



HAL
open science

Méthodes de Monte Carlo en Vision Stéréoscopique

Julien Senegas

► **To cite this version:**

Julien Senegas. Méthodes de Monte Carlo en Vision Stéréoscopique. Traitement du signal et de l'image [eess.SP]. École Nationale Supérieure des Mines de Paris, 2002. Français. NNT: . tel-00005637

HAL Id: tel-00005637

<https://pastel.hal.science/tel-00005637>

Submitted on 5 Apr 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité : Géostatistique

présentée et soutenue publiquement par

Julien SENEGAS

le vendredi 13 Septembre 2002

**METHODES DE MONTE CARLO EN
VISION STEREOSCOPIQUE
APPLICATION A L'ETUDE DE MODELES
NUMERIQUES DE TERRAIN**

Jury :

Xavier Guyon	Rapporteur
Hervé Le Men	Examinateur
Philippe Nonin	Directeur de thèse
Michel Schmitt	Directeur de thèse
Georges Stamon	Président
Josiane Zerubia	Rapporteur



THÈSE

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité : Géostatistique

présentée et soutenue publiquement par

Julien SENEGAS

le vendredi 13 Septembre 2002

**METHODES DE MONTE CARLO EN
VISION STEREOSCOPIQUE
APPLICATION A L'ETUDE DE MODELES
NUMERIQUES DE TERRAIN**

Jury :

Xavier Guyon	Rapporteur
Hervé Le Men	Examineur
Philippe Nonin	Directeur de thèse
Michel Schmitt	Directeur de thèse
Georges Stamon	Président
Josiane Zerubia	Rapporteur

*quid videat, nescit; sed quod videt, uritur illo,
atque oculos idem, qui decipit, incitat error.
credula, quid frustra simulacra fugacia captas?
quod petis, est nusquam; quod amas, avertere, perdes!
ista re percussae, quam cernis, imaginis umbra est:
nil habet ista sui; tecum venitque manetque;
tecum discedet, si tu discedere possis!*

Que voit-il donc? Il l'ignore; mais ce qu'il voit l'embrase, et la même erreur qui abuse ses yeux excite leur convoitise. Crédule enfant, à quoi bon ces vains efforts pour saisir une fugitive apparence? L'objet de ton désir n'existe pas! Celui de ton amour, détourne-toi, et tu le feras disparaître. Cette ombre que tu vois, c'est le reflet de ton image. Elle n'est rien par elle-même, c'est avec toi qu'elle est apparue, qu'elle persiste, et ton départ la dissiperait, si tu avais le courage de partir!

Ovide, *Les Métamorphoses*, III/430-436

Je tiens tout d'abord à exprimer mes remerciements et ma sympathie à l'ensemble du jury.

Georges Stamon a été le témoin tout aussi bien du commencement, en tant que rapporteur du projet de thèse, que de la fin, en tant que président du jury. Je le remercie de sa gentillesse et de sa bienveillance.

Josiane Zerubia et Xavier Guyon, par leurs commentaires précis et critiques, ont grandement contribué à l'amélioration du manuscrit. Je tiens à leur dire combien j'ai apprécié les discussions que j'ai eues avec eux ainsi que la disponibilité dont ils ont fait preuve. Je remercie tout particulièrement Josiane Zerubia de m'avoir reçu à l'INRIA de Sophia Antipolis et invité à participer à son séminaire.

Par deux fois, Hervé Le Men a accepté de venir assister à une présentation de mes travaux. Je lui suis reconnaissant de cet intérêt et des remarques constructives qu'il m'a formulées.

J'en profite également pour saluer mes deux directeurs de thèse, Philippe Nonin et Michel Schmitt, à la fois juges et parties. Ils ont bien voulu me laisser toute liberté pour conduire mon travail tout en m'apportant aux moments opportuns leur appui.

Sans le soutien, scientifique, technique et financier d'Istar, cette thèse n'aurait pas eu lieu. Je remercie tout le personnel qui m'a aidé et assisté, notamment à travers le fondateur de la société, Laurent Renouard.

La majeure partie de ce travail a été réalisée au Centre de Géostatistique de l'École des Mines de Paris. C'est peu dire que j'ai bénéficié des moyens mis à ma disposition et des conseils de chacun. Au-delà des rapports purement professionnels, j'aimerais dire combien j'ai pris plaisir à passer ces trois années avec tous les membres du centre. Je profite également de ces quelques lignes pour renouveler mes remerciements et toute mon estime à Christian Lajaunie et Christian Lantuéjoule. Je ne sais combien de fois j'ai frappé à leur porte pour leur soumettre des questions mal formulées, mais je sais qu'à chaque fois cette porte s'est ouverte et que mes pensées s'en sont trouvées grandement éclaircies.

Ces trois années, je les ai partagées avec beaucoup d'autres. La chance a bien voulu qu'un beau matin de septembre, Sarah et Marta viennent partager le bureau dans lequel je travaillais. J'ai été à coup sûr le grand gagnant de cette affaire! Simon a été pour moi un compagnon et une inspiration, je lui dois beaucoup. Une pensée chaleureuse pour Eric! Salut à tous les autres que j'ai croisés : Nicolas, Laurent, David, Omar, Christophe, Thomas, Sergio, Valery, Anne, Nicolas (un autre), et ceux que j'oublie certainement...

Table des matières

1	Introduction	1
1.1	Problématique générale	2
1.2	Le point de vue bayésien	3
1.3	Vers les méthodes de Monte Carlo	4
1.4	Sur le choix du modèle	5
1.5	Apports et organisation générale	5
2	Préliminaires	7
2.1	Vision stéréoscopique	7
2.1.1	Introduction : cas du modèle projectif linéaire parallèle . .	8
2.1.2	Géométrie épipolaire et rectification	10
2.1.3	Le cas des capteurs <i>pushbroom</i>	13
2.1.4	Définition de la disparité	17
2.1.5	La mise en correspondance : calcul de la disparité	20
2.1.6	Reconstitution	25
2.2	Etude de l'incertitude associée à un MNT : survol bibliographique	27
2.2.1	Validation <i>a posteriori</i> d'un MNT	29
2.2.2	Etude de la mise en correspondance	32
3	Modélisation bayésienne	39
3.1	Un problème d'échantillonnage	39
3.1.1	Loi <i>a posteriori</i> , relation de Bayes	42
3.1.2	Echantillonnage et méthodes de Monte Carlo	43

3.1.3	Bilan	44
3.2	Choix du modèle	45
3.2.1	Modèle <i>a priori</i>	45
3.2.2	Modèle de vraisemblance	47
3.2.3	Sur l'estimation des paramètres	51
3.3	Randomisation des paramètres, formulation bayésienne	53
3.3.1	Paramètres du modèle <i>a priori</i>	54
3.3.2	Paramètres du terme de vraisemblance	56
3.3.3	Alternative : l'algorithme EM stochastique	58
3.4	Mise en œuvre	59
3.4.1	Modèle <i>a priori</i>	60
3.4.2	Modèles de vraisemblance	64
3.4.3	Estimation par l'algorithme EM stochastique	69
3.5	Conclusion	72
4	Simulations spatiales	75
4.1	Simulations par chaînes de Markov	76
4.1.1	Définition	76
4.1.2	Conditions d'ergodicité	77
4.1.3	Méthodes de Monte Carlo	78
4.1.4	Sur l'échantillonnage d'importance	80
4.2	Algorithmes de simulation	83
4.2.1	Echantillonneur de Gibbs	84
4.2.2	Algorithmes du type Metropolis-Hastings	87
4.2.3	Metropolis-Hastings multiple	95
4.3	Illustration	103
5	Cas d'étude	115
5.1	Présentation des outils statistiques	115
5.2	Extrait de zone urbaine	117
5.2.1	Estimation de la disparité	120

5.2.2	Probabilité de dépassement de seuil	124
5.2.3	Intervalle de confiance	132
5.2.4	Bilan	134
5.3	Extrait de zone montagneuse	138
5.3.1	Probabilité de dépassement de seuil	140
5.3.2	Quelques remarques	143
6	Une alternative	145
6.1	Retour sur le modèle <i>a priori</i>	145
6.1.1	Quelle hypothèse de stationnarité?	145
6.1.2	Vers une solution	147
6.2	Illustration	148
6.2.1	Probabilité de dépassement de seuil	149
6.2.2	Sensibilité aux paramètres	150
6.2.3	Performances relatives	154
6.3	Exemples d'application	155
6.3.1	Altitude de vol	155
6.3.2	Zone de visibilité d'un point	156
6.4	Bilan	156
7	Conclusion générale et perspectives	163
A	Simulation conditionnelle d'un vecteur gaussien	169
B	Gibbs par blocs dans le cas gaussien	171
C	Modèles de Gibbs et champs de Markov	173
C.1	Modèles de Gibbs	173
C.2	Lois conditionnelles	174
C.3	Champs de Markov	174
C.4	Champs de Markov gaussiens	175
D	Le cas gaussien dans Metropolis-Hastings multiple	177

Notations

Vision stéréoscopique	
I	image
O, Π	centre optique et plan image d'une caméra
\mathcal{R}	repère
M, N	points physiques
m, n	points images
(x, y, z)	coordonnées d'un point physique
(u, v)	coordonnées d'un point image
f, b	distance focale, distance entre centres optiques
F, P, Q, R, T	matrices
D, d	disparité
$C_{u,v}(d)$	critère de corrélation au pixel (u, v) pour une disparité d
Probabilité, statistiques	
$(E, \mathcal{E}, \mu), (F, \mathcal{F}, \nu)$	espaces probabilisés
$P(A)$	probabilité de l'événement A
$\pi, \pi(z)$	loi de probabilité, sa densité
$\pi_Z(\cdot y), \pi_Z(z y)$	loi conditionnelle de Z à $Y = y$, sa densité
$L(y z)$	vraisemblance de y sachant z
K, q	noyaux de transition d'une chaîne de Markov
$E_\pi(g)$	Espérance sous la loi π de la fonction g
Var, Cov	variance, covariance
W, Y, Z, η, τ	fonctions aléatoires
W, Y, Z	vecteurs aléatoires
w, y, z	réalisation des vecteurs aléatoires W, Y, Z
C, γ	fonction de covariance, variogramme
C, Q	matrice de covariance, matrice de covariance inverse
m, σ	moyenne, écart-type
S_l	moyenne arithmétique à l termes

Lois	
$\mathcal{N}(m, \sigma), \phi(z)$	loi normale de moyenne m et d'écart-type σ , sa densité
$\Gamma(\alpha, \beta)$	loi Gamma d'indice α et de paramètre β
$D(\gamma_1, \dots, \gamma_l)$	loi de Dirichlet de paramètres $\gamma_1, \dots, \gamma_l$
Divers	
$\alpha, \beta, \gamma, \delta, \epsilon, \kappa, \theta, \mu, \rho$	paramètres
T, S	ensembles de sites
g, r, ψ	fonctions réelles
Symboles	
1_A	fonction indicatrice de l'ensemble A
δ_A	mesure de Dirac relative à l'ensemble A
$\nabla, \frac{\partial}{\partial u}$	gradient, opérateur de dérivé partielle
∂A	voisinage de l'ensemble A
\propto	proportionnel à
$\sim \pi$	de loi π
\star	produit de convolution
I_n	matrice identité de taille n
C^\top	transposé de la matrice C
\hat{d}	estimation de la variable d
\mathbb{R}	ensemble des réels

Chapitre 1

Introduction

Le mois de mai 2002 a vu la mise sur orbite du satellite SPOT 5 (Satellite Pour l'Observation de la Terre) avec pour objectif "de constituer en cinq ans un modèle numérique de terrain mondial de 30 à 50 millions de km² correspondant au tiers des terres émergées"¹. Un Modèle Numérique de Terrain (MNT) est une carte où la variable exprimée est l'altitude au point considéré. Si un point de la surface est représenté par ses coordonnées $(x, y, z)^\top$, il s'agit donc dans une description mathématique de l'application $(x, y) \rightarrow z(x, y)$. Les utilisateurs de ces bases de données cartographiques sont multiples : l'armée est un des acteurs historiques de ce marché, mais il faut également citer l'aviation civile, les collectivités locales qui désirent gérer l'aménagement de leur territoire, et un certain nombre d'activités économiques de haute technologie comme les télécommunications qui ont besoin de données récentes pour optimiser la position de leurs relais de téléphonie mobile, les systèmes de navigation embarqués...

L'imagerie stéréoscopique constitue avec le radar une des principales technologies permettant de recueillir des données utiles à la reconstitution 3D et de calculer automatiquement des MNTs. Le principe général, utilisé en photogrammétrie et en vision par ordinateur, est relativement simple et s'appuie sur une analogie avec le système visuel humain : à partir de deux vues "planes" d'une scène, il est possible de reconstituer sous certaines conditions une information tri-dimensionnelle. Dès lors, en utilisant des images satellitaires ou aériennes recouvrant de grandes zones géographiques (120 km par 60 km avec une résolution de 5 m pour SPOT 5), des modèles numériques de terrain peuvent être produits de manière automatique et en un temps relativement court. Dans des pays comme la France ou les Etats-Unis, l'intérêt des données issues de l'imagerie satellitaire ou aérienne est principalement d'actualiser les bases de données géographiques déjà existantes, que l'on s'intéresse par exemple à l'évolution des milieux urbains ou à la gestion de l'environnement ; mais en Asie ou en Amérique du Sud, bien souvent, aucune donnée n'est disponible

¹Jean-Marc Nasr, PDG de SPOT Image, Libération, 3 mai 2002

et ces nouvelles techniques sont donc essentielles pour constituer en un temps limité un catalogue important.

1.1 Problématique générale

ISTAR (Imagerie STéréoscopique Appliquée au Relief), société française créée en 1988 par des chercheurs issus de l'INRIA ², est un des principaux acteurs mondiaux sur le marché des bases de données géographiques numériques. Elle utilise actuellement principalement deux types de données : des images obtenues par les capteurs embarqués sur les satellites SPOT ou LANDSAT et des images aériennes haute résolution produites par une caméra spécialement développée par le centre d'études spatiales allemand, le DLR. A partir de ces images sont produits de manière quasiment automatique des MNTs, à différentes résolutions selon le capteur utilisé, et des orthoimages (images corrigées des déformations induites par le relief).

Le travail présenté ici répond à une problématique soumise par ISTAR : *comment chiffrer l'incertitude associée à un modèle numérique de terrain calculé à partir d'un couple stéréoscopique donné?* Cette question est en effet primordiale dès que le MNT est utilisé à des fins industrielles ou stratégiques : il faut pouvoir assurer la fiabilité d'un système ou la validité d'une décision en tenant compte des erreurs possibles. Notons que ce problème nécessite également de définir les outils appropriés pour décrire l'incertitude. Il s'agit par exemple de calculer un intervalle de confiance de l'altitude du terrain à un niveau de risque fixé.

La difficulté principale de l'étude réside dans l'absence d'information autre que le couple stéréoscopique ; en particulier, on renonce à recourir à des mesures directes de l'altitude, car dans de nombreux cas, celles-ci ne sont pas disponibles. A partir d'un même couple stéréoscopique, dans notre cas d'images SPOT, le problème est donc de calculer le MNT correspondant et d'en estimer la précision.

Le problème de la vision stéréoscopique a suscité un grand nombre de travaux, qui ont abouti au développement d'un nombre tout aussi grand d'algorithmes. De manière générale, la résolution du problème est scindée en trois étapes distinctes : la calibration, qui consiste à calculer les paramètres géométriques du système stéréoscopique, la mise en correspondance, dont l'objet est d'identifier les couples de points images provenant d'un même point physique, puis la triangulation spatiale, à savoir le calcul en lui-même des coordonnées spatiales des points de la scène observée. Calibration et triangulation reposent essentiellement sur des relations géométriques dépendant de la nature du capteur utilisé, et sont actuellement bien maîtrisées. En revanche, la mise en correspondance est toujours un domaine actif de recherche. C'est à cette étape que nous allons nous intéresser par la suite. De fait, actuellement, aucun des algorithmes de calcul n'est en mesure de donner

²voir la thèse du fondateur de la société [90]

une estimation précise de la fiabilité de la mise en correspondance. En réalité, ce problème est rarement abordé en tant que tel : on se contente généralement de chiffrer la précision d'un algorithme à partir d'un jeu de données de référence.

1.2 Le point de vue bayésien

Revenons au problème de la mise en correspondance et considérons deux acquisitions d'une même scène. Un point physique, en l'absence d'occlusions, est projeté dans chacune des deux images : les deux projections forment un couple de points homologues. La mise en correspondance consiste alors à déterminer l'ensemble des couples de points homologues à partir des deux images recueillies. Il s'agit donc d'un problème inverse dont la résolution dépend fortement de la qualité de l'information disponible, à savoir la radiométrie des images du couple. Les erreurs d'appariement ne dépendent pas du système d'observation, mais du contenu du couple et donc de la nature de la scène observée ; elles ne sont pas systématiques. Elles ont en revanche un caractère aléatoire : non pas que le phénomène étudié en lui-même soit aléatoire (cela n'aurait que peu de sens d'avancer que l'élévation d'un terrain est le résultat d'une expérience aléatoire), mais parce que l'information utilisée est bruitée et par nature ambiguë. Il faut bien comprendre que dans le contexte de la reconstruction 3D, il existe une incertitude intrinsèque au couple stéréoscopique, une partie de l'information n'étant pas accessible. C'est cet aspect que nous proposons de traiter.

Adoptons un point de vue bayésien et reformulons le problème. On dispose d'une information de nature radiométrique, le couple stéréoscopique, notée Y par la suite. A partir d'une réalisation particulière y de ce couple, on désire reconstituer la géométrie Z de la scène observée. La connaissance que l'on peut avoir de Z à partir de y est entièrement décrite par la probabilité *a posteriori* de Z sachant $Y = y$. En d'autres termes, parmi toutes les configurations Z possibles, on s'intéresse à celles qui se réalisent lorsque y est observé. L'incertitude quant à l'estimation de Z à partir de Y est contenue dans cette probabilité *a posteriori*.

Soit un espace probabilisé (E, \mathcal{E}, μ) , dans lequel la variable aléatoire Z prend ses valeurs. Notons $\pi_Z(\cdot|y)$ la loi *a posteriori* de Z sachant $Y = y$. Si nous notons \hat{z} une estimation de la géométrie, calculée à partir d'un algorithme quelconque, il est possible de calculer par exemple la probabilité de dépassement d'un certain seuil s :

$$\begin{aligned} P(|Z - \hat{z}| \geq s) &= E_{\pi_Z(\cdot|y)}(1_{|Z - \hat{z}| \geq s}) \\ &= \int 1_{|z - \hat{z}| \geq s} \pi_Z(z|y) dz \end{aligned} \quad (1.1)$$

Inversement, si l'on se fixe un risque α , le problème peut être de déterminer l'in-

tervalle de confiance $[z_{\text{inf}}; z_{\text{sup}}]$ associé, i.e. tel que :

$$P(Z \in [z_{\text{inf}}; z_{\text{sup}}]) \geq 1 - \alpha \quad (1.2)$$

Nous venons donc de donner une formulation mathématique au problème initial. Le cadre est un cadre probabiliste, bayésien au sens où nous nous intéressons à la loi *a posteriori* $\pi_Z(\cdot|y)$ de la scène Z connaissant le couple stéréoscopique $Y = y$. On pourrait également utiliser le terme de loi conditionnelle. Ce cadre probabiliste nous permet de quantifier de manière simple l'incertitude relative à Z .

1.3 Vers les méthodes de Monte Carlo

Si la formulation du problème repose sur le choix de la distribution $\pi_Z(z|y)$, sa résolution implique le calcul d'intégrales multiples. Or, la dimension du problème est telle (la taille d'une scène SPOT est de l'ordre de $36 \cdot 10^6$ pixels) qu'un calcul direct ne peut être mis en œuvre. Une solution bien connue est celle apportée par les méthodes dites de Monte Carlo, largement utilisées autant en analyse numérique qu'en calcul de probabilité. Elles consistent à générer des réalisations de la loi $\pi_Z(\cdot|y)$ et à remplacer le calcul de l'intégrale par une moyenne arithmétique. Pour une loi π et une fonction g π -mesurable, on a en effet :

$$\begin{aligned} E_{\pi}(g) &= \int g(z)\pi(z)dz \\ &\approx \frac{1}{l} \sum_{k=1,l} g(z_i) \end{aligned} \quad (1.3)$$

où $z_i, i = 1, l$ sont des réalisations de la loi π .

Encore faut-il pouvoir simuler la loi désirée... Deux approches existent. Pour certaines lois, il existe un algorithme de simulation directe, par exemple par inversion de fonction de répartition ou par acceptation / rejet. C'est le cas notamment des modèles gaussiens, pour lesquels un grand nombre d'algorithmes fréquemment utilisés en géostatistique existent. En revanche, lorsque la loi ne peut se déduire d'une loi aisément simulable, ce qui est souvent le cas lorsque l'on considère des lois conditionnelles, il semble que seules des méthodes itératives permettent d'apporter une réponse. Il s'agit principalement de générer une chaîne de Markov admettant pour loi stationnaire la loi à simuler. Parmi les algorithmes les plus utilisés, citons en particulier l'échantillonneur de Gibbs [44] et l'algorithme de Metropolis-Hastings [78] [57].

En statistique, le problème de l'étude de lois de probabilité par simulations markoviennes et intégrales de Monte Carlo, couramment désigné par l'expression Markov

chain Monte Carlo (MCMC), est un des domaines de recherche les plus actifs, surtout depuis le début des années 1990 où ces méthodes ont fait leur apparition en analyse bayésienne. Ce qui apparaissait initialement comme un problème propre au domaine de la vision par ordinateur fait donc en réalité référence à une problématique beaucoup plus générale et nécessitant un important développement probabiliste.

1.4 Sur le choix du modèle

Nous sommes passés un peu vite sur la formulation du problème pour déjà nous intéresser à sa résolution. Or, une grande partie du travail porte sur le choix du modèle probabiliste $\pi_Z(\cdot|y)$ susceptible de décrire le paradigme de la vision stéréoscopique. Si l'approche proposée ici et l'utilisation des méthodes MCMC sont tout à fait générales, le modèle choisi est forcément spécifique et tourné vers l'application considérée. Notamment, le choix de travailler avec des images SPOT a une incidence sur la nature des scènes observées : à cette résolution, seules les constructions urbaines importantes et les variations sensibles du relief peuvent être distinguées.

Il est primordial de comprendre que l'utilisation d'un modèle est forcément subjective. Il est notamment toujours beaucoup plus aisé d'en réfuter que d'en mettre en évidence la validité. C'est pourquoi nous utilisons dans la suite l'expression *choix* du modèle, afin de signaler qu'il s'agit là d'un parti pris non négligeable. Cette subjectivité du modèle ne rend d'ailleurs que plus nécessaire les étapes de validation des résultats. Enfin, pour paraphraser l'adage matheronien *Il n'y a pas de probabilité en soi, il n'y a que des modèles probabilistes* [76], gardons à l'esprit que les calculs de probabilité qui sont l'objet de ce travail ne sont valables que dans le cadre du modèle choisi, et n'ont aucune signification propre.

1.5 Apports et organisation générale

Du travail qui est rapporté ici nous pouvons dégager deux apports originaux. Le premier est plutôt d'ordre conceptuel : il s'agit de la formulation mathématique du problème posé en termes d'échantillonnage d'une loi de probabilité. Cette approche, qui apparaît naturelle dès que l'on adopte une description bayésienne du problème, est cependant relativement nouvelle en vision par ordinateur et exige un changement de point de vue. Il y a en effet un pas important à franchir dans l'appréhension du réel entre une vision déterministe, où la solution est forcément unique, et une vision probabiliste, où c'est au contraire un ensemble de solutions que l'on envisage. Même si les méthodes bayésiennes sont employées couramment en vision par ordinateur, elles restent avant tout un outil de calcul servant à esti-

mer la géométrie (ou toute autre propriété) de la scène. Autrement dit, bien qu'on utilise la loi de probabilité pour formuler le problème, bien souvent n'est retenue qu'une caractéristique de celle-ci (sa moyenne ou son maximum). La multiplicité des solutions continue de gêner... Or c'est justement cette multiplicité qui est au cœur du problème qui nous intéresse.

Le second apport est plus théorique et concerne le développement d'algorithmes de simulation par chaînes de Markov. En particulier, nous proposons une extension de l'algorithme de Metropolis-Hastings en adoptant un point de vue multidimensionnel et développons plusieurs algorithmes dans un cadre gaussien. Certains aspects font également appel aux méthodes d'échantillonnage d'importance pour diminuer le temps de calcul. Ces algorithmes ont un intérêt tout à fait général dans le domaine des statistiques spatiales. Soulignons également que certaines parties de ce travail ont fait l'objet de publications [105, 106, 107, 108, 109, 110].

L'organisation de ce travail est la suivante. Au chapitre 2, nous rappelons les principes de la vision stéréoscopique, en précisant notamment la définition de la disparité utilisée par la suite. Dans un second temps, nous dressons un état de l'art des approches proposées pour quantifier l'incertitude en vision stéréoscopique. Au chapitre 3, nous montrons comment une analyse bayésienne du problème, nécessitant l'utilisation des méthodes MCMC, permet de donner une réponse à la problématique générale. Puis, nous décrivons différents modèles probabilistes et nous intéressons notamment à l'estimation des paramètres de ces modèles. L'étude des algorithmes de simulation est menée au chapitre 4 : nous introduisons rapidement le principe de la simulation par chaînes de Markov et le calcul d'intégrales de Monte Carlo, et décrivons divers algorithmes de simulation. Une étude comparative de ces différents algorithmes est menée. L'application de ces méthodes à l'étude de couples stéréoscopiques SPOT est présentée au chapitre 5. Enfin, nous examinons au chapitre 6 quelques pistes pour proposer une solution pratique et transposable dans le cadre de la production automatique de MNTs.

Chapitre 2

Préliminaires

2.1 Vision stéréoscopique

Nous rappelons ici brièvement les différentes étapes de la reconstitution géométrique d'une scène à partir d'un couple d'images stéréoscopiques. Nous considérons tout d'abord un modèle particulier de capteur : le modèle projectif linéaire. Puis, nous abordons l'étude des capteurs *pushbroom*, qui sont utilisés sur SPOT. L'introduction de la géométrie épipolaire permet alors de définir de manière générale la disparité, et parmi les méthodes de calcul, nous présentons les algorithmes reposant sur un critère de corrélation. Enfin, nous indiquons rapidement les principes de la reconstruction d'une scène à partir de la carte de disparité.

Un système stéréoscopique est composé de deux caméras observant une même scène depuis deux positions différentes (figure 2.1). Par référence au système optique humain, l'une est généralement appelée caméra gauche et l'autre caméra droite, mais nous préférons par la suite les dénominations 1 et 2. Chaque caméra est symbolisée par son centre optique O_j (éventuellement en mouvement, comme dans le cas des capteurs *pushbroom*) et son plan image Π_j . On attache à chaque caméra un repère $\mathcal{R}_j = (O_j, U_j, V_j, Z_j)$ tel que les axes U_j et V_j soient parallèles au plan Π_j et l'axe Z_j lui soit perpendiculaire. On appelle distance focale, notée f , la distance entre le centre optique et le plan image. On note m_j la projection d'un point physique M sur le plan Π_j . Les coordonnées de m_j dans \mathcal{R}_j sont $(u_j, v_j, f)^\top$. L'image I_j produite par la caméra j est constituée par l'ensemble des points image m_j . Dans la suite, un point image m_j est repéré par ses coordonnées (u_j, v_j) dans I_j . Enfin, le couple stéréoscopique est la réunion des deux images I_1 et I_2 .

Dans cette configuration et en l'absence d'occlusions, un point physique M se projette dans le plan image de chaque caméra. Les deux projections m_1 et m_2 forment un couple de points *homologues*. Connaissant un repère de l'espace

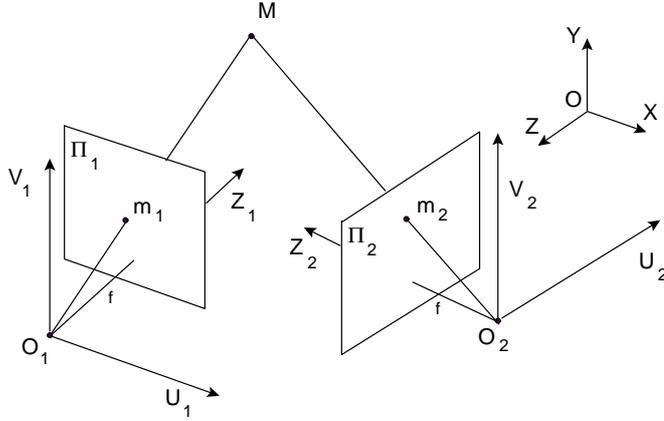


FIG. 2.1 – Vision stéréoscopique : un point physique M se projette dans le plan image Π_j , $j = 1, 2$, de chaque caméra ; les deux projections m_1 et m_2 forment un couple de points homologues.

$\mathcal{R} = (O, X, Y, Z)$, le problème qui se pose en vision stéréoscopique est de calculer les coordonnées $(x, y, z)^\top$ du point M dans \mathcal{R} à partir des coordonnées $(u_1, v_1)^\top$ et $(u_2, v_2)^\top$ de ses projections m_1 et m_2 . On procède généralement en deux étapes successives [34] :

- Connaissant le point m_1 de l'image I_1 , quel point m_2 de l'image I_2 lui correspond ? Il s'agit donc de déterminer l'ensemble des couples de points homologues. C'est l'objet de la *mise en correspondance*.
- Connaissant le couple de points homologues (m_1, m_2) et leurs coordonnées $((u_1, v_1)^\top, (u_2, v_2)^\top)$, quelles sont les coordonnées du point physique M correspondant dans le repère \mathcal{R} ? Il s'agit de l'étape de *triangulation*.

2.1.1 Introduction : cas du modèle projectif linéaire parallèle

Dans le modèle projectif linéaire ou sténopé, le centre optique O de la caméra est fixe. L'image d'un point physique M est définie par l'intersection entre le rayon (OM) et le plan Π . Les distorsions optiques sont négligées.

Considérons tout d'abord un système stéréoscopique particulièrement simple : il est constitué de deux caméras projectives linéaires dont les plans image Π_1 et Π_2 sont confondus, comme représenté à la figure 2.2. Il s'agit du modèle projectif linéaire parallèle. Les deux caméras ont même distance focale f . Les repères \mathcal{R}_1 et \mathcal{R}_2 ne diffèrent que par la position des deux origines O_1 et O_2 , les axes (O_1U_1) , (O_2U_2) , (O_1V_1) , (O_2V_2) et (O_1Z_1) , (O_2Z_2) étant parallèles deux à deux. Enfin, O_2 est situé sur l'axe (O_1U_1) , et l'on note b son abscisse, qui est également la distance

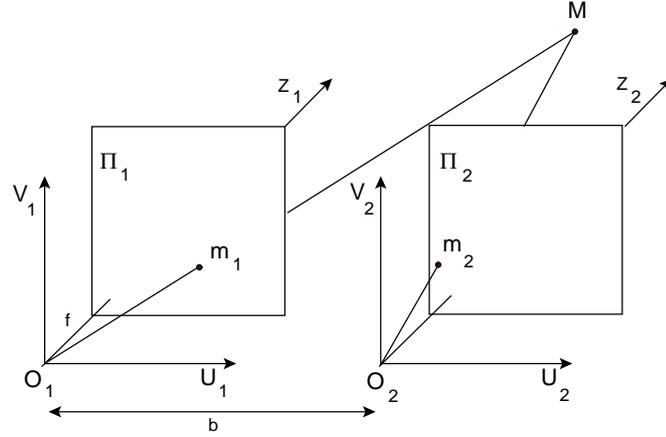


FIG. 2.2 – Système stéréoscopique avec deux caméras projectives linéaires parallèles.

entre les deux centres optiques (*baseline*).

On choisit le repère $\mathcal{R} = (O, X, Y, Z) = (O_1, U_1, V_1, Z_1)$ comme repère de la scène, et on considère un point physique M de coordonnées $(x, y, z)^\top$ dans \mathcal{R} . Ses projections dans les plans image Π_1 et Π_2 sont notées respectivement m_1 , de coordonnées $(u_1, v_1, f)^\top$ dans \mathcal{R}_1 , et m_2 , de coordonnées $(u_2, v_2, f)^\top$ dans \mathcal{R}_2 . La figure 2.3 représente une vue d'en haut de la figure 2.2. En appliquant le théorème de Thalès dans les triangles rectangles d'hypoténuse O_1M et O_2M respectivement (en rouge sur la figure), on a l'égalité entre les rapports :

$$\frac{f}{z} = \frac{u_1}{x} = \frac{u_2}{x - b} \quad (2.1)$$

On obtient par un raisonnement identique :

$$\frac{f}{z} = \frac{v_1}{y} = \frac{v_2}{y} \quad (2.2)$$

Il y a donc égalité entre les ordonnées : $v_1 = v_2$, et de la relation 2.1 on déduit :

$$\frac{f}{z} = \frac{u_1 - u_2}{b} \quad (2.3)$$

Si l'on définit la disparité d comme la différence entre les deux abscisses :

$$d = u_1 - u_2, \quad (2.4)$$

la relation précédente devient :

$$z = \frac{fb}{d} \quad (2.5)$$

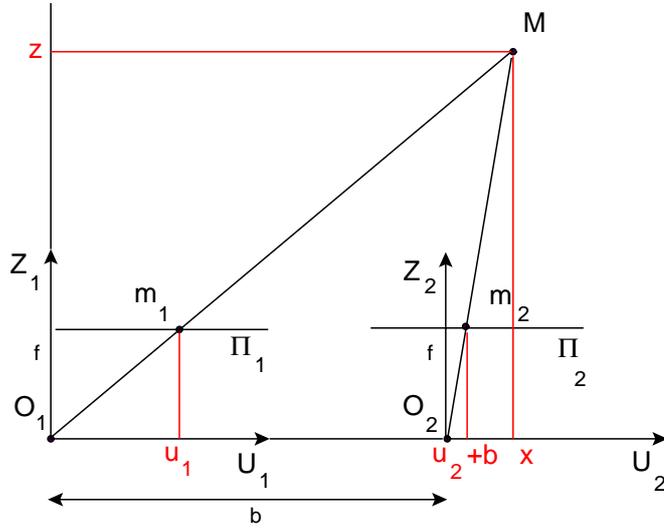


FIG. 2.3 – Vue d'en haut (selon l'axe (O_1V_1)) de la figure 2.2.

