(-U_D(\zeta)) < 1 \quad (6.22)$$

Donc la condition de stabilité locale pour le "Candy" modèle complet :

$$\lambda(\zeta/S) < M \quad (6.23)$$

est accomplie.

## 6.3 Résultats

Nous avons choisi d'étudier le comportement des paramètres sur un fragment d'image SPOT (cf. Figure 6.6). Nous avons choisi cette image à cause de sa complexité : la plupart des hypothèses caractéristiques des réseaux routiers y sont présentes.

Nous souhaitons analyser le comportement de la méthode en faisant varier la température  $H$  et le paramètre  $\omega_c$  qui contrôle le prolongement du réseau. Les autres paramètres





FIG. 6.6 – Image SPOT T1, utilisée comme échantillon d’entraînement

$\omega_f, \omega_o, \omega_r, \omega_t$  ont été gardés inchangés.

Nous avons décidé d'étudier l'influence de ces paramètres, car la pratique a montré que c'est eux qui influence le plus la qualité de la détection. Des résultats sont montrés sur les Figures 6.7(a,b et c) et 6.8(a,b,c).

Nous observons que plus  $\omega_c$  il est important, moins nous avons de fausses alarmes. Par contre, dans ce cas, le réseau n'est pas complet.

La température initiale  $H$  a une influence différente. A hautes températures beaucoup de segments libres, mal alignés ou en interaction de type répulsion sont acceptés. Cela permet, de trouver des réseaux plus complets dans l'image, mais toujours avec l'inconvénient de d'avoir des fausses alarmes (cf. Figures 6.7 et 6.8).

Nous avons utilisé des segments avec une épaisseur fixée à un pixel. L'écart entre les masques a été fixé aussi à  $\alpha = 1$ . Cela nous permet de détecter des réseaux dont l'épaisseur est comprise entre 1 et 3 pixels. Les algorithmes ont effectué  $10^6$  itérations pour obtenir les résultats présentes dans ce manuscrit. La configuration initiale est donnée par l'image vide, c'est à dire ne contenant aucun segment. Nous sommes donc dans le cadre complètement non-supervisé, une fois les paramètres fixés. L'extraction des structures principales du réseau nécessite environ 15-30 minutes sur une station de travail Sun Sparc Ultra 10 (300 MHz). Pour compléter le réseau, beaucoup plus de temps est nécessaire.

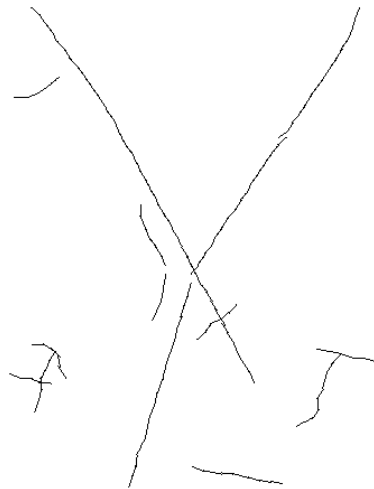
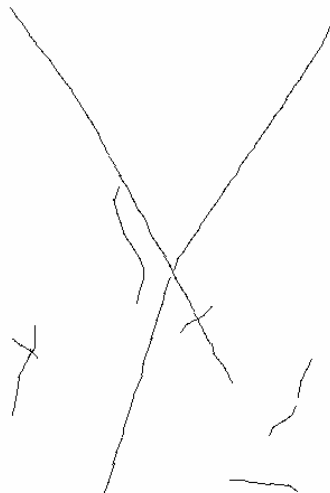
L'étape suivante a consisté à tester les mêmes paramètres sur des extraits de la même image SPOT. Les résultats sont montrés sur les Figures 6.9 et 6.10.

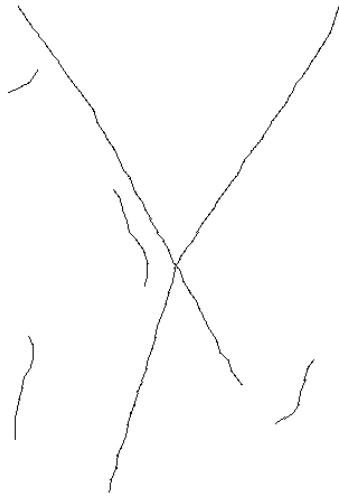
Avec les mêmes paramètres nous avons lancé la détection sur un fragment d'image aérienne fournie par l'IGN (cf. Figure 6.11). Comme sur les images SPOT, les routes sont "blanches". Par conséquent, nous avons utilisé dans le terme d'énergie d'attache aux données la condition de crête (cf. Eq. (6.15)).

Nous avons souhaité tester nos algorithmes sur des réseaux routiers beaucoup plus complexes. Sur la Figure 6.12, nous avons un fragment d'une image SPOT <sup>1</sup> contenant le réseau routier urbain de la ville de Mexico. Nous remarquons que, dans ce cas les routes présentes dans l'image sont plus "noires", donc nous devons utiliser le terme vallée (cf. Eq.(6.16))

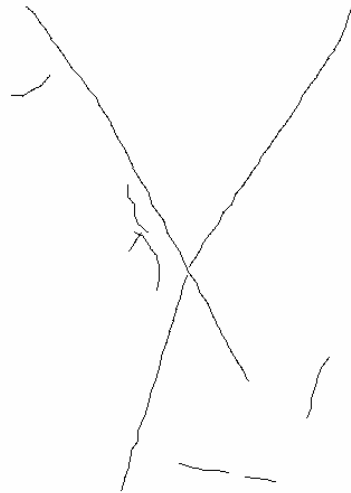
---

<sup>1</sup>image fournie par le CNES (projet ISIS) et par le projet RADAR-PNTS de GDR-ISIS

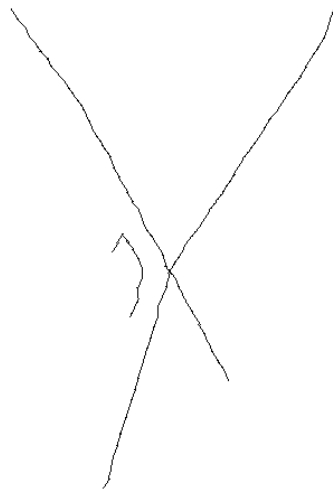
a) Extraction des routes :  $\omega_c = 2.5$ b) Extraction des routes :  $\omega_c = 5.0$ c) Extraction des routes :  $\omega_c = 7.5$ FIG. 6.7 – Détection des routes sur l'image T1 avec comme température initiale  $H = 100.0$



a) Extraction des routes :  $\omega_c = 2.5$



b) Extraction des routes :  $\omega_c = 5.0$



c) Extraction des routes :  $\omega_c = 7.5$

FIG. 6.8 – Détection des routes sur l'image T1 avec comme température initiale  $H = 50.0$



FIG. 6.9 – Extraction des routes : a) images SPOT T5 ; b) réseau routier détecté



FIG. 6.10 – Extraction des routes : a) images SPOT T3 ; b) réseau routier détecté

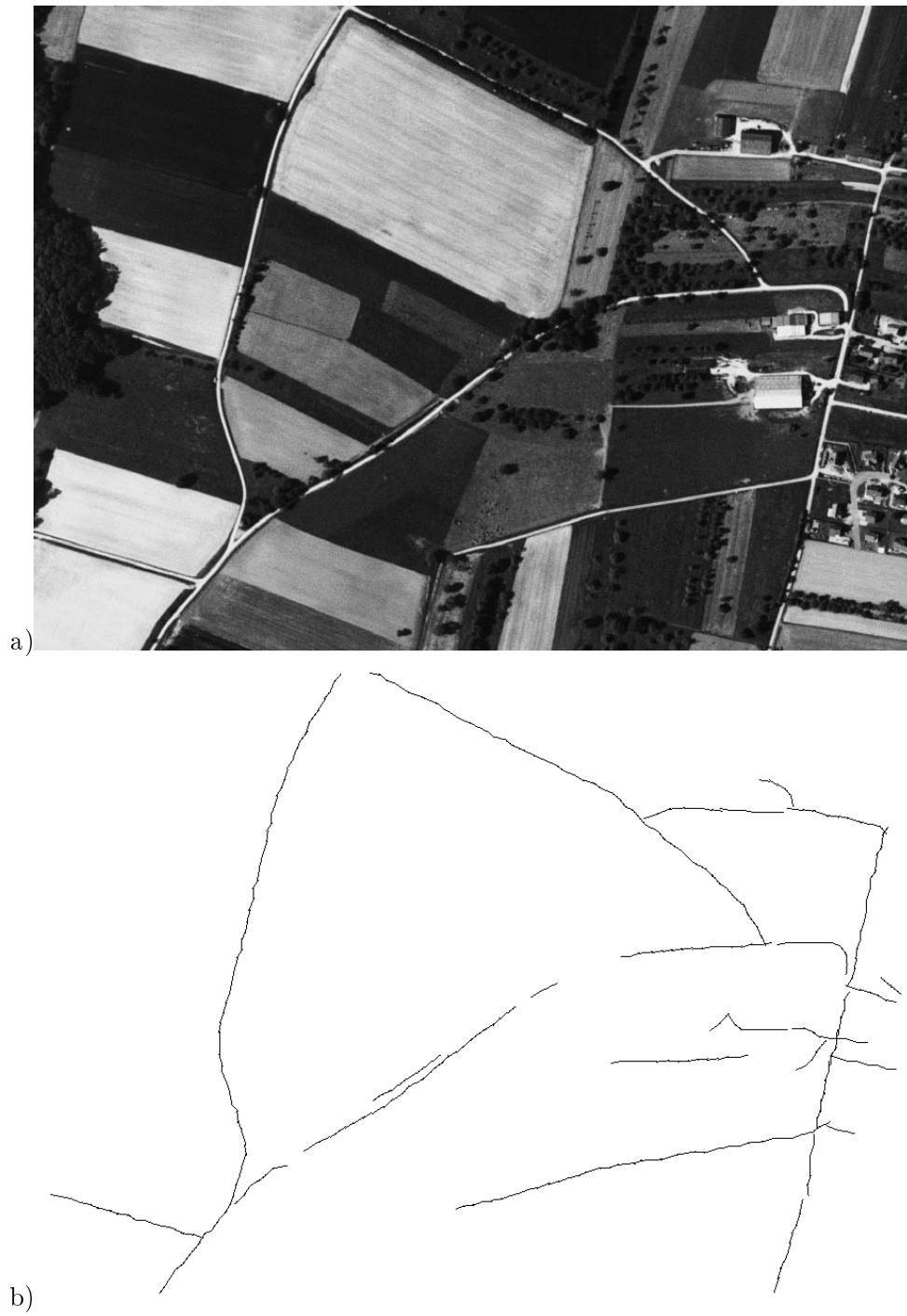


FIG. 6.11 – Extraction des routes sur une image aérienne : a) image initiale ; b) réseau routier détecté

dans le terme d'attache aux données. Le résultat de l'extraction est montré sur la Figure 6.13.