En remplaçant cette valeur dans l'équation 2.1, on obtient une expression de l'abscisse x :

$$x = \frac{u_1}{d}b \quad (2.6)$$

De même, l'ordonnée y s'écrit :

$$y = \frac{v_1}{d}b \quad (2.7)$$

Le système d'équations 2.5-2.7 montre que les coordonnées dans l'espace du point physique M s'expriment à partir des caractéristiques du système stéréoscopique (la distance focale f et la distance entre les deux centres optiques b), des coordonnées (u_1, v_1) et de la disparité d . Ce sont ces relations qui sont à la base du calcul des propriétés géométriques d'une scène à partir d'un couple d'images stéréoscopiques. Il apparaît notamment qu'une fois les paramètres de la prise de vue connus, les coordonnées spatiales d'un point de l'image $(u_1, v_1)^\top$ s'expriment uniquement en fonction de la disparité d . Ceci justifie le calcul de la carte de disparité, qui est abordé plus loin.

2.1.2 Géométrie épipolaire et rectification

Nous nous intéressons maintenant au modèle sténopé général, et supposons que la position et l'orientation des caméras est quelconque. Nous allons montrer que le lieu des points homologues à un point donné est une droite.

Considérons le système stéréoscopique de la figure 2.4. Le centre optique O_1 se projette dans le plan Π_2 au point E_2 : cette projection définit l'épipole de la

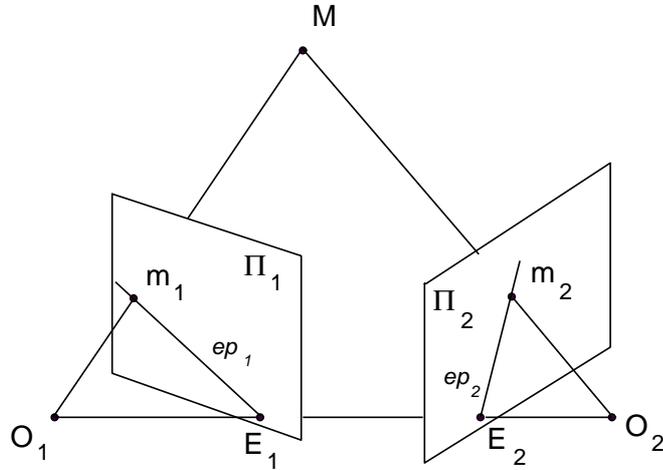


FIG. 2.4 – Géométrie épipolaire : E_1 et E_2 sont les épipoles, O_1MO_2 est le plan épipolaire défini par M , et ep_1 et ep_2 sont les droites épipolaires.

caméra 2 par rapport à la caméra 1. L'image de la droite (O_1M) dans Π_2 est la droite (E_2m_2) , notée ep_2 . On construit réciproquement l'épipole E_1 , projection de O_2 dans Π_1 , et la droite épipolaire ep_1 de m_2 dans Π_1 . Comme O_1 , m_1 et M sont alignés, c'est aussi l'image dans Π_2 de la droite (O_1m_1) , que l'on appelle *droite épipolaire* de m_1 dans le plan image Π_2 .

On vient donc de montrer que pour tout point m_1 de Π_1 son homologue m_2 se situe sur la droite épipolaire ep_2 : autrement dit, pour m_1 fixé, lorsque M décrit la droite (O_1m_1) , l'image m_2 est située sur la droite ep_2 . Ainsi, la recherche d'un point homologue peut être réduite à une recherche sur une droite. Remarquons que les droites épipolaires ep_1 et ep_2 sont les intersections du plan O_1MO_2 , appelé *plan épipolaire* défini par M , avec les plans image Π_1 et Π_2 . De plus, pour un plan image donné, toutes les droites épipolaires se coupent à l'épipole.

Lorsque les plans Π_1 et Π_2 sont parallèles à la droite (O_1O_2) , les épipoles sont projetés à l'infini, et les droites épipolaires deviennent parallèles. C'est le cas du modèle projectif linéaire parallèle de la figure 2.2, où l'on a vu que deux points homologues ont même ordonnée : les droites épipolaires correspondent exactement aux lignes de l'image.

L'utilisation de la géométrie épipolaire montre que la recherche des points homologues dans le modèle projectif linéaire peut être réduite à une recherche le long des droites épipolaires. L'étape de *rectification* consiste à reprojeter les images du couple sur un plan particulier afin que les droites épipolaires correspondent aux lignes de l'image et que l'on puisse se ramener au cas du modèle projectif linéaire parallèle. Après rectification, deux points homologues ont même ordonnée, ce qui simplifie grandement le problème de la mise en correspondance. La rectification est une étape qui ne dépend que de la géométrie du système stéréoscopique, et

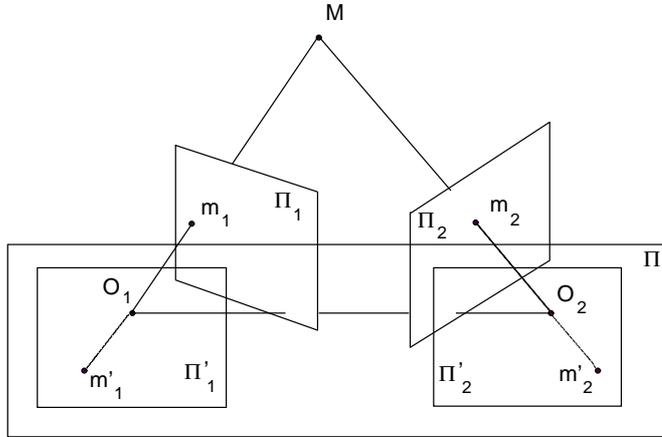


FIG. 2.5 – Rectification d’un couple stéréoscopique : chaque plan image Π_j , $j = 1, 2$, est projeté sur un même plan de rectification Π' , parallèle à la droite (O_1O_2) joignant les deux centres optiques. Dans les nouveaux plans Π'_1 et Π'_2 , les images m'_1 et m'_2 ont même ordonnée.

qui peut donc être menée sans aucune information sur la scène observée.

Montrons tout d’abord qu’une telle projection existe. Pour cela, considérons un plan de rectification Π' parallèle à la droite formée par les deux centres optiques (O_1O_2) . Un tel plan est représenté à la figure 2.5. La projection Π'_j du plan image Π_j , $j = 1, 2$, dans ce plan est obtenue en traçant pour chaque point m_j de Π_j l’intersection de la droite (O_jm_j) avec le plan Π' . Tout se passe donc comme si on observait le point M avec deux caméras placées en O_1 et O_2 , ayant chacune pour plan image le plan Π' . On est ainsi ramené au cas du modèle projectif linéaire parallèle : les épipoles sont projetés à l’infini et les droites épipolaires sont toutes parallèles aux lignes de l’image. Deux points homologues dans les nouvelles images ont donc bien même ordonnée.

Reste à choisir le plan de rectification, ou, plus généralement, les matrices de projection pour chaque caméra. La condition imposée est l’égalité entre les deux ordonnées après projection. Devernay [29] montre que l’ensemble des paires de matrices de rectification est une famille à 9 paramètres (variété de dimension 9). Il existe donc une grande liberté pour le choix de la rectification, et celui-ci dépend principalement de l’application visée. Si l’étape suivante est la mise en correspondance, on peut choisir une paire de rectifications qui minimise les déformations [56]. Si au contraire un plan dans la scène joue un rôle particulier (comme par exemple le sol pour le déplacement d’un robot), on peut avoir intérêt à rectifier par rapport à ce plan [29]. Notons que la rectification d’un couple stéréoscopique implique un rééchantillonnage des images initiales.

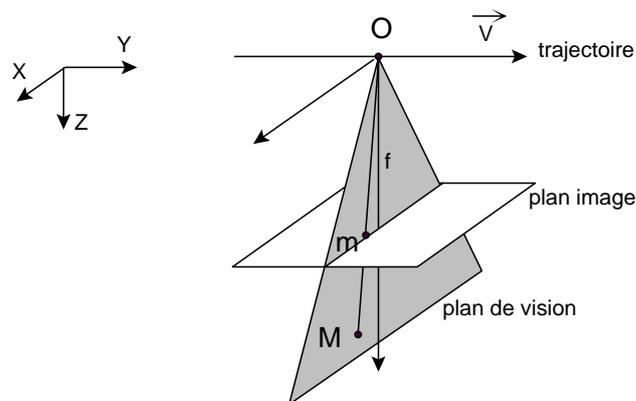


FIG. 2.6 – Capteur de type *pushbroom* : à un instant donné, seuls les points physiques M situés dans le plan de vision sont observés.

2.1.3 Le cas des capteurs *pushbroom*

Les relations établies précédemment sont valables pour des capteurs de type projectif linéaire : l'ensemble de l'image est acquis en une fois, l'espace étant projeté sur le plan image par rapport au centre optique. Dans le cas des capteurs *pushbroom*, le système optique est composé d'une barrette de détecteurs en mouvement. A chaque instant, seuls les points physiques situés dans le plan défini par le centre optique et la droite contenant la barrette sont vus par le système. Ce plan définit le *plan de vision*. Le rayonnement reçu pendant un certain temps Δt et provenant des points contenus dans le plan de vision est enregistré. Le capteur étant en mouvement, à l'instant suivant, une autre partie de l'espace est observée, et ainsi de suite. Ainsi, à intervalles de temps réguliers, le système enregistre une ligne de pixels ; l'ensemble de ces enregistrements constitue l'image. Contrairement au modèle sténopé, il n'y a pas unicité du centre optique, chaque ligne a un centre optique différent.

Le principe d'acquisition est représenté à la figure 2.6. Le capteur est symbolisé par son centre optique O en déplacement à la vitesse V ; la barrette de détecteurs, située à l'intersection entre le plan de vision et le plan image, est située à une distance f du centre optique.

Le capteur panchromatique utilisé sur SPOT est un exemple de capteur *pushbroom*. Chaque barrette est composée de 6 000 détecteurs CCD (*Charge Coupled Device*). Le temps d'acquisition pour chaque ligne de pixels est de 1.504 ms, si bien que 6 000 lignes sont enregistrées en 9 s, formant ainsi une image panchromatique SPOT. La zone au sol correspondante est approximativement un carré de 60 km de côté, chaque pixel ayant une résolution spatiale au sol approximative de 10 m.

Dans la suite, nous utilisons le modèle du capteur *pushbroom* linéaire proposé par [52]. On choisit tout d'abord comme repère de l'espace le repère \mathcal{R}_0 associé au

capteur à l'instant initial. Les hypothèses du modèle linéaire sont les suivantes :

- le capteur se déplace selon une droite à vitesse constante $V = (V_x, V_y, V_z)^\top$,
- l'orientation reste constante pendant la durée d'acquisition.

Compte tenu du déplacement, très complexe, du satellite, ces hypothèses ne sont pas vérifiées en pratique dans le cas de SPOT. Cependant, sur l'ensemble d'une image SPOT, ces approximations sont raisonnables et les résultats obtenus avec ce modèle sont très proches de ceux donnés par un modèle plus complet [52].

A la différence du modèle projectif linéaire, la dimension temporelle joue un rôle, puisque la position du capteur change pendant l'acquisition. Un point M de l'espace de coordonnées $(x_0, y_0, z_0)^\top$ dans le repère \mathcal{R}_0 est observé lorsqu'il se trouve dans le plan de vision, c'est-à-dire à l'instant t tel que $V_y t = y_0$. A cet instant, la position du capteur est $(\frac{V_x}{V_y} y_0, y_0, y_0 \frac{V_z}{V_y})^\top$, et la t ième ligne est enregistrée ; l'ordonnée du point image m vaut donc :

$$v = \frac{y_0}{V_y} \quad (2.8)$$

En notant p_u le décalage d'origine le long des abscisses et en utilisant le triangle rectangle passant par les points O et M et contenant la perpendiculaire au plan image, on obtient la relation suivante :

$$\frac{z_0 - y_0 \frac{V_z}{V_y}}{f} = \frac{x_0 - y_0 \frac{V_x}{V_y}}{u - p_u} \quad (2.9)$$

En notant w un facteur d'échelle, la projection sur le plan image est décrite par le système suivant :

$$\begin{pmatrix} v \\ wu \\ w \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ f & 0 & p_u \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{V_x}{V_y} & 0 \\ 0 & \frac{1}{V_y} & 0 \\ 0 & -\frac{V_z}{V_y} & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \quad (2.10)$$

Dans un système de coordonnées orthonormal quelconque, le point M a pour coordonnées $(x, y, z)^\top$ qui sont reliées aux coordonnées $(x_0, y_0, z_0)^\top$ par la relation :

$$\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = R \left[\begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \right] = (R| - RT) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2.11)$$

où R est une matrice de rotation, $T = (T_x, T_y, T_z)^\top$ un vecteur de translation et $(R| - RT)$ désigne la matrice 3×4 dont les 3 premières colonnes sont les colonnes de R et la dernière colonne le vecteur $-RT$.

Finalement, le modèle du capteur *pushbroom* linéaire peut être résumé par la relation algébrique :

$$\begin{pmatrix} v \\ wu \\ w \end{pmatrix} = P \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2.12)$$

où P est une matrice à 11 degrés de liberté [52] : 4 dans la première ligne et 7 dans les 2 dernières. Les points visibles par le capteur sont ceux pour lesquels $w > 0$. Notons que P n'est pas homogène au sens où la matrice $P' = kP$ avec $k \neq 0$ ne lui est pas équivalente. Seules les 2 dernières lignes de P sont homogènes. Pour résumer, l'image d'un point physique M est projective dans la direction des abscisses et orthographique dans celle des ordonnées.

Les images d'un couple stéréoscopique obtenu à partir de capteurs *pushbroom* ne sont généralement pas acquises simultanément. En effet, il faut attendre que le satellite repasse au-dessus de la zone observée pour réaliser une seconde image. Généralement, celle-ci est obtenue en inclinant légèrement la barrette de capteurs lorsque le satellite se trouve sur une orbite voisine de celle correspondant au premier passage. Le satellite SPOT par exemple observe une même zone tous les 3 jours [1], mais bien souvent le délai entre l'acquisition des deux images du couple est supérieur, car, entre autres, les conditions météorologiques doivent être favorables (couverture nuageuse faible) dans les deux cas. Ceci explique en partie les importantes différences radiométriques que l'on peut observer entre deux images d'un couple SPOT (cf. chapitre 3 pour des exemples). D'autres satellites, comme SPOT 5, ont plusieurs barrettes de capteurs embarquées : l'une regarde vers l'avant, l'autre vers l'arrière et une troisième à la verticale. Ainsi, sur la même orbite, le satellite peut acquérir un couple d'images stéréoscopiques quasi-simultanées. Ce dispositif a l'avantage de limiter les différences radiométriques entre les images du couple.

Intéressons-nous maintenant à la géométrie épipolaire dans le cas de capteurs *pushbroom*. Comme le centre optique se déplace durant l'acquisition, la notion d'épipoles pour un système stéréoscopique de capteurs *pushbroom* n'est pas pertinente. Il s'ensuit que la géométrie épipolaire est totalement différente de celle générée par des capteurs de type projectif. Rappelons que nous avons défini la droite épipolaire associée à un point m_1 de l'image 1 comme la projection de la droite (Mm_1) dans l'image 2. Nous avons notamment vu que pour le modèle projectif linéaire la projection d'une telle droite est aussi une droite. Qu'en est-il dans le cas du modèle de capteur *pushbroom* linéaire? Considérons deux capteurs *pushbroom* linéaires modélisés par les matrices de projection P_1 et P_2 . Les images d'un point de coordonnées $(x, y, z)^\top$ ont pour coordonnées $(u_1, v_1)^\top$ dans l'image 1 et $(u_2, v_2)^\top$ dans l'image 2 telles que :

$$\begin{pmatrix} v_1 \\ w_1 u_1 \\ w_1 \end{pmatrix} = P_1 \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad \begin{pmatrix} v_2 \\ w_2 u_2 \\ w_2 \end{pmatrix} = P_2 \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2.13)$$

On montre alors qu'il existe une matrice homogène F [52] telle que :

$$(u_2, u_2v_2, v_2, 1)F \begin{pmatrix} u_1 \\ u_1v_1 \\ v_1 \\ 1 \end{pmatrix} = 0 \quad (2.14)$$

Fixons un point de l'image 1 $(u_1, v_1)^\top$ et notons $(\alpha, \beta, \gamma, \delta)^\top = F(u_1, u_1v_1, v_1, 1)^\top$. Les coordonnées $(u_2, v_2)^\top$ de son homologue dans l'image 2 vérifient l'équation d'une hyperbole :

$$\alpha u_2 + \beta u_2v_2 + \gamma v_2 + \delta = 0 \quad (2.15)$$

Les courbes épipolaires dans le modèle *pushbroom* linéaire sont donc des hyperboles.

D'un point de vue strictement géométrique, il n'existe pas de projection qui permette d'obtenir des images rectifiées dont les lignes correspondent aux courbes épipolaires. Une méthode consiste à utiliser une approximation polynômiale. Tout d'abord, les images sont reprojctées sur l'ellipsoïde de référence. Cette reprojction nécessite la connaissance de la trajectoire du satellite, et permet de tenir compte des mouvements parasites qui l'écartent d'une trajectoire rectiligne. On élimine ainsi les distorsions haute fréquence, et l'on se retrouve dans un cas proche du modèle linéaire étudié précédemment. Dans le cas de couples stéréoscopiques obtenus avec SPOT, les trajectoires des capteurs lors des deux acquisitions sont de plus pratiquement parallèles. On peut alors appliquer à chaque image une transformation polynômiale Q_i , $i = 1, 2$, de \mathbb{R}^2 dans \mathbb{R}^2 de telle sorte que deux points homologues soient envoyés sur la même ligne :

$$(u'_1, v'_1)^\top = Q_1(u_1, v_1)^\top \quad (u'_2, v'_2)^\top = Q_2(u_2, v_2)^\top \quad (2.16)$$

Les transformations Q_1 et Q_2 sont déterminées par moindres carrés à l'aide de couples de points homologues identifiés manuellement dans les images. Soient $(u_1^i, v_1^i)^\top, (u_2^i, v_2^i)^\top$, $i = 1, l$, ces couples de points, on cherche alors les paramètres de Q_1 et Q_2 qui minimisent :

$$\sum_{i=1}^l \|Q_1(u_1^i, v_1^i)^\top - Q_2(u_2^i, v_2^i)^\top\|^2 \quad (2.17)$$

où $\|(u, v)^\top\| = \sqrt{u^2 + v^2}$. On obtient ainsi un couple d'images rectifiées où la disparité résiduelle aux points de calage est pratiquement nulle.

Pour d'autres développements sur la géométrie des capteurs *pushbroom*, en particulier ceux embarqués sur les satellites SPOT, et les approximations polynômiales pour calculer les courbes épipolaires, on peut se reporter à [64, 1].

On a représenté à la figure 2.7 un couple stéréoscopique rectifié d'images panchromatiques obtenues par le satellite SPOT sur la région d'Aix-Marseille

(France). Des extraits de ce couple sont étudiés dans les chapitres suivants. Il a l'avantage de réunir à la fois des zones urbaines et péri-urbaines et des zones au relief plus accidenté.

2.1.4 Définition de la disparité

Après avoir explicité la géométrie épipolaire dans le cas des capteurs sténopé et “pusbroom”, revenons au problème général et considérons un couple (obtenu par un capteur quelconque) d'images stéréoscopiques rectifiées I_1 et I_2 . Comme il a été vu précédemment, deux points homologues m_1 , de coordonnées $(u_1, v_1)^\top$, et m_2 , de coordonnées $(u_2, v_2)^\top$, ont même ordonnée, soit $v_1 = v_2$. On peut alors définir la disparité d comme la différence entre les abscisses après rectification. Si l'on choisit l'image 1 comme référence, la disparité en tout point de coordonnées $(u_1, v_1)^\top$ de l'image 1 dont l'homologue a pour coordonnées $(u_2, v_2)^\top$ dans l'image 2 s'écrit :

$$d_1(u_1, v_1) = u_2 - u_1 \quad (2.18)$$

On peut définir symétriquement la disparité par rapport à l'image 2 ; on a ainsi pour tout point de coordonnées $(u_2, v_2)^\top$ de l'image 2 :

$$d_2(u_2, v_2) = u_1 - u_2 \quad (2.19)$$

Etant définies dans des systèmes de coordonnées différents, les disparités d_1 et d_2 ne sont pas les opposées l'une de l'autre ; la bonne relation est la suivante :

$$d_1(u, v) = -d_2(u + d_1(u, v), v), \quad \forall (u, v)^\top \in I_1 \quad (2.20)$$

A tout pixel de l'image I_1 , on peut ainsi faire correspondre une valeur de la disparité, à condition que le pixel homologue existe, c'est-à-dire en l'absence d'occlusions. On peut donc construire une carte de disparité, superposable à I_1 , qui permet de calculer les coordonnées des pixels homologues dans l'image 2. Symétriquement, on peut construire une autre carte de disparité, superposable à l'image 2. D'après la remarque précédente, ces deux cartes ne se déduisent pas immédiatement l'une de l'autre, et ce d'autant plus que chaque carte n'est connue qu'en des points de coordonnées entières, qui correspondent aux pixels.

La définition que nous venons de proposer présente un désavantage certain : elle n'est pas symétrique, et l'on peut se demander quelle carte de disparité utiliser en pratique, puisqu'elles ne sont pas strictement équivalentes. Il est possible en réalité de définir de manière totalement symétrique la disparité, en se plaçant dans un repère convenablement choisi. Considérons pour cela la figure 2.8, représentant deux capteurs de centres optiques O_1 et O_2 et de plans image Π_1 et Π_2 que l'on suppose confondus (ou ce qui revient au même on travaille avec les images

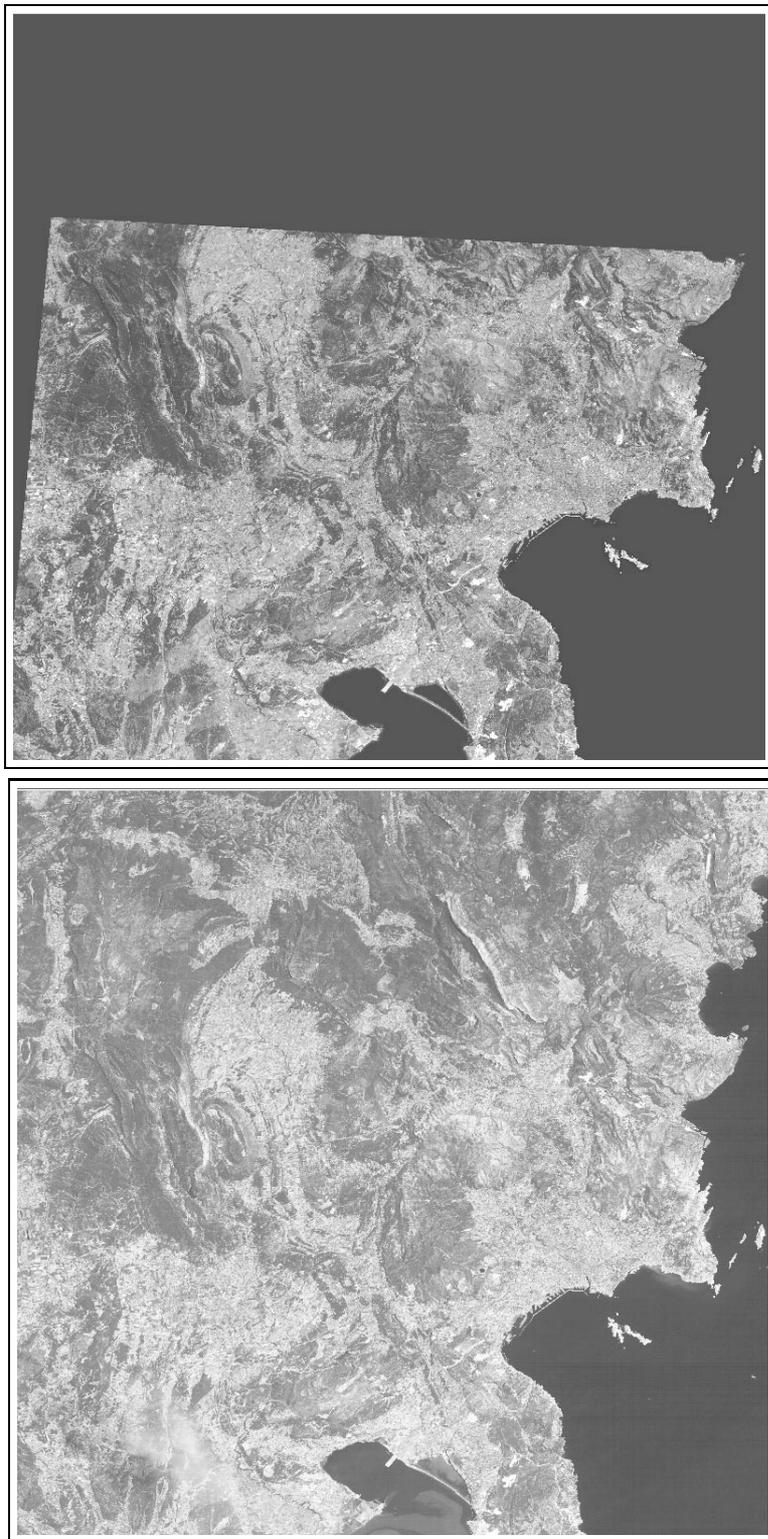


FIG. 2.7 – Couple stéréoscopique rectifié d’images panchromatiques obtenues par le satellite SPOT sur la région d’Aix-Marseille (France).

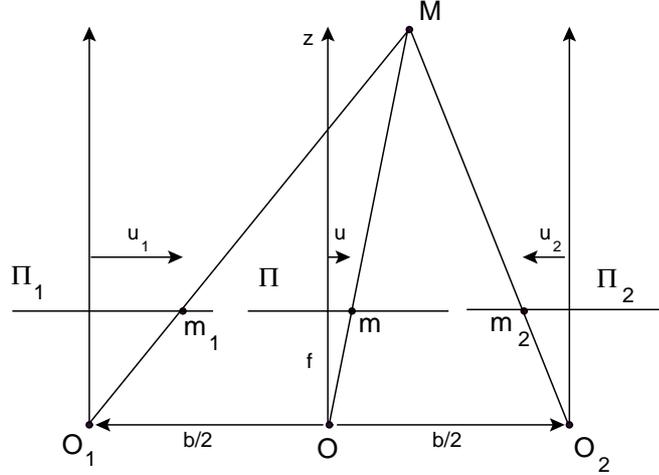


FIG. 2.8 – Définition de l’image cyclopéenne. Elle se construit en imaginant un capteur de centre optique O placé à mi-distance des centres optiques O_1 et O_2 .

rectifiées). On imagine alors un capteur dont le centre optique O est placé au milieu du segment $[O_1O_2]$, de longueur b , et dont le plan image Π est confondu avec les plans Π_1 et Π_2 . L’image ainsi obtenue est appelée l’image *cyclopéenne*. Un point physique M se projette en m_1 dans Π_1 , m_2 dans Π_2 et m dans Π . Ces 3 points ont même ordonnée, et ont pour abscisses respectivement u_1 , u_2 et u . En utilisant les triangles MOO_1 et MOO_2 et les droites parallèles (m_1m_2) et (O_1O_2) , on obtient la relation suivante :

$$\frac{b/2}{b/2 - u_1 + u} = \frac{b/2}{b/2 - u + u_2} \quad (2.21)$$

Ainsi, on a l’égalité suivante entre les différences d’abscisses :

$$u - u_2 = -(u - u_1) = \frac{u_1 - u_2}{2} \quad (2.22)$$

On peut donc définir une carte de disparité relative à l’image cyclopéenne, en posant pour tout pixel de coordonnées $(u, v)^\top$ dans le repère attaché à Π [6] :

$$d(u, v) = u - u_2 = -(u - u_1) \quad (2.23)$$

En revanche, la carte de disparité définie de manière symétrique dans l’image cyclopéenne n’est superposable à aucune des 2 images, et a un système de coordonnées propre.

Pour une discussion plus détaillée sur la définition de la disparité, en particulier à partir des angles de vision d’un point, le lecteur est renvoyé à [72].

Dans la suite, nous définissons toujours la disparité par rapport à l’image 1 ; ainsi un couple de points homologues est parfaitement identifié par les coordonnées $(u, v)^\top$ du pixel de l’image 1 et la valeur de la disparité $d(u, v)$ en ce point.

2.1.5 La mise en correspondance : calcul de la disparité

Les calculs précédents ont montré que les coordonnées dans l'espace d'un point physique M se déduisent à partir des coordonnées des points image m_1 et m_2 , et tout particulièrement de la disparité, et des paramètres des deux caméras. En pratique cependant, on dispose seulement des deux images et il faut donc déterminer les couples de points homologues. La disparité s'obtient alors par différence des coordonnées des points homologues. Nous considérons ci-après un couple d'images rectifiées (I_1, I_2) .

Traditionnellement, il existe deux approches au problème de la mise en correspondance, dont on peut trouver un exposé complet dans [30]. La première, qui n'est pas traitée ici, consiste à mettre en correspondance des attributs remarquables des images, telles des droites ou les contours d'objets. Cette méthode permet d'obtenir un calcul assez précis de la disparité, peu sensible au bruit des images, mais nécessite une étape d'interpolation pour obtenir une estimation de la disparité en tout point.

La seconde approche utilise directement l'intensité enregistrée en chaque pixel, et cherche à déterminer en chaque point d'une image son correspondant dans l'autre image. Cette méthode repose sur l'hypothèse que localement les deux images du couple sont semblables : pour un couple de points homologues $(u, v)^\top$ et $(u + d(u, v), v)^\top$, on doit vérifier en effet que les valeurs des pixels sont proches, soit :

$$I_1(u, v) \approx I_2(u + d(u, v), v) \quad (2.24)$$

Il doit donc être possible de reconnaître l'homologue d'un pixel donné. Pour cela un critère de similitude est utilisé, et le pixel homologue est choisi comme celui qui maximise ce critère.

Considérons un couple d'images rectifiées. On cherche à calculer la carte de disparité associée à l'image 1. Pour un pixel $(u, v)^\top$ et une valeur d de la disparité donnés, appelons $C_{u,v}(d)$ la valeur du critère de similitude entre les pixels $(u, v)^\top$ et $(u + d, v)^\top$. L'estimation de la disparité en $(u, v)^\top$ est alors [102] :

$$\hat{d}(u, v) = \arg \max_d C_{u,v}(d) \quad (2.25)$$

Il existe plusieurs critères de similitude utilisés, selon que l'on choisisse comme attribut directement l'intensité ou une quantité déduite à l'aide d'un opérateur, tel la convolution avec le Laplacien d'une gaussienne, ou encore une combinaison des deux. Généralement, le critère est fondé sur le calcul d'une distance quadratique et l'utilisation d'un voisinage du pixel d'intérêt. En effet, l'information contenue dans l'intensité d'un pixel pris isolément est trop faible, d'où la nécessité d'utiliser des filtres de convolution ou des voisinages.

Un critère communément utilisé est le calcul d'un coefficient de corrélation $C_{u,v}(d)$

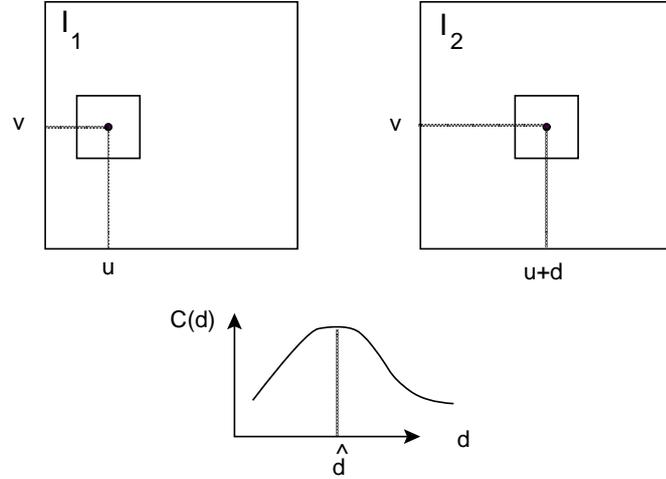


FIG. 2.9 – Calcul du critère de corrélation au pixel $(u, v)^\top$: on considère une fenêtre centrée sur le pixel $(u, v)^\top$ (haut) et on cherche le déplacement d qui maximise la corrélation $C_{u,v}(d)$ (bas).

sur une fenêtre T , de taille $|T|$ [34] :

$$C_{u,v}(d) = \frac{1}{|T|} \sum_{(u',v') \in T} \frac{I_2(u + u' + d, v + v') - m_2(u + d, v)}{\sigma_2(u + d, v)} \times \frac{I_1(u + u', v + v') - m_1(u, v)}{\sigma_1(u, v)} \quad (2.26)$$

où m_j et σ_j , $j = 1, 2$, désignent respectivement la moyenne et l'écart type de l'intensité I_j , calculés sur la fenêtre T :

$$m_j(u, v) = \frac{1}{|T|} \sum_{(u',v') \in T} I_j(u + u', v + v'), \quad j = 1, 2 \quad (2.27)$$

$$\sigma_j^2(u, v) = \frac{1}{|T|} \sum_{(u',v') \in T} (I_j(u + u', v + v') - m_j(u, v))^2, \quad j = 1, 2 \quad (2.28)$$

Le principe du calcul de la corrélation sur une fenêtre T est représenté à la figure 2.9. Implicitement, une hypothèse de faible variation de la disparité à l'intérieur de la fenêtre de corrélation T est faite. Celle-ci revient à supposer que les surfaces observées sont approximativement parallèles au plan de rectification. Dans le cas d'un capteur situé à une distance importante de la scène observée, cette hypothèse n'est pas trop contraignante, bien que mise en défaut aux extrémités d'objets. Une extension possible consiste à faire varier la taille de la fenêtre de corrélation afin de mieux tenir compte des fortes variations de la disparité [60, 71].

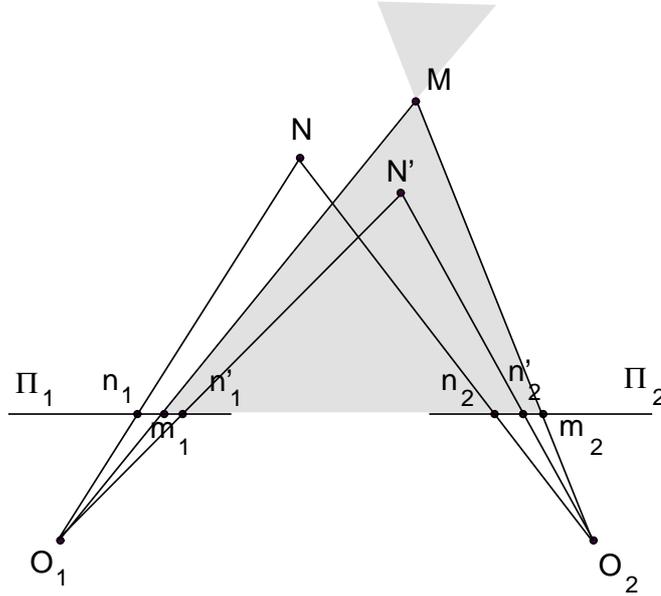


FIG. 2.10 – Contrainte d'ordre : pour un point M de l'espace donné, l'ordre avec un autre point N est respecté si N n'appartient pas au cône grisé. Il est inversé dans le cas contraire.

Remarquons que la contrainte épipolaire est implicitement utilisée dans cette formulation car la disparité est directement exprimée sous forme d'un scalaire, les points homologues appartenant à la même ligne. Généralement, on utilise une contrainte supplémentaire, appelée contrainte d'ordre, pour réduire l'espace de recherche. Cette contrainte implique que l'ordre des points soit préservé d'une image à l'autre. Considérons pour cela la figure 2.10. Soit un point M de l'espace fixé, se projetant respectivement en m_1 et m_2 . A partir des centres optiques O_1 et O_2 , on peut tracer le cône passant par M (en grisé sur la figure 2.10). On voit alors que pour tout point N' appartenant à ce cône, l'ordre des images est inversé : si n'_1 apparaît à droite de m_1 dans l'image 1, alors n'_2 apparaît à gauche de m_2 dans l'image 2, et vice-versa. En revanche, si le point N appartient au cône complémentaire, l'ordre des images est conservé. L'utilisation de la contrainte d'ordre revient à supposer que les configurations du premier type, où le point N est à l'intérieur du cône, ne sont pas rencontrées. Dans le cas de capteurs très éloignés de la scène observée, cette hypothèse est raisonnable car le cône critique est pratiquement fermé. Plus généralement, cela revient à supposer que les deux points M et N appartiennent à la même surface physique, et que cette surface est relativement régulière [34].

La contrainte d'ordre peut s'exprimer à l'aide de la disparité. Il suffit d'écrire que pour deux points d'une même ligne de l'image 1 $(u, v)^\top$ et $(u', v)^\top$, les points

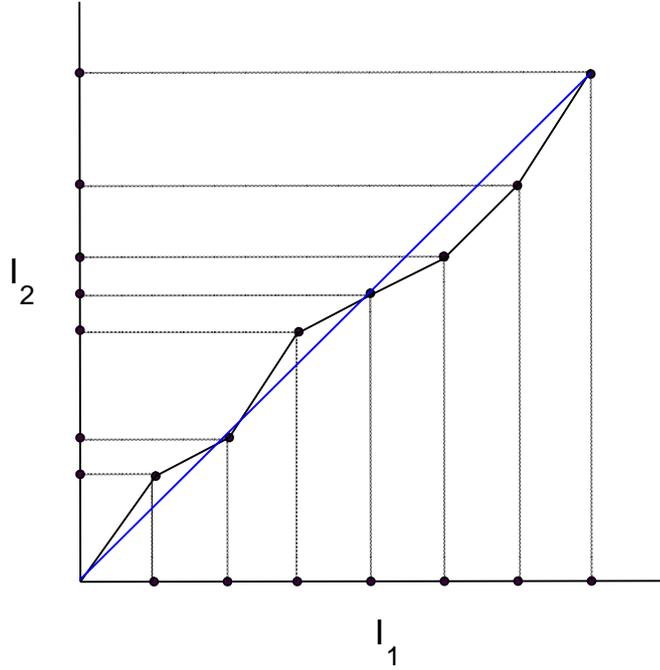


FIG. 2.11 – La recherche des points homologues pour une ligne du couple peut être vue comme la recherche d’un chemin minimal dans un graphe. Chaque nœud du chemin correspond à la mise en correspondance de 2 pixels, un coût lui est attribué en fonction de la corrélation. Les chemins qui respectent la contrainte d’ordre sont les chemins croissants. La distance entre le chemin et la bissectrice (en bleu) est égale à la disparité.

homologues dans l’image 2 sont dans le même ordre, soit :

$$\frac{u' + d(u', v) - u - d(u, v)}{u' - u} > 0, \quad \forall u' \neq u \quad (2.29)$$

On obtient une condition sur la dérivée de la disparité :

$$\frac{\partial d(u, v)}{\partial u} > -1 \quad (2.30)$$

Cette contrainte impose une relation entre les disparités d’une même ligne, si bien que le calcul de la disparité est généralement résolu pour toute une ligne à la fois. La recherche des points homologues peut alors être vue comme la recherche d’un chemin minimal dans un graphe. Considérons le schéma de la figure 2.11, qui représente le calcul de la disparité pour une ligne du couple (I_1, I_2) . Pour chacune des deux images, on fait correspondre aux pixels de la ligne analysée un axe. Ainsi, pour cette ligne, chaque couple de points supposés homologues peut être représenté par un nœud dans le plan. Une solution au problème de la mise

en correspondance d'une ligne du couple correspond donc au chemin reliant ces nœuds. Chaque nœud se voit attribué un coût en fonction de la valeur du critère utilisé. La solution optimale est celle qui minimise le coût total. Ce problème d'optimisation est généralement résolu à l'aide d'un algorithme de programmation dynamique, de type Viterbi [81, 34, 47]. Remarquons que le respect de la contrainte d'ordre revient à ne choisir que les chemins croissants du graphe.

Fua [40] a proposé une extension de l'algorithme de corrélation qui permet de répondre à un certain nombre de difficultés :

- asymétrie du calcul de la disparité,
- inadéquation dans le cas d'occlusions,
- manque de robustesse de l'estimation.

L'idée, assez simple, consiste à remarquer qu'un point m_2 de l'image I_2 est l'homologue d'un point m_1 de l'image I_1 si et seulement si m_1 est l'homologue de m_2 . Ainsi, dans la procédure d'optimisation de l'image I_1 vers I_2 , on ne retient un appariement (m_1, m_2) que si la procédure d'optimisation symétrique de I_2 vers I_1 donne l'appariement (m_2, m_1) . Dans le cas contraire, on laisse la valeur inconnue. La carte de disparité obtenue peut être complétée par interpolation. Cette méthode permet notamment de repérer les zones de forte déformation géométrique, voire d'occlusion. La carte de disparité calculée reste asymétrique dans le sens où elle est définie dans le repère d'une des images, mais son calcul est symétrique. Notamment, les cartes de disparité gauche et droite sont parfaitement équivalentes. L'algorithme permet de réaliser une bijection entre 2 sous-ensembles des images I_1 et I_2 . C'est cet algorithme qui est utilisé dans les applications des chapitres suivants.

Dans une image numérique, l'information est quantifiée : seules des coordonnées entières de points sont disponibles. Ainsi, le résultat de la mise en correspondance par corrélation est une carte de disparité à valeurs dans les entiers relatifs. Il est possible d'obtenir cependant une estimation subpixellique de la disparité, en approchant la courbe de corrélation $C_{u,v}(d)$ par une fonction donnée. Une telle méthode est décrite dans [29], avec pour fonction d'approximation une parabole. Supposons qu'en $(u, v)^\top$ l'estimation entière de la disparité soit \hat{d}_0 . On trace alors la parabole passant par les points $(\hat{d}_0 - 1, C_{u,v}(\hat{d}_0 - 1))$, $(\hat{d}_0, C_{u,v}(\hat{d}_0))$, et $(\hat{d}_0 + 1, C_{u,v}(\hat{d}_0 + 1))$. La correction subpixellique s'écrit :

$$\hat{d} = \hat{d}_0 + \frac{1}{2} \frac{C_{u,v}(\hat{d}_0 + 1) - C_{u,v}(\hat{d}_0 - 1)}{C_{u,v}(\hat{d}_0) - C_{u,v}(\hat{d}_0 + 1) + C_{u,v}(\hat{d}_0) - C_{u,v}(\hat{d}_0 - 1)} \quad (2.31)$$

Dans le cas du traitement de couples stéréoscopiques d'images SPOT, les algorithmes de calcul de disparité reposant sur la maximisation de la corrélation donnent généralement de bons résultats, car à cette résolution les occlusions et les discontinuités de la disparité sont assez rares. Parmi les limites de l'algorithme, citons entre autres le lissage des variations fortes de disparité en raison du caractère spatial du critère, le traitement indépendant des lignes successives de l'image

qui empêche de tenir compte de contraintes spatiales à deux dimensions et la sensibilité à l'information radiométrique.

Récemment, une étude comparative [104] portant sur une vingtaine d'algorithmes et utilisant une même série de couples stéréoscopiques (principalement orientée vers des applications en vision par ordinateur) a été menée et a tenté de faire le point sur les performances respectives de ces différents algorithmes. Ce travail montre que le problème du calcul de la disparité n'est pas encore totalement résolu. Il apparaît que les méthodes de minimisation d'énergie à partir de coupures de graphes [61, 63, 62], qui permettent de calculer des minima locaux forts dans un cadre assez général (notamment en tenant compte de contrainte spatiale bi-dimensionnelles et en traitant de manière explicite les occlusions) et en un temps de calcul réduit, sont parmi les plus efficaces et les plus prometteuses.

L'algorithme de calcul développé par Istar utilise un critère de corrélation et une méthode de programmation dynamique avec respect de la contrainte d'ordre. De plus, la contrainte de symétrie de l'optimum est utilisée, ainsi qu'une estimation subpixelique de la disparité. La figure 2.12 montre la carte de disparité du couple stéréoscopique d'Aix-Marseille (figure 2.7) calculée à l'aide de l'algorithme de corrélation d'Istar. Les points en rouge (dans la zone de recouvrement des deux images du couple) correspondent aux valeurs de la disparité non estimées, car ne vérifiant pas la contrainte de symétrie.

2.1.6 Reconstitution

L'étape de reconstitution consiste à déterminer les coordonnées d'un point physique M à partir des coordonnées de ses points image m_1 et m_2 . Il ne s'agit pas ici de décrire en détails ce processus, mais juste de donner un aperçu simplifié de la méthode.

Notons $z = (x, y, z)^\top$ les coordonnées d'un point physique dans le référentiel choisi, et $u = (u_1, v_1, u_2, v_2)^\top$ les coordonnées des points image correspondants. Soit θ le vecteur de paramètres d'un système stéréoscopique, et F la fonction de localisation du système, i.e :

$$F(\theta, z) = u \tag{2.32}$$

La première étape, dite de calibration, a pour but d'estimer le vecteur de paramètres θ . Pour cela, on utilise une méthode d'estimation par moindres carrés à partir de points d'appui. Soit $(z_i, i = 1, l)$ un ensemble de points de coordonnées connues pour lesquels on connaît également les coordonnées u_i des points image (par exemple par repérage sur une carte), et Q la matrice de précision correspondante (matrice de covariance inverse des erreurs de repérage, en général diagonale).

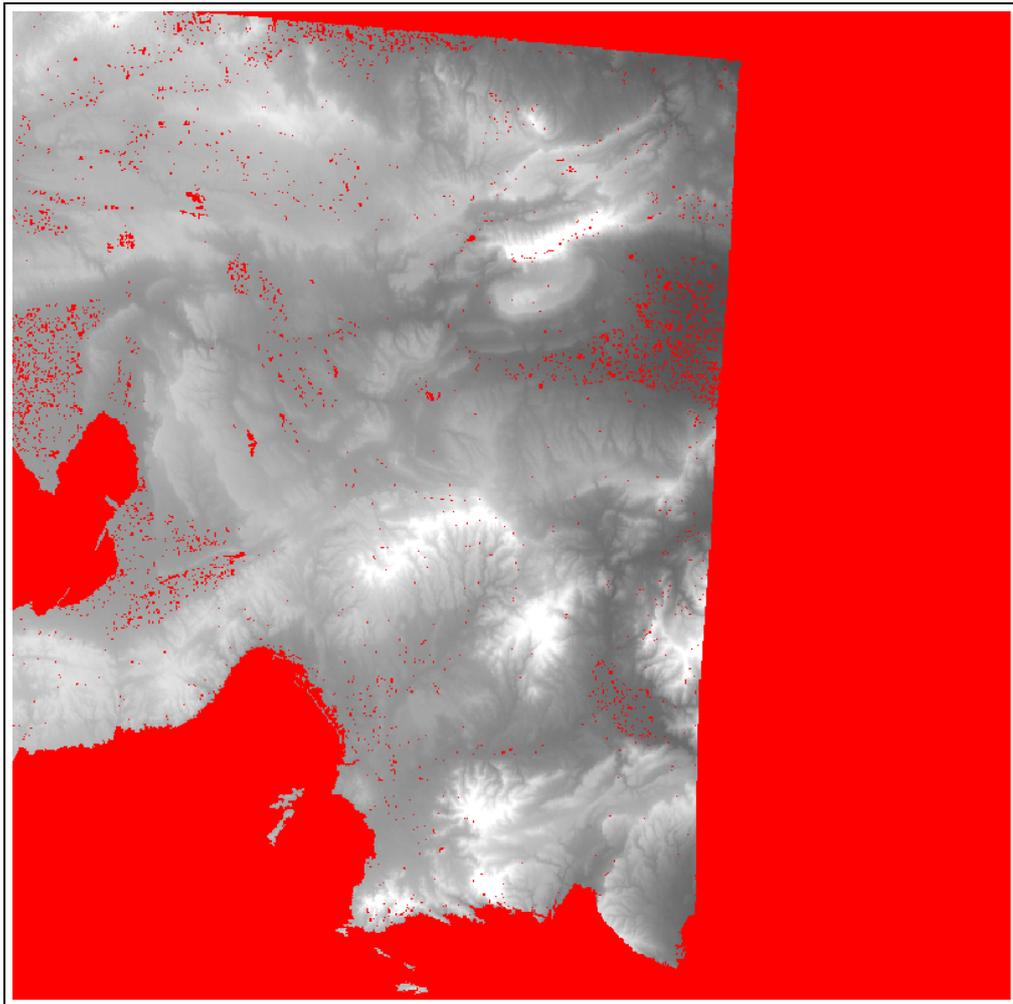


FIG. 2.12 – Carte de disparité du couple de la figure 2.7 calculée par corrélation.

Soit $r(\theta)$ le vecteur des résidus associé à θ :

$$r(\theta) = \begin{pmatrix} F(\theta, z_1) - u_1 \\ \dots \\ F(\theta, z_l) - u_l \end{pmatrix} \quad (2.33)$$

L'estimateur par moindres carrés de θ est donné par la relation :

$$\hat{\theta} = \arg \min_{\theta} r(\theta)^T Q r(\theta) \quad (2.34)$$

Une fois déterminé le vecteur de paramètres θ , il suffit d'inverser la relation 2.32 pour déterminer les coordonnées z d'un point physique. C'est l'étape de spatio-triangulation. En notant cette fois-ci $r(z)$ le vecteur de résidus associés à z :

$$r(z) = F(\hat{\theta}, z) - u, \quad (2.35)$$

la solution au problème de spatio-triangulation est celle qui minimise $r(z)^T r(z)$. Sous réserve de régularité de r , ce problème a une solution unique qui vérifie :

$$\nabla r(z)^T r(z) = 0 \quad (2.36)$$

et vers laquelle la suite :

$$z_{k+1} = z_k - (\nabla r(z_k)^T \nabla r(z_k))^{-1} \nabla r(z_k)^T r(z_k) \quad (2.37)$$

converge. On utilise ainsi une méthode de gradient pour résoudre le problème.

A la figure 2.13, on a reproduit le MNT de la région d'Aix-Marseille calculé à partir de la carte de disparité de la figure 2.12.

2.2 Etude de l'incertitude associée à un MNT : survol bibliographique

L'étude de l'incertitude liée à la reconstruction de la géométrie d'une scène, ou dans notre cas de l'élévation d'un terrain, n'est pas nouvelle. En ce qui concerne les MNTs, cette analyse s'inscrit dans le contexte plus large de l'utilisation de systèmes d'informations géographiques, qui a connu un développement important durant la dernière décennie. Les utilisateurs de ces systèmes se trouvent confrontés à des informations diverses comme l'élévation du terrain, la localisation du réseau hydrographique ou du réseau routier, le plan d'occupation des sols, informations elles-mêmes dérivées de données d'origines variées. Naturellement, ces utilisateurs ont exprimé des exigences en matière de qualité des données géographiques, notamment sur deux points : quelle est la nature des données utilisées et quelle est la

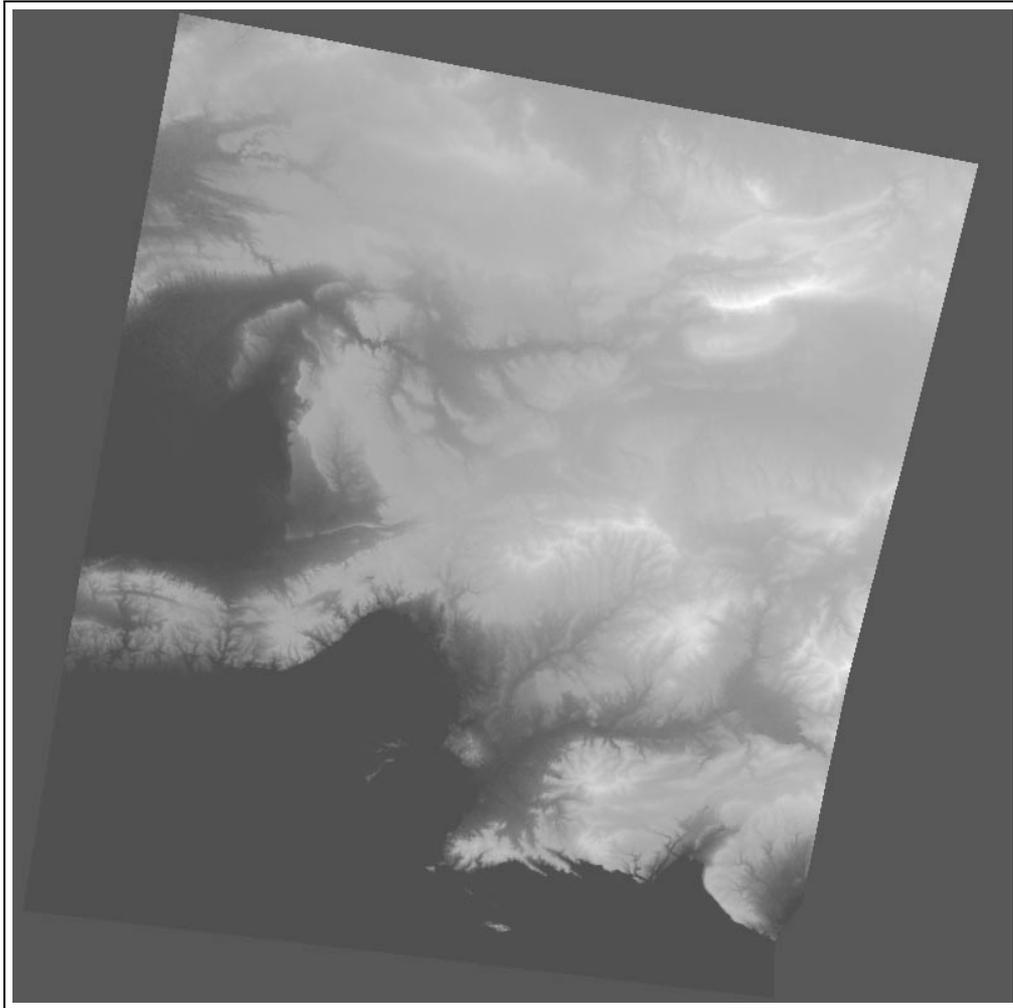


FIG. 2.13 – Modèle numérique de terrain calculé à partir de la carte de disparité de la figure 2.12

qualité de l'information fournie [115] ? On comprend alors qu'une réflexion se soit mise en place sur la définition même des critères utilisés pour spécifier la qualité de cette information [69, 37].

Cependant, les contextes dans lesquels les analyses des erreurs d'un MNT sont menées varient fortement d'une étude à l'autre. Un premier groupe de travaux s'intéresse à la précision d'un MNT sans se référer à la méthode ni aux données utilisées pour son calcul. Généralement, il s'agit de s'appuyer sur des mesures de référence et d'étudier la structure de l'erreur obtenue par différence. Dans d'autres cas, il s'agit de tester la cohérence de la surface calculée, et éventuellement d'éliminer les points aberrants. Un second ensemble de travaux, qui nous intéresse particulièrement, est plus spécifiquement dédié à l'analyse des erreurs propres aux techniques de stéréoscopie. Ils offrent une analyse détaillée des algorithmes de calcul de la disparité, souvent dans le but de proposer un algorithme plus performant. Les études sur la précision des MNTs sans référence à un algorithme de calcul particulier font généralement partie de la littérature de télédétection et de photogrammétrie, alors que l'analyse des algorithmes de stéréoscopie relève plutôt du domaine de la vision par ordinateur.

2.2.1 Validation *a posteriori* d'un MNT

Quelle que soit la méthode de construction d'un MNT, il est toujours possible d'analyser les erreurs d'altitude à partir de points de référence, lorsque ceux-ci sont disponibles. Ceux-ci peuvent être des points de contrôle, pour lesquels l'altitude a été mesurée sur le terrain, ou éventuellement des points d'un MNT de résolution plus fine. Au point de coordonnées $(x, y)^\top$, l'erreur d'altitude est définie comme la différence :

$$e(x, y) = z(x, y) - \hat{z}(x, y) \quad (2.38)$$

où $z(x, y)$ est la mesure de référence et $\hat{z}(x, y)$ est l'estimation donnée par le MNT. Si l'on dispose de plusieurs données de référence $z(x_\alpha, y_\alpha)$, $\alpha = 1, l$, on peut calculer la moyenne et l'écart type de l'erreur sur l'ensemble du jeu de données :

$$m_e = \frac{1}{l} \sum_{i=1}^l e(x_i, y_i) \quad (2.39)$$

$$\sigma_e^2 = \frac{1}{l} \sum_{i=1}^l e^2(x_i, y_i) - m_e^2 \quad (2.40)$$

Généralement, la moyenne m_e est appelée biais ou erreur systématique, dû par exemple à des problèmes de calibration des appareils, tandis que l'écart type σ_e est supposé décrire des fluctuations "aléatoires" [115]. Remarquons qu'il n'y a pas besoin de recourir à la théorie statistique, et notamment à des hypothèses de stationnarité spatiale, pour procéder à une telle analyse descriptive. Il s'agit

simplement de choisir des critères susceptibles de décrire l'erreur et de pouvoir qualifier la précision du MNT. D'autres critères fréquemment utilisés sont l'erreur quadratique moyenne et les erreurs maximales et minimales :

$$RMSE = \sqrt{\frac{1}{l} \sum_{i=1}^l e^2(x_i, y_i)} \quad (2.41)$$

$$e_{\max} = \max_{i \leq l} e(x_i, y_i) \quad (2.42)$$

$$e_{\min} = \min_{i \leq l} e(x_i, y_i) \quad (2.43)$$

Ces quantités sont par exemple fournies avec les MNTs produits par le British Ordnance Survey [83]. Elles sont également utilisées comme critères par le U.S. Geological Survey, qui préconise l'emploi d'au moins 20 données de référence et une erreur quadratique moyenne maximale de 7m [118]. Dans [70], l'écart type est utilisé comme critère pour comparer des MNTs construits à partir de contours et d'autres à partir de grilles, et évaluer l'apport de points d'appui dans le calcul du MNT.

Approche géostatistique

Il est possible d'utiliser les données de référence pour corriger l'estimation du MNT. On peut notamment adopter une approche géostatistique, qui consiste à interpréter l'élévation comme une fonction aléatoire $Z(x, y)$, dont le MNT $\hat{z}(x, y)$ représente une première estimation. Cette estimation peut alors être corrigée en krigeant les données supplémentaires $Z(x_i, y_i), i = 1, l$. Cela suppose l'étude préalable de la structure spatiale, à partir du variogramme expérimental [24]. La nouvelle estimation est alors donnée par les équations de krigeage :

$$Z^*(x, y) = \hat{z}(x, y) + \sum_{i=1}^l \lambda_i (Z(x_i, y_i) - \hat{z}(x_i, y_i)) \quad (2.44)$$

où $\lambda_i, i = 1, l$, sont les poids de krigeage. La variance d'estimation s'exprime également en fonction des poids de krigeage.

Si l'on s'intéresse non pas à l'altitude, mais à des variables dérivées, telles que la pente ou le chemin de plus courte distance entre deux points, on peut étudier par une méthode de Monte Carlo l'influence de l'incertitude relative au MNT sur ces variables. Cela suppose la génération de simulations conditionnelles de $Z(x, y)$, en utilisant un modèle stochastique et les données de référence $Z(x_i, y_i), i = 1, l$. Pour chaque simulation générée, on calcule alors la variable d'intérêt, pour laquelle on obtient ainsi un ensemble de simulation conditionnelles.

C'est par exemple l'approche suivie dans [36], où est analysée l'influence des erreurs du MNT sur le calcul de la région visible à partir d'un point. Ce travail met

bien en évidence la pertinence et l'intérêt des simulations stochastiques, mais on peut regretter que l'algorithme de simulation utilisé soit à la fois peu performant et critiquable d'un point de vue théorique. D'autres applications des méthodes de simulations stochastiques sont présentées dans [35] et dans [67]. Cependant, dans ces cas, aucune donnée de référence n'est utilisée; il s'agit de simulations non conditionnelles, dont le seul but est d'examiner l'influence de la variance et de la corrélation spatiale de l'erreur sur les résultats.

Ces méthodes sont très proches des méthodes purement géostatistiques utilisées pour construire des surfaces topographiques ou bathymétriques [21, 19, 24]. Dans ce cas, on dispose de mesures de l'élévation (ou de la profondeur) en certains points. L'estimation de la surface se fait alors par krigeage des points de donnée, après étude de la structure spatiale. Avec ce type de données, la difficulté réside dans l'existence de non-stationnarités, et en pratique des modèles intrinsèques ou plus généralement intrinsèques d'ordre k , $k \geq 1$, doivent être utilisés. Afin d'étudier la variabilité de variables déduites de la surface topographique, telles que le volume au-dessus d'un niveau [19], il est possible de générer des simulations conditionnelles de la surface. Les méthodes mentionnées précédemment pour mesurer l'incertitude d'un MNT à partir de données de référence diffèrent en ce que les données utilisés pour construire le MNT et celles utilisées pour étudier la variabilité ne sont pas les mêmes. En particulier, lors des simulations, le MNT original est utilisé comme terme de dérive auquel est ajouté un résidu simulé, supposé stationnaire.

L'utilisation de points de référence pour tester la validité d'un MNT, souvent considérée comme la seule méthode valable, est cependant peu réaliste en pratique, car l'obtention de tels points de référence est problématique : c'est justement pour éviter d'effectuer une campagne de terrain que l'on utilise des images satellitaires ou aériennes. Quant à l'utilisation d'un MNT de résolution spatiale plus forte, il existe peu de chances pour qu'un tel jeu de données soit disponible, sinon pourquoi avoir procédé à une nouvelle acquisition de qualité moindre? D'un point de vue plus théorique, on peut également avancer que la donnée de points de référence n'apporte une réelle information qu'en ces points, et que des hypothèses statistiques de stationnarité sont nécessaires pour étendre cette information au reste de la zone étudiée. Mais de telles hypothèses doivent être justifiées en se référant explicitement aux données initiales qui ont permis d'établir le MNT et à la méthode de calcul. C'est donc plutôt par rapport à ces données initiales que la variabilité du résultat doit être étudiée. Nous retenons cependant l'intérêt des méthodes de simulations stochastiques conditionnelles, qui nous semblent particulièrement pertinentes, en particulier quand il s'agit d'étudier une variable dérivée du MNT.

Approche fractale

Un autre type d'approches consiste à tester la cohérence du MNT respectivement à certains critères. Un de ces critères est le comportement fractal de la surface topographique. L'analyse des surfaces topographiques en termes de fractales, dont on peut trouver une description dans [123], revient en fait à supposer que la surface peut être modélisée par un processus gaussien à deux dimensions de variogramme linéaire ou puissance (ou processus brownien). Un tel variogramme a pour expression analytique :

$$\gamma(h_x, h_y) = \sigma^2(h_x^2 + h_y^2)^{\alpha/2} \quad (2.45)$$

Si l'on représente cette fonction en échelle logarithmique, l'équation précédente s'écrit :

$$\log(\gamma(h_x, h_y)) = \log \sigma^2 + \alpha \log \left(\sqrt{h_x^2 + h_y^2} \right) \quad (2.46)$$

La dimension fractale se déduit du paramètre α et vaut $3 - \frac{\alpha}{2}$.

Dans [87], il est proposé de tester l'auto-similarité de la surface fractale en calculant deux estimations de la dimension fractale : une à partir de couples de points proches, entre 1 et 5 pixels, et une autre à partir de points plus éloignés, typiquement entre 10 et 30 pixels. Une différence significative peut éventuellement être révélatrice d'un biais dû à une interpolation des données. Bien sûr, ce test n'est valide que si la surface ne présente pas réellement de changements de structure spatiale. En réalité, l'hypothèse d'auto-similarité qui sous-tend l'approche fractale est une hypothèse forte, qui n'est pas vérifiée en pratique, ne serait-ce que parce que l'altitude est bornée [86]. Il convient donc d'être vigilant quant à l'emploi d'un tel critère.

Une approche semblable consiste à tester l'isotropie de la surface à courtes distances, soit par un calcul de la dimension fractale dans différentes directions, soit par comparaison directe des variogrammes [87, 15]. L'hypothèse sous-jacente est qu'à courtes distances l'existence d'une anisotropie est révélatrice d'un biais dans le calcul du MNT. Là encore, l'efficacité du test dépend grandement de la pertinence de cette hypothèse.

Bien que l'analyse fractale des modèles numériques de terrain ne semble plus consister aujourd'hui un axe de recherche porteur, on peut retenir cependant l'intérêt de la modélisation des surfaces topographiques à l'aide de processus gaussiens non stationnaires, de variogramme linéaire ou puissance.

2.2.2 Etude de la mise en correspondance

Dans la reconstitution de la géométrie d'une scène à partir d'un couple d'images stéréoscopiques, l'incertitude est due à des erreurs ou à des approximations à chaque étape du processus décrit à la section 2.1 : calibration, rectification, mise

en correspondance et triangulation. Cependant, les natures de ces erreurs sont très différentes. En effet, les étapes de calibration, de rectification et de triangulation reposent sur des relations géométriques bien maîtrisées. Même lorsque celles-ci ne sont que des approximations des véritables contraintes géométriques, les erreurs induites par ces approximations peuvent être contrôlées et il est possible d'en donner l'ordre de grandeur. Ainsi, la précision de l'estimation par moindres carrés des paramètres du système à partir de points de contrôle dépend du nombre de points utilisés, sous l'hypothèse que les équations de la géométrie sont adaptées. Il est toujours possible d'augmenter ce nombre pour obtenir des estimations plus précises. De même, un certain nombre de travaux se sont intéressés aux problèmes dus à la quantification des images : en effet, une image n'est pas une information continue, et les coordonnées des objets représentés dans un pixel ne sont connues qu'avec une précision de l'ordre de la moitié de l'intervalle d'échantillonnage. Cependant, on peut montrer que ces erreurs ne dépendent que des paramètres du système stéréoscopique [13, 101]. En particulier, il est possible de les diminuer en modifiant ces paramètres : ainsi, il est bien connu qu'une augmentation du produit bf (*baseline* \times distance focale) permet d'obtenir des estimations plus précises [101].

Les erreurs issues de la mise en correspondance sont d'une tout autre nature, et ne peuvent pas être quantifiées *a priori*. En effet, celles-ci dépendent de l'information contenue dans les images, ce qui rend beaucoup plus délicate leur analyse. Les problèmes propres aux méthodes de mise en correspondance par corrélation sont depuis longtemps bien identifiées [55] : bruit d'acquisition, manque de contraste dans certaines régions, déformations géométriques induites par le relief et présence d'ambiguïtés du fait de la ressemblance entre plusieurs objets voisins ou de l'existence de motifs périodiques. Dans ces conditions, l'identification de couples de points homologues est problématique : soit un pixel et son homologue n'apparaissent plus identiques (conséquence du bruit ou des déformations), le cas extrême étant rencontré en présence d'occlusions où l'homologue n'existe pas, soit le pixel homologue ne peut plus être distingué des pixels voisins (manque de contraste ou périodicité).

Retour sur les algorithmes de corrélation

Les remarques précédentes expliquent pourquoi le problème de la mise en correspondance, et donc celui du calcul de la disparité, a été l'objet d'un nombre important de travaux. Cependant, la réponse a le plus souvent consisté à proposer des algorithmes moins sensibles à ces difficultés, plutôt qu'à traiter le problème de l'incertitude en tant que tel.