FIG. 6.12 – Image SPOT de la ville de Mexico

Nous avons également testé l'algorithme proposé sur une image radar de type RSO <sup>2</sup>. Le fragment d'image étudié ( cf. Figure 6.14a ) contient le réseau urbain de la ville de Mexico. Nous avons lancé les programmes en gardant l'hypothèse gaussienne dans les tests

---

<sup>2</sup>image fournie par l'ESA





FIG. 6.13 – Réseau routier détecté

utilisés pour l'attache aux données. Il est bien connu que la distribution gaussienne n'est pas la plus souhaitable dans le cas des images radar. Néanmoins, le résultat de l'extraction sur la Figure 6.14b montre le potentiel de notre méthode : une grande partie du réseau est détectée, avec peu de fausses alarmes.

## 6.4 Schéma de descente de température

L'estimateur des réseaux linéiques dans les images a été calculé en utilisant un algorithme de type recuit simulé fondé sur la dynamique RJMCMC construite pour simuler le "Candy" modèle. Le point clef dans la convergence de l'algorithme est le schéma de refroidissement de la température.

[Lieshout, 1994] prouve l'existence d'un tel schéma pour la simulation des processus ponctuels, mais le schéma n'est pas utilisé en pratique. L'optimum global est approché par une réalisation du modèle à température constante. La température est maintenue assez basse tout au long de la simulation (par exemple  $H = 1.0$ ).

Nous avons décidé d'utiliser un schéma mixte. Nous commençons à des températures assez hautes pour favoriser l'apparition des segments les moins probables dans l'image. Pour cette première étape, nous utilisons un des schémas de refroidissement les plus lents :

$$H_{n_{iter}} = \frac{H_0}{\log(1 + n_{iter})} \quad (6.24)$$

avec  $H_0$  la température initiale et  $n_{iter}$  le nombre d'itérations.

Quand la température descend au dessous d'un certain Seuil ( $H < 4.0$ ), nous appliquons un schéma de descente polynomial :

$$H_{n_{iter}+1} = c \times H_{n_{iter}} \quad (6.25)$$

avec  $c = 0.975$ .

En utilisant ce schéma le refroidissement va se produire assez rapidement. Donc les segments ayant une probabilité importante vont être ajoutés à la configuration plus tôt. Quand la température est proche de zéro, par exemple  $H = 0.5$ , nous laissons l'algorithme tourner avec une température constante.

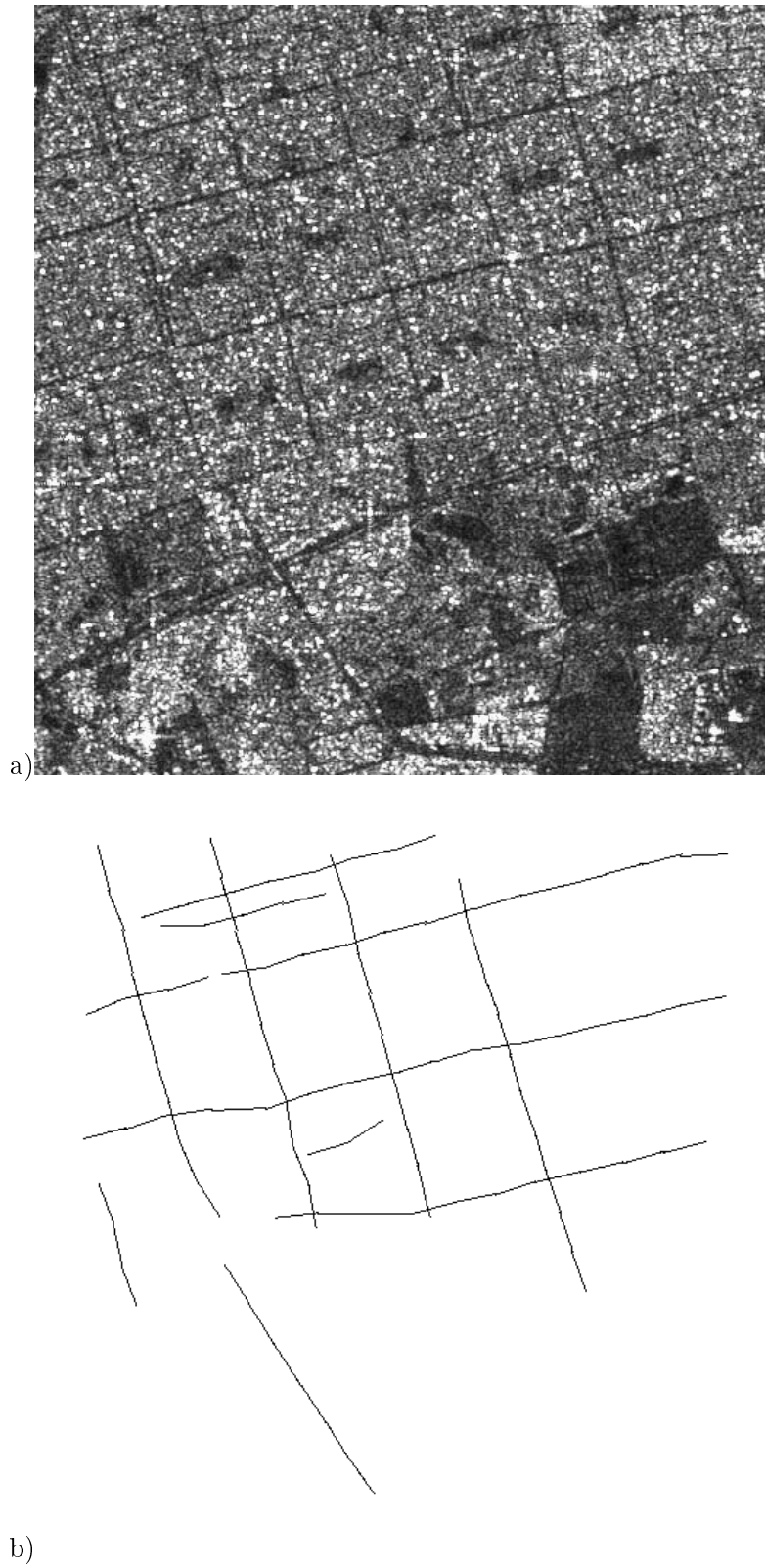


FIG. 6.14 – Extraction des routes sur une image radar ERS : a) image initiale ; b) réseau routier détecté

Ce schéma est bien évidemment une heuristique. Mais par rapport à l'heuristique utilisé dans [Lieshout, 1994] elle a l'avantage de mettre en évidence le rôle de la température dans un algorithme de type recuit simulé.

## 6.5 Validation des résultats

### 6.5.1 Généralités

Les résultats présentés sur les images SPOT et sur l'image aérienne ont été obtenus avec le même jeu de paramètres utilisé pour l'apprentissage. Pour l'image radar, nous avons renforcé le "Candy" modèle pour essayer de compenser le fait que le terme d'attache aux données n'était pas parfaitement adapté à ce type d'image.

Il est toujours délicat d'apprécier la qualité des résultats obtenus, surtout dans le cas où l'on est pas en possession des vérités de terrain. L'apprentissage des paramètres est purement subjectif quand nous ne connaissons pas exactement ce qui est route et ce qui ne l'est pas dans l'image. Nous avons dû nous contenter d'une appréciation uniquement visuelle des résultats obtenus. Mais dans la plupart des cas, les images sur lesquelles nous avons testé notre méthode contenaient des routes qui étaient facilement observables.

Les résultats que nous montrons dans ce manuscrit sont les meilleurs, dans le sens où nous avons dû faire un compromis entre l'exhaustivité du réseau et un nombre important de fausses alarmes. Généralement, le poids du compromis est déterminé par le type d'application.

Nous pouvons remarquer, par rapport aux considérations précédentes, que le réseau détecté par la méthode proposée est assez complet avec très peu de fausses alarmes. Nous pouvons notamment éliminer les fausses alarmes causées par la présence d'objets ayant une radiométrie similaire à celle d'une route (par exemple les petites maisons de la Figure 6.9). Le prolongement du réseau se fait toujours dans la bonne direction. Les interruptions dans le réseau peuvent être comblées si elles sont plus petites que la taille maximale d'un segment (cf. Figure 6.11).

### 6.5.2 Fôrets galeries

Les fôrets galeries sont des objets géographiques représentant des rivières couvertes par les arbres des deux rives qui en forment des galeries. Nous trouvons souvent les fôrets ga-

leries dans les pays d'Afrique.

Le BRGM a mis à notre disposition une image SPOT <sup>3</sup> représentant des forêts galeries ainsi que la vérité terrain correspondante (cf. Figure 6.15).

Pour détecter les réseaux des forêts galeries, nous avons dû modifier les paramètres du "Candy" modèle, par rapport aux paramètres utilisés pour l'extraction de routes. Nous avons relâché les pénalisations sur la répulsion et l'alignement, car les rivières sont beaucoup plus sinueuses que les routes.

Les résultats de l'extraction, obtenus en modifiant le paramètre du prolongement de réseau  $\omega_c$ , sont montrés dans la Figure 6.16. Nous pouvons observer, que nous avons toujours le même compromis à faire par rapport à la vérité terrain : exhaustivité du réseau versus fausses alarmes.

Faute de temps, nous n'avons pas pu approfondir cette étude. La prochaine étape serait de chiffrer ce compromis. Car il est du à plusieurs facteurs. Un de plus importants est le fait que la vérité de terrain est construite pixel par pixel, alors que le réseau que nous détectons est approximé par des segments.

Nous avons donc montré par diverses simulations que le "Candy" modèle peut être utilisé pour l'extraction de réseaux linéïques ayant des topologies différentes comme les rivières ou les routes.

---

<sup>3</sup>images fournies par le BRGM dans le cadre du projet MAJOR soutenu par la DGA

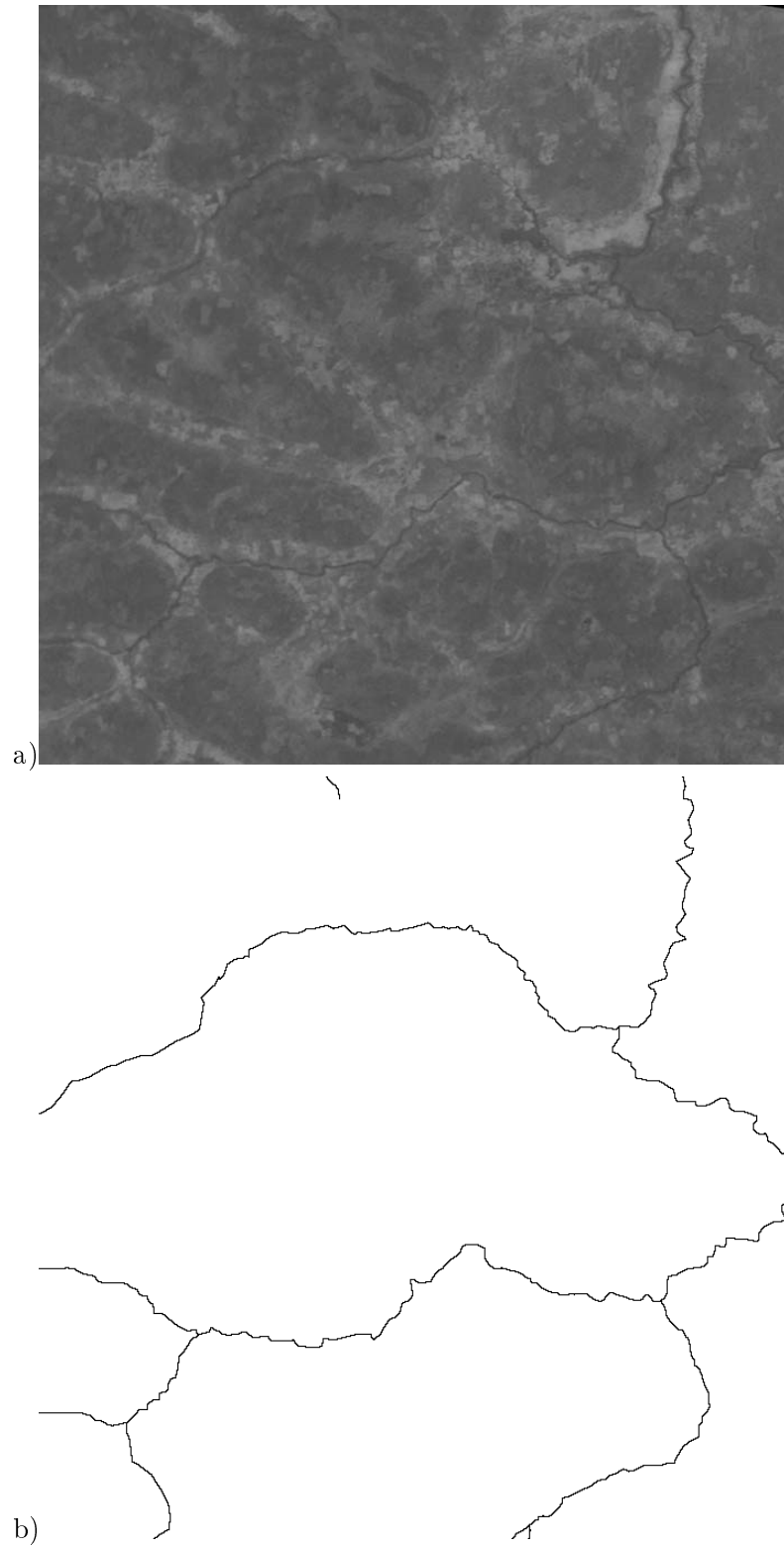


FIG. 6.15 – Extraction des forets galeries : a) image SPOT ; b) vérité de terrain

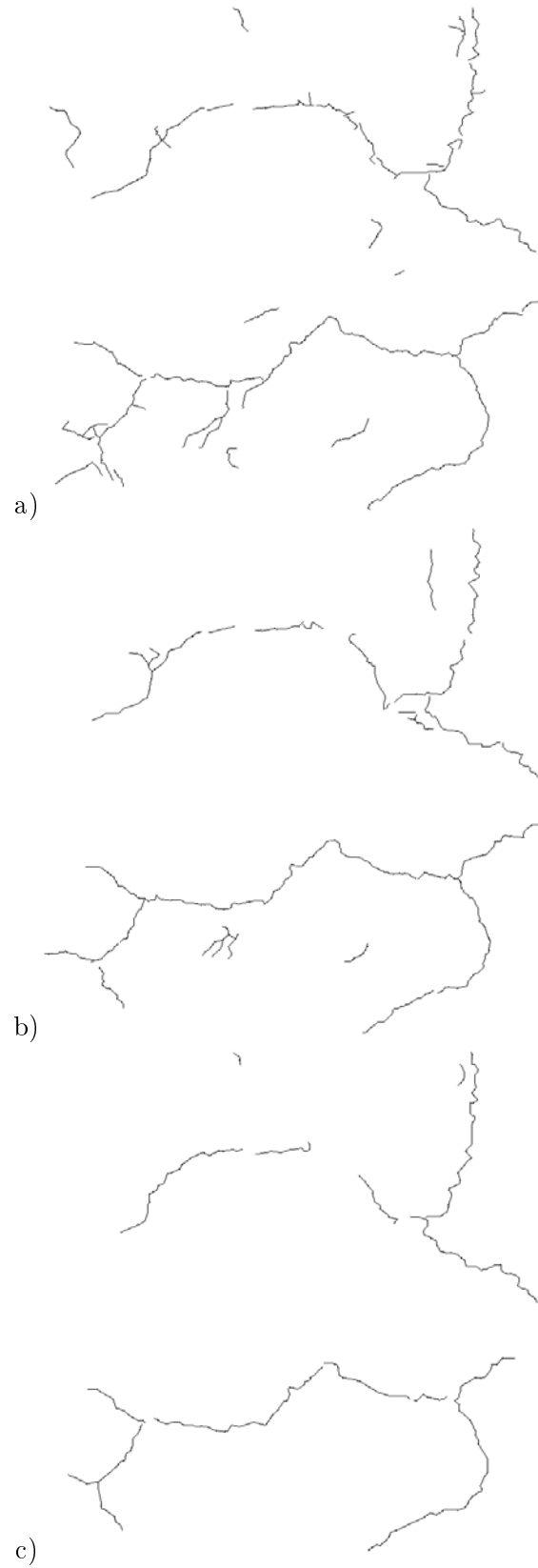


FIG. 6.16 – Résultat de l'extraction : a)  $\omega_c = 2.5$ ; b)  $\omega_c = 5.0$ ; c)  $\omega_c = 7.5$





## Chapitre 7

# Estimation des paramètres

---

Estimer les paramètres d'un modèle utilisé en traitement d'image veut dire que l'on est capable de construire des méthodes complètement non-supervisées.

Nous présentons dans ce chapitre une méthode d'estimation des paramètres du "Candy" modèle, utilisant le Maximum de Vraisemblance faisant appel à une méthode de type MCMC.

Le chapitre commence par la présentation du formalisme de construction du Maximum de Vraisemblance. Ensuite, nous expliquons les modifications que nous avons apportés sur le "Candy" modèle afin que l'estimation des paramètres soit possible.

L'estimateur de Maximum de Vraisemblance est obtenu en utilisant une méthode d'optimisation locale qui s'appuie sur le principe du gradient à pas optimal.

Des résultats et quelques considérations générales sont présentés à la fin de ce chapitre.

---



## 7.1 Fonction de vraisemblance

Soit  $f_\omega(s)$  la densité non-normalisée d'un processus ponctuel,  $c(\omega)$  la constante de normalisation,  $\omega \in V$  les paramètres du modèle et  $V$  le sous-espace des paramètres. Suivant [Geyer et Thompson, 1992], [Geyer, 1999], la fonction de log vraisemblance pour une observation  $s$  s'écrit :

$$l(\omega) = \log f_\omega(s) - \log c(\omega) \quad (7.1)$$

Comme nous cherchons uniquement à maximiser la log-vraisemblance, nous pouvons la considérer à une constante près et modifier légèrement sa définition :

$$l(\omega) = \log \frac{f_\omega(s)}{f_\psi(s)} - \log \frac{c(\omega)}{c(\psi)} \quad (7.2)$$

où  $\psi$  est un paramètre fixé.

Le terme relatif au rapport des constantes de normalisation peut ne pas être calculable analytiquement. Mais, nous avons :

$$\begin{aligned} \frac{c(\omega)}{c(\psi)} &= \frac{1}{c(\psi)} \int_S f_\omega(s) \mu(ds) \\ &= \int_S \frac{f_\omega(s)}{f_\psi(s)} \frac{f_\psi(s)}{c(\psi)} \mu(ds) \\ &= \int_S \frac{f_\omega(s)}{f_\psi(s)} p_\psi(s) \mu(ds) \\ &= \mathbb{E}_\psi \left( \frac{f_\omega(S)}{f_\psi(S)} \right) \end{aligned} \quad (7.3)$$

ce qui montre la possibilité d'approximer le calcul de la log vraisemblance par des méthodes de type Monte Carlo.

Soit  $S_1, S_2, \dots, S_n$  des échantillons simulés à partir de  $f_\psi(s)$ . Alors, la log vraisemblance qui s'écrit :

$$l(\omega) = \log \frac{f_\omega(s)}{f_\psi(s)} - \log \left( \mathbb{E}_\psi \frac{f_\omega(S)}{f_\psi(S)} \right) \quad (7.4)$$

peut être approximée par :

$$l_n(\omega) = \log \frac{f_\omega(s)}{f_\psi(s)} - \log \left( \sum_{i=1}^n \frac{f_\omega(S_i)}{f_\psi(S_i)} \right) \quad (7.5)$$

La valeur  $\hat{\omega}_n$  qui maximise l'équation (7.5) est une approximation de l'estimateur  $\hat{\omega}$  qui maximise l'équation (7.4).

Soit le gradient de l'équation (7.5) :

$$\nabla l_n(\omega) = \nabla \log f_\omega(s) - \frac{\sum_{i=1}^n \frac{\nabla f_\omega(S_i) f_\omega(S_i)}{f_\omega(S_i) f_\psi(S_i)}}{\sum_{i=1}^n \frac{f_\omega(S_i)}{f_\psi(S_i)}} \quad (7.6)$$

Nous reconnaissons une situation d'échantillonnage pondéré, avec les poids d'échantillonnage données par :

$$w_{n,\omega,\psi}(s) = \frac{\frac{f_\omega(s)}{f_\psi(s)}}{\sum_{i=1}^n \frac{f_\omega(S_i)}{f_\psi(S_i)}} \quad (7.7)$$

Le gradient de la fonction de log vraisemblance devient alors :

$$\nabla l_n(\omega) = \nabla \log f_\omega(s) - \mathbb{E}_{n,\omega,\psi} \nabla \log f_\omega(S) \quad (7.8)$$

avec :

$$\mathbb{E}_{n,\omega,\psi} \nabla \log f_\omega(S) = \sum_{i=1}^n \nabla \log f_\omega(S_i) w_{n,\omega,\psi}(S_i) \quad (7.9)$$

l'approximation de Monte Carlo de  $\mathbb{E}_\omega \nabla \log f_\omega(S)$ .

De la même façon, nous avons le hessien de la fonction de log - vraisemblance :

$$\nabla^2 l_n(\omega) = \nabla^2 \log f_\omega(s) - \mathbb{E}_{n,\omega,\psi} \nabla^2 \log f_\omega(S) - \text{Var}_{n,\omega,\psi} \nabla \log f_\omega(S) \quad (7.10)$$

et

$$\begin{aligned} \text{Var}_{n,\omega,\psi} \nabla \log f_\omega(S) &= \mathbb{E}_{n,\omega,\psi} \nabla \log f_\omega(S) \nabla \log f_\omega(S)^T \\ &\quad - [\mathbb{E}_{n,\omega,\psi} \nabla \log f_\omega(S)] [\mathbb{E}_{n,\omega,\psi} \nabla \log f_\omega(S)]^T \end{aligned}$$

avec  $\text{Var}_{n,\omega,\psi} \nabla \log f_\omega(S)$  la matrice de covariance.

En utilisant une méthode itérative d'optimisation locale, nous estimons les paramètres du modèle :

$$\omega_{k+1} = \omega_k - [\nabla^2 l_n(\omega_k)]^{-1} \nabla l_n(\omega_k) \quad (7.11)$$

Nous avons donc tous les éléments pour pouvoir estimer les paramètres d'un processus ponctuel dont on connaît la densité de probabilité à une constante près. Deux conditions sont encore nécessaires : nous devons échantillonner suffisamment longtemps pour atteindre la convergence avant la construction des poids d'échantillonnage, et  $\omega_0$  la valeur initiale des paramètres ne doit pas être trop éloignée de la solution.