Ainsi, pour réduire la sensibilité de l'algorithme de corrélation au bruit des images et limiter les erreurs dues aux fortes variations de disparité à l'intérieur de la fenêtre de corrélation, une procédure d'adaptation locale de la taille de la fenêtre a

été proposée par Kanade et Okutomi dans [60]. Dans cette approche, un modèle stochastique de disparité est proposé : à l'intérieur de la fenêtre de corrélation, la disparité est décrite par un processus gaussien de variogramme linéaire. Kanade et Okutomi montrent que localement, la distribution conditionnelle de la disparité prend la forme d'une distribution gaussienne, dont la moyenne et l'écart type dépendent de la valeur de la disparité au centre de la fenêtre, des intensités des images et de leurs dérivées, ainsi que de la forme de la fenêtre de corrélation. Ils proposent alors un algorithme de calcul itératif optimisant la taille de la fenêtre de corrélation de manière à minimiser la variance d'estimation. Cet optimum est un compromis local entre une faible taille de fenêtre de corrélation pour tenir compte des variations fortes de disparité et une grande taille pour pallier à la présence de bruit. L'algorithme proposé fournit également une valeur de l'écart type comme mesure de l'incertitude. L'approche est limitée en ce sens que cet écart type traduit principalement l'incertitude liée à la variabilité de la disparité à l'intérieur de la fenêtre, mais ne rend pas compte de celle due à l'ambiguïté de l'information radiométrique. Cependant, cette méthode présente l'intérêt d'analyser la variabilité de la disparité en termes de distribution conditionnelle au couple stéréoscopique. Le modèle fait intervenir un terme de vraisemblance décrivant l'écart entre les deux images après correction de la disparité :

$$I_1(u, v) = I_2(u + d(u, v), v) + \eta(u, v) \quad (2.47)$$

où le bruit $\eta(u, v)$ est un processus gaussien non corrélé :

$$\eta(u, v) \sim \mathcal{N}(0, \sigma_\eta^2) \quad (2.48)$$

La structure spatiale de la disparité repose sur une hypothèse d'accroissements gaussiens de variance proportionnelle à la distance (soit un modèle intrinsèque de variogramme linéaire) :

$$D(u + h_u, v + h_v) - D(u, v) \sim \mathcal{N}(0, \sigma_p^2 \sqrt{h_u^2 + h_v^2}) \quad (2.49)$$

où le facteur σ_p est constant à l'intérieur de la fenêtre de corrélation.

Approche bayésienne

Une approche purement statistique du problème consiste à envisager le calcul de la disparité en termes d'estimation bayésienne. On peut notamment se référer à [114] pour une présentation assez générale des méthodes bayésiennes en vision. Cette approche repose sur le calcul de la distribution *a posteriori* (ou conditionnelle) de la disparité D connaissant le couple stéréoscopique $Y = (I_1, I_2)$. La densité de cette distribution, notée $\pi_D(d|y)$, s'exprime à l'aide de la relation de Bayes :

$$\pi_D(d|y) = \frac{\pi_Y(y|d)}{\pi_Y(y)} \pi_D(d) \quad (2.50)$$

$\pi_D(d)$ est le terme de loi *a priori*, tandis que $\pi_Y(y|d)$, noté généralement $L(y|d)$, est le terme de vraisemblance. La densité marginale du couple $\pi_Y(y)$ est une constante du problème (on travaille à couple fixé). La connaissance de $\pi_D(d|y)$ à un facteur multiplicatif près étant suffisante, le modèle est souvent spécifié sous la forme :

$$\pi_D(d|y) \propto L(y|d)\pi_D(d) \quad (2.51)$$

L'estimation bayésienne procède donc en deux temps : il convient tout d'abord de spécifier le modèle de l'équation 2.51, puis de calculer l'estimateur retenu. En général, celui-ci est le maximum *a posteriori* (MAP) :

$$\hat{d} = \arg \max_d \pi_D(d|y) \quad (2.52)$$

Il est souvent plus pratique de faire intervenir le log des distributions, si bien que l'équation précédente devient :

$$\hat{d} = \arg \max_d (\log L(y|d) + \log \pi_D(d)) \quad (2.53)$$

Sorti du contexte bayésien, cette expression est parfaitement similaire à l'expression générale 2.25 donnée au paragraphe 2.1. Le critère à optimiser est la somme $\log L(y|d) + \log \pi_D(d)$: le premier terme traduit l'écart entre les deux images à disparité fixée, tandis que le second terme peut être interprété comme un terme de régularisation. Le terme de loi *a priori* introduit une contrainte spatiale forte : la solution du problème d'optimisation 2.53 est globale (i.e. relative à l'ensemble du couple), contrairement aux solutions obtenues par corrélation, où chaque ligne est traitée indépendamment des autres. L'avantage de l'approche bayésienne est qu'elle permet de décrire l'incertitude liée à l'estimation : en effet, celle-ci est décrite par la donnée de la loi *a posteriori*, qui décrit les écarts au MAP et leur probabilité d'occurrence.

Un des travaux les plus aboutis en analyse bayésienne appliquée à la vision stéréoscopique est certainement celui apporté par Belhumeur [5]. Le modèle qu'il développe est plutôt orienté vers des applications en vision par ordinateur, où la scène représentée est typiquement l'intérieur d'une pièce meublée. Les problèmes qu'il aborde, en particulier celui des occlusions, diffèrent ainsi quelque peu de ceux rencontrés en imagerie SPOT, mais la méthode employée est elle parfaitement généralisable. Notons qu'il choisit de travailler dans l'image cyclopéenne pour définir la disparité, mais ce choix n'a pas de réelle incidence sur le modèle qu'il utilise. Le modèle de vraisemblance $L(y|d)$ repose sur une hypothèse de bruit additif gaussien non corrélé, identique à 2.47. L'apport essentiel consiste en l'analyse du terme de loi *a priori*. Cette analyse procède en plusieurs étapes, chaque étape correspondant à un degré de complexité supplémentaire de la scène observée : modélisation d'une surface, modélisation des contours d'un objet, et modélisation des points de non-dérivabilité d'une surface.

Le modèle *a priori* est construit à partir d'un processus mono-dimensionnel (à

savoir le long des lignes), étendu heuristiquement à deux dimensions. Une même surface est décrite par un processus gaussien de variogramme linéaire (mouvement brownien). Les discontinuités de disparité aux contours des objets sont modélisées par des sauts d'amplitude uniforme, l'emplacement du saut étant lui-même modélisé par un processus de Poisson. Soit $N(h)$ l'indicatrice de comptage d'un intervalle de longueur h associée au processus de Poisson, la différence des disparités entre les points $(u, v)^\top$ et $(u + h, v)^\top$ suit le modèle :

$$D(u + h, v) - D(u, v) = \sqrt{h}\sigma Z + \sum_{j=1}^{N(h)} W_j \quad (2.54)$$

où Z est une variable normale centrée réduite, et W_j est une variable uniforme sur un intervalle fixé $[-\Delta, \Delta]$. Z et les W_j sont mutuellement indépendants. Enfin, afin de prendre en compte l'existence de ruptures de pente, le modèle de l'équation 2.54 est utilisé pour modéliser à la fois une première approximation de la dérivée de la disparité et le résidu entre l'intégrale de cette approximation et la disparité. On peut émettre quelques réserves sur l'emploi simultané d'un processus modélisant les ruptures de pente et d'un processus gaussien de variogramme linéaire modélisant les surfaces. En effet, les réalisations d'un processus gaussien de variogramme linéaire présentent presque partout des ruptures de pente, puisque celui-ci est non dérivable. Si on désire prendre en compte explicitement ces ruptures de pente à l'aide d'un processus ponctuel, une structure spatiale plus régulière pour les surfaces est certainement plus appropriée.

La disparité est alors estimée à partir du MAP, calculé par une méthode de programmation dynamique en deux étapes : tout d'abord, une approximation mono-dimensionnelle du modèle est utilisée (qui revient à supposer que les disparités d'une ligne à l'autre sont indépendantes), puis l'optimum ainsi calculé est corrigé en tenant compte de la structure bi-dimensionnelle.

Bien que cette approche présente une modélisation explicite de l'incertitude associée au calcul de la disparité, elle n'est utilisée que dans un but d'estimation, et aucun critère de qualité de cette estimation n'est proposé à ce stade.

Il existe un certain nombre de travaux reprenant une analyse bayésienne, et s'appuyant sur des modèles similaires. Le modèle de vraisemblance repose généralement sur une hypothèse de bruit additif gaussien non corrélé [114, 18, 26], la seule exception rencontrée étant un modèle gaussien perturbé proposé par [103], dont l'expression est :

$$L(y|d) \propto \prod_{u,v} (1 - \epsilon) \exp\left(-\frac{(I_1(u, v) - I_2(u + d(u, v), v))^2}{2\sigma_l^2}\right) + \epsilon \quad (2.55)$$

où ϵ est un paramètre permettant de modéliser l'existence d'un résidu non gaussien, soit que le point homologue n'existe pas (occlusion), soit que la différence d'intensité soit anormalement élevée (par exemple en cas de réflexion spéculaire).

Dans tous les cas, le modèle *a priori* suppose une distribution gaussienne de la disparité. Certains travaux [114, 18, 102] ont privilégié une approche par champs de Markov, relativement populaires en analyse d'images depuis l'article des frères Geman [44] et l'utilisation de l'échantillonneur de Gibbs associé à une méthode de recuit simulé. Se reporter à l'annexe C pour une introduction aux champs de Markov. Leur intérêt réside notamment dans l'expression particulièrement simple des lois conditionnelles. Dans le cas d'un champ de Markov d'ordre 1 à deux dimensions, pour lequel le voisinage d'un point est limité aux 4 plus proches voisins, la matrice de covariance inverse est fortement creuse, seulement 5 diagonales sont non nulles. Les calculs sont donc très rapides, ce qui explique en partie l'emploi fréquent de ce type de modèles.

Il nous paraît cependant moins intuitif de modéliser la matrice de covariance inverse, plutôt que la matrice de covariance elle-même. De plus, bien souvent, pour des raisons de limitation du temps de calcul, des voisinages réduits sont utilisés (souvent limités aux 4 pixels adjacents, aux 8 plus proches voisins). Or, ces modèles sont quelque peu limités, et la structure spatiale résultante peut parfois exhiber de manière assez surprenante de longues portées de corrélation [8]. Surtout, la modélisation par champs de Markov repose simplement sur une analyse des voisinages, sans aucune notion de distance entre les points. Ceci peut se révéler problématique dès que l'on a besoin de détruire les relations de voisinage, soit que l'on procède à un échantillonnage de la grille initiale ou qu'un effet de support intervient. En conclusion, la modélisation par des fonctions aléatoires nous semble plus souple. Remarquons que dans aucun des travaux cités précédemment, une étude sur le choix du modèle *a priori* n'est menée : dans tous les cas, la structure spatiale est postulée, qu'il s'agisse d'une hypothèse de type mouvement brownien ou de champ de Markov d'ordre 1 ou 2. Seuls les paramètres sont éventuellement estimés à l'aide des données.

Méthodes de Monte Carlo

Les travaux cités précédemment portent principalement sur la définition d'un modèle mathématique, les méthodes de calcul étant généralement analytiques. Plus récemment, et notamment en estimation de la structure d'une scène à partir de séquences d'images, certains auteurs ont proposé l'utilisation de méthodes de Monte Carlo par chaînes de Markov couplées à des modèles bayésiens. Le but est d'estimer la distribution *a posteriori* (et non uniquement le MAP) des paramètres de la scène en construisant une chaîne de Markov qui admet pour loi stationnaire la loi cible ; les états de la chaîne permettent alors de calculer par intégrale de Monte Carlo l'espérance de toute fonctionnelle, comme une moyenne, un écart type ou un histogramme.

Dans [27], un algorithme de type Metropolis-Hastings est utilisé pour estimer la distribution d'une rotation à partir de deux images bruitées d'un hexagone. Le

calcul permet notamment de mettre en évidence l'influence de l'information *a priori* sur la scène : lorsque la correspondance entre les sommets de l'hexagone d'une image à l'autre est connue, la distribution *a posteriori* est unimodale : seul le bruit est source d'incertitude. En revanche, lorsque la correspondance est inconnue, toutes les rotations modulo 60° sont équivalentes, et la distribution *a posteriori* est multimodale. D'autres exemples d'application sont présentés dans [38, 31].

A la différence des méthodes d'estimation bayésienne, qui reposent sur une procédure d'optimisation, les méthodes de Monte Carlo ont explicitement pour objet l'analyse de la variabilité de l'estimation, par le biais du calcul de la loi *a posteriori*. Elles s'inscrivent donc directement dans la problématique que nous nous sommes fixée.

Chapitre 3

Modélisation bayésienne

3.1 Un problème d'échantillonnage

Au chapitre 2, la disparité a été définie d'un point de vue géométrique (paragraphe 2.1.4) : étant donné un point physique M et ses images m_1 et m_2 , de coordonnées $(u_1, v_1)^\top$ et $(u_2, v_2)^\top$ en géométrie épipolaire (avec $v_1 = v_2$), on appelle disparité la différence entre les abscisses u_1 et u_2 , soit :

$$d(u_1, v_1) = u_2 - u_1 \quad (3.1)$$

Cette définition correspond à l'approche directe du problème où le point physique M est connu, et ses images m_1 et m_2 parfaitement définies. Autrement dit, l'ensemble des couples de points homologues entre les images I_1 et I_2 est totalement identifié.

En pratique, la seule information dont on dispose est la donnée du couple d'images (I_1, I_2) , et il faut à partir de cette information reconstituer l'ensemble des couples de points homologues, ou, ce qui est équivalent, calculer la carte de disparité. La difficulté du problème réside dans la relation entre l'information disponible et la variable à estimer : il n'existe pas en effet de bijection entre la radiométrie et la disparité. Autrement dit, l'estimation de la disparité à partir d'un couple d'images stéréoscopiques est un *problème inverse*. A ce type de problèmes est associée une solution définie par une procédure d'optimisation : on se fixe un critère de vraisemblance dont on cherche l'optimum. En vision stéréoscopique, on utilise la propriété de similarité que doivent vérifier les deux images à disparité fixée pour construire un tel critère. Ainsi, si $C_{u,v}(d)$ est une mesure de la similarité entre I_1 et I_2 au point $(u, v)^\top$ pour une disparité d , une estimation de la disparité est :

$$\hat{d}(u, v) = \arg \max_d C_{u,v}(d) \quad (3.2)$$

On est donc passé d'une définition purement géométrique 3.1 à la résolution d'un problème inverse par une procédure d'optimisation 3.2.



FIG. 3.1 – Extrait du couple stéréoscopique d’Aix-Marseille. Exemples de différence radiométrique importante due à des problèmes de réflexion, ici sur des bâtiments.

Comme dans tout problème inverse, la qualité de l’estimation dépend principalement de l’information contenue dans les données, ici le couple stéréoscopique, et non de l’algorithme de calcul utilisé. Ainsi, la mise en correspondance suppose que pour tout point de l’image de référence, on peut reconnaître son homologue dans l’autre image. Cette hypothèse est vérifiée seulement si les deux conditions suivantes sont réalisées :

- les deux points homologues ont une radiométrie semblable,
- dans une même image, on peut distinguer un pixel de ses voisins.

La première condition peut être infirmée lorsque le bruit lié aux capteurs est significatif, lorsque l’hypothèse lambertienne est fortement violée (cas des réflexions spéculaires par exemple) ou lorsque la scène observée a changé entre les deux acquisitions (problème de diachronisme, résolu sur le satellite SPOT 5 qui dispose de 3 capteurs observant dans des directions différentes et pouvant acquérir des images stéréoscopiques quasiment simultanées). Mais elle est également mise en défaut si les déformations géométriques induites par la prise de vue sont importantes, auquel cas les deux scènes observées ne sont pas rigoureusement identiques. Le cas extrême est rencontré en cas d’occlusion, où un élément de la scène n’est visible que dans une seule image. Les figures 3.1 à 3.3 illustrent ces différentes situations.

La seconde condition revient à supposer que l’information radiométrique est discriminante. Par exemple, les zones uniformes et peu texturées, comme des zones d’ombre, des champs, des forêts ou des surfaces recouvertes de neige, sont problématiques (des exemples sont donnés aux figures 3.3 et 3.4). Il en est de même dans le cas de motifs périodiques.

Remarquons que si on peut songer à réduire l’incertitude due aux problèmes de



FIG. 3.2 – Extrait du couple stéréoscopique d’Aix-Marseille. Exemples de différence radiométrique importante due au changement de la scène (diachronisme), ici dans le cas de champs.

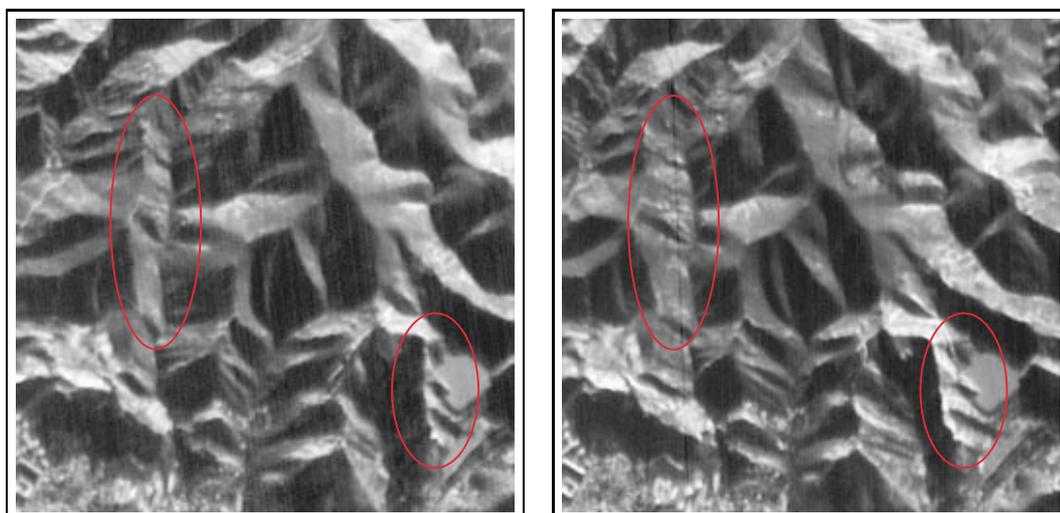


FIG. 3.3 – Extrait du couple stéréoscopique d’Aix-Marseille. Exemples de déformations géométriques importantes, ici dans une région montagneuse. A noter également la présence d’ombres dues au relief.

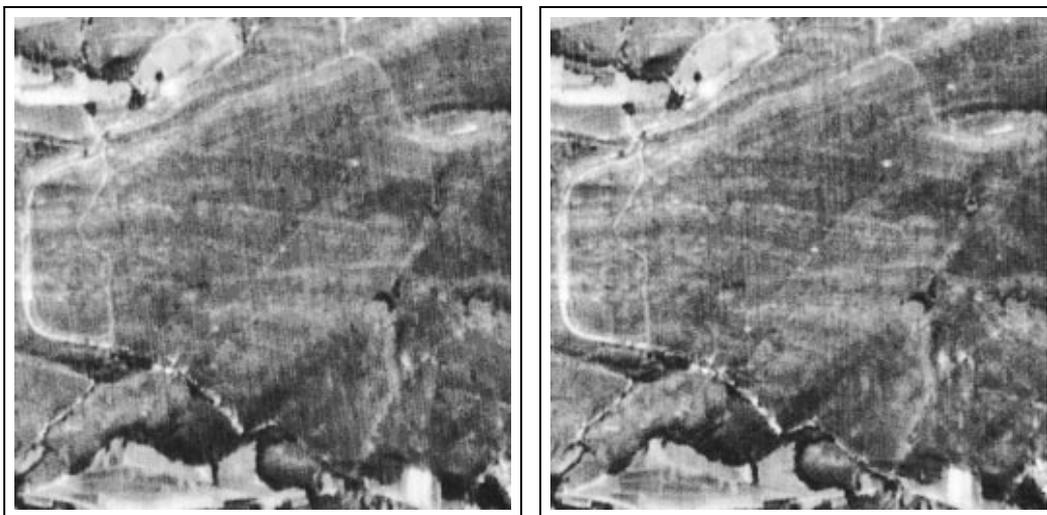


FIG. 3.4 – Extrait du couple stéréoscopique d’Aix-Marseille. Exemple de zone uniforme peu texturée (végétation).

bruit ou aux déformations géométriques, en améliorant la qualité des capteurs ou en modifiant le système de prise de vue, les ambiguïtés présentes dans le couple stéréoscopique sont directement liées au contenu de la scène observée. L’incertitude qu’elles génèrent est donc intrinsèque au problème de la vision stéréoscopique.

3.1.1 Loi *a posteriori*, relation de Bayes

Dans un cadre bayésien, l’estimation de la disparité peut être abordée en termes de loi *a posteriori* : parmi les différentes valeurs de la disparité, on s’intéresse à celles qui se réalisent lorsque l’on observe le couple stéréoscopique $Y = (I_1, I_2)$. Cet ensemble est entièrement décrit par la distribution conditionnelle ou *a posteriori* $\pi_D(\cdot|y)$ de la disparité D^1 connaissant le couple stéréoscopique $Y = y$. Ainsi, chaque carte de disparité se voit attribuer une pondération selon sa probabilité d’occurrence connaissant le couple Y .

L’application de méthodes bayésiennes en imagerie, en particulier à des problèmes de restauration [44, 8, 17, 12, 3, 49], est relativement ancienne ; ce n’est que plus récemment qu’elles ont fait leur apparition en vision par ordinateur [114, 38], et notamment en vision stéréoscopique [18, 5, 6, 102].

Comme énoncé au paragraphe 2.2.2, la relation de Bayes permet d’obtenir une expression de la densité *a posteriori*, $\pi_D(d|y)$:

$$\pi_D(d|y) \propto L(y|d)\pi_D(d) \quad (3.3)$$

¹La notation avec une majuscule fait référence à la variable aléatoire par opposition à la réalisation, notée avec une minuscule.

$\pi_D(d)$ est la densité de la loi *a priori*, ou marginale, de la disparité, tandis que $L(y|d)$, appelée vraisemblance, est la densité conditionnelle du couple stéréoscopique à disparité fixée. La constante de proportionnalité, égale à la loi marginale du couple, peut être obtenue par intégration, mais est en réalité sans grande importance. L'intérêt de la relation de Bayes est d'exprimer le problème inverse en fonction du problème direct par l'intermédiaire de la vraisemblance. Il est donc plus aisé d'exprimer la forme de $\pi_D(d|y)$ à partir de l'expression 3.3.

Dans les exemples étudiés, la disparité est définie sur une grille de taille $n = n_x \times n_y$. Posons alors $E = \mathbb{R}^n$; en tant que fonction de d , les densités π_D , L et $\pi_D(\cdot|y)$ sont définies sur E . La mesure de référence sur cet espace est la mesure de Lebesgue. On choisit comme tribu \mathcal{E} la tribu des boréliens. Le problème qui se pose maintenant est l'étude de la loi $\pi_D(\cdot|y)$.

3.1.2 Echantillonnage et méthodes de Monte Carlo

D'un point de vue pratique, l'étude de la loi *a posteriori* $\pi_D(\cdot|y)$ se résume principalement au calcul d'espérance sous $\pi_D(\cdot|y)$ de fonctionnelles g , soit :

$$\mathbb{E}_{\pi_D(\cdot|y)}(g) = \int_E g(z)\pi_D(z|y)dz \quad (3.4)$$

Cette formulation permet notamment de calculer tous les moments de la loi $\pi_D(\cdot|y)$. Pour $g(z) = z$, on obtient le moment d'ordre 1, que l'on identifie ici à l'espérance conditionnelle de D sachant $Y = y$:

$$\int_E z\pi_D(z|y)dz = \mathbb{E}(D|Y = y) \quad (3.5)$$

De même, la variance de D conditionnellement à $Y = y$ se calcule en prenant $g(z) = z^2$:

$$\int_E z^2\pi_D(z|y)dz - \left(\int_E z\pi_D(z|y)dz \right)^2 = \text{Var}(D|Y = y) \quad (3.6)$$

De plus, la probabilité d'un événement quelconque $A \in \mathcal{E}$ se calcule également par intégrale en considérant la fonction caractéristique de A :

$$\pi_D(A|y) = \int_A \pi_D(z|y)dz = \int_E 1_A(z)\pi_D(z|y)dz \quad (3.7)$$

Par exemple, fonction de répartition et fonction de répartition complémentaire en un point $(u, v)^\top$ s'obtiennent en calculant pour différentes valeurs seuil s les probabilités $P(D(u, v) \leq s)$ et $P(D(u, v) > s)$:

$$P(D(u, v) \leq s) = \int_{z \leq s} \pi_{D(u,v)}(z|y)dz \quad (3.8)$$

$$P(D(u, v) > s) = \int_{z > s} \pi_{D(u,v)}(z|y)dz \quad (3.9)$$

où $\pi_{D(u,v)}(z|y)$ est la densité marginale au point $(u, v)^\top$, qui s'obtient en intégrant $\pi_D(z|y)$ sur le sous-ensemble complémentaire à $(u, v)^\top$.

Dans le cas du traitement d'images, la dimension n de l'espace considéré (plusieurs milliers) est telle que le calcul numérique d'intégrales du type 3.4 n'est pas envisageable, d'autant plus que la densité $\pi_D(d|y)$ ne se factorise pas en termes de dimension moindre. Une méthode courante de résolution consiste alors à générer un ensemble de réalisations de loi $\pi_D(\cdot|y)$ et de calculer une moyenne. En effet, d'après la loi des grands nombres, si $\{z_i, i = 1, l\}$ est un échantillon de $\pi_D(\cdot|y)$, la moyenne arithmétique des $\{g(z_i), i = 1, l\}$ est une approximation de l'intégrale 3.4 :

$$S_l(g) = \frac{1}{l} \sum_{i=1}^l g(z_i) \xrightarrow{l \rightarrow +\infty} E_{\pi_D(\cdot|y)}(g) \quad (3.10)$$

On parle alors d'intégrale de Monte Carlo. Les méthodes de Monte Carlo sont couramment utilisées en analyse numérique [66] pour calculer des intégrales, et, couplées à la simulation par chaînes de Markov (cf. [45] pour un exposé très complet), ont redonné tout leur intérêt aux approches bayésiennes en permettant de résoudre les calculs souvent complexes. L'approximation dans l'équation 3.10 ne dépend que de la taille l de l'échantillon utilisé, et non de la dimension n de l'espace d'intégration. Une étude plus détaillée de ces méthodes est présentée au chapitre 4.

3.1.3 Bilan

Le calcul de la disparité à partir d'un couple stéréoscopique peut être interprété comme une procédure d'optimisation dans un problème inverse. A ce type de problèmes est associée une incertitude intrinsèque qui dépend directement de la nature de l'information, et non de l'algorithme de calcul. Le caractère nécessairement incomplet et ambigu de l'information radiométrique du couple constitue ainsi une sorte de barrière en terme de précision qu'aucun algorithme ne peut franchir. Le problème de l'incertitude ne peut donc être résolu en proposant un nouvel algorithme de calcul de la disparité, mais en étant abordé directement. Dans un contexte bayésien, cette incertitude est décrite par la loi *a posteriori* $\pi_D(\cdot|y)$ de la disparité D connaissant le couple stéréoscopique $Y = y$. L'étude quantitative de cette incertitude s'exprime au moyen d'intégrales sous la densité $\pi_D(d|y)$.

Par le biais des méthodes de Monte Carlo, ce calcul d'intégrales peut être ramené à un problème d'échantillonnage. Il s'agit de construire un ensemble de simulations $\{z_i, i = 1, l\}$ de loi $\pi_D(\cdot|y)$.

Le problème, tel que nous venons de le reformuler, peut ainsi être scindé en trois étapes distinctes :

- choix du modèle bayésien :

- un terme de loi *a priori*, qui spécifie la structure de la disparité, et notamment sa structure spatiale,
- un terme de vraisemblance, qui décrit la relation entre les 2 images du couples à disparité fixée,
- choix d’un algorithme de simulation,
- étude de la loi $\pi_D(.|y)$ à partir de fonctionnelles, dont on calcule par Monte Carlo l’intégrale.

Remarquons que l’approche proposée ne s’appuie aucunement sur l’algorithme de calcul de la disparité (de type corrélation dans notre cas), celui-ci est mené indépendamment. Ainsi, nous avons dissocié le problème de l’estimation de la disparité et celui de l’analyse de l’incertitude qui lui est associée, qui nécessitent chacun des méthodes de résolution spécifiques.

On peut objecter qu’*in fine* c’est l’incertitude relative au MNT, et non à la carte de disparité, qui importe. En fait, cela n’a pas d’incidence sur la méthode proposée, car le MNT est une fonction mathématique de la disparité. Pour chaque carte de disparité simulée, il suffit donc de calculer le MNT correspondant et d’effectuer l’intégrale de Monte Carlo exprimée en fonction de l’altitude. Cependant, dans la suite, par commodité, nous travaillons toujours sur les cartes de disparité.

Le choix du modèle probabiliste est traité au paragraphe suivant, tandis que l’aspect algorithmique des simulations, qui implique la construction d’une chaîne de Markov de loi stationnaire $\pi_D(.|y)$, est traité au chapitre 4. Les études pratiques de cartes de disparité sont abordées au chapitre 5.

3.2 Choix du modèle

Nous nous inspirons ici principalement des modèles bayésiens exposés au paragraphe 2.2 : modèle *a priori* gaussien, et hypothèse de bruit d’acquisition gaussien. Les modèles que nous présentons ici n’ont pas de portée générale et dépendent entièrement du type d’images utilisées. Dans la suite, c’est uniquement aux couples stéréoscopiques d’images panchromatiques SPOT que nous nous intéressons : celles-ci sont caractérisées par une résolution spatiale de l’ordre de 10 m au sol [48] (pour les générations antérieures à SPOT 5).

3.2.1 Modèle *a priori*

Le choix du modèle *a priori* dépend principalement du type d’application. Il est évident que les modèles utilisés en vision par ordinateur (comme le mouvement brownien proposé par Belhumeur [6]), où la scène observée est typiquement une pièce meublée ou un environnement urbain, ne peuvent être utilisés avec le même succès lorsqu’il s’agit de reconstituer la surface du globe terrestre. A la résolu-

tion des images SPOT, la carte de disparité apparaît relativement régulière, les microstructures étant moyennées spatialement. Subsistent cependant les discontinuités majeures, telles celles induites par des falaises, des lignes de crête, certains bâtiments, etc.

Compte tenu de ces remarques et de l'étude bibliographique présentée au paragraphe 2.2, on peut adopter un cadre gaussien pour décrire la disparité, le modèle étant entièrement spécifié par la moyenne et la covariance. Nous préférons la formulation en termes de fonctions aléatoires, définies en tout point de l'espace, à l'approche par champs de Markov, discrets par essence. En effet, la définition d'un champ de Markov repose sur la définition d'un voisinage (cf. annexe C), la notion de distance entre les points disparaissant. Or, celle-ci nous semble indispensable lorsque l'on travaille non plus sur l'image, mais sur la disparité, qui est assimilable à une surface topographique. Notamment, cette notion de distance est primordiale dès que l'on désire travailler à plusieurs échelles, ou ne calculer la disparité qu'en certains points particuliers.

Le problème réside principalement dans la formulation d'une hypothèse de stationnarité sur la fonction aléatoire D . Si le modèle stationnaire d'ordre 2 (moyenne constante et covariance stationnaire) a l'avantage de la simplicité, il est peu adapté dès que le champ étudié (i.e. la taille des images) est réduit (typiquement en deçà de la portée). Une hypothèse intrinsèque, où seuls les accroissements de la disparité sont supposés stationnaires, peut alors s'avérer plus appropriée. On peut également songer à envisager le phénomène comme la somme d'un terme basse fréquence (la "tendance générale") et d'un terme haute fréquence (des "variations locales"). Cette modélisation se traduit en géostatistique par l'introduction d'une dérive, qui est soit fournie par une variable externe ("dérive externe"), ou estimée à partir des données (cadre des fonctions aléatoires intrinsèques d'ordre k). Nous renvoyons à [73, 74, 121, 122, 76, 24] pour une discussion des hypothèses de stationnarité.

Pour résumer, trois choix se présentent en ce qui concerne la structure spatiale de la fonction aléatoire gaussienne D :

- modèle stationnaire d'ordre 2 : moyenne constante et covariance stationnaire,
- modèle stationnaire avec dérive : moyenne non constante et covariance stationnaire,
- modèle intrinsèque : défini par les accroissements de la variable, de moyenne nulle et de covariance stationnaire.

En notant C_p la matrice de covariance, $|C_p|$ son déterminant et m_p la moyenne (constante ou non), le modèle *a priori* s'écrit dans les 2 premiers cas :

$$\pi_D(d) \propto |C_p|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(d - m_p)^\top C_p^{-1}(d - m_p)\right) \quad (3.11)$$

Dans l'hypothèse intrinsèque, les accroissements du type $D(u+h_u, v+h_v) - D(u, v)$ sont stationnaires en (u, v) pour tout (h_u, h_v) . Posons donc pour (u_0, v_0) fixé $W(u, v) = D(u + u_0, v + v_0) - D(u, v)$. Si D est une fonction aléatoire gaussienne

intrinsèque, W est lui-même une fonction aléatoire gaussienne de covariance stationnaire C_p , et le vecteur W a pour densité :

$$\pi_W(w) \propto |C_p|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}w^\top C_p^{-1}w\right) \quad (3.12)$$

On retrouve la même expression qu'en 3.11, mais pour la variable W , et avec une moyenne nulle. Notons que la covariance C_p de W s'exprime en fonction du variogramme γ_p de D : $C_p(h_u, h_v) = \gamma_p(h_u - u_0, h_v - v_0) + \gamma_p(h_0 + u_0, h_v + v_0) - 2\gamma_p(h_u, h_v)$.

L'important problème du choix de l'hypothèse de stationnarité et des paramètres, notamment ceux de la covariance, est abordé plus loin.

La modélisation à partir de fonctions aléatoires gaussiennes permet une certaine flexibilité de part le choix du modèle de covariance : ainsi, un modèle dérivable à l'origine (comme le modèle cubique ou stable pour certaines valeurs des paramètres) est bien adapté à la description de surfaces régulières, tandis que des modèles seulement continus à l'origine (sphérique, exponentiel...) sont préférables dans le cas de fortes irrégularités de la surface.

Dans d'autres applications, notamment en vision par ordinateur ou en vision stéréoscopique haute résolution, les contours des objets de la scène (les bâtiments par exemple) jouent un rôle essentiel. Il devient alors nécessaire de modéliser explicitement les discontinuités de la carte de disparité, par exemple à l'aide d'un processus de Poisson [6]. Plus récemment, des modèles stochastiques géométriques qui prennent explicitement en compte la forme des objets, à l'aide de polygones par exemple, ont été utilisés dans des problématiques similaires : extraction de réseaux routiers [112] ou de bâtiments [42], reconstitution de vues en trois dimensions de bâtiments [31]. Ces modèles sont intéressants car ils utilisent réellement l'information dont on dispose *a priori* sur la forme des objets que l'on observe. En revanche, ils nécessitent une librairie assez riche de "briques" de base, capable de décrire efficacement l'ensemble des configurations rencontrées.

3.2.2 Modèle de vraisemblance

Le terme de vraisemblance décrit la relation entre les 2 images du couple après correction de la disparité. Nous pouvons donc commencer par proposer un modèle de formation pour chaque image, puis à partir de cette formulation construire un modèle pour le couple stéréoscopique.