## 7.2 Modification du "Candy" modèle

Soit  $f(S)$  la densité du "Candy" modèle mise sous la forme d'un processus ponctuel de Gibbs :

$$f(S) \propto \exp\left(\sum_{i=1}^5 N_i \omega_i\right) = \exp(\langle \vec{N}, \vec{\omega} \rangle) \quad (7.12)$$

$\vec{\omega}$  est le vecteur des paramètres :

$$\begin{aligned} \omega_1 &= \omega_c = \log g_{22} \\ \omega_2 &= \omega_f = \log g_{21} \\ \omega_3 &= \omega_o = \log h_o \\ \omega_4 &= \omega_r = \log h_r \\ \omega_5 &= \omega_t = \log \beta \end{aligned} \quad (7.13)$$

$\omega_c$  contrôle le prolongement du réseau.  $\omega_f, \omega_o, \omega_r$  pénalisent respectivement les segments libres, les segments mal orientés et les segments en interaction de type répulsion.  $\omega_t$  contrôle la densité de segments dans la configuration.

$\vec{N}$  est le vecteur des caractéristiques du modèle :

$$\begin{aligned} N_1 &= N_c \\ N_2 &= N_f \\ N_3 &= N_o \\ N_4 &= N_r \\ N_5 &= N_t \end{aligned} \quad (7.14)$$

$N_c$  est le nombre de segments avec une seule extrémité connectée,  $N_f$  est le nombre de segments libres,  $N_o$  est le nombre de paires de segments mal orientés,  $N_r$  est le nombre de paires de segments en répulsion, et  $N_t$  est le nombre total de segments dans la configuration  $S$ .

Le terme  $\omega_t$  agit sur tous les segments indifféremment de l'état des segments et du type d'interaction : répulsion ou attraction. Le terme  $\omega_c$  pénalise tout prolongement du réseau, indépendamment de la façon d'interagir d'un segment avec les autres segments.

La seule configuration état qui est uniquement pénalisée par  $\omega_t$  est quand un segment est connecté aux deux extrémités et correctement aligné par rapport aux autres segments. Si un segment est libre avec ou sans interaction, ajouter le terme  $\omega_t$  à l'énergie est une

pénalisation de plus. Il en est de même pour le terme  $\omega_c$ .

Dans le cas de l'extraction de routes, cette sur-paramétrisation est utile dans le sens où, tout en contrôlant la densité des segments dans l'image, nous souhaitons prolonger le réseau dans la direction où les données sont les meilleures.

Pour l'estimation des paramètres, il est nécessaire de réduire cette redondance entre les paramètres. Nous souhaitons que chaque paramètre agisse dans des situations qui ne se superposent pas, afin que les paramètres soient plus "orthogonaux".

Un segment a deux états : libre ou connecté. Il peut être en interaction de type répulsion ou d'attraction. S'il est dans une interaction de type attraction, il peut être bien ou mal orienté.

Pour mieux préciser l'action de chaque paramètre en fonction de l'état d'un segment et de ses interactions avec le reste de la configuration, nous avons décidé de modifier le "Candy" modèle comme suit :

$$\begin{aligned}
 \omega_1 &= \omega_f = -\log g_{21} \\
 \omega_2 &= \omega_o = -\log h_o \\
 \omega_3 &= \omega_r = -\log h_r \\
 \omega_4 &= \omega_a = -\log h_a
 \end{aligned}
 \tag{7.15}$$

avec  $h_a < 1$ .

Le dernier terme est une pénalisation appliquée aux segments qui se trouvent dans une interaction de type attraction et qui vérifient les conditions d'alignement. Les interactions sont définies de la même façon que pour le "Candy" modèle initial. Les caractéristiques d'une configuration sont :

$$\begin{aligned}
 N_1 &= N_f \\
 N_2 &= N_o \\
 N_3 &= N_r \\
 N_4 &= N_a
 \end{aligned}
 \tag{7.16}$$

avec  $N_a$  le nombre de paires de segments qui sont en interaction d'attraction et qui sont bien alignés.

Il est très facile de prouver que le modèle reste stable malgré cette modification. Pour simuler ce modèle, nous utilisons la même méthode que pour le "Candy" modèle. Il n'y

a aucun changement au niveau de la dynamique et des taux d'acceptation. La chaîne de Markov qui simule un tel modèle a les mêmes propriétés d'irréductibilité, de récurrence et d'ergodicité que la chaîne construite pour échantillonner la loi du "Candy" modèle initial.

Sur la Figure 7.1, nous montrons quelques réalisations du nouveau modèle pour différents paramètres.

### 7.3 Log-vraisemblance du "Candy" modèle

Soit  $f_\omega(S)$  la densité de probabilité non-normalisée de "Candy" modèle modifié :

$$f(S) \propto \exp\left(\sum_{i=1}^4 N_i(S)\omega_i\right) \propto \exp(\langle N(S), \omega \rangle) \quad (7.17)$$

avec le vecteur des caractéristiques de la réalisation  $S$  :

$$N(S) = [N_f(S), N_o(S), N_r(S), N_a(S)] \quad (7.18)$$

et le vecteur des paramètres :

$$\omega(S) = [\omega_f(S), \omega_o(S), \omega_r(S), \omega_a(S)] \quad (7.19)$$

Soit  $S_d$  une réalisation du modèle. Le vecteur des caractéristiques du modèle est :

$$N(S_d) = [N_f(S_d), N_o(S_d), N_r(S_d), N_a(S_d)] \quad (7.20)$$

Nous souhaitons estimer les paramètres du modèle,  $\hat{\omega}$ , qui maximisent la fonction de log-vraisemblance.

La forme de la densité du modèle nous donne :

$$\begin{aligned} \log f_\omega(s) &= \langle N(s), \omega \rangle \\ \nabla \log f_\omega(s) &= N(s) \\ \nabla^2 \log f_\omega(s) &= 0 \end{aligned} \quad (7.21)$$

et les poids d'échantillonnage :

$$w_{n,\omega,\psi}(s) = \frac{\exp\langle N(s), \omega - \psi \rangle}{\sum_{i=1}^n \exp\langle N(S_i), \omega - \psi \rangle} \quad (7.22)$$

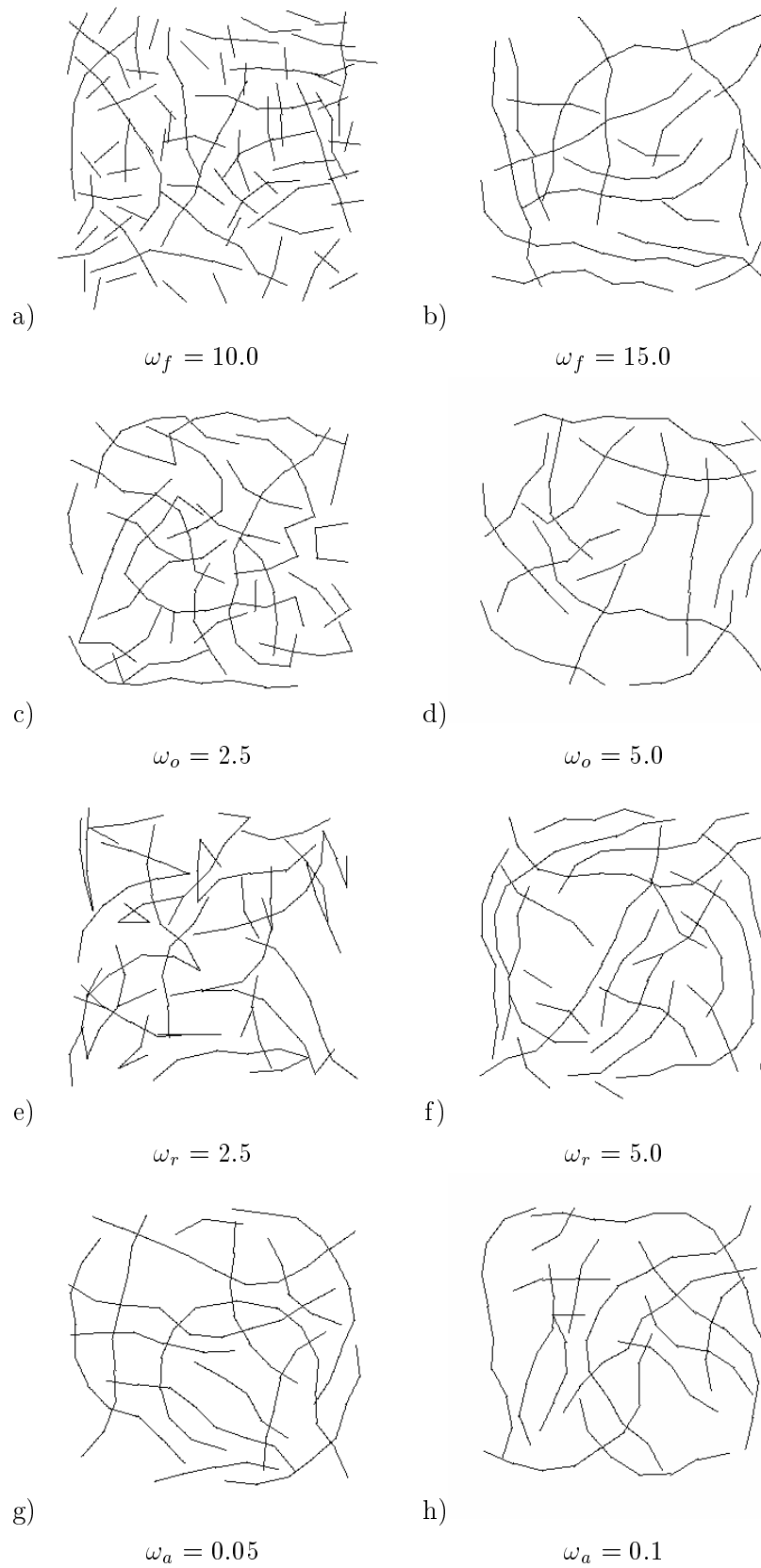


FIG. 7.1 – Réalisations du "Candy" modèle modifié



Nous obtenons pour la fonction de log-vraisemblance :

$$l_n(\omega) = \langle N(S_d), \omega - \psi \rangle - \log \left( \frac{1}{n} \sum_{i=1}^n \exp^{\langle N(S_i), \omega - \psi \rangle} \right) \quad (7.23)$$

et le gradient devient :

$$\nabla l_n(\omega) = N(S_d) - \mathbb{E}_{n,\omega,\psi} N(S) = N(S_d) - \sum_{i=1}^n N(S_i) w_{n,\omega,\psi}(S_i) \quad (7.24)$$

ainsi que le Hessien :

$$\nabla^2 l_n(\omega) = -\text{Var} N(S) \quad (7.25)$$

## 7.4 Algorithme d'estimation

L'algorithme d'estimation des paramètres que nous présentons ci-après s'appuie sur le principe des méthodes de gradient à pas optimal [Ciarlet, 1994] :

$$\begin{cases} l_n(\omega_k + \rho(\omega_k) \nabla l_n(\omega_k)) = \max_{\rho \in \mathbb{R}} l_n(\omega_k + \rho \nabla l_n(\omega_k)) \\ \omega_{k+1} = \omega_k + \rho(\omega_k) \nabla l_n(\omega_k) \end{cases} \quad (7.26)$$

$\rho(\omega_k)$  étant le pas optimal. Sa valeur est trouvée en effectuant une optimisation unidimensionnelle de la fonction de vraisemblance.

Le fait que nous calculions le gradient de la fonction de vraisemblance et que nous calculions aussi le pas d'optimisation nous démarque des méthodes présentés dans [Younes, 1988], [Moyeed et Baddeley, 1991].

La différence principale par rapport à l'algorithme décrit dans [Descombes et al., 1999a] réside dans la non-utilisation du hessien. Dans la plupart des cas nous souhaitons avoir des configurations avec très peu de segments libres, mal orientés ou en interaction de type répulsion. Ceci peut rendre le hessien mal conditionné.

ALGORITHME :

1. Soit une réalisation  $S_d$ , caractérisée par  $N_d$

2. Initialiser  $\hat{\omega}_0$  et  $k = 0$

3. Echantillonner la loi  $f_{\omega_k}(S)$

4. Calculer le gradient de la fonction de log-vraisemblance
5. Pour chaque direction du gradient  $\Delta_i$ , calculer les intervalles  $I_k^i = [\omega_k^i - \lambda\Delta_i, \omega_k^i + \lambda\Delta_i]$ , avec  $\lambda$  de valeur fixée
6. Calculer la nouvelle valeur  $\omega_{k+1}$  en maximisant la log-vraisemblance dans chaque intervalle, utilisant la méthode de recherche par section d'or
7. Si  $\|\omega_{k+1} - \omega_k\| > T$ , alors  $k = k + 1$  et aller à l'étape 3.  $T$  est une valeur fixée a priori.

Pour choisir la valeur  $\lambda$  nécessaire pour calculer les intervalles  $I_i$ , nous nous sommes appuyés sur des heuristiques issues des simulations. Chaque composante  $\omega^i$  du vecteur des paramètres peut être encadrée par l'intervalle  $[\omega_{\min}^i, \omega_{\max}^i]$ .

Par exemple, l'intervalle correspondant à la pénalisation des segment libres est compris entre  $[10, 25]$  : pour  $\omega_f < 10$ , le nombre de segments libres  $N_f$  est très important dans une configuration, ce qui alourdit les calculs et qui est assez éloigné des configurations que l'on souhaite déterminer. Pour  $\omega_f > 25$ , la valeur moyenne de  $N_f$  tend vers zéro, donc nous ne pouvons pas estimer avec précision les valeurs des paramètres qui dépassent cette valeur moyenne. Il est évident, que ces limites doivent être fixées par rapport aux topologies du modèle que l'on souhaite obtenir.

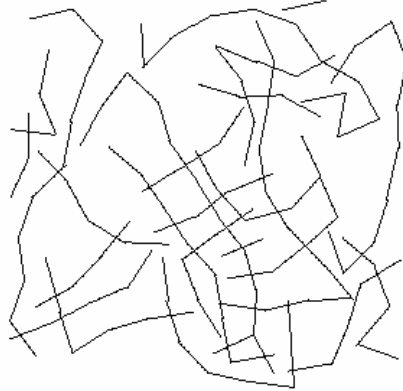
En conclusion,  $\lambda$  est fixé de telle façon que  $I_k^i \subset [\omega_{\min}^i, \omega_{\max}^i]$ .

## 7.5 Résultats

Nous montrons sur la Figure 7.2 une réalisation du modèle à estimer et ses paramètres. Le tableau montre les valeurs des statistiques suffisantes de la réalisation, les moyennes des statistiques suffisantes et les paramètres avec lesquelles le modèle a été simulé.

Le modèle a été simulé pendant  $10^6$  itérations. Les statistiques suffisantes ont été prélevées toutes les  $10^3$  itérations. L'algorithme d'estimation des paramètres contient les mêmes mouvements que ceux utilisés dans la construction du modèle. Les mouvements ont également le même poids dans la dynamique. Pour calculer les moyennes des statistiques suffisantes, nous avons simulé le modèle pendant  $1.5 \times 10^5$  itérations avec une période de

chauffage de  $2.5 \times 10^5$  itérations. Ces valeurs ont été choisies par rapport aux courbes sur l'évolution des moyennes des statistiques suffisantes présentées dans le chapitre étudiant la dynamique RJMCMC.



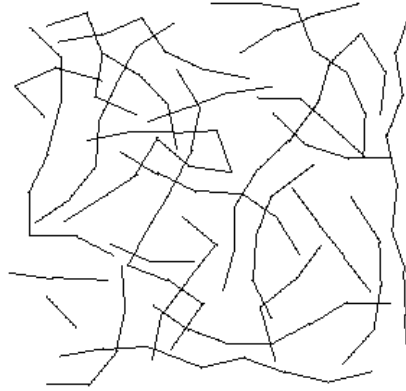
Statistiques suffisantes	Moyennes des statistiques	Paramètres
$N_f = 2$	$\overline{N_f} = 2.0030$	$\omega_f = 15.0$
$N_o = 24$	$\overline{N_o} = 24.5490$	$\omega_o = 2.5$
$N_r = 4$	$\overline{N_r} = 2.8340$	$\omega_r = 5.0$
$N_a = 75$	$\overline{N_a} = 67.8230$	$\omega_a = 0.25$

FIG. 7.2 – Réalisation du modèle à estimer

Pour tester l'algorithme, nous avons décidé dans un premier temps, de considérer comme données les moyennes des statistiques suffisantes. Le premier test (cf. Figure 7.3) a été réalisé à partir d'une initialisation avec la solution. Des résultats avec une initialisation différente de la solution sont montrés sur les Figures 7.4 et 7.5. Dans les tableaux de résultats, nous montrons les données, les conditions initiales, les moyennes des statistiques suffisantes et les paramètres estimés. Nous présentons également une réalisation du modèle avec les paramètres estimés.

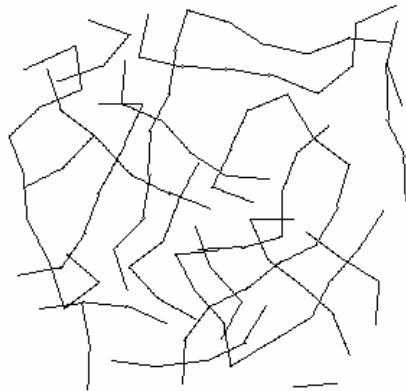
Utiliser comme données les moyennes des statistiques suffisantes n'est pas toujours possible. Souvent, nous devons pouvoir estimer les paramètres d'un modèle à partir d'un seul échantillon.

Les résultats d'une telle estimation sont montrés dans les Figures 7.6, 7.7 et 7.8.



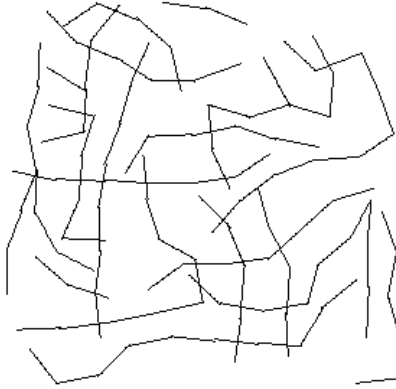
Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$\overline{N}_f(S_d) = 2.0030$	$\omega_f = 15.0$	$\overline{N}_f = 1.9990$	$\hat{\omega}_f = 15.0277$
$\overline{N}_o(S_d) = 24.5490$	$\omega_o = 2.5$	$\overline{N}_o = 24.5473$	$\hat{\omega}_o = 2.4831$
$\overline{N}_r(S_d) = 2.8340$	$\omega_r = 5.0$	$\overline{N}_r = 2.8255$	$\hat{\omega}_r = 4.9951$
$\overline{N}_a(S_d) = 67.8230$	$\omega_a = 0.25$	$\overline{N}_a = 67.8241$	$\hat{\omega}_a = 0.2723$

FIG. 7.3 – Estimation des paramètres à partir des moyennes des statistiques suffisantes



Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$\overline{N}_f(S_d) = 2.0030$	$\omega_f = 12.0$	$\overline{N}_f = 1.99786$	$\hat{\omega}_f = 14.9684$
$\overline{N}_o(S_d) = 24.5490$	$\omega_o = 6.5$	$\overline{N}_o = 24.5457$	$\hat{\omega}_o = 2.4725$
$\overline{N}_r(S_d) = 2.8340$	$\omega_r = 1.5$	$\overline{N}_r = 2.8438$	$\hat{\omega}_r = 5.0235$
$\overline{N}_a(S_d) = 67.8230$	$\omega_a = 0.1$	$\overline{N}_a = 67.8216$	$\hat{\omega}_a = 0.2540$

FIG. 7.4 – Estimation des paramètres à partir des moyennes des statistiques suffisantes



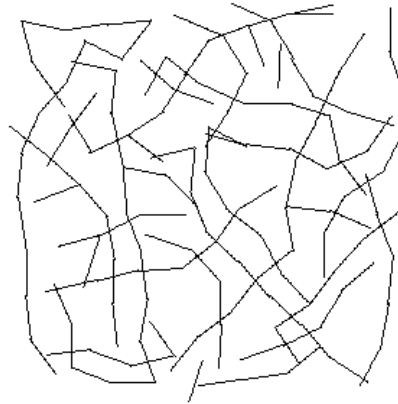
Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$\overline{N}_f(S_d) = 2.0030$	$\omega_f = 18.0$	$\overline{N}_f = 2.2133$	$\hat{\omega}_f = 15.0476$
$\overline{N}_o(S_d) = 24.5490$	$\omega_o = 8.5$	$\overline{N}_o = 25.6466$	$\hat{\omega}_o = 2.4870$
$\overline{N}_r(S_d) = 2.8340$	$\omega_r = 3.5$	$\overline{N}_r = 2.8699$	$\hat{\omega}_r = 4.9479$
$\overline{N}_a(S_d) = 67.8230$	$\omega_a = 0.9$	$\overline{N}_a = 65.4733$	$\hat{\omega}_a = 0.2659$

FIG. 7.5 – Estimation des paramètres à partir des moyennes des statistiques suffisantes

La différence entre les résultats de ces deux expériences s'explique par le fait qu'il existe un certain écart entre les statistiques suffisantes et les moyennes des statistiques suffisantes. Cet écart existe aussi dans le cas de l'estimation des paramètres pour les champs de Markov en traitement d'image. Mais dans ce dernier cas, cet écart a moins d'influence sur les paramètres [Descombes et al., 1999a].

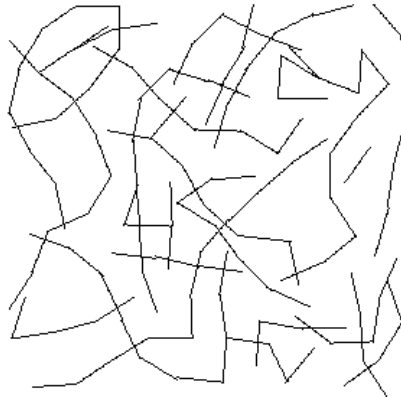
La fonction de log-vraisemblance est convexe. Pour être sûr d'avoir atteint la convexité, c'est à dire la convergence de la chaîne simulée, nous devons échantillonner longtemps. La méthode est robuste si les conditions initiales ne sont pas très éloignées de la solution recherchée [Geyer, 1999], [Descombes et al., 1999a]. Dans le cas où nous sommes trop éloignés des conditions initiales, nous devons réinitialiser l'algorithme et ré-échantillonner le modèle, ce qui est coûteux en temps de calcul. Nous observons donc l'utilité pratique de posséder des critères de convergence fiables pour la résolution d'un tel problème.