La modélisation traditionnelle d'une image (i.e. de son intensité) consiste généralement à considérer séparément l'étape d'acquisition proprement dite (fonction de transfert du capteur) et l'étape d'enregistrement (le codage sous forme digitale). On peut tout d'abord admettre que l'intensité correspondant à un pixel est le résultat d'une moyenne spatiale, modélisée par une convolution entre l'intensité

ponctuelle de la scène et une fonction de pondération propre au capteur (*point spread function*). La moyenne obtenue est elle-même transformée pour donner l'image calibrée (par exemple une valeur entre 0 et 255). Enfin, on peut considérer l'existence d'un bruit venant corrompre l'acquisition. C'est l'approche suivie par exemple par les frères Geman dans [44]. Si on note I l'intensité initiale, ζ la fonction modélisant le capteur, ψ la transformation due au codage et que l'on suppose le bruit η additif, l'intensité observée I_{obs} s'écrit :

$$I_{\text{obs}} = \psi(I \star \zeta) + \eta \quad (3.13)$$

où \star désigne un produit de convolution.

En vision stéréoscopique, la modélisation s'enrichit du fait que l'on dispose de deux acquisitions de la même scène. Si l'on accepte l'hypothèse lambertienne (la quantité de lumière reçue par le capteur en provenance d'un point physique ne dépend pas de l'angle d'observation [34, 48]), l'intensité I est la même pour les deux images I_1 et I_2 . Si, de plus, les caractéristiques des capteurs sont identiques, on a pour un point physique de coordonnées $(x, y, z)^\top$:

$$I_1(u_1, v_1) = \psi(I(x, y, z) \star \zeta) + \eta_1(u_1, v_1) \quad (3.14)$$

$$I_2(u_2, v_2) = \psi(I(x, y, z) \star \zeta) + \eta_2(u_2, v_2) \quad (3.15)$$

Les deux équations précédentes traduisent le fait que les points images $(u_1, v_1)^\top$ et $(u_2, v_2)^\top$ sont homologues. En géométrie épipolaire, on a donc $v_1 = v_2$ et $u_2 = u_1 + d(u_1, v_1)$. En notant alors $(u, v) = (u_1, v_1)$, on obtient immédiatement :

$$I_1(u, v) - I_2(u + d(u, v), v) = \eta_{u,v} \quad (3.16)$$

où $\eta_{u,v} = \eta_1(u, v) - \eta_2(u + d(u, v), v)$.

Suivant cette description, le modèle du couple fait intervenir uniquement l'intensité résiduelle η des images. Choisir un modèle de vraisemblance revient donc à choisir un modèle pour l'intensité résiduelle η , soit :

$$L(y|d) = \pi_H(\eta) \quad (3.17)$$

Au paragraphe 3.4, on étudie un extrait du couple stéréoscopique d'Aix-Marseille (figure 3.5). La figure 3.6 représente notamment l'intensité résiduelle obtenue à partir du couple et de la carte de disparité calculée par corrélation. On observe principalement un comportement de type bruit blanc, mais aussi des motifs ordonnés (bords de bâtiments, route...) issus du couple original, et quelques points aberrants dus à des différences d'intensité fortes entre les deux images du couple.

Modèles gaussiens

On suppose ici que le résidu η suit une distribution gaussienne, qui est spécifiée par sa moyenne m_l et sa matrice de covariance C_l , soit :

$$L(y|d) = \pi_H(\eta) \propto |C_l|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\eta - m_l)^\top C_l^{-1}(\eta - m_l)\right) \quad (3.18)$$

Se pose alors la question de la structure spatiale du résidu η , et donc du modèle pour C_l .

Une approche habituelle, que l'on retrouve dans [18] et dans une certaine mesure dans [26, 6] (avec la différence que ces deux derniers modélisent explicitement les occlusions en ajoutant un terme spécifique), est de considérer un bruit gaussien non corrélé. La matrice C_l^{-1} de l'équation 3.18 est alors diagonale, et si l'on note σ_l^2 la variance, la fonction de vraisemblance s'écrit :

$$L(y|d) = \pi_H(\eta) \propto \frac{1}{\sigma_l^n} \exp\left(-\frac{1}{2\sigma_l^2} \sum_{u,v} (\eta_{u,v} - m_l)^2\right) \quad (3.19)$$

Ce modèle est acceptable lorsque l'intensité résiduelle ne présente que peu de structure spatiale. C'est généralement vrai dans le cas d'images obtenues avec le satellite SPOT (cf. figure 3.6), où les déformations géométriques sont assez faibles, et les occlusions rares.

Dans certaines situations, le modèle de bruit blanc pour le résidu η s'avère trop simpliste. Ainsi, lorsque les déformations géométriques sont importantes, la différence d'intensité entre les deux images, même après correction de la disparité, présente des motifs ordonnés, qui correspondent aux contours des objets de la scène (cf. figure 3.6). On constate alors que le terme résiduel η est spatialement corrélé.

On peut essayer de tenir compte de cette structure spatiale par un choix approprié de la covariance C_l . Cependant, le terme de vraisemblance doit être évalué fréquemment ; on a donc intérêt à choisir un modèle où la matrice inverse C_l^{-1} est creuse (nulle en dehors de certaines diagonales) afin de diminuer le temps de calcul. Un exemple est un modèle de champ markovien gaussien de voisinage réduit aux 4 plus proches voisins (ordre 1). De tels modèles sont utilisés couramment en analyse d'images [44, 8, 53, 68, 120]. L'intérêt des champs de Markov gaussiens est que l'on modélise directement la matrice de covariance inverse, le choix de la taille de voisinage revenant à spécifier le nombre de diagonales non nulles.

Notons ∂_{uv} l'ensemble des couples $(u+1, v)$, $(u-1, v)$, $(u, v+1)$, $(u, v-1)$ formant le voisinage du pixel (u, v) . Le modèle choisi pour η s'exprime en fonction des lois conditionnelles (cf. annexe C) :

$$\begin{aligned} \pi_H(\eta_{u,v} | \eta_{s,t}, (s,t) \neq (u,v)) &= \pi_H(\eta_{u,v} | \eta_{s,t}, (s,t) \in \partial_{uv}) \\ &= \frac{\sqrt{|\partial_{uv}|}}{\sqrt{2\pi\sigma_l}} \exp\left(-\frac{|\partial_{uv}|}{2\sigma_l^2} \left(\eta_{u,v} - m_l - \frac{\rho_l}{|\partial_{uv}|} \sum_{(s,t) \in \partial_{uv}} (\eta_{s,t} - m_l)\right)^2\right) \end{aligned} \quad (3.20)$$

$|\partial_{uv}|$ désigne le cardinal de ∂_{uv} : 4 en général, sauf sur les bords (3) et dans les coins (2). ρ_l est un paramètre de dépendance spatiale. Il est possible de tenir compte d'une anisotropie en introduisant deux coefficients ρ_{l_x} et ρ_{l_y} distincts suivant que l'on considère des voisins situés sur la même ligne ou sur la même colonne. Remarquons que dans ce modèle, la variance conditionnelle, égale à $\frac{\sigma_l^2}{|\partial_{uv}|}$, est proportionnelle au nombre de voisins. Il est également important de noter que la spécification du modèle est entièrement dépendante du pas de discrétisation utilisée pour calculer la disparité. Si l'on change de pas, l'hypothèse markovienne risque de ne plus être valable, ou du moins faut-il calculer les paramètres ρ_l et σ_l correspondants.

La loi spatiale se déduit des lois conditionnelles (les détails sont présentés en annexe C). On obtient alors l'expression suivante pour le terme de vraisemblance :

$$\begin{aligned}
L(y|d) = \pi_H(\eta) \propto & \exp\left(-\frac{1}{2\sigma_l^2} \sum_{u,v} |\partial_{uv}| (\eta_{u,v} - m_l)^2\right) \\
& - \frac{\rho_l}{\sigma_l^2} \sum_{u,v} (\eta_{u,v} - m_l)(\eta_{u+1,v} - m_l) \\
& - \frac{\rho_l}{\sigma_l^2} \sum_{u,v} (\eta_{u,v} - m_l)(\eta_{u,v+1} - m_l) \quad (3.21)
\end{aligned}$$

où les sommes sont prises sur les indices faisant partie du champ de définition.

Dans la suite, on utilise autant que possible l'expression générale 3.18 pour étudier ces deux modèles. Le premier modèle (équation 3.19) est appelé GMF-0, et le second (équation 3.21) GMF-1 (pour *Gaussian Markov Field* d'ordre 0 et 1 respectivement).

Modèle de mélange de distribution

Les modèles gaussiens précédents ne permettent pas de décrire le cas de résidu aberrant, i.e. de différences fortes d'intensité entre deux points homologues. En effet, dès que les surfaces ne sont plus lambertiennes (existence de réflexions spéculaires par exemple), ou lorsque les conditions d'éclairage varient entre les deux acquisitions, on ne peut plus supposer que l'intensité initiale I est la même pour les 2 images I_1 et I_2 . Ce phénomène est accentué lorsque les deux images sont prises à des dates espacées (ce qui peut être le cas pour les capteurs embarqués sur les satellites SPOT antérieurs à SPOT 5), auquel cas certains détails (les champs par exemple) ont changé. Une manière de prendre en compte ces valeurs aberrantes est de supposer que le résidu est composé de 2 types de points : le premier correspond à un bruit d'acquisition, comme dans les deux modèles précédents, tandis que le second est un terme parasite, rendant compte des fortes valeurs. Ce type de modèles est appelé modèle de mélange.

Le bruit d'acquisition suit une loi gaussienne, comme précédemment (équation

3.18). Pour le terme parasite, on peut supposer que n'importe quelle valeur de la forme $I_1(u_1, v_1) - I_2(u_2, v_2)$ peut apparaître. Ce terme suit donc la loi de la différence des images. En supposant pour simplifier que pour ces points, les intensités I_1 et I_2 sont indépendantes et suivent chacune une loi gaussienne, la loi de la différence est elle-même gaussienne, avec pour moyenne la différence des moyennes et pour variance la somme des variances. Sous ces hypothèses, le résidu peut être modélisé par un mélange de 2 lois gaussiennes ϕ_1 et ϕ_2 .

Notons p_{l_1} la proportion de points de loi ϕ_1 , et p_{l_2} celle des points de loi ϕ_2 , avec la condition $p_{l_1} + p_{l_2} = 1$. La densité du mélange pour un pixel est obtenue comme moyenne pondérée de ϕ_1 et ϕ_2 . Comme nous supposons de plus l'indépendance spatiale (ce qui est certes simplificateur), le modèle de vraisemblance prend l'expression :

$$L(y|d) = \pi_H(\eta) \propto \prod_{u,v} p_{l_1} \phi_1(\eta_{u,v}) + p_{l_2} \phi_2(\eta_{u,v}) \quad (3.22)$$

Ce modèle est proche du modèle utilisé dans [102] et présenté au paragraphe 2.2.2, dans lequel la loi gaussienne est perturbée par un terme de loi uniforme. Le choix d'une loi gaussienne pour modéliser le terme parasite est assez bien adapté dans le cas des images SPOT. Notons enfin que l'estimation des proportions p_{l_1} et p_{l_2} est délicate, car la classification des points du résidu dans l'une ou l'autre des catégories est inconnue. Ce problème est traité aux paragraphes 3.2.3 et 3.3.

Dans la suite, on utilise l'abréviation MIXT (en référence à *mixture*) pour désigner ce modèle de mélange.

3.2.3 Sur l'estimation des paramètres

Si le cadre général est fixé, reste le problème de l'estimation des paramètres du modèle. Nous proposons dans ce paragraphe d'adopter une démarche pragmatique qui a le mérite d'être simple à mettre en œuvre. Dans le cas particulier qui nous intéresse, on dispose en effet déjà d'une estimation de la disparité, qui a été calculée par corrélation. Notons \hat{d} cette estimation. Nous proposons d'utiliser \hat{d} pour estimer les paramètres du modèle.

Intéressons-nous tout d'abord au modèle *a priori*. A partir de la carte \hat{d} , il est possible de calculer une moyenne et une covariance expérimentale (ou éventuellement un variogramme expérimental), et donc d'ajuster un modèle (c'est l'approche classique de la géostatistique, cf. [24]). De plus, l'étude expérimentale de \hat{d} permet de préciser l'hypothèse de stationnarité : en traçant le variogramme expérimental, on peut valider l'hypothèse de stationnarité d'ordre 2 ou au contraire opter pour un modèle intrinsèque. Le cas échéant, l'utilisation d'une dérive peut être envisagée.

Considérons maintenant l'estimation des paramètres du terme de vraisemblance. Ce problème est délicat, car la donnée du couple (I_1, I_2) seule ne permet

pas de calculer le résidu η , et donc d'inférer les paramètres. La démarche précédente peut cependant s'appliquer : à partir de la disparité \hat{d} , on peut estimer un résidu selon la relation $\hat{\eta}_{u,v} = I_1(u,v) - I_2(u + \hat{d}(u,v), v)$. Les paramètres sont alors estimés par maximum de vraisemblance.

Ainsi, dans le cas du modèle GMF-0 (équation 3.19), on obtient les estimateurs suivants :

$$\hat{m}_l = \frac{1}{n} \sum_{u,v} \hat{\eta}_{u,v} \quad (3.23)$$

$$\hat{\sigma}_l^2 = \frac{1}{n} \sum_{u,v} (\hat{\eta}_{u,v} - \hat{m}_l)^2 \quad (3.24)$$

L'estimation des paramètres du modèle GMF-1 (équation 3.21) n'est pas immédiate car l'estimation de la constante de normalisation dans 3.21, qui fait intervenir σ_l et ρ_l , n'est pas triviale. On peut commencer par choisir comme estimateur de m_l la moyenne d'échantillonnage, soit :

$$\hat{m}_l = \frac{1}{n} \sum_{u,v} \hat{\eta}_{u,v} \quad (3.25)$$

Cet estimateur n'est plus le maximum de vraisemblance (il n'est pas non plus de variance minimale). Cependant, pour n grand, l'approximation est raisonnable, et permet de simplifier le problème en éliminant σ_l et ρ_l .

Pour estimer le paramètre de dépendance spatiale, on peut recourir à une méthode, dite de codage, proposée par Besag dans [7, 11] et qui consiste à utiliser directement les lois conditionnelles données par l'équation 3.20. Supposons que l'on fixe un pixel sur deux de sorte à former un damier d'échecs. Notons T cet ensemble, et \bar{T} son complémentaire. Conditionnellement aux points de \bar{T} , les points de T sont mutuellement indépendants, de densité :

$$\pi_H(\eta_{u,v}, (u,v) \in T | \eta_{s,t}, (s,t) \in \bar{T}) = \prod_{(u,v) \in T} \pi_H(\eta_{u,v} | \eta_{s,t}, (s,t) \in \partial_{uv}) \quad (3.26)$$

On peut donc associer à cette configuration un estimateur de ρ_l , dépendant uniquement des points de T . De même, on construit un second estimateur à partir des points de \bar{T} . En incluant dans T et dans \bar{T} seulement les points ayant 4 voisins (i.e. on exclue les bords de la grille), leur expression est :

$$\hat{\rho}_{l_T} = 4 \frac{\sum_{(u,v) \in T} (\hat{\eta}_{u,v} - \hat{m}_l) \left(\sum_{(s,t) \in \partial_{uv}} \hat{\eta}_{s,t} - 4\hat{m}_l \right)}{\sum_{(u,v) \in T} \left(\sum_{(s,t) \in \partial_{uv}} \hat{\eta}_{s,t} - 4\hat{m}_l \right)^2} \quad (3.27)$$

$$\hat{\rho}_{l_{\bar{T}}} = 4 \frac{\sum_{(u,v) \in \bar{T}} (\hat{\eta}_{u,v} - \hat{m}_l) \left(\sum_{(s,t) \in \partial_{uv}} \hat{\eta}_{s,t} - 4\hat{m}_l \right)}{\sum_{(u,v) \in \bar{T}} \left(\sum_{(s,t) \in \partial_{uv}} \hat{\eta}_{s,t} - 4\hat{m}_l \right)^2} \quad (3.28)$$

$$\hat{\rho}_l = \frac{\hat{\rho}_{l_T} + \hat{\rho}_{l_{\bar{T}}}}{2} \quad (3.29)$$

où $|T|$ et $|\overline{T}|$ désignent les cardinaux des ensembles T et \overline{T} respectivement. Puis, une fois ρ_l et m_l estimés, on peut déduire aisément un estimateur de σ_l :

$$\begin{aligned}\hat{\sigma}_l^2 &= \frac{1}{n} \sum_{u < n_x, v} (\hat{\eta}_{u,v} - \hat{m}_l)^2 + (\hat{\eta}_{u+1,v} - \hat{m}_l)^2 - 2\hat{\rho}_l(\hat{\eta}_{u,v} - \hat{m}_l)(\hat{\eta}_{u+1,v} - \hat{m}_l) + \\ &\quad \frac{1}{n} \sum_{u, v < n_y} (\hat{\eta}_{u,v} - \hat{m}_l)^2 + (\hat{\eta}_{u,v+1} - \hat{m}_l)^2 - 2\hat{\rho}_l(\hat{\eta}_{u,v} - \hat{m}_l)(\hat{\eta}_{u,v+1} - \hat{m}_l)\end{aligned}\tag{3.30}$$

Intéressons-nous pour finir au modèle MIXT, dont les paramètres sont les moyennes m_{l_1} et m_{l_2} et les écarts types σ_{l_1} et σ_{l_2} des 2 densités gaussiennes ϕ_1 et ϕ_2 , ainsi que les proportions p_{l_1} et p_{l_2} . Même lorsque l'on dispose d'un estimateur $\hat{\eta}$ du résidu, l'estimation des paramètres du modèle de mélange par maximum de vraisemblance n'est pas possible [119]. Ce problème est en réalité un problème à variables latentes : si l'on connaissait la variable de classification W_l telle que $W_l(u, v) = i$ si le point $(u, v)^\top$ a pour densité ϕ_i , $i = 1, 2$, l'estimation serait beaucoup plus simple. Mais ce n'est pas le cas.

Nous proposons la démarche suivante, purement empirique. D'après la description du modèle, les points de densité ϕ_1 sont majoritaires. En première approximation, on peut donc estimer les paramètres de ϕ_1 à partir des équations 3.23 et 3.24, en utilisant tous les points du résidu $\hat{\eta}$. Comme indiqué précédemment, les paramètres de ϕ_2 sont estimés à partir des statistiques des images I_1 et I_2 : la moyenne \hat{m}_{l_2} est égale à la différence des moyennes, et la variance $\hat{\sigma}_{l_2}^2$ est égale à la somme des variances. Restent les proportions p_{l_1} et p_{l_2} . Remarquons qu'un point suit la distribution i avec la vraisemblance ϕ_i . Pour chaque point $(u, v)^\top$, on peut donc calculer le rapport $\frac{\phi_1(\eta_{u,v})}{\phi_2(\eta_{u,v})}$: s'il est supérieur à 1, le point fait partie de la première population. On calcule ainsi l'estimateur \hat{p}_{l_1} , puis $\hat{p}_{l_2} = 1 - \hat{p}_{l_1}$. Une méthode d'estimation plus rigoureuse reposant sur l'introduction explicite de la variable de classification est présentée au paragraphe 3.3.

3.3 Randomisation des paramètres, formulation bayésienne

Revenons au problème de l'estimation des paramètres. L'approche proposée au paragraphe 3.2.3 reposant sur l'utilisation de la carte de disparité calculée par corrélation présente certains inconvénients. Cet estimateur de la disparité n'a pas en effet les bonnes propriétés classiques d'un estimateur statistique. En particulier, on peut imaginer que la structure locale de la disparité est biaisée par l'utilisation de la fenêtre de corrélation qui a tendance à introduire une dépendance spatiale artificielle. De plus, si l'on adopte un point de vue bayésien, il est incorrect d'utiliser à deux reprises l'information initiale (le couple (I_1, I_2)) : une première fois pour

choisir le modèle, une seconde fois pour contraindre les simulations. Il convient en effet dans une telle formulation de tenir compte explicitement du manque d'information sur les paramètres, et de les traiter comme des inconnues devant être estimées. Une solution consiste donc à randomiser les paramètres du modèle, et à les simuler en même temps que la disparité.

Notons Θ_p les paramètres (randomisés) de la loi *a priori*, et Θ_l ceux relatifs au terme de vraisemblance. La loi *a posteriori* qu'il convient alors de simuler s'écrit :

$$\pi_{D, \Theta_p, \Theta_l}(d, \theta_p, \theta_l | y) \propto L(y | \theta_l, d) \pi_D(d | \theta_p) \pi_{\Theta_p}(\theta_p) \pi_{\Theta_l}(\theta_l) \quad (3.31)$$

où $\pi_{\Theta_p}(\theta_p)$ et $\pi_{\Theta_l}(\theta_l)$ désignent les densités des loi *a priori* des vecteurs Θ_p et Θ_l (on a supposé légitimement l'indépendance *a priori* entre Θ_p et Θ_l).

Anticipons quelque peu sur le chapitre 4 : la simulation d'une telle loi se fait aisément à l'aide de l'échantillonneur de Gibbs (cf. paragraphe 4.2.1). Le principe est de simuler tour à tour chacune des variables D , Θ_p et Θ_l selon les lois conditionnelles :

$$\pi_D(d | y, \theta_p, \theta_l) \propto L(y | \theta_l, d) \pi_D(d | \theta_p) \quad (3.32)$$

$$\pi_{\Theta_p}(\theta_p | d, \theta_l, y) \propto \pi_D(d | \theta_p) \pi_{\Theta_p}(\theta_p) \quad (3.33)$$

$$\pi_{\Theta_l}(\theta_l | y, d, \theta_p) \propto L(y | \theta_l, d) \pi_{\Theta_l}(\theta_l) \quad (3.34)$$

$\pi_D(d | y, \theta_p, \theta_l)$ correspond à la densité $\pi_D(d | y)$ de l'équation 3.3 dans laquelle on a simplement explicité les paramètres du modèle θ_p et θ_l . Les densités $\pi_{\Theta_p}(\theta_p | d, \theta_l, y)$ et $\pi_{\Theta_l}(\theta_l | y, d, \theta_p)$ sont les densités des lois conditionnelles des paramètres. On remarque notamment que $\pi_D(d | y, \theta_p, \theta_l)$ dépend bien entièrement de θ_p , θ_l et y , mais $\pi_{\Theta_l}(\theta_l | y, d, \theta_p)$ ne dépend que de d et y , et $\pi_{\Theta_p}(\theta_p | d, \theta_l, y)$ de d . Ceci est dû au choix du modèle hiérarchique 3.31. Pour alléger les notations, nous notons donc par la suite les densités conditionnelles des lois de Θ_p et Θ_l $\pi_{\Theta_p}(\theta_p | d)$ et $\pi_{\Theta_l}(\theta_l | y, d)$ respectivement. Nous allons expliciter ces densités pour chacun des modèles du paragraphe 3.2.

3.3.1 Paramètres du modèle *a priori*

Pour ce qui est du terme *a priori*, nous supposons que le modèle de covariance est connu et nous nous intéressons uniquement aux deux paramètres que sont la moyenne m_p et la variance σ_p^2 , C_p désignant la matrice de covariance normalisée. Il est en fait plus commode de faire intervenir la précision $a_p = \frac{1}{\sigma_p^2}$, et c'est donc cette variable que nous choisissons comme paramètre, la simulation de a_p étant équivalente à celle de σ_p^2 . Nous supposons donc que les paramètres de forme de la covariance et la portée sont fixés. L'estimation bayésienne de ces paramètres est en effet très lourde d'un point de vue numérique, car elle nécessite à chaque

itération de l'échantillonneur de Gibbs le calcul de la matrice inverse C_p^{-1} . Le vecteur θ_p est donc égal à (m_p, a_p) . Dans le cadre de l'échantillonneur de Gibbs, on renouvelle séquentiellement la moyenne m_p et la précision a_p selon les lois conditionnelles de densités $\pi_{M_p}(m_p|d, a_p)$ et $\pi_{A_p}(a_p|d, m_p)$.

Pour le terme de moyenne, on peut choisir un *a priori* gaussien de moyenne μ_p et de variance κ_p^2 . La distribution conditionnelle admet alors pour densité :

$$\begin{aligned}\pi_{M_p}(m_p|d, a_p) &\propto \pi_D(d|m_p, a_p)\pi_{M_p}(m_p) \\ &\propto \exp\left(-\frac{a_p}{2}(d - m_p)^\top C_p^{-1}(d - m_p)\right) \exp\left(-\frac{(m_p - \mu_p)^2}{2\kappa_p^2}\right)\end{aligned}\quad (3.35)$$

Sous ces hypothèses, la densité $\pi_{M_p}(m_p|d, a_p)$ est une densité de loi gaussienne, de moyenne $\tilde{\mu}_p$ et de variance $\tilde{\kappa}_p^2$, avec :

$$\tilde{\mu}_p = \mu_p \frac{\tilde{\kappa}_p^2}{\kappa_p^2} + a_p \tilde{\kappa}_p^2 (1 \dots 1) C_p^{-1} d \quad (3.36)$$

$$\frac{1}{\tilde{\kappa}_p^2} = \frac{1}{\kappa_p^2} + a_p (1 \dots 1) C_p^{-1} (1 \dots 1)^\top \quad (3.37)$$

A chaque itération de l'échantillonneur, la moyenne m_p peut donc être renouvelée directement en générant une réalisation d'une variable gaussienne de moyenne $\tilde{\mu}_p$ et de variance $\tilde{\kappa}_p^2$. Le choix d'une loi gaussienne comme *a priori* est principalement justifié par le fait que sous cette hypothèse la loi conditionnelle $\pi_{M_p}(\cdot|d, a_p)$ a une expression simple (gaussienne). Eventuellement, on peut faire tendre la variance *a priori* κ_p^2 vers l'infini, ce qui revient à prendre un *a priori* non informatif.

Intéressons-nous maintenant à la précision a_p . Sous une hypothèse de loi *a priori* Gamma d'indice α_p et de paramètre 1, la loi conditionnelle de a_p admet pour densité :

$$\begin{aligned}\pi_{A_p}(a_p|d, m_p) &\propto \pi_D(d|a_p, m_p)\pi_{A_p}(a_p) \\ &\propto a_p^{\frac{n}{2}} \exp\left(-\frac{a_p}{2}(d - m_p)^\top C_p^{-1}(d - m_p)\right) a_p^{\alpha_p - 1} \exp(-a_p)\end{aligned}\quad (3.38)$$

On reconnaît la densité d'une loi Gamma d'indice $\tilde{\alpha}_p$ et de paramètre $\tilde{\beta}_p$, soit :

$$\pi_{A_p}(a_p|d, m_p) \propto a_p^{\tilde{\alpha}_p - 1} \exp(-\tilde{\beta}_p a_p) \quad (3.39)$$

avec :

$$\tilde{\alpha}_p = \alpha_p + \frac{n}{2} \quad (3.40)$$

$$\tilde{\beta}_p = \frac{1}{2}(d - m_p)^\top C_p^{-1}(d - m_p) + 1 \quad (3.41)$$

Une étape de l'échantillonneur de Gibbs pour a_p revient donc à générer une variable Gamma d'indice $\tilde{\alpha}_p$ et de paramètre $\tilde{\beta}_p$, ce qui peut être fait directement.

Là encore, le choix d'une loi Gamma comme *a priori* est justifié par l'expression sous cette hypothèse de la loi conditionnelle $\pi_{A_p}(\cdot|d, m_p)$. L'hyperparamètre α_p n'a que peu d'influence dès que n est grand.

Notons cependant que la simulation de m_p et a_p nécessite le calcul de la forme quadratique associée à C_p^{-1} . Dès que la taille n du vecteur considéré est importante (typiquement supérieure à 100), ceci n'est envisageable que si la matrice C_p est facilement inversible et son inverse C_p^{-1} est creuse.

3.3.2 Paramètres du terme de vraisemblance

A première vue, le traitement des paramètres du terme de vraisemblance est délicate car la donnée seule du couple y ne permet pas de calculer l'intensité résiduelle η et donc d'accéder aux paramètres du modèle. Cependant, dans le cadre de l'échantillonneur de Gibbs, le problème est grandement simplifié car les paramètres du modèle sont envisagés à couple y et disparité d fixés. On peut donc à chaque itération de l'échantillonneur calculer le résidu selon la relation $\eta_{u,v} = I_1(u, v) - I_2(u + d(u, v), v)$. La densité de la loi conditionnelle s'écrit alors :

$$\pi_{\Theta_l}(\theta_l|y, d) = L(y|d, \theta_l)\pi_{\Theta_l}(\theta_l) \quad (3.42)$$

$$\propto \pi_H(\eta|\theta_l)\pi_{\Theta_l}(\theta_l) \quad (3.43)$$

Dans cette équation, la densité $\pi_H(\eta|\theta_l)$ correspond à la densité $\pi_H(\eta)$ de l'équation 3.17 dans laquelle on a explicité les paramètres θ_l .

Modèles gaussiens

Nous nous intéressons tout d'abord aux deux modèles de bruit gaussien GMF-0 et GMF-1, et là encore, seulement à l'estimation de la moyenne m_l et de la variance σ_l^2 (on note dès lors C_l la covariance normalisée). Le problème est rigoureusement similaire à celui traité au paragraphe 3.3.1. Le vecteur de paramètres θ_l est égal à (m_l, a_l) avec $a_l = \frac{1}{\sigma_l^2}$, et on choisit comme lois *a priori* une loi gaussienne $\mathcal{N}(\mu_l, \kappa_l^2)$ pour la moyenne m_l et une loi Gamma $\Gamma(\alpha_l, 1)$ pour la précision a_l .

La loi conditionnelle de la moyenne m_l , $\pi_{M_l}(\cdot|y, d, a_l)$, est une loi gaussienne, de moyenne $\tilde{\mu}_l$ et de variance $\tilde{\kappa}_l$, avec :

$$\tilde{\mu}_l = \mu_l \frac{\tilde{\kappa}_l^2}{\kappa_l^2} + a_l \tilde{\kappa}_l^2 (1 \dots 1) C_l^{-1} \eta \quad (3.44)$$

$$\frac{1}{\tilde{\kappa}_l^2} = \frac{1}{\kappa_l^2} + a_l (1 \dots 1) C_l^{-1} (1 \dots 1)^\top \quad (3.45)$$

La loi conditionnelle de la précision a_l , $\pi_{A_l}(a_l|y, d, m_l)$, est une loi Gamma, d'indice $\tilde{\alpha}_l$ et de paramètre $\tilde{\beta}_l$, avec :

$$\tilde{\alpha}_l = \alpha_l + \frac{n}{2} \quad (3.46)$$

$$\tilde{\beta}_l = \frac{1}{2}(\eta - m_l)^\top C_l^{-1}(\eta - m_l) + 1 \quad (3.47)$$

Rappelons que dans le modèle GMF-0, la matrice inverse C_l^{-1} est diagonale (c'est la matrice identité), tandis que dans le modèle GMF-1, elle contient des termes non diagonaux, sur les 4 diagonales qui correspondent au voisinage d'ordre 1.

Modèle de mélange

Nous nous intéressons maintenant à la simulation conjointe des paramètres du modèle de mélange de l'équation 3.22. Le vecteur de paramètres devient $\theta_l = (m_{l_1}, a_{l_1}, p_{l_1}, m_{l_2}, a_{l_2}, p_{l_2})$.

Nous reprenons ici la méthode exposée dans [92]. Les modèles *a priori* pour les moyennes m_{l_1} et m_{l_2} et les précisions a_{l_1} et a_{l_2} sont les mêmes que précédemment ; pour les proportions p_{l_1} et p_{l_2} , il est commode d'utiliser une distribution de Dirichlet $D(\gamma_{l_1}, \gamma_{l_2})$ dont la densité s'écrit :

$$\pi_{P_l}(p_{l_1}, p_{l_2}) \propto p_{l_1}^{\gamma_{l_1}-1} p_{l_2}^{\gamma_{l_2}-1}, \quad p_{l_1} + p_{l_2} = 1 \quad (3.48)$$

Supposons que la classification du résidu soit connue, i.e. on dispose d'une variable catégorielle W_l telle que $W_l(u, v) = i$ si le point $(u, v)^\top$ admet la densité ϕ_i , $i = 1, 2$. Avec cette information supplémentaire, le modèle MIXT 3.22 prend une expression plus simple :

$$L(y|d, w_l) = \pi_H(\eta|w_l) \propto \prod_{i=1,2} \prod_{w_l(u,v)=i} a_{l_i}^{\frac{1}{2}} \exp\left(-\frac{a_{l_i}}{2}(\eta_{u,v} - m_{l_i})^2\right) \quad (3.49)$$

La simulation du vecteur de paramètres Θ_l peut alors être menée en complétant la structure du modèle avec la donnée non observée W_l , ce qui implique la simulation conjointe de la variable catégorielle W_l à l'aide de l'échantillonneur de Gibbs. On commence par simuler Θ_l connaissant y , d et w_l , puis W_l connaissant y , d et θ_l (puis d connaissant y , θ_l et w_l). Les lois conditionnelles correspondantes sont alors :

$$\pi_{M_{l_i}}(m_{l_i}|y, d, a_{l_i}, w_l) \propto \prod_{w_l(u,v)=i} \exp\left(-\frac{a_{l_i}}{2}(\eta_{u,v} - m_{l_i})^2\right) \exp\left(-\frac{(m_{l_i} - \mu_{l_i})^2}{2\kappa_{l_i}^2}\right), \quad (3.50)$$

$$\pi_{A_{l_i}}(a_{l_i}|y, d, m_{l_i}, w_l) \propto \prod_{w_l(u,v)=i} a_{l_i}^{\frac{1}{2}} \exp\left(-\frac{a_{l_i}}{2}(\eta_{u,v} - m_{l_i})^2\right) a_{l_i}^{\alpha_{l_i}-1} \exp(-a_{l_i}), \quad (3.51)$$

pour $i = 1, 2$, et

$$\pi_{P_l}(p_{l_1}, p_{l_2} | w_l) \propto p_{l_1}^{\gamma_{l_1}-1} p_{l_2}^{\gamma_{l_2}-1} \prod_{w_l(u,v)=1} p_{l_1} \prod_{w_l(u,v)=2} p_{l_2}. \quad (3.52)$$

Notons alors n_1 le nombre de points de loi ϕ_1 à un instant donné, et n_2 le nombre de point de loi ϕ_2 ($n_1 + n_2 = n$). L'algorithme de simulation est le suivant [92] :

1. $m_{l_i} \sim \mathcal{N}(\tilde{\mu}_{l_i}, \tilde{\kappa}_{l_i}^2)$, $i = 1, 2$, avec

$$\tilde{\mu}_{l_i} = \mu_{l_i} \frac{\tilde{\kappa}_{l_i}^2}{\kappa_{l_i}^2} + a_{l_i} \tilde{\kappa}_{l_i}^2 \sum_{w_l(u,v)=i} \eta_{u,v}, \quad \frac{1}{\tilde{\kappa}_{l_i}^2} = \frac{1}{\kappa_{l_i}^2} + a_{l_i} n_i, \quad (3.53)$$

2. $a_{l_i} \sim \Gamma(\tilde{\alpha}_{l_i}, \tilde{\beta}_{l_i})$, $i = 1, 2$, avec

$$\tilde{\alpha}_{l_i} = \alpha_{l_i} + \frac{n_i}{2}, \quad \tilde{\beta}_{l_i} = \frac{1}{2} \sum_{w_l(u,v)=i} (\eta_{u,v} - m_{l_i})^2 + 1, \quad (3.54)$$

3. $(p_{l_1}, p_{l_2}) \sim D(\gamma_{l_1} + n_1, \gamma_{l_2} + n_2)$,
4. $w_l(u, v) = i$, $i = 1, 2$, avec la probabilité

$$p_{l_i}(u, v) = \frac{p_{l_i} a_{l_i}^{\frac{1}{2}} \exp\left(-\frac{a_{l_i}}{2} (\eta_{u,v} - m_{l_i})^2\right)}{\sum_{j=1,2} p_{l_j} a_{l_j}^{\frac{1}{2}} \exp\left(-\frac{a_{l_j}}{2} (\eta_{u,v} - m_{l_j})^2\right)}, \quad (3.55)$$

5. $n_i = \sum_{w(u,v)=i} 1$, $i = 1, 2$.