Ce travail reste à être complété en ajoutant un terme d'attache aux données. C'est à dire passer au cas des données incomplètes. Cette dernière étape permettra le passage à des



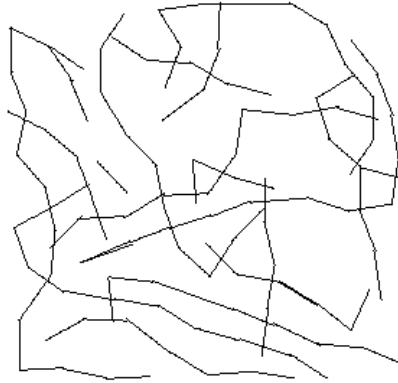
Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$N_f(S_d) = 2$	$\omega_f = 15.0$	$\overline{N_f} = 2.0099$	$\hat{\omega}_f = 14.7949$
$N_r(S_d) = 24$	$\omega_o = 2.5$	$\overline{N_o} = 24.5333$	$\hat{\omega}_o = 2.5101$
$N_o(S_d) = 4$	$\omega_r = 5.0$	$\overline{N_r} = 4.1433$	$\hat{\omega}_r = 4.6762$
$N_a(S_d) = 75$	$\omega_a = 0.25$	$\overline{N_a} = 74.7066$	$\hat{\omega}_a = 0.1825$

FIG. 7.6 – Estimation des paramètres à partir des statistiques suffisantes



Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$N_f(S_d) = 2$	$\omega_f = 12.0$	$\overline{N_f} = 2.0283$	$\hat{\omega}_f = 14.8863$
$N_r(S_d) = 24$	$\omega_o = 6.5$	$\overline{N_o} = 24.0051$	$\hat{\omega}_o = 2.5041$
$N_o(S_d) = 4$	$\omega_r = 1.5$	$\overline{N_r} = 3.9934$	$\hat{\omega}_r = 4.7568$
$N_a(S_d) = 75$	$\omega_a = 0.1$	$\overline{N_a} = 75.0031$	$\hat{\omega}_a = 0.1942$

FIG. 7.7 – Estimation des paramètres à partir des statistiques suffisantes



Données	Conditions initiales	Moyennes des statistiques	Paramètres estimés
$N_f(S_d) = 2$	$\omega_f = 18.0$	$\overline{N_f} = 1.9931$	$\omega_f = 14.7721$
$N_r(S_d) = 24$	$\omega_o = 8.5$	$\overline{N_o} = 23.99746$	$\omega_o = 2.4976$
$N_o(S_d) = 4$	$\omega_r = 3.5$	$\overline{N_r} = 3.9908$	$\omega_r = 4.6536$
$N_a(S_d) = 75$	$\omega_a = 0.9$	$\overline{N_a} = 75.0006$	$\omega_a = 0.1602$

FIG. 7.8 – Estimation des paramètres à partir des statistiques suffisantes

techniques complètement non-supervisées.





# Conclusion et perspectives

Le travail que nous avons mené propose une nouvelle manière de résoudre des problèmes classiques en traitement d'image comme l'extraction de réseaux linéiques, par l'adoption d'un cadre théorique peu utilisé jusqu'à présent dans ce domaine : les processus ponctuels marqués.

Nous avons donc continué la voie ouverte par des pionniers : [Lieshout, 1994], [Storvik, 1994], [Rue et Husby, 1998], [Pievatolo et Green, 1998]. L'idée qui est à la base de ces travaux donne un point de départ sur la possibilité d'interpréter l'image comme une collection d'objets et non plus seulement comme un ensemble de valeurs numériques.

A partir de cette idée, nous avons créé une solution de toutes pièces, par un cheminement classique : étude bibliographique, prise en main de l'appareil théorique, modélisation du problème, étude du modèle proposé, présentation et analyse des résultats.

Dans ce qui suit, nous allons essayer de présenter un certain nombre d'améliorations et de perspectives pratiques et théoriques qui peuvent être envisagées.

En ce qui concerne les résultats obtenus, une validation quantitative est nécessaire. Il existe une différence essentielle entre notre solution et celles proposées par d'autres méthodes, dans le sens où la méthode étudiée dans ce manuscrit est complètement non-supervisée, une fois les paramètres du modèle fixés, c'est à dire qu'aucune intervention de l'utilisateur (définition de points de contrôle, d'amorces, ...) n'est nécessaire. Une autre différence importante est la souplesse de "Candy" modèle à être adapté pour l'extraction des réseaux linéiques avec des topologies différentes. Nous pensons que nous avons montré aussi des exemples sur des scènes réalistes où les méthodes classiques mentionnées dans la littérature ne pourraient pas aboutir. Nous faisons ici référence aux méthodes de suivi de routes ainsi qu'aux méthodes qui s'appuient sur un raffinement d'une détection utilisant des opérateurs bas-niveau, méthodes qui ont beaucoup des difficultés à surmonter les obstacles

sur les routes et qui sont très sensibles au bruit dans l'image.

Néanmoins, tester différentes méthodes sur les mêmes images, accompagnées par des vérités de terrain, nous permettra dans le futur de chiffrer et d'évaluer la capacité réelle de la méthode proposée pour extraire des réseaux linéïques. Nous pensons notamment à comparer notre méthode avec celle proposée par [Tupin, 1997], méthode très performante qui obtient des résultats excellents sur des images radar, donc très bruitées. La méthode précitée a été spécialement conçue pour des images radar, donc pour effectuer une comparaison réelle, il faudrait aussi que nous adaptions notre terme d'attache aux données à ce type d'images. Pour construire une attache aux données qui sera indépendante du type d'image, nous pensons notamment à utiliser pour l'attache aux données un test de type Kolmogoroff-Smirnov, test qui n'effectue aucune hypothèse sur la loi de probabilité suivie par les niveaux de gris des pixels dans l'image.

Nous pensons également qu'il faut analyser l'apport qui peut être amené par l'utilisation de plusieurs bandes spectrales dans la construction du terme d'attache aux données. Pour éliminer les difficultés de détection intra-urbaines rencontrées dans les images haute résolution, une segmentation de ces régions peut être nécessaire [Lorette et al, 2000]. Une telle segmentation pourrait également être utile à la transformation du "Candy" modèle en processus non-homogène.

L'étude de l'estimation des paramètres du "Candy" modèle nous a mené à modifier et à reparamétriser celui-ci. Ceci est dû à notre façon d'appréhender le problème : nous avons cherché à obtenir les résultats sur des données réelles, avant de faire une analyse plus minutieuse de l'ancien modèle <sup>1</sup>. Il est donc nécessaire de construire un détecteur de réseaux linéïques utilisant le "Candy" modèle modifié, en lui ajoutant par exemple le terme d'attache aux données développé dans ce manuscrit. Ensuite, il faudra étudier la possibilité de faire l'estimation des paramètres dans le cas de données incomplètes. C'est seulement à ce moment là, que l'on possèdera une méthode réellement non-supervisée.

L'estimation des paramètres ne peut être effectuée que si la fonction de vraisemblance a atteint la convexité, c'est à dire que si la distribution de la chaîne de Markov qui simule le modèle a convergé vers sa distribution d'équilibre.

Nous avons donc besoin de critères fiables qui attestent de la convergence de la chaîne.

---

<sup>1</sup>... et pourtant c'est moi qui ai écrit la première page du mémoire ...

Nous espérons trouver une réponse à ce problème et améliorer la convergence de la dynamique en utilisant une simulation exacte [Kendall et Møller, 1999].

Un facteur déterminant dans l'accélération de la convergence de la dynamique est le choix des mouvements. C'est toujours un problème ouvert, mais si la proposition des mouvements en fonction de l'état de la configuration reste un problème très délicat à mettre en œuvre, nous pouvons essayer de construire des mouvements qui ont plus de chance d'être acceptés que d'autres, comme la naissance et mort d'un segment correctement orienté, connexion et déconnexion des extrémités de segments qui se trouvent à une distance plus proche que la longueur minimale d'un segment ou bien le déplacement d'une extrémité où plusieurs segments sont connectés. Si nous n'avons pas envisagé ces améliorations pendant la thèse, c'est que l'écriture du taux d'acceptation de ces mouvements est loin d'être triviale.

Pour accélérer les algorithmes simulant le "Candy" modèle et améliorer la convergence, il faut étudier la possibilité de paralléliser ce type d'algorithmes. Nous pensons que ce point est d'un grand intérêt pour la simulation des processus ponctuels en général.

Les résultats et les simulations que nous avons présentés ne prennent pas en considération l'épaisseur du segment. Nous pensons qu'il est peut-être intéressant d'enrichir le "Candy" modèle avec un terme concernant l'épaisseur : privilégier la connexion entre les segments de même épaisseur. Cela peut nous aider à classifier par exemple les éléments du réseau routier (par exemple : route nationale, autoroute, etc.).

Le "Candy" modèle est un processus ponctuel homogène. Une direction intéressante à étudier peut être le cas des réseaux linéiques en tant que processus ponctuels non-homogènes [Stoyan et Stoyan, 1998] : cela peut permettre de séparer, par exemple, le réseau routier intra-urbain du réseau routier inter-urbain, ou bien de détecter dans les images médicales des régions avec une forte vascularisation.

Le développement de "Candy" modèle en trois dimensions a un intérêt particulier pour des applications comme la simulation des processus de fissures, ou l'extraction des vaisseaux sanguins.

Nous allons finir cette thèse en énonçant une des perspectives les plus utopiques, mais qui nous est très chère : utilisant le cadre de la théorie des processus ponctuels, mettre en œuvre des modèles qui prennent en compte plusieurs objets [Descombes et al., 2001]. Si cela

est possible, nous pourrions effectuer des tâches comme la segmentation des images satellitaires en utilisant des contraintes de type : "une route mène toujours à une ville", ou bien "une route passe au bord des champs, pas en plein milieu", et ainsi de suite. Nous croyons que le formalisme des processus ponctuels peut nous amener vers une vraie modélisation de scène.

Voilà quelques points, quelques idées qui peuvent être prises en compte si l'on souhaite améliorer, développer ce modèle, ou bien en créer d'autres. Ce travail est plus que perfectible et loin d'être fini, mais il montre également que la théorie des processus ponctuels marqués peut être un outil théorique à prendre en considération pour résoudre des problèmes d'une réelle importance en traitement d'images.

# Bibliographie

- [Antoine, 2000] F. Antoine, "Extraction de Routes par Analyse Multi-Echelle", *Rapport de stage d'option scientifique, Ecole Polytechnique, INRIA*, 2000.
- [Azencott, 1992] R. Azencott, "Markov Field Approach : Parameter Estimation by Qualitative Boxes", Cours : Ecole d'été des Houches, 1992.
- [Baddeley et van Lieshout, 1993] A. J. Baddeley, M. N. M. van Lieshout, " Stochastic geometry models in high-level vision ", In *Statistics and Images, Volume 1*, K. V. Mardia and G. K. Kanji (Eds.) *Advances in Applied Statistics*, a supplement to *Journal of Applied Statistics* 20 : 231-256. Abingdon : Carfax, 1993.
- [Barzohar et Cooper, 1996] M. Barzohar, D. B. Cooper, "Automatic Finding of Main Roads in Aerial Images by Using Geometric-Stochastic Models and Estimation", *IEEE Transactions on PAMI* vol. 18, pp. 707-721, July 1996.
- [Bogess, 1994] J. E. Bogess "Using Artificial Neural Networks to Identify Roads in Satellite Images", Proceedings of the World Congress on Neural Networks - San Diego, pp. 410-415, 1994.
- [Bordes, 1997] G. Bordes, "Interprétation d'Images Aériennes par une Base de Données Cartographiques", Thèse de doctorat, Université de Marne-la-Vallée, 1997.
- [Brooks et Giudici, 1998] S. P. Brooks, P. Giudici, "Convergence Assessment for Reversible Jump MCMC Simulations". In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds), *Bayesian Statistics*, 6, Oxford University Press, 1998.
- [Brooks, 1998] S. P. Brooks, "Quantitative Convergence Assessment for Markov Chain Monte Carlo via Cusums", *Statistics and Computing*, 8, pp. 267-274, 1998.
- [Brooks et Roberts, 1998] S. P. Brooks, G. Roberts, "Assessing Convergence of Markov Chain Monte Carlo Algorithms", *Statistics and Computing*, 8, pp. 319-335.
- [Ciarlet, 1994] P. G. Ciarlet, "Introduction à l'Analyse Numérique Matricielle et à l'Optimisation", Masson, Paris, 1994.
- [Cocquerez et Philipp (eds.), 1995] J.-P. Cocquerez, S. Philipp, "Analyse d'images : filtrage et segmentation", Masson, Paris, 1995.

- [Dacunha et al., 1982] D. Dacunha-Castelle, M. Duffo, "Probabilités et Statistiques", Masson, Paris, 1982.
- [Descombes et al., 1999a] X. Descombes, R. D. Morris, J. Zerubia, M. Berthod, "Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood", *IEEE Trans. on IP*, pp. 954-963, 8, Juillet 1999.
- [Descombes et al., 1999b] X. Descombes, R. Stoica, J. Zerubia, "Two Markov Point Processes for Simulating Line Networks", *Invited paper : IEEE International Conference on Image Processing* Kobe, Japan, 1999.
- [Descombes et al., 2001] X. Descombes, R. Stoica, L. Garcin, J. Zerubia, "A RJMCMC Algorithm for Object Processes", dans *Monte Carlo Methods and Applications*, à paraître en 2001.
- [Deriche et Faugeras, 1995] R. Deriche, O. Faugeras, "Les EDP en Traitement des Images et Vision par Ordinateur", *Rapport de Recherche INRIA*, No. 3239, Novembre 1995.
- [Dhéréte et Desachy, 1999] P. Dhéréte, J. Desachy, "Data Fusion for Linear Geographic Feature Matching on SPOT Images", *Workshop "3D Geospatial Data Production : Meeting Application Requirements"*, Bulletin de la SFPT, No. 153, Paris, Avril 1999.
- [Eberly et al., 1994] D. Eberly, R. Gardner, B. Morse, S. Pizer, C. Scarlach, "Ridges for Image Analysis", *Journal of Mathematical Imaging and Vision*, 4, pp. 353-373, 1994.
- [Figueiredo et al., 1997] M. A. T. Figueiredo, J. M. N. Leitao, A. K. Jain, "Adaptive Parametrically Deformable Contours", *Lecture Notes in Computer Science (1223)* M. Pelillo, E. R. Hancock (Eds.) : Energy Minimisation Methods in Computer Vision and Pattern Recognition, Springer, Venice, Italy, 1997.
- [Fischler et al., 1981] M. A. Fischler, J. M. Tenenbaum, H. C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique", *Computer Graphics and Image Processing*, vol. 15, pp. 201-223, 1981.
- [Fua et Leclerc, 1990] P. Fua, Y. G. Leclerc, "Model Driven Edge Detection", *Machine Vision and Applications*, 3, pp. 45-56, 1990.
- [Geman et Jedynek, 1996] D. Geman, B. Jedynek, "An Active Testing Model for Tracking Roads in Satellite Images", *IEEE Trans. on PAMI*, vol. 18, pp. 1-14, 1996.
- [Germain et Réfrégier, 1996] O. Germain, P. Réfrégier, "Optimal snake-based segmentation of a random luminance target on a spatially disjoint background", *Optics Letters*, 15(22), pp. 1845-1847, 1996.
- [Geyer et Møller, 1994] C. J. Geyer, J. Møller, "Simulation and likelihood inference for spatial point process", *Scandinavian Journal of Statistics, Series B*, 21, 359-373, 1994.

- 
- [Geyer et Thompson, 1992] C. J. Geyer, E. A. Thompson, "Constrained Monte Carlo Maximum Likelihood for Dependent Data", *Journal of the Royal Statistical Society, Series B*, 56, 657-699, 1992.
- [Geyer, 1999] C. J. Geyer, "Likelihood Inference for Spatial Point Processes". In O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M van Lieshout, editors, *Stochastic Geometry, Likelihood and Computation*, Chapman and Hall, London, 1999.
- [Giudici et Roberts, 1998] P. Giudici, G. Roberts, "On the Automatic Choice of Reversible Jumps". In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M Smith (Eds), *Bayesian Statistics*, 6, Oxford University Press, 1998.
- [Graffigne et Herlin, 1989] C. Graffigne, I. Herlin, "Modélisation de réseaux pour l'Imagerie Satellite SPOT, 7eme Congrès de Reconnaissance des Formes et Intelligence Artificielle, Paris, pp. 833-842, 1989.
- [Green, 1995] P. Green, "Reversible Jump MCMC Computation and Bayesian Model Determination", *Biometrika*, 82, pp. 711-732, 1995.
- [Green, 1996] P. Green, "MCMC in image analysis". In W. R. Gilks, S. Richardson and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pp. 381-399, Chapman and Hall, London, 1996.
- [Green et Mira, 1999] P. Green, A. Mira, "Delayed Rejection in Reversible Jump Metropolis-Hastings", *Mathematics Research Report* No. S-01-99, University of Bristol, 1999. Submitted to *Biometrika*.
- [Guerin, 1996] P. Guerin, "Apport des Cartes Topographiques pour l'Analyse de Scène en Imagerie Aérienne : Application à la Détection du Réseau Routier", Thèse de doctorat, Université de Marne-la-Vallée, 1996.
- [Häggström et al., 1999] O. Häggström, M.N.M. van Lieshout, J. Møller, "Characterization Results and Markov Chain Monte Carlo Algorithms Including Exact Simulation for Some Spatial Point Processes", *Bernoulli*, 5, pp. 641-659, 1999.
- [Hansen et al., 2000] M.B. Hansen, J. Møller, F.Aa. Tøgersen, "Bayesian contour detection in a time series of ultrasound images through dynamic deformable template models", *Research Report R-00-2007*, Department of Mathematical Sciences, Aalborg University, 2000. Submitted.
- [Haralick, 1983] R. Haralick, "Ridges and Valleys on Digital Images", *Computer Vision, Graphics, and Image Processing* 22, pp. 28-38, 1983.
- [Imberty et Descombes, 2000] M. Imberty, X. Descombes "Simulation de processus objets : Etude de faisabilité pour une application à la segmentation" *Rapport de Recherche INRIA* No.3881, Février 2000.

- [Jedynak, 1995] B. Jedynak, "Modèles Stochastiques et Méthodes Déterministes pour Extraire les Routes des Images de la Terre Vues du Ciel", Thèse de doctorat, Université de Paris-Sud, 1995.
- [Kendall et Møller, 1999] W. S. Kendall, J. Møller, "Perfect Metropolis-Hastings Simulation of Locally Stable Point Processes", *Research Report No. 347, Department of Statistics, University of Warwick* 1999.
- [Kendall et Thonnes, 1997] W. S. Kendall, E. Thonnes, "Perfect Simulation in Stochastic Geometry", *Research Report No. 323, Department of Statistics, University of Warwick*, 1997.
- [Lieshout, 1994] M. N. M. van Lieshout, "Stochastic annealing for nearest-neighbour point process with application to object recognition", *Advances in Applied Probability*, 26, pp. 281-300, 1994.
- [Lieshout et al. (eds.),1999] O. E. Barndoff-Nielsen, W. S. Kendall and M. N. M van Lieshout, editors, *Stochastic Geometry, Likelihood and Computation*, Chapman and Hall, London, 1999.
- [Lieshout, 2000] M. N. M. van Lieshout, "Markov Point Process and Their Applications", Imperial College Press, London, 2000.
- [Lorette et al, 2000] A. Lorette, X. Descombes, J. Zerubia. "Texture Analysis through a Markovian Modelling and Fuzzy Classification : Application to Urban Area Extraction from Satellite Images" *International Journal of Computer Vision* Vol. 36, no. 3, pp. 219-234, 2000.
- [Lusternik et Sobolev, 1989] L. Lusternik, V. Sobolev, "Précis d'Analyse Fonctionnelle", Editions Mir Moscou, 1989.
- [Mallat, 1989] S. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation", *IEEE Trans. on PAMI*, 11, pp. 674-693, Juillet 1989.
- [Mangin, 1994] F. Mangin "Amélioration de la Détection de Contours en Imagerie Artificielle par un Modèle Coopératif Multi-résolution", Thèse de Doctorat, Université de Nice Sophia-Antipolis, 1994.
- [Merlet et Zerubia, 1996] N. Merlet, J. Zerubia, "New Prospects in Line Detection by Dynamic Programming", *IEEE Trans. on PAMI* vol. 18, pp. 426-431, Avril 1996.
- [Merlet et Zerubia, 1997] N. Merlet, J. Zerubia, "Integration of Global Information for Roads Detection in Satellite Images", *Rapport de Recherche INRIA* No. 3239, Août 1997.
- [Meyn et Tweedie, 1993] S. P. Meyn, R. L. Tweedie, "Markov Chains and Stochastic Stability", Springer-Verlag, London, 1993.



- 
- [Møller,1999] J. Møller, "Likelihood Inference for Spatial Point Processes". In O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M van Lieshout, editors, *Stochastic Geometry, Likelihood and Computation*, Chapman and Hall, London, 1999.
- [Moyeed et Baddeley, 1991] R. A. Moyeed, A. J. Baddeley, "Stochastic Approximation of the MLE for a Spatial Point Pattern", *Scandinavian Journal of Statistics*, vol. 18, pp. 39-50, 1991.
- [Neuenschwander et al., 1997] W. M. Neuenschwander, P. Fua, L. Iverson, G. Szekely, O. Kubler, "Ziplock Snakes", *International Journal of Computer Vision*, 25(3), pp. 191-201, 1997.
- [Pievatolo et Green, 1998] A. Pievatolo, P. J. Green, "Boundary Detection Through Dynamic Polygons", *Journal of the Royal Statistical Society, Series B*,60, pp. 609-626,1998.
- [Poli et Vali, 1997] R. Poli, G. Valli, "An algorithm for real-time vessel enhancement and detection", *Computer Methods and Programs in Biomedicine*, 52, pp. 1-22, 1997
- [Propp et Wilson, 1996] J. G. Propp, D. B. Wilson "Exact Sampling with Coupled Markov Chains and applications to Statistical Mechanics" *Random Structures and Algorithms*, 9, pp. 223-252,1996.
- [Rellier et al., 2000] G. Rellier, X. Descombes, J. Zerubia, "Deformation of a Cartographic Road Network on a SPOT Satellite Image", *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 736-739, Vancouver, Canada, Septembre 2000.
- [Réfrégier et al., 1997 (a)] P. Réfrégier, O. Germain, T. Gaidon "Optimal snake segmentation of target and background with independent Gamma density probabilities application to speckled and preprocessed images", *Optics Communications*, 137, pp. 382-388, 1997.
- [Réfrégier et al., 1997 (b)] P. Réfrégier, O. Germain, C. Chesnaud, T. Gaidon, "Segmentation par contour actif probabiliste", *GRETSI - Grenoble*, pp. 1459-1462, 1997.
- [Ripley, 1977] B. D. Ripley, "Modelling Spatial Patterns", *Journal of the Royal Statistical Institute, Series B*, 39, pp. 172-212, 1977.
- [Ripley et Kelly, 1977] B. D. Ripley, F. P. Kelly, "Markov Point Processes", *J. London Math. Soc.*, 15, 188-192, 1977.
- [Ripley, 1979] B. D. Ripley, "Simulating Spatial Patterns : Dependent Sample from a Multivariate Density", *Journal of the Royal Statistical Society, Series C*, 28, pp. 109-112, 1979.
- [Robert, 1996] C. Robert, "Méthodes de Monte Carlo par Chaînes de Markov", Ed. Economica, 1996.