Ici encore, pour n grand, le choix des hyperparamètres $\mu_{l_i}, \kappa_{l_i}, \alpha_{l_i}, \beta_{l_i}, \mu_{l_i}, \kappa_{l_i}$, $i = 1, 2$, n'a que peu d'influence.

3.3.3 Alternative : l'algorithme EM stochastique

L'algorithme EM stochastique [16, 32, 33] est très proche dans l'esprit de la méthode de randomisation que nous venons de présenter : il s'agit simplement de remplacer l'étape de simulation du paramètre par une étape d'estimation par maximum de vraisemblance. L'idée centrale, qui est de compléter le problème en ajoutant et simulant la variable manquante, à savoir la disparité (et la variable de classification dans le problème de mélange), demeure. Alors que l'approche par simulation essaie plutôt de prendre en compte l'incertitude associée au paramètre, l'algorithme EM se situe dans une perspective d'estimation.

La mise en œuvre de cet algorithme est semblable à l'approche séquentielle de l'algorithme de Gibbs. Notons $\hat{\theta}_p$ et $\hat{\theta}_l$ les estimations courantes des paramètres θ_p et θ_l . On commence par simuler la disparité D conditionnellement au couple y et aux valeurs $\hat{\theta}_p$ et $\hat{\theta}_l$. Puis, connaissant la disparité d et le couple y , on actualise les valeurs de $\hat{\theta}_p$ et $\hat{\theta}_l$ par maximisation des vraisemblances $\pi_D(d|\theta_p)$ et $L(y|\theta_l, d)$. D'où le schéma suivant :

1. $D|y, \hat{\theta}_p, \hat{\theta}_l \sim L(y|d, \hat{\theta}_l)\pi_D(d|\hat{\theta}_p)$,
2. $\hat{\theta}_p = \arg \max_{\theta_p} \pi_D(d|\theta_p)$,
3. $\hat{\theta}_l = \arg \max_{\theta_l} L(y|\theta_l, d)$,

L'expression des densités $\pi_D(d|\theta_p)$ et $L(y|\theta_l, d)$ a été donnée au paragraphe 3.2 pour les différents modèles considérés. Chaque fois que l'on est capable de calculer le maximum de vraisemblance (ou une forme approchée), on peut donc mettre en œuvre cet algorithme.

Pour le modèle *a priori* gaussien, on retrouve les mêmes limitations que précédemment : l'estimation des paramètres de forme de la covariance suppose l'inversion répétée de la matrice de covariance, ceci n'étant envisageable que dans certains cas particuliers. Pour les modèles de vraisemblance gaussiens, on peut reprendre les expressions des estimateurs des équations 3.23 à 3.30, où le résidu approché $\hat{\eta}_{u,v} = I_1(u, v) - I_2(u + \hat{d}(u, v), v)$ est remplacé par le résidu simulé $\eta_{u,v} = I_1(u, v) - I_2(u + d(u, v), v)$.

Pour le modèle de mélange, il faut également simuler la variable de classification W_l , comme indiqué au paragraphe 3.3.2. On obtient alors les estimateurs suivants :

$$n_i = \sum_{w_l(u,v)=i} 1, i = 1, 2 \quad (3.56)$$

$$\hat{m}_{l_i} = \frac{1}{n_i} \sum_{w_l(u,v)=i} \eta_{u,v}, i = 1, 2 \quad (3.57)$$

$$\hat{\sigma}_{l_i}^2 = \frac{1}{n_i} \sum_{w_l(u,v)=i} (\eta_{u,v} - \hat{m}_{l_i})^2, i = 1, 2 \quad (3.58)$$

$$\hat{p}_{l_i} = \frac{n_i}{n_1 + n_2}, i = 1, 2 \quad (3.59)$$

et toujours $w_l(u, v) = i, i = 1, 2$, avec la probabilité

$$p_{l_i}(u, v) = \frac{p_{l_i} a_{l_i}^{\frac{1}{2}} \exp\left(-\frac{a_{l_i}}{2}(\eta_{u,v} - m_{l_i})^2\right)}{\sum_{j=1,2} p_{l_j} a_{l_j}^{\frac{1}{2}} \exp\left(-\frac{a_{l_j}}{2}(\eta_{u,v} - m_{l_j})^2\right)}.$$

En pratique, on peut estimer avec cet algorithme les paramètres dans une première phase de la simulation, puis les fixer à leurs valeurs estimées pour la suite.

3.4 Mise en œuvre

Nous proposons dans ce paragraphe d'appliquer les méthodes précédentes pour estimer les paramètres du modèle stochastique. Nous considérons pour cela un

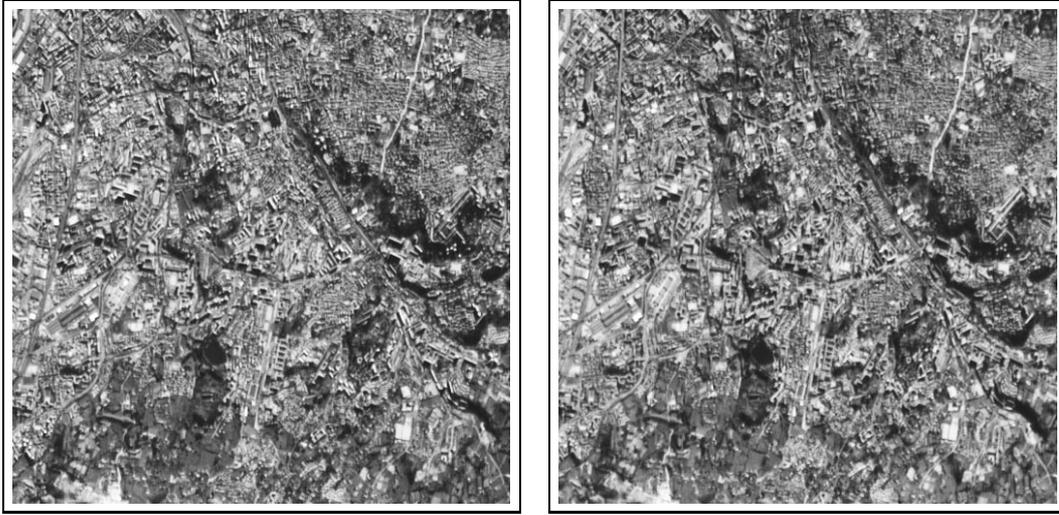


FIG. 3.5 – Extrait du couple stéréoscopique d’Aix-Marseille.

extrait du couple stéréoscopique d’Aix-Marseille (figure 3.5). Nous avons dans un premier temps calculé une carte de disparité à l’aide de l’algorithme de corrélation d’Istar (figure 3.6), puis l’intensité résiduelle entre les deux images du couple (figure 3.6).

3.4.1 Modèle *a priori*

Etude variographique de la disparité

Le variogramme expérimental (omnidirectionnel) correspondant à la carte de disparité 3.6 est représenté à la figure 3.7. Il exhibe un comportement proche du linéaire, avec cependant une légère inflexion vers 200 pixels. Cette structure laisse plutôt suggérer une hypothèse de stationnarité intrinsèque, avec comme modèle de variogramme un variogramme linéaire. Une surface gaussienne de variogramme linéaire présente une propriété d’auto-similarité (modèle invariant par changement d’échelle) ou plutôt d’auto-affinité [24]. Cette propriété n’est pas à confondre avec le comportement “irrégulier” de la surface, dû au caractère non-dérivable du variogramme à l’origine.

Cependant, les premiers calculs effectués avec un modèle linéaire ont montré que ce modèle n’était pas adapté dans cet exemple : la structure spatiale des cartes de disparité obtenues différait sensiblement de la carte de disparité calculée par corrélation (fortes variations aux grandes distances), et l’algorithme de simulation par Monte Carlo présentait d’importants problèmes de convergence (taux de transition proche de 0). Il semble ici que ce soit la propriété d’auto-similarité,

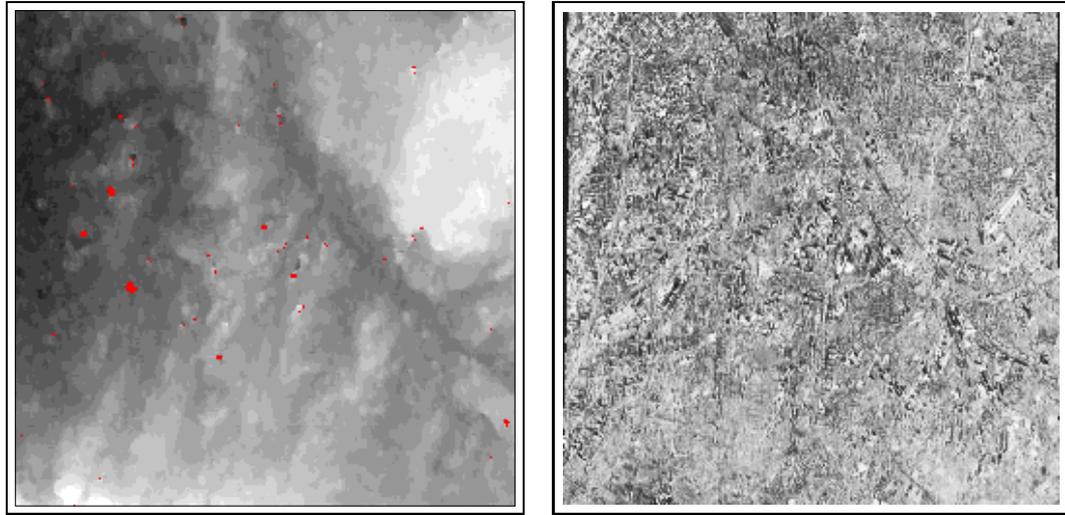


FIG. 3.6 – Carte de disparité du couple de la figure 3.5 calculée par corrélation (gauche) et intensité résiduelle du couple obtenue à partir de cette carte de disparité (droite).

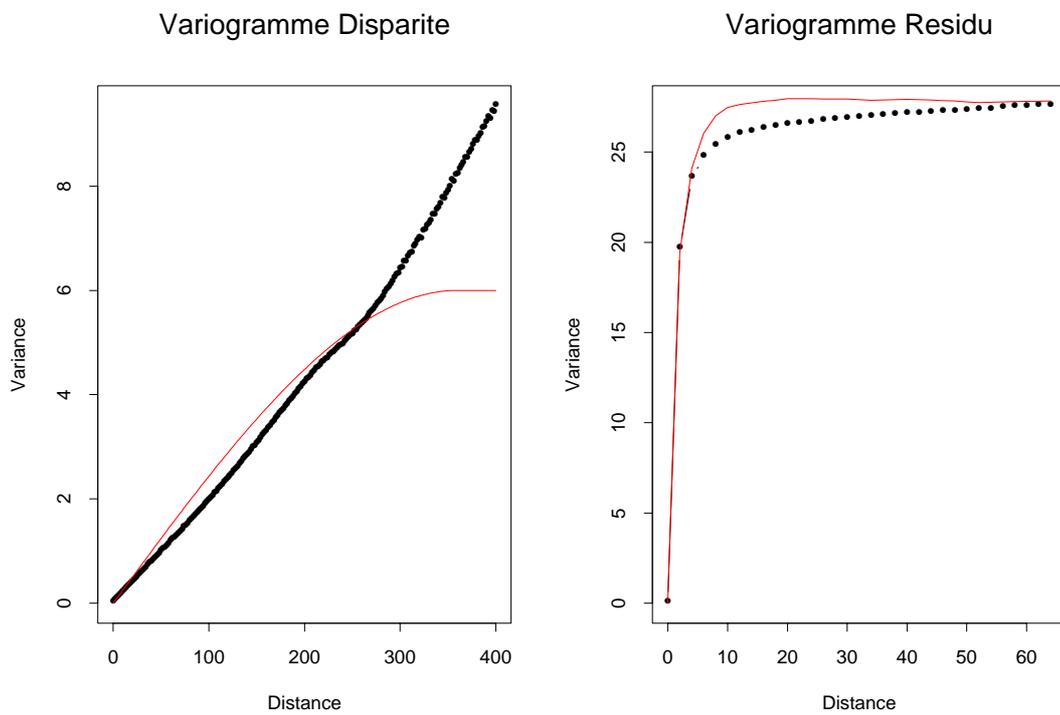


FIG. 3.7 – Gauche : variogramme de la carte de disparité (courbe noire) et modèle estimé (courbe rouge). Droite : Variogramme du résidu (courbe noire) et variogramme d’une simulation du modèle GMF-1 (courbe rouge).

et notamment le caractère non borné du variogramme, qui pose problème. Nous avons donc choisi de travailler en hypothèse de stationnarité d'ordre 2, avec pour covariance un modèle sphérique de portée égale à 360 pixels et de pallier égal à 6 (représenté également à la figure 3.7). Remarquons que le variogramme sphérique présente un comportement linéaire à l'origine, et donc le comportement irrégulier de la surface est conservé. La portée utilisée est assez élevée par rapport à la taille de la carte de disparité considérée (512×512). Il est d'ailleurs possible que le comportement non stationnaire du variogramme expérimental soit dû à un problème d'inférence et non à une véritable non stationnarité, le support de l'échantillon étudié étant réduit par rapport à la portée du phénomène.

Estimation par Monte Carlo

Comme précisé au paragraphe 3.3.1, l'estimation par Monte Carlo de la variance du modèle *a priori* nécessite l'inversion de la matrice de covariance normalisée C_p . Compte tenu de la taille du vecteur z simulé (512×512), un calcul exact n'est pas envisageable. Cependant, on obtient une très bonne approximation en utilisant un sous-vecteur de taille réduite pour le calcul de la forme quadratique, ce qui permet de réduire la taille de la matrice de covariance et donc de calculer dans de bonnes conditions son inverse. Dans cet exemple, nous n'avons conservé pour les calculs que 256 composantes de z , échantillonnées régulièrement, en tirant aléatoirement et de manière uniforme à chaque itération l'origine de la grille d'échantillonnage, afin d'éviter des biais éventuels.

Nous rappelons que la simulation de ce paramètre nécessite la simulation conjointe de la disparité, mais nous nous intéressons dans ce paragraphe uniquement à l'estimation de la variance. De plus, les paramètres de forme et la portée de la covariance, estimés précédemment, sont maintenus fixes.

La figure 3.8 représente les valeurs de la variance obtenues au cours des 50 000 premières itérations de la simulation, ainsi que l'évolution de l'estimation par Monte Carlo de ce paramètre. La chaîne exhibe un comportement stationnaire, et l'intégrale de Monte Carlo converge rapidement (5 000 itérations suffisent pour obtenir une bonne estimation). La valeur estimée (4.28) est inférieure à la valeur ajustée (6) précédemment ; ceci n'est pas très surprenant compte tenu du caractère très subjectif de l'ajustement proposé.

La simulation par l'échantillonneur de Gibbs de la variance du modèle *a priori* nécessite d'estimer la forme quadratique associée à la matrice de covariance inverse. On obtient une bonne approximation de cette forme quadratique en utilisant une sous-matrice de la matrice de covariance, qui peut être inversée une fois pour toutes. Cette méthode permet d'inférer un paramètre essentiel et délicat à estimer en pratique.

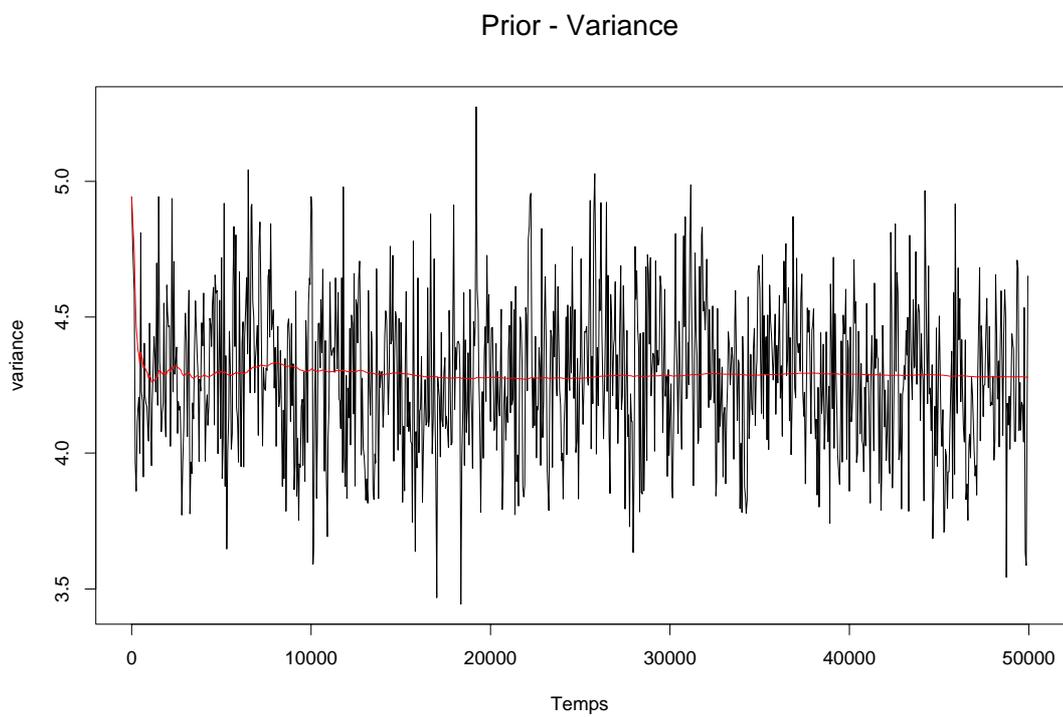


FIG. 3.8 – Une simulation markovienne de la variance du modèle *a priori*. La courbe rouge représente l'évolution en fonction du nombre d'itérations de l'estimation de ce paramètre par intégrale de Monte Carlo.

3.4.2 Modèles de vraisemblance

Estimation directe

A partir du résidu de la figure 3.6, et conformément à la méthodologie exposée au paragraphe 3.2.3, nous avons estimé directement les paramètres du terme de vraisemblance sous les trois hypothèses de modèle (GMF-0, GMF-1 et MIXT). Le tableau 3.1 donne les valeurs des paramètres des deux modèles gaussiens GMF-0 et GMF-1. Notons que dans le cas du modèle GMF-1, le paramètre σ_l ne corres-

	\hat{m}_l	$\hat{\sigma}_l$	$\hat{\rho}_l$
GMF-0	-5.26	5.33	—
GMF-1	-5.26	8.55	0.92
GMF-1*	-5.21	9.64	0.33

TAB. 3.1 – Estimation directe des paramètres du terme de vraisemblance pour les modèles GMF-0 et GMF-1. * estimation à partir d'un sous-ensemble (un pixel sur quatre dans chaque direction).

pond pas à l'écart type de l'image au sens statistique, ce qui explique les différences importantes de valeur entre les 2 modèles. On se rappelle notamment que le paramètre σ_l dépend fortement du paramètre de dépendance spatiale ρ_l . Nous avons également calculé les paramètres du modèle GMF-1 correspondant à un échantillonnage de l'image résiduelle initiale, obtenu en ne conservant qu'un pixel sur quatre dans chaque direction (cette configuration est utilisée dans le paragraphe suivant). Comme attendu, le paramètre de dépendance spatiale diminue fortement dans ce cas, et la valeur correspondante de σ_l augmente, conformément à l'équation 3.30.

Afin de tester la pertinence du modèle GMF-1 pour tenir compte de la structure spatiale du résidu, nous avons calculé le variogramme du terme résiduel et celui d'une simulation d'un champ aléatoire gaussien d'ordre 1 de même taille et de paramètres correspondant à ceux du tableau 3.1. Le résultat est présenté à la figure 3.7.

Les deux variogrammes, notamment en ce qui concerne le comportement à l'origine (linéaire) et la portée spatiale, sont semblables. Cette comparaison montre que le modèle GMF-1 permet de tenir compte avec une bonne approximation de la structure spatiale du terme résiduel et que la formulation en termes de champ de Markov est relativement bien appropriée.

Les valeurs des paramètres estimées sous l'hypothèse du modèle MIXT sont présentées au tableau 3.2. Les deux distributions gaussiennes diffèrent surtout par la valeur de leur écart type : comme attendu, la seconde distribution, qui correspond au terme parasite, exhibe un écart type bien plus élevé et permet

	\hat{m}_{l_1}	$\hat{\sigma}_{l_1}$	\hat{p}_{l_1}	\hat{m}_{l_2}	$\hat{\sigma}_{l_2}$	\hat{p}_{l_2}
MIXT	-5.26	5.33	0.923	-5.38	14.60	0.077

TAB. 3.2 – Estimation directe des paramètres du terme de vraisemblance pour le modèle MIXT

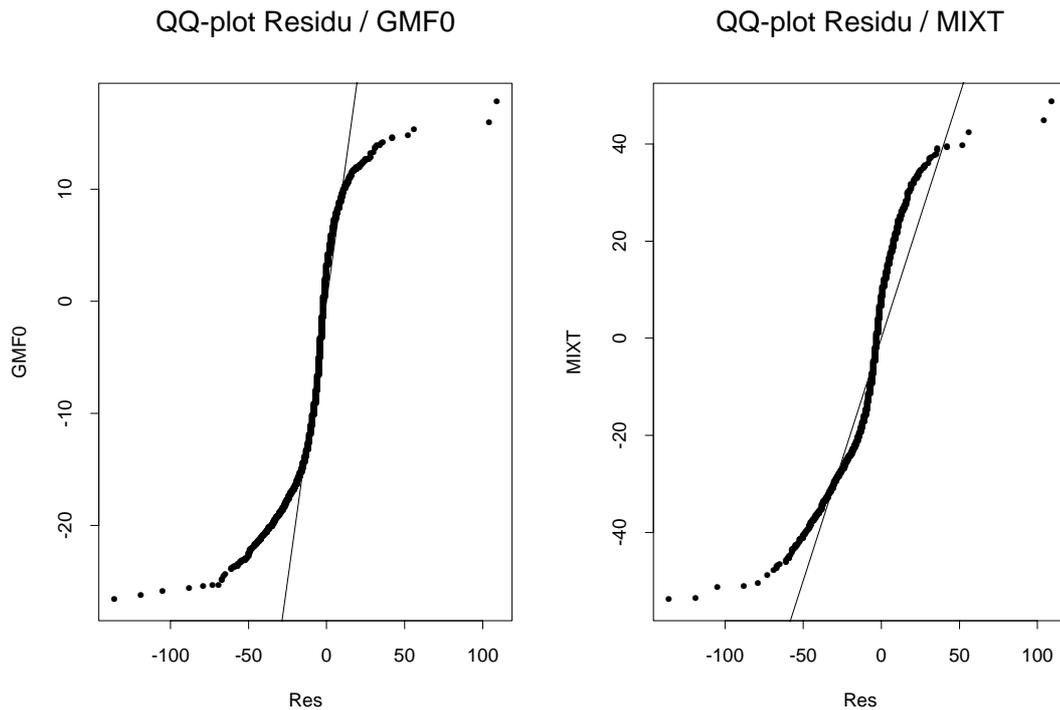


FIG. 3.9 – Graphe quantile-quantile entre le résidu et respectivement le modèle GMF-0 (gauche) et le modèle MIXT (droite).

donc de tenir compte de valeurs résiduelles anormalement élevées. La proportion de points du second type est faible (7.7%).

Afin de juger de la pertinence de ce modèle, nous avons comparé l'histogramme du résidu aux histogrammes du modèle GMF-0 et du modèle MIXT. Le résultat, sous forme de graphe quantile-quantile, est présenté à la figure 3.9. Le premier graphe quantile-quantile, correspondant au modèle GMF-0, met clairement en évidence l'existence de valeurs fortes qui violent le caractère gaussien. Le modèle MIXT, par comparaison, permet de mieux tenir compte de ces valeurs, cependant la partie centrale de l'histogramme est moins bien rendue. De plus, les queues d'histogramme du résidu restent plus importantes, ce qui laisse penser que d'autres modèles que les modèles gaussiens seraient plus appropriés pour modéliser le terme parasite du modèle MIXT.

Estimation par Monte Carlo

Nous adoptons maintenant un point de vue purement bayésien, selon l’approche exposée au paragraphe 3.3, et considérons les paramètres du modèle comme des variables aléatoires pouvant être estimées par une procédure de Monte Carlo. Nous rappelons que ceci suppose la simulation conjointe de la disparité et des paramètres du modèle, selon la méthode de l’échantillonneur de Gibbs. Le problème de la simulation de la disparité, abordée aux chapitres 4 et 5, est supposé résolu. Nous nous intéressons ici uniquement aux paramètres. Dans chacun des 3 cas, nous avons procédé à 50 000 itérations de l’échantillonneur de Gibbs et avons calculé la moyenne de chaque paramètre le long de la chaîne correspondante. Cependant, pour limiter le temps de calcul, nous avons utilisé un échantillonnage de la carte de disparité et du terme résiduel, obtenu en ne conservant qu’un pixel sur 4 dans chaque direction. Cela n’a que peu d’influence pour les modèles GMF-0 et MIXT, qui supposent le résidu non corrélé spatialement. En revanche, pour le modèle GMF-1, il convient d’utiliser le paramètre de dépendance spatiale ρ_l adéquat (troisième ligne du tableau 3.1). Nous rappelons que ce paramètre est maintenu fixe.

Les simulations obtenues dans le cas des deux modèles gaussiens GMF-0 et GMF-1 sont représentées à la figure 3.10, tandis que le tableau 3.3 résume les valeurs estimées par intégrale de Monte Carlo. Les courbes d’évolution des pa-

	\bar{m}_l	$\bar{\sigma}_l$
GMF-0	-5.27	4.56
GMF-1	-5.27	8.99

TAB. 3.3 – Estimation par Monte Carlo des paramètres du terme de vraisemblance pour les modèles GMF-0 et GMF-1 (à partir d’un sous-ensemble, avec un pixel sur quatre conservé dans chaque direction).

ramètres mettent clairement en évidence l’existence d’une période transitoire, de convergence, de l’ordre de 10 000 itérations. Après cette première phase, chaque chaîne semble exhiber un comportement stationnaire.

Si l’on compare les valeurs obtenues par Monte Carlo et les estimations du paragraphe précédent, on constate que les estimations de la moyenne sont très proches, en revanche, les valeurs de l’écart type obtenues par Monte Carlo sont sensiblement plus faibles. Ceci n’est pas surprenant, car dans l’approche directe, ce n’est qu’une estimation de l’intensité résiduelle qui est utilisée, et donc l’estimation de l’écart type est corrompue par la présence de valeurs aberrantes, généralement plus élevées, dues à une mise en correspondance erronée.

Intéressons-nous pour finir au modèle MIXT (figure 3.11 et tableau 3.4). La situation est assez différente des deux cas précédents. En effet, les estimations

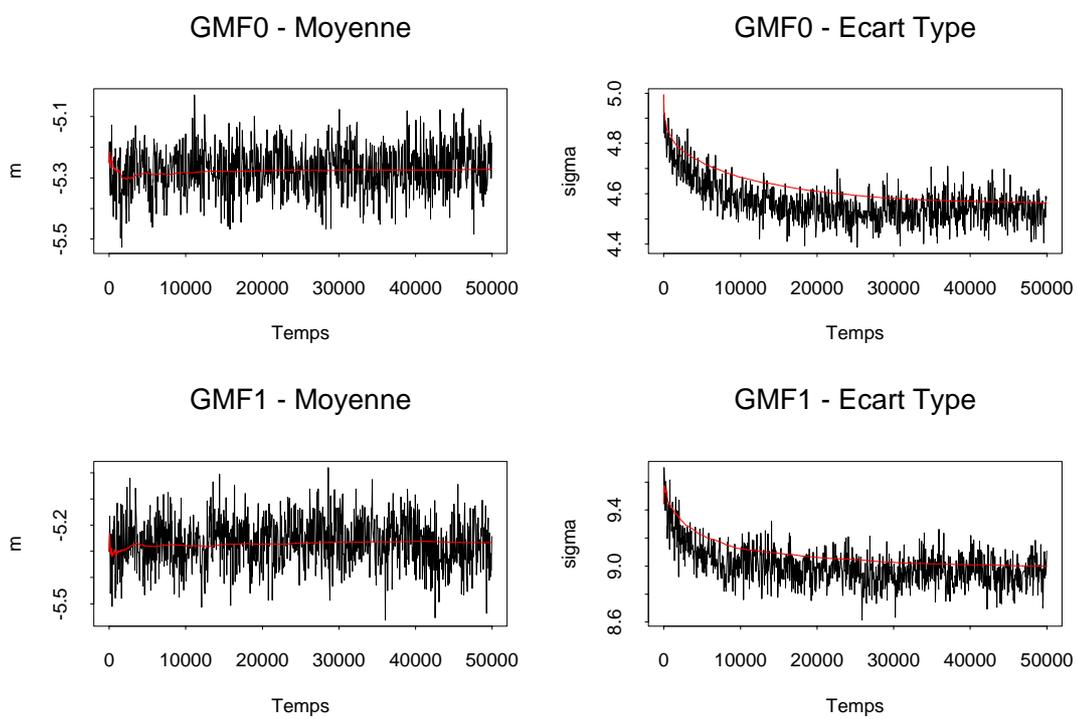


FIG. 3.10 – Une simulation markovienne des paramètres des modèles GMF-0 et GMF-1. La courbe rouge représente l'évolution en fonction du nombre d'itérations de l'estimation du paramètre par intégrale de Monte Carlo.

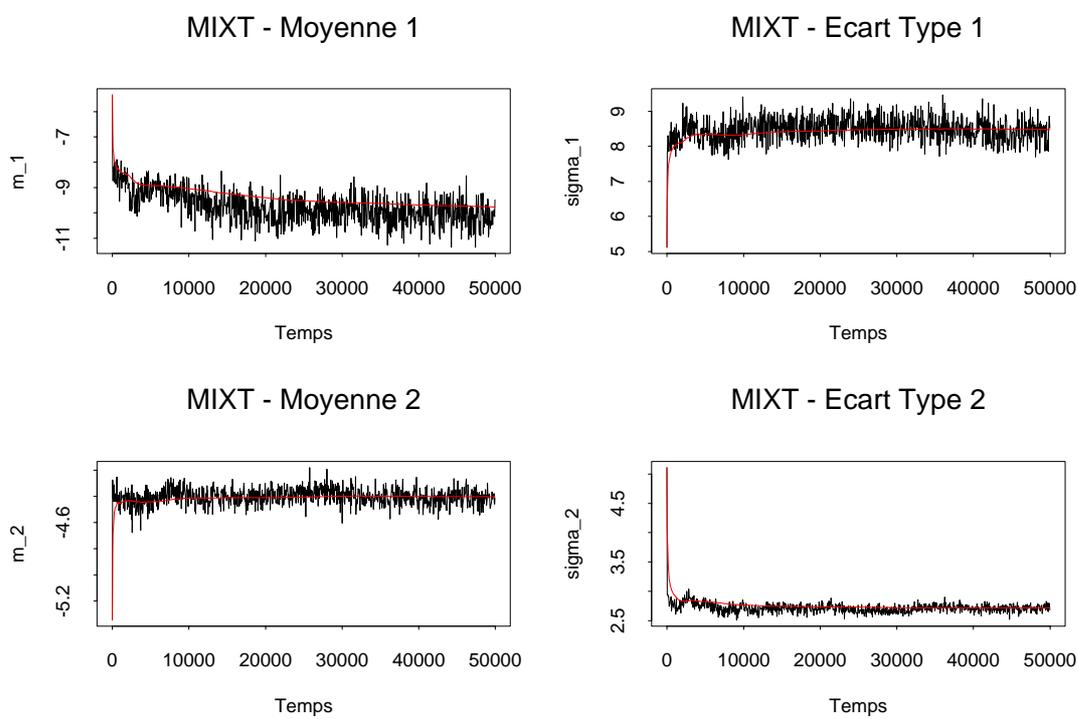


FIG. 3.11 – Une simulation markovienne des paramètres du modèle MIXT. La courbe rouge représente l'évolution en fonction du nombre d'itérations de l'estimation du paramètre par intégrale de Monte Carlo.

	\bar{m}_{l_1}	$\bar{\sigma}_{l_1}$	\bar{p}_{l_1}	\bar{m}_{l_2}	$\bar{\sigma}_{l_2}$	\bar{p}_{l_2}
MIXT	-9.76	8.48	0.16	-4.40	2.71	0.84

TAB. 3.4 – Estimation par Monte Carlo des paramètres du terme de vraisemblance pour le modèle MIXT (à partir d’un sous-ensemble, avec un pixel sur quatre conservé dans chaque direction).

par Monte Carlo convergent vers des valeurs très différentes de celles obtenues précédemment et utilisées pour l’initialisation. On constate même un basculement entre les 2 distributions, la distribution 1 exhibant l’écart type le plus élevé et étant minoritaire. Deux raisons peuvent expliquer ce phénomène. Tout d’abord, la procédure d’estimation directe à partir de l’intensité résiduelle que nous avons proposée repose sur des hypothèses approximatives, selon lesquelles les paramètres de chaque distribution peuvent être estimés séparément, à partir des images du couple d’une part et du terme résiduel d’autre part, les proportions du mélange étant estimées *in fine*. De plus, la formulation même du problème est délicate : on comprend en effet que le problème de l’approximation d’une distribution donnée par un mélange de distributions gaussiennes n’a pas de solution stable, plusieurs jeux de paramètres très différents pouvant donner des résultats proches. Remarquons pour finir que les paramètres relatifs à la distribution minoritaire (1) ont une variabilité beaucoup plus grande que ceux de la distribution majoritaire (2). En effet, ceux-ci sont estimés à partir d’un nombre beaucoup plus faible (facteur 5) de points.

3.4.3 Estimation par l’algorithme EM stochastique

Pour conclure cette étude, nous proposons de comparer rapidement les estimations données par intégrale de Monte Carlo et par l’algorithme EM stochastique. Nous considérons par exemple l’estimation des paramètres du modèle MIXT, dont nous avons souligné que l’inférence était délicate.

Des illustrations des résultats obtenus sont représentées aux figures 3.12 et 3.13. Dans chacun des cas, on constate une période de convergence, suivie de fluctuations dans un régime stationnaire. Comme attendu, ces fluctuations sont plus importantes lorsque le paramètre est simulé (méthode de Monte Carlo). Elles subsistent cependant dans le cas de l’algorithme EM stochastique du fait de la simulation en parallèle de la disparité (et de la variable de classification). Le résultat intéressant est que les deux méthodes donnent des estimations très proches des paramètres : on observe en effet une égalité presque parfaite des valeurs moyennes.

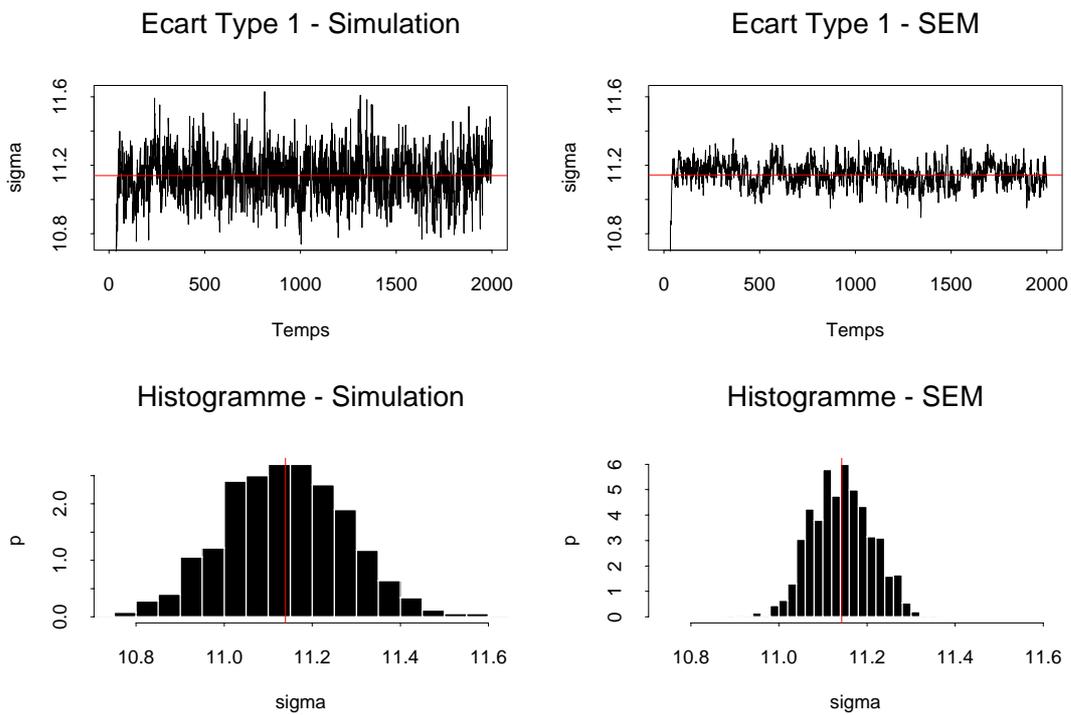


FIG. 3.12 – Estimation de l'écart type de la distribution 1 du modèle de vraisemblance MIXT. A gauche : résultats obtenus par Monte Carlo, à droite : résultats obtenus à l'aide de l'algorithme EM stochastique. Sont représentés dans chaque cas une simulation du paramètre (figures du haut) et l'histogramme correspondant (figures du bas). En rouge : moyenne du paramètre.

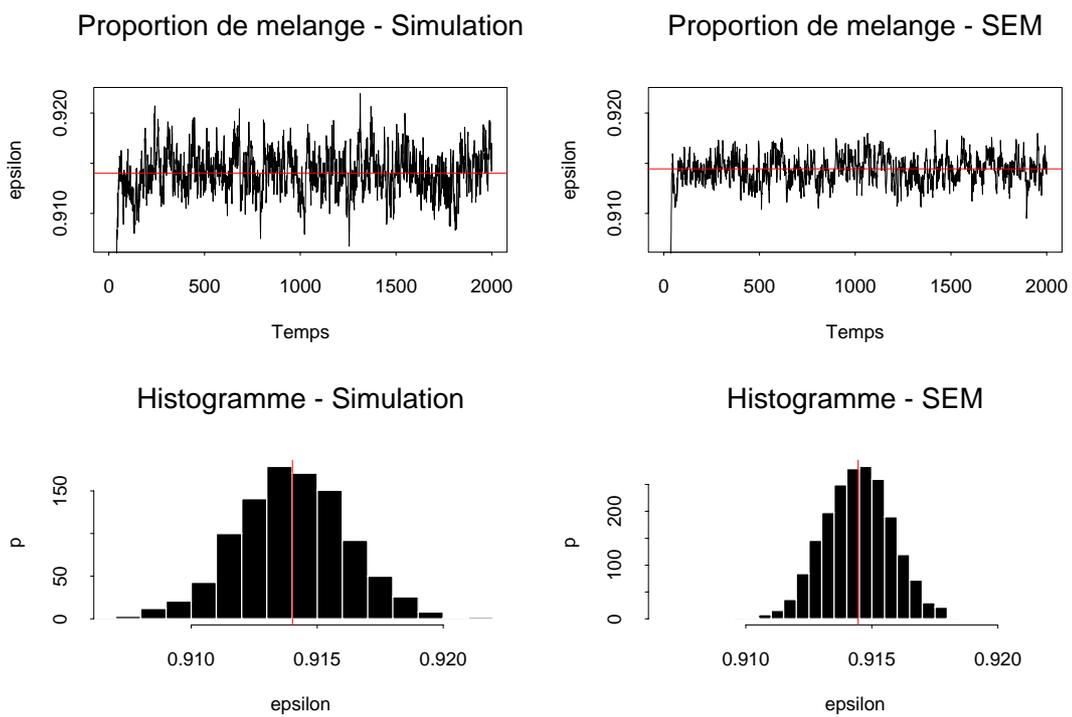


FIG. 3.13 – Estimation de la proportion de mélange du modèle de vraisemblance MIXT. A gauche : résultats obtenus par Monte Carlo, à droite : résultats obtenus à l'aide de l'algorithme EM stochastique. Sont représentés dans chaque cas une simulation du paramètre (figures du haut) et l'histogramme correspondant (figures du bas). En rouge : moyenne du paramètre

3.5 Conclusion

L'approche bayésienne que nous avons proposée permet une formulation mathématique simple du problème posé, cependant elle suppose que l'on puisse choisir un modèle stochastique approprié pour décrire disparité et couple stéréoscopique. Si l'on désire de plus développer une méthode relativement automatisée, il convient d'opter pour une classe de modèles susceptibles de décrire des situations variées et dont les paramètres peuvent être inférés aisément.

Pour ce qui est du modèle *a priori*, le cadre des fonctions aléatoires gaussiennes présente un avantage triple :

- champ d'application large de part le choix de la covariance spatiale,
- facilité d'inférence des paramètres à partir d'un échantillon, seules covariance spatiale et moyenne devant être estimées,
- mise en œuvre algorithmique aisée, un grand nombre d'algorithmes de simulation ayant été développés.

Le problème de l'inférence mérite cependant d'être précisé. Il nous semble notamment illusoire de vouloir estimer l'ensemble des paramètres de la covariance spatiale, en particulier paramètre de forme et portée, de la disparité à partir du couple stéréoscopique seulement, notamment au travers d'une approche bayésienne reposant sur les méthodes de Monte Carlo. Comme évoqué plusieurs fois précédemment, ceci requiert en effet de calculer l'inverse de la matrice de covariance, ce qui n'est pas envisageable dans le cas général. La méthode que nous avons proposée reposant sur la carte de disparité calculée par corrélation est en revanche une solution pouvant être facilement mise en œuvre. Il est possible d'ajuster ensuite l'estimation de la variance en utilisant l'échantillonneur de Gibbs.

Pour le modèle de vraisemblance, le cadre choisi est quelque peu plus restrictif. A partir du modèle de base que représente le modèle de bruit blanc gaussien, nous avons proposé deux extensions : d'un côté, la prise en compte de corrélations spatiales à travers un modèle de champ de Markov gaussien d'ordre 1, de l'autre la modélisation de valeurs fortes à l'aide d'un modèle de mélange. Ces trois modèles permettent de décrire des situations assez variées, le modèle markovien d'ordre 1 étant notamment bien adapté pour décrire la structure spatiale du résidu.

L'étude précédente a montré l'intérêt des méthodes d'estimation des paramètres. Maximum de vraisemblance (ou approximation du maximum de vraisemblance) et Monte Carlo donnent des résultats comparables pour les 2 modèles gaussiens, tandis que les paramètres du modèle MIXT sont plus délicats à inférer. L'approche purement bayésienne où les paramètres sont randomisés est certainement plus rigoureuse, cependant elle soulève des problèmes de convergence certains, notamment dus à l'estimation de la variance du terme de vraisemblance. En effet, l'estimation par Monte Carlo converge vers une valeur plus faible que celle estimée à partir du seul résidu, et donc la distribution *a posteriori* de la disparité se resserre autour du maximum de vraisemblance. La simulation de la disparité

devient donc de plus en plus délicate, et le taux de transition de la chaîne de Markov diminue fortement. Puisqu'*in fine* le but est d'étudier la carte de disparité calculée par corrélation, il paraît raisonnable d'utiliser le modèle estimé à partir de cette carte ; la vraisemblance des simulations générées est ainsi du même ordre de grandeur que la vraisemblance de la carte calculée par corrélation.

Remarquons enfin que du point de vue de l'estimation des paramètres, les approches par intégrale de Monte Carlo et algorithme EM stochastique donnent des résultats équivalents. Cette dernière est peut-être plus simple à mettre en œuvre, mais la première est plus proche dans l'esprit de l'étude de l'incertitude à proprement parler.

Chapitre 4

Simulations spatiales

Soit Z une fonction aléatoire définie sur \mathbb{R}^d et à valeurs dans un espace mesurable (F, \mathcal{F}) muni de la mesure de référence ν . Typiquement, $\mathbb{R}^d = \mathbb{R}^2$, $F = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$ la tribu des boréliens et ν est la mesure de Lebesgue. En pratique, Z ne sera considérée qu'en un ensemble fini S de points, de cardinal n , et on notera Z la restriction de Z à S . Une réalisation $z = (z_1, \dots, z_n)$ de Z est un élément de $E = F^n$, dont la tribu associée est $\mathcal{E} = \mathcal{F}^{\otimes n}$, elle-même munie de la mesure $\mu = \nu^{\otimes n}$. On note π la loi de Z sur E . On suppose qu'elle admet une densité $\pi(z)$ continue par rapport à μ . Le problème que l'on se pose est celui de la simulation de Z selon la loi π .

Il peut être utile dans certains cas de considérer des restrictions de la loi π à un sous-espace. Soit un sous-ensemble T de S . z_T désigne alors la restriction de z à T et z^T celle au complémentaire de T dans S , de sorte que $z = (z_T, z^T)$. On notera de même $(E_T, \mathcal{E}_T, \mu_T)$ et $(E^T, \mathcal{E}^T, \mu^T)$ les espaces probabilisés correspondants.

La seule restriction que l'on s'impose est de pouvoir calculer la densité $\pi(z)$ à une constante de normalisation près. En particulier, on peut calculer pour deux états z et z' de E le rapport $\frac{\pi(z')}{\pi(z)}$. La constante de normalisation peut s'obtenir par intégration de $\pi(z)$ sur E , mais dans la plupart des cas, l'intégrale n'est pas calculable, et est à vrai dire sans grand intérêt.

Eventuellement, si l'on dispose d'une loi auxiliaire $\tilde{\pi}$ définie sur E , de densité $\tilde{\pi}(z)$ telle que $(\tilde{\pi}(z) = 0) \Rightarrow (\pi(z) = 0)$, on utilisera la décomposition suivante :

$$\pi(z) = r(z)\tilde{\pi}(z) \tag{4.1}$$

où $r(z) = \frac{\pi(z)}{\tilde{\pi}(z)}$ si $\tilde{\pi}(z) \neq 0$ et constant sinon. En pratique, $\tilde{\pi}$ pourra désigner une loi selon laquelle on sait générer des simulations, et dont on se sert pour construire des simulations de π . C'est notamment l'approche utilisée en échantillonnage d'importance, où le ratio r sert de pondérateur dans l'intégration de Monte Carlo (cf. paragraphe 4.1.4).

Dans le cas où π désigne une loi conditionnelle, il existe une décomposition

naturelle de π . Notons Y l'information dont on dispose sur Z , le problème étant de générer des simulations conditionnelles de Z à $Y = y$. En utilisant la relation de Bayes, cette loi conditionnelle admet pour densité :

$$\pi_Z(z|y) = \frac{\pi_Y(y|z)}{\pi_Y(y)} \pi_Z(z) \quad (4.2)$$

Il suffit alors de poser $\pi(z) = \pi_Z(z|y)$, $r(z) = \frac{\pi_Y(y|z)}{\pi_Y(y)}$ et $\tilde{\pi}(z) = \pi_Z(z)$. Ces quantités sont bien des fonctions de z seulement car y est fixé. On retrouve ainsi naturellement la décomposition proposée à l'équation 4.1, où π est une loi conditionnelle, ou loi *a posteriori*, r peut être interprété comme une fonction de vraisemblance, et $\tilde{\pi}$ est la loi marginale, ou *a priori*.

Nous proposons dans la suite un certain nombre d'algorithmes de simulation. Le cas gaussien, pour lequel des algorithmes de simulation directe sont disponibles, est traité en annexe A. Nous envisageons ici le cas général, pour lequel une solution consiste à construire une chaîne de Markov de loi stationnaire π . Le chapitre est organisé comme suit : nous commençons par rappeler les principes de la simulation par chaînes de Markov et de l'intégrale de Monte Carlo. Nous présentons notamment une application aux simulations spatiales des méthodes d'échantillonnage d'importance. Puis, les principaux algorithmes de simulation, à savoir l'échantillonneur de Gibbs et les algorithmes du type Metropolis-Hastings, sont étudiés. L'apport principal de cette étude est un nouvel algorithme de simulation qui généralise à plusieurs dimensions l'algorithme de Metropolis-Hastings. Enfin, un cas d'étude permet d'illustrer le comportement des différents algorithmes proposés.

4.1 Simulations par chaînes de Markov

Ce paragraphe rassemble quelques éléments de théorie sur les chaînes de Markov à valeurs dans un espace général E (non nécessairement fini ou dénombrable). Nous nous inspirons principalement des exposés donnés dans [116], [54] et [50], la référence étant l'ouvrage de Meyn et Tweedie [79], complété par [117].

4.1.1 Définition

Une chaîne de Markov $\mathcal{Z} = (Z_k)_{k \geq 0}$ sur E est une suite de variables aléatoires définie par la loi de probabilité de son état initial Z_0 et son noyau de transition $K : E \times \mathcal{E} \rightarrow [0, 1]$:

$$P(Z_{k+1} \in A | Z_k = z) = K(z, A), \quad \forall A \in \mathcal{E} \quad (4.3)$$

A z fixé dans E , $K(z, \cdot)$ est une probabilité sur \mathcal{E} , et à A fixé dans \mathcal{E} , $K(\cdot, A)$ est une fonction mesurable.

De même, on note K^l le noyau correspondant à l transitions : $P(Z_{k+l} \in A | Z_k = z) = K^l(z, A)$. On a immédiatement la relation suivante :

$$K^{l+1}(z, A) = \int_E K^l(z, dy)K(y, A), \quad \forall l \geq 1 \quad (4.4)$$

Afin d'étudier la convergence de la chaîne, introduisons la norme de la variation totale entre deux lois μ_1 et μ_2 [79] :

$$\|\mu_1 - \mu_2\| = 2 \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)| \quad (4.5)$$

Une question primordiale est la suivante : à quelle condition existe-t-il une loi π telle que $K^l(z, \cdot) \rightarrow \pi(\cdot)$ pour la norme de la variation totale ? Si une telle propriété est vérifiée, on dit que la chaîne est *ergodique* et qu'elle admet π pour loi stationnaire. L'ergodicité implique nécessairement l'invariance de π sous le noyau de transition K , à savoir :

$$\int_E K(z, A)\pi(z)\mu(dz) = \pi(A), \quad \forall A \in \mathcal{E} \quad (4.6)$$

Autrement dit, partant d'un état quelconque tiré selon la loi π , la probabilité d'atteindre un ensemble A après une transition avec le noyau K est exactement $\pi(A)$, d'où la notion d'invariance. Une condition suffisante d'invariance est la condition de réversibilité suivante :

$$\int_A \pi(z)K(z, B)\mu(dz) = \int_B \pi(z)K(z, A)\mu(dz), \quad \forall A, B \in \mathcal{E}^2 \quad (4.7)$$

Pour un noyau réversible, la probabilité d'atteindre B partant de A est égale à la probabilité d'atteindre A partant de B . Si K admet une densité, cette égalité est équivalente à :

$$\pi(z)K(z, z') = \pi(z')K(z', z), \quad \forall z, z' \in E^2 \quad (4.8)$$

Dans l'étude des chaînes de Markov, deux questions prédominent :

1. Sous quelles conditions un noyau K de loi invariante π est-il ergodique ?
2. Quel est le comportement asymptotique de $\|K^l(z, \cdot) - \pi(\cdot)\|$?

Le première interrogation concerne des conditions nécessaires de convergence de la chaîne, tandis que le second point s'intéresse à la vitesse de convergence.

4.1.2 Conditions d'ergodicité

Deux propriétés caractérisent la convergence d'une chaîne de Markov : l'irréductibilité et la périodicité. L'irréductibilité garantit notamment que la chaîne visite bien tout l'espace nécessaire.

Soit ψ une probabilité sur \mathcal{E} . La chaîne Z de noyau de transition K est dite ψ -irréductible si pour tout élément A de \mathcal{E} :

$$\psi(A) > 0 \Rightarrow P(\tau_A < +\infty | Z_0 = z) > 0, \quad \forall z \in E \quad (4.9)$$

où τ_A est le temps de retour dans A : $\tau_A : \min\{k \geq 1, Z_k \in A\}$. L'irréductibilité exprime le fait que la chaîne atteigne tout ensemble de ψ -mesure non nulle en un temps fini avec une probabilité non nulle.

L'irréductibilité implique l'unicité de la loi invariante π si celle-ci existe [79] :

Théorème 1 *Soit Z une chaîne de Markov ψ -irréductible, de loi invariante π . Alors, la chaîne est π -irréductible et π est l'unique loi invariante de la chaîne.*

On peut montrer que l'irréductibilité garantit la convergence des moyennes empiriques vers l'espérance. Pour obtenir la convergence en variation totale, une condition supplémentaire d'apériodicité est nécessaire. On définit un cycle de longueur l pour une chaîne de noyau de transition K comme une famille d'ensembles disjoints $\{A_0, \dots, A_{l-1}\}$ telle que $K(z, A_j) = 1$ pour $j = i + 1 \pmod l$ et tout $z \in A_i$. Ainsi, partant de A_i la chaîne atteint A_j en une transition presque sûrement. La période de la chaîne est le plus grand entier l pour lequel un cycle de longueur l existe. La chaîne est dite apériodique si la période vaut 1.

Irréductibilité et apériodicité impliquent ergodicité [79] :

Théorème 2 *Soit Z un chaîne de Markov de noyau de transition K , irréductible, apériodique et de loi invariante π . Alors :*

$$\|K^l(z, \cdot) - \pi(\cdot)\| \rightarrow_{l \rightarrow +\infty} 0, \quad \text{pour } \pi\text{-p.s. tout } z \in E \quad (4.10)$$

La simulation d'une loi π par chaîne de Markov consiste à construire une chaîne de Markov de loi stationnaire π . La première phase de la simulation est une phase de convergence (au sens de la relation 4.10), dite de "burn-in", après laquelle on suppose que la chaîne a atteint son régime stationnaire. Après cette étape, les états de la chaîne sont alors retenus pour former un échantillon (non-indépendant) de loi π .

L'étude de la vitesse de convergence du noyau K^l vers π , bien que capitale, n'est pas abordée ici. Citons quelques références récentes sur le sujet [43, 14, 25, 77, 93].

4.1.3 Méthodes de Monte Carlo

Une des applications essentielles des chaînes de Markov est le calcul d'intégrales par des méthodes de Monte Carlo. Supposons que l'on veuille calculer l'espérance sous la loi π d'une fonction g telle que $\int_E |g(z)|\pi(z)\mu(dz) < +\infty$:

$$E_\pi(g) = \int_E g(z)\pi(z)\mu(dz) \quad (4.11)$$

La loi des grands nombres assure que la moyenne empirique d'un échantillon $\{g(z_i), i = 1, l\}$, où les z_i sont des réalisations indépendantes de la loi π , converge vers l'espérance :

$$S_l(g) = \frac{1}{l} \sum_{i=1}^l g(z_i) \xrightarrow{l \rightarrow +\infty} \mathbb{E}_\pi(g) \quad (4.12)$$

Ainsi, à l'aide d'un échantillon de π , on peut estimer toute probabilité relative à π :

$$\pi(A) = \int_A \pi(z) \mu(dz) \approx \frac{1}{l} \sum_{i=1}^l 1_A(z_i), \quad \forall A \in \mathcal{E} \quad (4.13)$$

Remarquons que l'approximation dans le calcul de l'intégrale peut être contrôlée. En effet, si les z_i sont indépendants, la variance de l'estimateur est :

$$\text{Var}(S_l(g)) = \frac{1}{l} \text{Var}_\pi(g) \quad (4.14)$$

La variance $\text{Var}_\pi(g)$, *a priori* inconnue, peut également être estimée par intégration de Monte Carlo, et ainsi de suite éventuellement.

A la différence du schéma précédent, lorsque l'on simule la loi π à l'aide d'une chaîne de Markov, les différents états ne sont plus nécessairement indépendants. Cependant, sous réserve d'ergodicité, la convergence de l'intégrale de Monte Carlo est préservée. En effet, si \mathcal{Z} est une chaîne de Markov irréductible et de loi invariante π , le théorème ergodique implique [94] :

$$S_l(g) = \frac{1}{l} \sum_{i=1}^l g(z_i) \xrightarrow{l \rightarrow +\infty} \mathbb{E}_\pi(g) \quad (4.15)$$

La variance de l'estimateur $S_l(g)$ fait cependant intervenir les corrélations entre les différents états de la chaîne :

$$\text{Var}(S_l(g)) = \frac{1}{l^2} \sum_{i=1}^l \sum_{h=-i+1}^{l-i} \text{Cov}(g(z_i), g(z_{i+h})) \quad (4.16)$$

Définissons alors la portée intégrale comme suit [51, 65] :

$$A(g) = \sum_{h=-\infty}^{+\infty} \frac{\text{Cov}(g(z_i), g(z_{i+h}))}{\text{Var}_\pi(g)} \quad (4.17)$$

La portée intégrale dépend directement du noyau de transition K de la chaîne. La variance de l'estimateur $S_l(g)$ peut alors être approché par :

$$\text{Var}(S_l(g)) \approx \frac{A(g)}{l} \text{Var}_\pi(g) \quad (4.18)$$

Plus la portée intégrale est grande, plus il faut augmenter la taille l de l'échantillon, donc le nombre d'itérations, afin d'obtenir un estimateur de la qualité de l'estimateur indépendant. Ceci explique pourquoi il faut généralement préférer des noyaux de transition qui garantissent un mélange important à l'intérieur de la chaîne. En pratique, on a intérêt à ne retenir qu'un état de la chaîne toutes les k itérations, afin de garder une taille d'échantillon raisonnable tout en diminuant la portée intégrale.

L'estimation de $\text{Var}(S_l(g))$ est d'une importance capitale en pratique. Elle permet notamment de comparer les vitesses de convergence des intégrales de Monte Carlo pour différents algorithmes. Une méthode courante consiste à employer des moyennes calculées par paquets ("batch means"). Soit $l = ks$ le nombre d'itérations, avec k assez grand. On peut alors calculer pour chaque cycle de longueur k :

$$S_k^i(g) = \frac{1}{k} \sum_{j=(i-1)k+1}^{ik} g(z_j), \quad \forall i = 1, s \quad (4.19)$$

Chaque $S_k^i(g)$ a approximativement pour variance $\text{Var}(S_k(g))$ et pour moyenne $S_l(g)$. Une estimation de $A(g)\text{Var}_\pi(g)$ est alors :

$$\frac{k}{s-1} \sum_{i=1}^s (S_k^i(g) - S_l(g))^2 \quad (4.20)$$

4.1.4 Sur l'échantillonnage d'importance

Avant d'aborder l'important problème de la simulation de la loi π , au paragraphe 4.2, nous présentons dans ce paragraphe un aperçu des techniques d'échantillonnage d'importance. Celles-ci font partie des méthodes de réduction de variance permettant d'améliorer les performances des algorithmes de Monte Carlo par chaînes de Markov [46]. On peut également se référer à [80] pour une utilisation dans des problèmes d'optimisation par recuit simulé.

Principe

Reprenons le calcul de l'intégrale 4.11. Par Monte Carlo, cette intégrale est approchée par la moyenne arithmétique $S_l(g)$ donnée par l'équation 4.12.

Utilisons la décomposition 4.1 : $\pi(z) = r(z)\tilde{\pi}(z)$ avec $\tilde{\pi}$ une loi définie sur E telle que $(\tilde{\pi}(z) = 0) \Rightarrow (\pi(z) = 0)$. L'intégrale 4.1 peut alors s'écrire :

$$\mathbb{E}_\pi(g) = \int_E g(z) \frac{\pi(z)}{\tilde{\pi}(z)} \tilde{\pi}(z) dz \quad (4.21)$$

$$= \mathbb{E}_{\tilde{\pi}}(gr) \quad (4.22)$$

Autrement dit, l'espérance de g sous π peut être vue comme l'espérance de gr sous $\tilde{\pi}$. On obtient un estimateur de $E_\pi(g)$ en générant un échantillon $\{\tilde{z}_i, i = 1, l\}$ de loi $\tilde{\pi}$:

$$\tilde{S}_l(g) = \frac{1}{l} \sum_{i=1}^l g(\tilde{z}_i)r(\tilde{z}_i) \quad (4.23)$$

En pratique cependant, les densités $\pi(z)$ et $\tilde{\pi}(z)$ ne sont connues qu'à une constante de proportionnalité a près. Notons $\lambda_i = ar(\tilde{z}_i) = a\frac{\pi(\tilde{z}_i)}{\tilde{\pi}(\tilde{z}_i)}$. On a :

$$\frac{1}{l} \sum_{i=1}^l \lambda_i \approx \int_E a \frac{\pi(z)}{\tilde{\pi}(z)} \tilde{\pi}(z) dz \quad (4.24)$$

$$\approx a \quad (4.25)$$

Finalement, un estimateur de l'intégrale 4.1 s'écrit :

$$\tilde{S}_l(g) = \frac{1}{\sum_{j=1}^l \lambda_j} \sum_{i=1}^l \lambda_i g(\tilde{z}_i) \quad (4.26)$$

Autrement dit, lorsque les densités ne sont connues qu'à un facteur de proportionnalité près, il convient de normaliser les pondérateurs λ_i avant de calculer l'intégrale de Monte Carlo.

Un des intérêts de l'échantillonnage d'importance est la réduction de variance de l'estimateur. Comparons en effet sous hypothèse d'indépendance des échantillons les variances d'estimation de $S_l(g)$ et de $\tilde{S}_l(g)$:

$$\text{Var}(S_l(g)) = \frac{1}{l} \int_E (g(z) - E_\pi(g)) \pi(z) dz \quad (4.27)$$

$$\text{Var}(\tilde{S}_l(g)) = \frac{1}{l} \int_E (g(z)r(z) - E_\pi(g)) \tilde{\pi}(z) dz \quad (4.28)$$

Pour chaque fonction g , il existe donc un choix de $\tilde{\pi}$ optimal au sens de la réduction de la variance. Il est facile de montrer que cet optimum est atteint lorsque le produit gr est constant, soit pour $\tilde{\pi} \propto g\pi$ [91]. Même si en pratique on ne sait simuler selon cette loi, ce résultat donne une idée d'un choix judicieux pour $\tilde{\pi}$.

Dans [46], il est proposé d'appliquer une technique d'échantillonnage d'importance pour choisir le noyau de transition d'une chaîne de Markov. Ainsi, au lieu de prendre un noyau de loi stationnaire π , on peut préférer échantillonner une loi de la forme $\tilde{\pi} \propto \pi^{\frac{1}{\beta}}$, avec $\beta > 1$, pour laquelle la masse est répartie plus uniformément. Ceci permet d'augmenter le mélange à l'intérieur de la chaîne, ce qui peut être intéressant pour visiter des modes isolés de π .

Réduction de l'espace d'intégration

Nous proposons d'appliquer une technique d'échantillonnage d'importance pour estimer une espérance sur E à partir de simulations générées dans un sous-espace E_T de E . Reprenons la décomposition de π sous la forme 4.1. Celle-ci peut également s'écrire :

$$\pi(z) = r(z)\tilde{\pi}(z) \quad (4.29)$$

$$= \frac{r(z)}{r_T(z_T)}\tilde{\pi}(z|z_T)r_T(z_T)\tilde{\pi}_T(z_T) \quad (4.30)$$

$$= \frac{r(z)}{r_T(z_T)}\tilde{\pi}(z|z_T)\pi_T(z_T) \quad (4.31)$$

$$(4.32)$$

où π_T et $\tilde{\pi}_T$ sont les lois marginales de π et $\tilde{\pi}$ sur E_T , r_T est le ratio entre π_T et $\tilde{\pi}_T$, et $\tilde{\pi}(\cdot|z_T)$ est la loi conditionnelle de z à z_T .

Cette décomposition suggère l'algorithme de simulation suivant :

1. pour $i = 1, l$, on simule z_T^i selon la loi π_T sur E_T (par exemple à l'aide d'une chaîne de Markov, cf. paragraphe 4.2),
2. pour tout $i \leq l$, on simule le vecteur z^i conditionnellement à z_T^i selon la loi $\tilde{\pi}(\cdot|z_T)$,
3. on calcule les pondérateurs :

$$\lambda_i = \frac{r(z^i)}{r_T(z_T^i)} \quad (4.33)$$

et l'espérance de g sous π est obtenue par échantillonnage d'importance selon l'équation 4.26.

Cet algorithme revient à utiliser la loi $\tilde{\pi}(\cdot|z_T)\pi_T(z_T)$ comme loi d'échantillonnage.

L'intérêt de cette décomposition réside dans une éventuelle réduction du temps de calcul. En effet, la simulation par chaîne de Markov nécessite la simulation d'un grand nombre d'états dont certains seulement sont conservés pour l'intégration de Monte Carlo, soit que l'on supprime les premiers états pendant la phase de "burn-in", soit que l'on ne conserve qu'une fraction des états simulés afin de diminuer les corrélations à l'intérieur de l'échantillon simulé. Or, souvent, c'est la phase de simulation qui est la plus coûteuse numériquement, le coût augmentant avec la taille du vecteur à simuler. Il existe donc un intérêt direct à réduire la taille du vecteur pour la simulation markovienne. De plus, plus la dimension de l'espace E est grande, plus les problèmes de convergence de l'intégrale de Monte Carlo deviennent critiques, car généralement le noyau de transition K ne permet que des transitions entre des états voisins et la chaîne "explore" lentement l'espace E . Le principe de cet algorithme est d'approcher la loi π par sa réduction π_T sur E_T ,

puis de corriger cette approximation grâce aux pondérateurs λ_i , l'intérêt étant que sur E_T le mélange est meilleur.

Dans une seconde étape, l'ensemble du vecteur z est généré, conditionnellement au sous-vecteur z_T . Cet algorithme suppose donc que l'on soit capable de simuler directement selon $\tilde{\pi}(\cdot|\cdot_T)$. Remarquons qu'il s'agit dans ce cas d'une simulation conditionnelle classique par la donnée d'un sous-ensemble du vecteur z . Dans le cas où $\tilde{\pi}$ est gaussienne notamment, cette étape peut être réalisée directement, par un algorithme du type conditionnement par krigeage (cf. annexe A).

Cet algorithme n'est réellement intéressant que si la loi π possède une structure spatiale assez forte (par exemple dans le cas gaussien si la portée de la covariance est longue) ; dans le cas contraire, il est possible que les pondérateurs soient dégénérés : tous voisins de 0, sauf quelques-uns qui "prennent" tout le poids. Lorsque le problème est la simulation d'une loi conditionnelle dans un cadre bayésien, où $\tilde{\pi}$ désigne la loi *a priori*, le choix du sous-ensemble T peut être guidé par l'étude de la structure spatiale de $\tilde{\pi}$. En première approximation et sous une hypothèse de stationnarité, celle-ci est décrite par sa covariance (éventuellement généralisée) $C(h)$ où h est la distance entre deux sites. Si la grille T est choisie régulière, de maille a , la covariance du vecteur Z_T est alors $C(\frac{h}{a})$, autrement dit la portée est divisée par a . On peut donc choisir la maille a en fonction de la portée de la covariance C , typiquement inférieure à la moitié de la portée, afin de conserver une information spatiale pertinente.