- [Robert et Casella, 1999] C. Robert, G. Casella, "Monte Carlo Statistical Methods", Springer-Verlag, 1999.
- [Ronfard, 1994] R. Ronfard, "Region-Based Strategies for Active Contour Models", *International Journal of Computer Vision*, 13(2), pp. 229-251, 1994.
- [Roux, 1992] M. Roux, "Recalage d'Images Multi-Sources. Application au Recalage d'une Image SPOT et d'une Carte", Thèse de doctorat, ENST, 1992.
- [Rue et Husby, 1998] H. Rue, O. K. Husby, "Identification of Partly Destroyed Objects using Dynamic Polygons" *Stat. Comp.*, 8, pp. 221-228, 1998.
- [Rue et Syverseen, 1998] H. Rue, A. R. Syverseen, "Bayesian Object Recognition with Baddeley's Delta Loss" *Adv. Appl. Prob.*, 30, pp. 64-84, 1998.
- [Rue et Hurn, 1999] H. Rue, M. Hurn " Bayesian Object Identification ", *Biometrika*, 3, pp. 649-660, 1999.
- [Ruelle, 1970] D. Ruelle, "Superstable Interactions in Classical Statistical Mechanics", *Commun. math. Phys.*, Springer-Verlag, 18, pp. 127-159, 1970.
- [Serendero, 1989] M.A. Serendero "Extraction d'Informations Symboliques en Imagerie SPOT : Réseaux de Communication et Agglomerations", Thèse de Doctorat, Université de Nice-Sophia Antipolis 1989.
- [Stoica et al., 2000a] R. Stoica, X. Descombes, J. Zerubia, "A Markov Point Process for Road Extraction in Remote Sensed Images", *Rapport de Recherche INRIA RR-3929*, Avril 2000.
- [Stoica et al., 2000b] R. Stoica, X. Descombes, J. Zerubia, "Road Extraction in Remote Sensed Images using a Stochastic Geometry Framework", *Proc. of International Workshop on Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, France, Juillet, 2000.
- [Stoica et al., 2000c] R. Stoica, X. Descombes, J. Zerubia, "A Gibbs Point Process for Road Extraction in Remotely Sensed Images", Soumis à *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Octobre 2000.
- [Storvik, 1994] G. Storvik, "A Bayesian Approach to Dynamic Contours through Stochastic Sampling and Simulated Annealing", *IEEE Trans. on PAMI* vol. 10, pp. 976-986, October 1994.
- [Stoyan et al., 1987] D. Stoyan, W. S. Kendall, J. Mecke, "Stochastic Geometry and its Applications", John Willey and Sons, 1987.
- [Stoyan et Stoyan, 1998] D. Stoyan, H. Stoyan, "Non-Homogeneous Gibbs Process Models for Forestry - A Case Study", *Biometrical Journal* 40, 5, pp. 521-531, 1998.

- [Tierney, 1994] L. Tierney, “Markov Chains for Exploring Posterior Distributions”, *Tech. Report. No. 560*, University of Minnesota, March, 1994.
- [Tupin, 1997] F. Tupin, “Reconnaissance de formes et analyse de scènes en imagerie radar à ouverture synthétique”, Thèse de doctorat, ENST, 1997.
- [Urago, 1996] S. Urago, “Groupements de contours en utilisant des modélisations markoviennes”, Thèse de doctorat, Université de Nice Sophia-Antipolis, 1996.
- [Wolfer et al., 1994] J. Wolfer, J. Robergé, T. Grace, “Robust Multispectral Road Classification in Landsat Thematic Mapper Imagery”, Proceedings of the World Congress on Neural Networks, San Diego, pp. 260-268, 1994.
- [Wu, 1987] Y. Wu, “Application de la programmation dynamique au recalage d’images”, Thèse de doctorat, ENST, 1987.
- [Younes, 1988] L. Younes, “Estimation and Annealing for Gibbsian Fields”, *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, vol. 24, pp. 269-294, 1988.
- [Yu et Mykland, 1998] B. Yu, P. Mykland, “Looking at Markov Samplers through CUSUM Path Plots : a Simple Diagnostic Idea”, *Statistics and Computing*, 8(3), pp. 275-286, 1998.



# Liste de Publications

## Journaux

- X. Descombes, R. Stoica, L. Garcin, J. Zerubia. A RJMCMC Algorithm for Object Processes. *Monte Carlo Methods and Applications*, A paraître en 2001.
- R. Stoica, X. Descombes, J. Zerubia. A Gibbs Point Process for Road Extraction in Remotely Sensed Images. Soumis à *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Octobre 2000.

## Conférences

- R. Stoica, X. Descombes, J. Zerubia. Road Extraction in Remote Sensed Images using a Stochastic Geometry Framework. *Proc. of International Workshop on Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, France, Juillet, 2000.
- X. Descombes, R. Stoica, J. Zerubia. Two Markov Point Processes for Simulating Line Networks. *Invited paper : IEEE International Conference on Image Processing* Kobe, Japan, 1999.

## Rapports de recherche

- R. Stoica, X. Descombes, J. Zerubia. A Markov Point Process for Road Extraction in Remote Sensed Images. *Rapport de Recherche INRIA RR-3929*, Avril 2000.

## Colloques avec résumé (sans actes)

- R. Stoica, X. Descombes, J. Zerubia. A Line Markov Point Process for Thin Network Extraction. *International Conference on Spatial Statistics in the Agro-, Bio-, and Geosciences* Organisateurs : J. Chadœuf, J. Møller, A. Penttinen et D. Stoyan, Freiberg, Allemagne, Juillet 2000.

# Voilà les gens !

## Thèse

Josiane, Xavier, Marc, Christophe, M. Schmitt, H. Le Men, M. Barlaud, M. N. M. van Lieshout, N. Baghdadi, Corinne, Guillaume, Oscar, Dédé, Caro, Seb, Moustapha, Fred, Laure, Anne, Laurent, Max Rzepka, Yossi, Gyorgy, Sandrine, Françoise, Marie-Laure, Nathalie, Valerie, Odile ... et tous les gens de l'INRIA qui disent bonjour.

## Antithèse

Irina, Barbu, Andrei, Olga, Mircea, Hugues, Martine, Guy, Vio, Remi, Henri, Jean, Pilou, Faustine, Patrick, Paul, Guillaume, Fred, Raph, Capucine, Jean, Anton, Isa, Thierry, Didier, Sylvie, Valoche, Laurent, Motorhead, Anthony, Faabb, Roger, Weber, Jeudi, Le Grand Blond, Cedric, Nath, Lena, Corinne, La Fraternité, Liana, Horia, Aura, Rodica, Gil, Ioana, Adrian, Mihai, Matei, Elena, Alain, Jules, Magali, Gilles, Hamou, Tania, Boris, Isa, Sophie, Theresa, Liz, Dan, Catherine et Patator ... et tous les gens avec qui j'ai fait le marché le dimanche.

## Synthèse

Hugues et Paul ...

... mais, mais c'est quand la Saint Thèse ?

- Demain.







---

## RÉSUMÉ

---

Les réseaux routiers, ou les réseaux hydrographiques, les vaisseaux sanguins ou bien les fissures dans les matériaux sont connus dans la communauté du traitement d'image sous le nom générique de réseaux linéiques. La théorie des processus ponctuels marqués est un cadre mathématique rigoureux qui donne la possibilité de modéliser l'image comme un ensemble d'objets en interaction.

Les deux idées principales qui ont motivé ce travail sont : ces réseaux sont approchés par de segments de droite connectés, et les réseaux linéiques dans une image sont la réalisation d'un processus ponctuel de Gibbs.

Le processus ponctuel qui modélise les réseaux comporte deux composantes. Le premier terme ("Candy" modèle) gère les états et les interactions entre segments : densité, connectivité, alignement et répulsion des segments. L'emplacement du réseau dans l'image est trouvé grâce au second terme, le terme d'attache aux données. Cette composante du modèle est construite à partir de tests d'hypothèses.

L'estimateur des réseaux dans l'image est donné par le minimum d'une fonction d'énergie de Gibbs. Pour trouver l'optimum global de cette fonction, nous mettons en œuvre un algorithme de type recuit simulé qui s'appuie, sur une dynamique de type Monte Carlo par Chaînes de Markov (MCMC) à sauts réversibles. Des résultats sont présentes sur des images aériennes, SPOT et RADAR (RSO).

Nous abordons ensuite deux de problèmes ouverts liés au "Candy" modèle, mais d'un intérêt théorique général : la convergence d'une dynamique de Monte Carlo à sauts réversibles, et l'estimation des paramètres des processus ponctuels. Une solution à ces problèmes pourrait ouvrir une nouvelle direction dans la recherche de méthodes non-supervisées en traitement d'image.

**Mots clefs :** réseaux linéiques, extraction de routes, géométrie stochastique, processus ponctuel marqué, processus ponctuel de Gibbs, "Candy" modèle, tests d'hypothèses, dynamique MCMC à sauts réversibles, estimation de paramètres, images aériennes et satellitaires

---

## ABSTRACT

---

Road or hydrographical networks, blood vessels or fissures in materials are all known by the image processing community under the general name of line networks. The theory of point processes is a rigorous mathematical framework which allows us to model an image as a set of interacting objects.

The two main ideas which are the basis of this work are : these networks can be considered as connected segments, and the line networks in an image are the realization of a Gibbs point process.

The point process used to model the networks has two components. The first one (Candy model) deals with the states and the interaction of the segments : density, connectivity, alignment, attraction and rejection. The location of the network is determined by the second component, the data term. This component is based on hypothesis tests. The network estimator is given by the minimum of a Gibbs energy. We build a simulated annealing algorithm in order to avoid local minima. This algorithm uses reversible jump Monte Carlo Markov Chain (RJMC) dynamics. Results are shown on aerial, SPOT and RADAR (SAR) images.

Finally, we start a study on two open problems related to the Candy model, but of general theoretical interest : the convergence of a RJMC dynamics, and parameter estimation related to point processes. A solution to these problems would give a new direction for the research of unsupervised methods in image processing.

**Keywords :** line networks, road extraction, stochastic geometry, marked point process, Gibbs point process, Candy model, hypothesis tests, reversible jump MCMC dynamics, parameter estimation, aerial and satellite images