En conclusion, l'algorithme que nous proposons permet d'employer les méthodes de simulation conditionnelle utilisées couramment en géostatistique dans des problèmes de simulation conditionnelle plus complexes. L'intérêt premier que nous y voyons est une amélioration de la convergence des intégrales de Monte Carlo, et donc un gain en temps de calcul. Cependant, la décomposition proposée peut également être mise en œuvre dans une approche multi-échelle, en particulier dans la phase de "burn-in", pour accélérer la convergence vers la loi stationnaire : la chaîne de Markov est tout d'abord simulée sur une grille de maille importante, puis on décroît progressivement la taille de la maille. A chaque changement de taille de la maille, une simulation conditionnelle "classique" est nécessaire.

4.2 Algorithmes de simulation

Le problème de la simulation d'une loi π est ramené à celui de la construction d'une chaîne de Markov de loi stationnaire π . En particulier, il s'agit de déterminer un noyau de transition K qui garantit l'invariance de π :

$$\int_E K(z, A)\pi(z)\mu(dz) = \pi(A), \quad \forall A \in \mathcal{E} \quad (4.34)$$

Par rapport au paragraphe précédente, le problème est donc inversé : la loi π est fixée, et on cherche K . Une fois K déterminé, reste à s'assurer que la chaîne est

bien ergodique : irréductible et apériodique.

4.2.1 Echantillonneur de Gibbs

L'échantillonneur de Gibbs a été introduit par les frères Geman en 1984 dans le contexte de la reconstruction d'images [44]. Cette méthode repose sur l'utilisation des lois conditionnelles totales, selon une terminologie employée dans [10], et d'un balayage de l'ensemble des sites.

Principe

On considère un sous-ensemble T de S . La loi π peut se décomposer selon le schéma :

$$\pi(z) = \pi_T(z_T|z^T)\pi^T(z^T) \quad (4.35)$$

Le principe de l'échantillonneur de Gibbs est de substituer la valeur en T par une variable simulée selon la loi conditionnelle $\pi_T(\cdot|z^T)$. On dit alors que l'on a *relaxé* T . A chaque étape, la transition de z vers z' est la suivante :

$$\begin{aligned} z'^T &= z^T \\ z'_T &\sim \pi_T(\cdot|z^T) \end{aligned} \quad (4.36)$$

Le noyau de transition associé a donc pour expression :

$$K(z, dz') = \pi_T(z'_T|z^T)\mu_T(dz'_T)\delta_{z^T}(dz'^T) \quad (4.37)$$

Il est aisé de vérifier que la loi π est bien invariante par cette transition. En effet, pour tout $A \in \mathcal{E}$:

$$\begin{aligned} \int_E \pi(z)K(z, A)\mu(dz) &= \int_E \pi(z_T, z^T) \int_A \pi_T(z'_T|z^T)\delta_{z^T}(dz'^T)\mu_T(dz'_T)\mu(dz) \\ &= \int_E \int_A \pi(z_T, z^T)\pi_T(z'_T|z^T)\delta_{z^T}(dz'^T)\mu_T(dz'_T)\mu(dz) \\ &= \int_E \int_A \pi(z'_T, z'^T)\pi_T(z_T|z'^T)\delta_{z'^T}(dz^T)\mu_T(dz_T)\mu(dz') \\ &= \int_A \pi(z'_T, z'^T) \int_E \pi_T(z_T|z'^T)\delta_{z'^T}(dz^T)\mu_T(dz_T)\mu(dz') \\ &= \int_A \pi(z')\mu(dz') \\ &= \pi(A) \end{aligned}$$

Cependant, ce noyau de transition n'assure pas l'irréductibilité de la chaîne, puisqu'une partie du vecteur n'est pas relaxée. C'est pour cela qu'il est nécessaire de

visiter séquentiellement l'ensemble des sites. On a effectué un cycle lorsque chacun des sites a été relaxé au moins une fois. Soit (T_1, \dots, T_p) une partition de S , le noyau de transition s'écrit :

$$K(z, dz') = \prod_{i=1}^p \pi_{T_i}(z'_{T_i} | z'_{T_1}, \dots, z'_{T_{i-1}}, z_{T_{i+1}}, \dots, z_{T_p}) \mu_{T_i}(dz'_{T_i}) \quad (4.38)$$

Au i -ème balayage, on relaxe le vecteur z_{T_i} suivant la loi conditionnelle appropriée. Chacune de ces transitions garantit l'invariance de π , il est aisé de vérifier en reprenant le calcul précédent qu'il en est de même pour le noyau composé 4.38.

Il existe plusieurs stratégies de balayage : la plus simple consiste à visiter les parties T_i dans un ordre fixe et déterminé. On peut d'ailleurs remarquer que les mêmes résultats d'ergodicité sont valables si certains sites sont visités plusieurs fois au cours d'un même cycle. De manière plus générale, il est possible de choisir aléatoirement les sites à relaxer. La convergence est assurée dès que le balayage choisi permet de visiter chaque site une infinité de fois [10]. On parle alors de balayage aléatoire, par rapport au balayage systématique. Une dernière variante est possible : on peut tirer à chaque cycle une origine de manière aléatoire, puis visiter les sites de manière déterministe à partir de cette origine.

La mise en œuvre la plus simple de l'échantillonneur de Gibbs consiste à ne relaxer qu'un site à la fois. Le noyau de transition de l'équation 4.38 s'écrit alors :

$$K(z, dz') = \prod_{i=1}^n \pi_i(z'_i | z'_1, \dots, z'_{i-1}, z_{i+1}, \dots, z_n) \nu(dz'_i) \quad (4.39)$$

Nous allons voir cependant par la suite que dans certaines situations, il peut être plus intéressant de regrouper les sites.

Pour résumer, remarquons que l'utilisation de l'échantillonneur de Gibbs ne requiert que la connaissance des lois conditionnelles totales. En revanche, il suppose que l'on puisse simuler selon ces lois conditionnelles, ce qui explique que cet algorithme ne soit employé que dans certains cas particuliers où celles-ci ont une expression simple. C'est notamment le cas des champs de Markov (cf. annexe C). Cependant, il est également possible d'utiliser l'échantillonneur de Gibbs en combinaison avec l'algorithme de Metropolis-Hastings (cf. paragraphe 4.2.2). Nous avons insisté dans cette présentation sur l'application de l'échantillonneur de Gibbs aux simulations spatiales. La même technique peut être utilisée lorsque le vecteur Z regroupe plusieurs variables différentes, typiquement l'état du système et les paramètres du modèle. On simule ainsi tour à tour conditionnellement chaque composante du modèle.

La figure 4.1 montre la trajectoire d'une chaîne générée par l'échantillonneur de Gibbs, dans le cas d'une loi bivariable. On remarque notamment que les transitions ne peuvent avoir lieu que le long des axes, puisqu'à chaque étape seule une coordonnée est relaxée. Ceci explique que dans certains cas particuliers, l'irréduc-

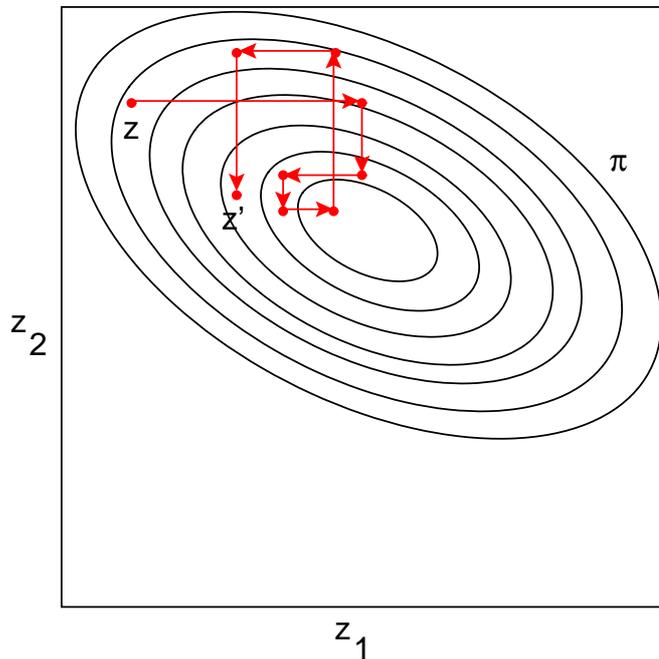


FIG. 4.1 – Transitions dans l’espace d’états générées par l’échantillonneur de Gibbs. La loi π , bivariable, est représentée en lignes de niveau. z' est l’état obtenu en partant de z après 4 cycles.

tibilité de la chaîne puisse être mise en défaut : c’est le cas notamment si l’espace E n’a pas la forme d’un espace produit.

Stratégies de groupement par blocs

Afin d’étudier l’intérêt de regrouper les sites à chaque relaxation, nous étudions le cas particulier d’une loi gaussienne. En effet, un certain nombre de résultats théoriques [98, 95, 97, 41], que nous reprenons, existent en ce qui concerne la vitesse de convergence de l’algorithme. Ces résultats permettent d’avoir une idée de l’impact d’un groupement par blocs dans des situations plus générales.

Les détails sont reportés en annexe B. Le résultat important est que l’échantillonneur de Gibbs peut être vu comme une itération stochastique de matrice de transition B :

$$Z' = BZ + W \tag{4.40}$$

où W est un vecteur gaussien. L’expression de la matrice B est donnée en annexe B.

Il existe un parallèle évident entre les itérations stochastiques de ce type et les itérations linéaires du point fixe, pour lesquelles on sait qu’une condition nécessaire

et suffisante de convergence est $\rho(B) < 1$, où $\rho(B)$ est le rayon spectral de la matrice B . On peut montrer que cette même condition est également nécessaire et suffisante dans le cas d'itérations stochastiques. De plus, le caractère symétrique défini positif de la matrice de covariance C entraîne automatiquement $\rho(B) < 1$. Une preuve de ces résultats est présentée dans [41].

Ainsi, il apparaît que le paramètre qui gouverne la vitesse de convergence de l'échantillonneur de Gibbs est le rayon spectral $\rho(B)$.

Dans [97] et [41], plusieurs stratégies de groupement par blocs et de visite des sites sont étudiées; voici en substance les principales conclusions en ce qui concerne les vitesses de convergence (ou ce qui est équivalent la valeur du rayon spectral) :

- le groupement par blocs l'emporte sur la stratégie consistant à relaxer chaque site séquentiellement [41],
- le groupement est d'autant plus efficace qu'il réunit les variables corrélées, de telle sorte que la corrélation entre des variables de deux blocs différents diminue [41],
- si la matrice de covariance inverse est tridiagonale par blocs, alors une visite déterministe des blocs est plus efficace qu'une visite aléatoire (théorème 5 et son corollaire 1 de [97]),
- si toutes les corrélations partielles de la matrice de covariance sont positives, alors une visite déterministe des blocs est plus efficace qu'une visite aléatoire (théorème 6 de [97]).

4.2.2 Algorithmes du type Metropolis-Hastings

Proposé par Metropolis en 1953 [78], l'algorithme a pris sa forme générale avec Hastings en 1970 [57]. Le principe est d'utiliser un noyau de transition auxiliaire pour proposer une transition, et d'accepter ou de rejeter cette transition selon une certaine probabilité. Il s'agit donc d'une méthode très générale, qui peut être déclinée sous plusieurs versions, selon le noyau de proposition utilisé.

Principe

Soit un noyau de transition $q(., .)$ défini sur $E \times E$, π -irréductible. $q(z, z')$ est la densité de probabilité de proposer le changement $z \rightarrow z'$. On suppose de plus que $(q(z, z') = 0) \Rightarrow (q(z', z) = 0)$.

A partir de ce noyau et de la loi cible π , le principe est de construire une probabilité d'acceptation a qui garantit la condition de réversibilité pour π . L'algorithme de Metropolis-Hastings (en abrégé MH) est alors composé de deux étapes :

1. on propose le changement $z \rightarrow z'$ suivant la densité $q(z, z')$,

2. on accepte ce changement avec la probabilité $a(z, z')$. Sinon, on conserve l'état z , soit $z' = z$.

Le noyau de transition résultant a pour expression :

$$K(z, dz') = q(z, z')a(z, z')\mu(dz') + \left(1 - \int_E q(z, y)a(z, y)\mu(dy)\right) \delta_z(dz') \quad (4.41)$$

Remarquons qu'en général, $K(z, \cdot)$ n'admet pas de densité et possède un point d'accumulation en z .

Afin de garantir la réversibilité de π sous K , il suffit de choisir a tel que :

$$\pi(z)q(z, z')a(z, z') = \pi(z')q(z', z)a(z', z), \quad \forall z' \neq z \quad (4.42)$$

En effet, on a alors pour tout $A, B \in \mathcal{E}^2$:

$$\begin{aligned} \int_A \pi(z)K(z, B)\mu(dz) &= \int_A \int_B \pi(z)q(z, z')a(z, z')\mu(dz)\mu(dz') \\ &\quad + \int_{A \cap B} \pi(z) \left(1 - \int_E \pi(z)q(z, y)a(z, y)\mu(dy)\right) \mu(dz) \\ &= \int_B \int_A \pi(z')q(z', z)a(z', z)\mu(dz')\mu(dz) \\ &\quad + \int_{B \cap A} \pi(z) \left(1 - \int_E \pi(z)q(z, y)a(z, y)\mu(dy)\right) \mu(dz) \\ &= \int_B \pi(z')K(z', A)\mu(dz') \end{aligned}$$

L'irréductibilité de K est elle-même assurée par l'irréductibilité de q . Une condition suffisante d'apériodicité est :

$$\mu \left(\left\{ z : 1 - \int_E q(z, y)a(z, y)\mu(dy) > 0 \right\} \right) > 0 \quad (4.43)$$

Ainsi, sur un ensemble de mesure non nulle, la probabilité de rester dans le même état est non nulle.

Citons deux choix de probabilité d'acceptation. La première, la plus répandue, est la dynamique de Metropolis :

$$a(z, z') = \min \left(1, \frac{\pi(z')q(z', z)}{\pi(z)q(z, z')} \right) \quad (4.44)$$

On peut montrer qu'elle est optimale parmi les choix possibles pour le critère de la variance d'estimation dans l'intégration de Monte Carlo [85].

Un autre choix possible est la dynamique de Barker :

$$a(z, z') = \frac{\pi(z')q(z', z)}{\pi(z)q(z, z') + \pi(z')q(z', z)} \quad (4.45)$$

Un certain nombre de résultats sur la convergence des algorithmes de type Metropolis-Hastings sont connus, nous renvoyons le lecteur à [98, 100].

Utilisation d'une loi auxiliaire

Reprenons tout d'abord la décomposition de la loi π de l'équation 4.1. Pour un noyau de transition $q(., .)$ donné, la probabilité d'acceptation dans Metropolis-Hastings s'écrit :

$$a(z, z') = \min \left(1, \frac{r(z')\tilde{\pi}(z')q(z', z)}{r(z)\tilde{\pi}(z)q(z, z')} \right), z \neq z' \quad (4.46)$$

On peut par exemple choisir un noyau qui est réversible pour la loi $\tilde{\pi}$, soit :

$$\tilde{\pi}(z)q(z, z') = \tilde{\pi}(z')q(z', z) \quad (4.47)$$

La probabilité d'acceptation s'exprime alors simplement en fonction de r :

$$a(z, z') = \min \left(1, \frac{r(z')}{r(z)} \right) \quad (4.48)$$

Cette remarque prend tout son sens dans le cas de lois conditionnelles où $\tilde{\pi}$ est la loi non conditionnelle ou marginale. Si on dispose déjà d'un noyau réversible pour la loi non conditionnelle (par exemple à la suite d'une itération MH), ce noyau peut être utilisé comme noyau de proposition pour la loi conditionnelle.

L'échantillonneur indépendant

Le principe est d'utiliser comme noyau de transition une loi $\tilde{\pi}$ fixée, et de proposer la transition indépendamment de l'état de la chaîne. Ainsi :

$$q(z, z') = \tilde{\pi}(z') \quad (4.49)$$

La probabilité d'acceptation s'écrit :

$$a(z, z') = \min \left(1, \frac{\pi(z')\tilde{\pi}(z)}{\pi(z)\tilde{\pi}(z')} \right) \quad (4.50)$$

En reprenant la décomposition 4.1, celle-ci ne dépend que de la fonction r :

$$a(z, z') = \min \left(1, \frac{r(z')}{r(z)} \right) \quad (4.51)$$

Dans le cas de la simulation d'une loi conditionnelle, un candidat naturel pour $\tilde{\pi}$ est la loi *a priori*. La transition est alors toujours acceptée si l'état proposé est plus vraisemblable que l'état courant, et rejetée suivant une probabilité qui ne dépend que du ratio des vraisemblances. Dans bien des cas, la vraisemblance a une expression relativement simple, et son calcul est très rapide. L'algorithme est ainsi facile à mettre en œuvre, et le temps de calcul d'une itération très court. En

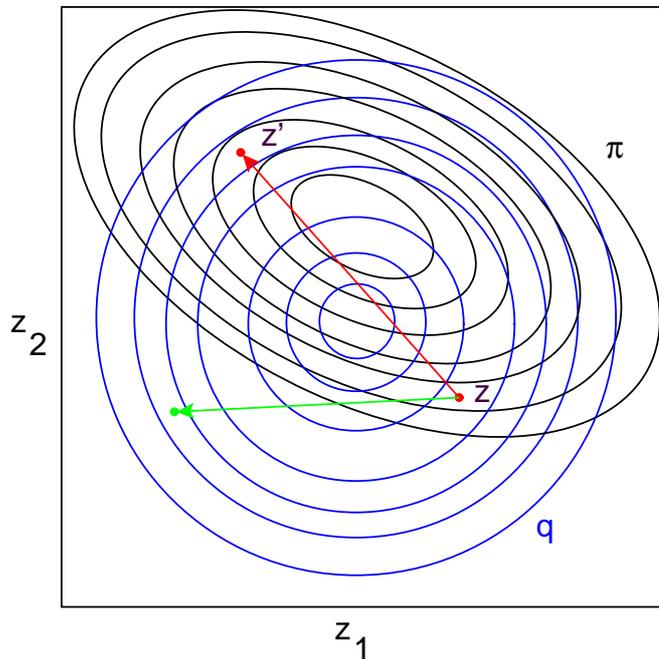


FIG. 4.2 – Transition dans l’espace d’états générée par l’échantillonneur indépendant. Le noyau de proposition $q = \tilde{\pi}$ est représenté en bleu, la loi cible π en noir. La transition rouge est acceptée (car $\pi(z')\tilde{\pi}(z) > \pi(z)\tilde{\pi}(z')$), en revanche la verte a une forte probabilité d’être rejetée.

revanche, la vitesse de convergence dépend directement de la qualité d’approximation de π par $\tilde{\pi}$. Le risque est de construire une chaîne qui reste bloquée très longtemps dans certains états de forte probabilité *a posteriori*.

Une transition d’une chaîne de Markov générée par un échantillonneur indépendant est représentée à la figure 4.2. Dans le cas idéal, la loi $\tilde{\pi}$ utilisée pour proposer les transitions charge toute la zone de l’espace d’états chargée par π .

Marche aléatoire

Dans une marche aléatoire, c’est l’amplitude du déplacement à partir de l’état courant qui suit une probabilité fixe. Le noyau de transition est symétrique :

$$q(z, z') = \tilde{\pi}(|z' - z|) \quad (4.52)$$

La probabilité d’acceptation prend dans ce cas une expression particulièrement simple :

$$a(z, z') = \min \left(1, \frac{\pi(z')}{\pi(z)} \right) \quad (4.53)$$

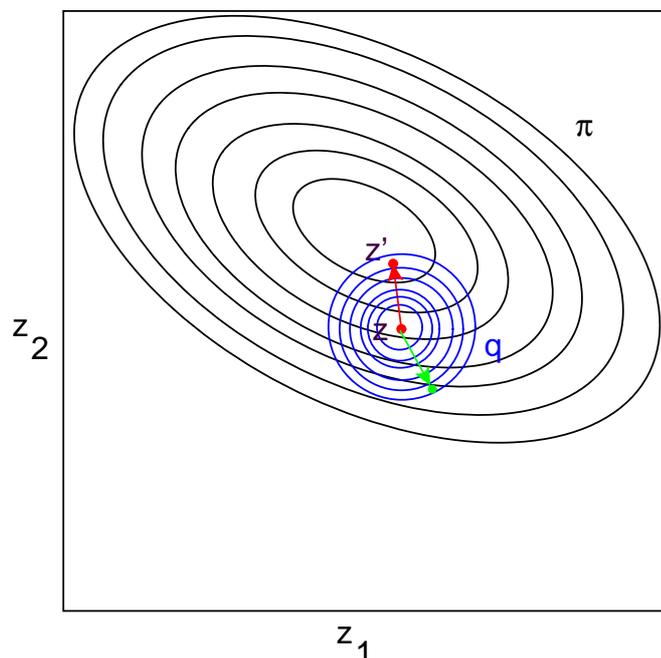


FIG. 4.3 – Transition dans l’espace d’états générée par une marche aléatoire. Le noyau de proposition q est représenté en bleu, la loi cible π en noir. La transition rouge est acceptée (car $\pi(z') > \pi(z)$), en revanche la verte a une forte probabilité d’être rejetée.

On peut par exemple choisir pour $\tilde{\pi}$ une loi gaussienne centrée en 0. Le comportement de l’algorithme dépend alors directement de la variance choisie. Si celle-ci est faible, les transitions sont fréquemment acceptées, mais les états successifs de la chaîne fortement corrélés, la chaîne visitant lentement E . Au contraire, une variance forte permet de proposer des transitions vers des états éloignés de l’état courant, mais le risque est de les rejeter fréquemment. Le choix de la variance est donc primordial. A la figure 4.3, on a représenté une transition de la chaîne générée par une marche aléatoire.

Diffusions de Langevin

Les deux algorithmes précédents sont particulièrement simples à mettre en œuvre. En revanche, un choix judicieux des lois de proposition nécessite idéalement une connaissance préalable de la forme de π . Pour des modèles plus complexes, notamment lorsque la taille du système est importante, cette connaissance fait souvent défaut. Une idée est donc d’utiliser une information sur la loi π pour construire le noyau de transition.

Supposons que π soit non nulle et différentiable et considérons l'équation de diffusion suivante :

$$dZ_t = dW_t + \frac{1}{2} \nabla \log \pi(Z_t) dt \quad (4.54)$$

où W est un mouvement brownien standard de dimension n . Lorsque π est relativement régulière, on peut montrer que la diffusion $(Z_t)_{t \geq 0}$ admet π comme loi stationnaire, de telle sorte que [99] :

$$\|K^t(z, \cdot) - \pi(\cdot)\| \rightarrow_{t \rightarrow \infty} 0 \quad (4.55)$$

pour π -p.s. tout $z \in E$, où $K^t(z, A) = P(Z_t \in A | Z_0 = z)$, $A \in \mathcal{E}$. Ainsi, une diffusion de Langevin est construite de telle sorte qu'elle admet π comme loi stationnaire.

Une mise en œuvre en temps discret de l'équation de diffusion 4.54 revient à proposer une transition selon une loi normale centrée en z auquel on ajoute le terme de dérive $\nabla \log \pi(z)$:

$$z' \sim \mathcal{N}\left(z + \frac{1}{2} h^2 \nabla \log \pi(z), h^2 I_n\right) \quad (4.56)$$

où h est le pas de discrétisation, et I_n la matrice identité sur E .

Dans [9], Besag fait cependant remarquer que la chaîne ainsi discrétisée n'assure qu'approximativement l'invariance de π . Une modification consiste à utiliser la transition 4.56 comme transition de proposition dans un algorithme MH et d'ajouter une étape d'acceptation / rejet. Le noyau admet alors pour densité :

$$q(z, z') = \phi\left(\frac{z' - z - \frac{1}{2} h^2 \nabla \log \pi(z)}{h}\right), \quad (4.57)$$

où ϕ désigne la densité d'une variable gaussienne standard, et la probabilité d'acceptation s'écrit :

$$a(z, z') = \min\left(1, \frac{\pi(z') \phi\left(\frac{z - z' - \frac{1}{2} h^2 \nabla \log \pi(z')}{h}\right)}{\pi(z) \phi\left(\frac{z' - z - \frac{1}{2} h^2 \nabla \log \pi(z)}{h}\right)}\right) \quad (4.58)$$

La connaissance de π à un facteur multiplicatif près est suffisante, puisque ce facteur constant disparaît dans le calcul du gradient de $\log \pi$. Le grand intérêt de cet algorithme est qu'une information sur la loi cible π est utilisée pour générer les transitions. L'algorithme se comporte comme une marche aléatoire centrée sur l'état courant auquel est ajouté un terme de dérive. Ainsi, à chaque itération, on procède tout d'abord à une descente de gradient, puis on visite aléatoirement l'espace avoisinant. Comme pour la marche aléatoire, le choix de l'écart type h est d'une grande influence.

Remarquons toutefois que l'algorithme de Langevin est beaucoup plus complexe

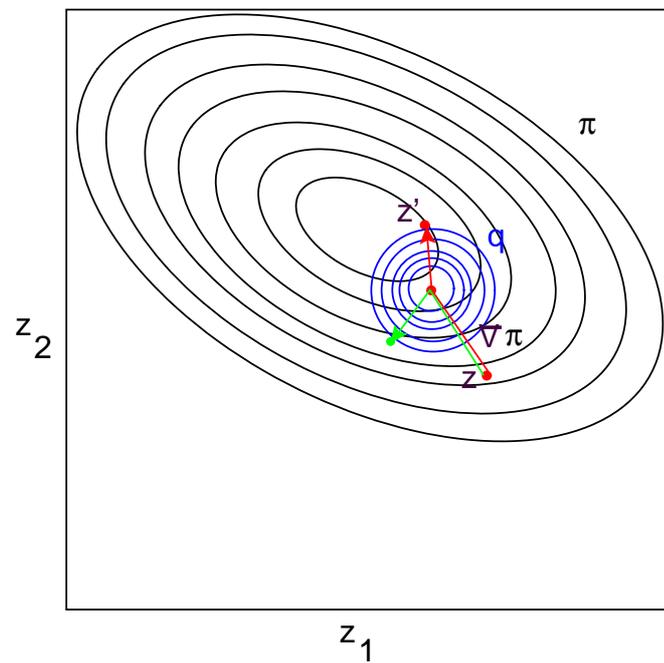


FIG. 4.4 – Transition dans l'espace d'états générée par une diffusion de Langevin. Le noyau de proposition q est représenté en bleu, la loi cible π en noir. Tout d'abord, on se déplace selon le gradient, puis on explore l'espace comme à l'aide d'une marche aléatoire. Dans cet exemple, les deux transitions (rouge et verte) ont de fortes chances d'être acceptées.

que les deux précédents : il nécessite à chaque itération deux calculs de gradient et deux évaluations de la loi π . Une transition générée par l'algorithme est représentée à la figure 4.4.

En pratique cependant, cet algorithme apparaît peu stable numériquement, le terme de dérive lié au gradient de $\log \pi$ pouvant exploser. Ce comportement est directement lié à la forme de la loi cible π . Une modification assez simple consiste à tronquer le gradient dès qu'il excède un seuil fixé M . Le terme de dérive prend alors la forme [96] :

$$\frac{1}{2} \frac{M \nabla \log \pi(z)}{\max(M, |\nabla \log \pi(z)|)} \quad (4.59)$$

On calcule alors la probabilité d'acceptation correspondante.

Des résultats sur les propriétés de convergence des algorithmes de type Langevin sont énoncés dans [113, 99, 96].

Gibbs-Metropolis

Lorsque l'espace E est de grande dimension, ou que le vecteur z est composé de variables distinctes (typiquement le vecteur d'état et les paramètres du modèle), il devient intéressant de scinder la simulation de z en plusieurs étapes. Autrement dit, on réduit la résolution d'un problème complexe à la résolution séquentielle de problèmes de complexité inférieure.

Le principe est de combiner une étape de l'échantillonneur de Gibbs et une étape d'acceptation / rejet du type Metropolis-Hastings. Considérons de nouveau une partition (T_1, \dots, T_p) de S . D'après l'échantillonneur de Gibbs, au i -ème balayage, il faut générer un vecteur z'_{T_i} selon la loi conditionnelle de densité :

$$\pi_{T_i}(z'_{T_i} | z^{T_i}) \quad (4.60)$$

Soit alors $q_{T_i}(\cdot, \cdot)$ un noyau défini sur $E_{T_i} \times E_{T_i}$. D'après Metropolis-Hastings, on génère un candidat z'_{T_i} suivant la densité $q_{T_i}(z_{T_i}, z'_{T_i})$, et cette transition est acceptée avec la probabilité :

$$a(z_{T_i}, z'_{T_i}) = \min \left(1, \frac{\pi_{T_i}(z'_{T_i} | z^{T_i}) q_{T_i}(z_{T_i}, z'_{T_i})}{\pi_{T_i}(z_{T_i} | z^{T_i}) q_{T_i}(z'_{T_i}, z_{T_i})} \right) \quad (4.61)$$

L'étape de Metropolis-Hastings garantit bien que la transition élémentaire a pour loi stationnaire $\pi_{T_i}(z'_{T_i} | z^{T_i})$, et en balayant séquentiellement les blocs T_i l'échantillonneur de Gibbs assure que la chaîne admette π pour loi stationnaire [10].

Pour le choix des noyaux q_{T_i} , $i = 1, p$, on peut par exemple considérer des marches aléatoires ou des échantillonneurs indépendants définis sur E_{T_i} .

4.2.3 Metropolis-Hastings multiple

Nous proposons de généraliser l'algorithme de Metropolis-Hastings en considérant à chaque itération non pas une transition possible mais tout un ensemble de transitions possibles.

Principe

Considérons un noyau de transition de densité $q(\cdot|\cdot)$ défini sur $E^p \times E$, pour un entier p fixé, tel que :

$$q(z'_1, \dots, z'_p | z) = q(z'_{\sigma(1)}, \dots, z'_{\sigma(p)} | z) \quad (4.62)$$

pour toute permutation σ de $\{1, \dots, p\}$. Nous appelons un tel noyau de transition un noyau multiple. Nous attirons l'attention sur le fait que z'_i ne désigne plus une composante du vecteur z' , mais un élément de E .

A chaque itération, on propose tout d'abord un ensemble de transitions $\{z'_1, \dots, z'_p\}$ selon la loi $q(\cdot|z)$. Notons $z'_0 = z$ et $\Omega_p(z) = \{z'_0, z'_1, \dots, z'_p\}$. Pour tout $i = 0, p$, on peut calculer la probabilité $\pi(z'_i)$ et construire sur $\Omega_p(z)$ un système de probabilité $a = (a(z, z'_i))_{i=0,p}$ de la manière suivante :

$$a(z, z'_i) = \frac{\pi(z'_i)q(z'^i, z|z'_i)}{\sum_{j=1}^p \pi(z'_j)q(z'^j, z|z'_j) + \pi(z)q(z'_1, \dots, z'_p|z)}, \quad i = 1, p \quad (4.63)$$

$$a(z, z) = \frac{\pi(z)q(z'_1, \dots, z'_p|z)}{\sum_{j=1}^p \pi(z'_j)q(z'^j, z|z'_j) + \pi(z)q(z'_1, \dots, z'_p|z)} \quad (4.64)$$

où $z'^i = \{z'_1, \dots, z'_{i-1}, z'_{i+1}, \dots, z'_p\}$, la notation sous forme d'ensemble étant justifiée du fait de l'hypothèse 4.62.

On choisit alors z' dans $\Omega_p(z)$ selon le système de probabilité $a : z' = z'_i$ avec la probabilité $a(z, z'_i)$. Le noyau de transition correspondant $K(z, dz')$ s'écrit :

$$K(z, dz') = \int_{z'_1, \dots, z'_p} \left(\sum_{i=0}^p a(z, z'_i) \delta_{z'_i}(dz') \right) q(z'_1, \dots, z'_p | z) \mu(dz'_1) \dots \mu(dz'_p) \quad (4.65)$$

Montrons que le noyau ainsi construit est réversible pour π . On a :

$$\begin{aligned} \pi(z)K(z, dz') &= \int_{z'_1, \dots, z'_p} a(z, z) \pi(z) q(z'_1, \dots, z'_p | z) \mu(dz'_1) \dots \mu(dz'_p) \delta_z(dz') \\ &\quad + \sum_{i=1}^p \int_{z'^i} a(z, z') \pi(z) q(z'^i, z' | z) \mu^i(dz'^i) \mu(dz') \end{aligned}$$

Le premier terme est évidemment symétrique en z et z' . De plus, chaque terme de la somme est égal à :

$$\int_{z^i} \frac{\pi(z')q(z^i, z|z')\pi(z)q(z^i, z|z)}{\sum_{j \neq i} \pi(z'_j)q(z^{i,j}, z', z|z'_j) + \pi(z')q(z^i, z|z') + \pi(z)\pi(z^i, z'|z)} \mu^i(dz^i)$$

où $z^{i,j} = \{z'_k, k \neq i, k \neq j\}$. L'hypothèse d'invariance de q par des permutations 4.62 assure que le dénominateur est symétrique en z et z' . Le numérateur l'est également ; d'où finalement :

$$\int_A \pi(z)K(z, B) = \int_B \pi(z)K(z, A), \quad \forall A, B \in \mathcal{E}^2$$

Cette démarche généralise bien l'algorithme de Metropolis-Hastings, mais seulement pour la dynamique de Barker. En effet, supposons que $p = 1$, alors le système de probabilité a se réduit à :

$$\begin{aligned} a(z, z'_1) &= \frac{\pi(z'_1)q(z|z'_1)}{\pi(z)q(z'_1|z) + \pi(z'_1)q(z|z'_1)} \\ a(z, z) &= 1 - a(z, z'_1) \end{aligned}$$

Un exemple immédiat de loi multiple vérifiant la relation 4.62 est le cas d'indépendance conditionnelle :

$$q(z'_1, \dots, z'_p|z) = \prod_{i=1}^p q(z'_i|z) \quad (4.66)$$

Ce cas particulier n'est cependant pas très intéressant en pratique, car le coût de calcul est directement proportionnel à la taille de l'ensemble généré.

Noyau symétrique

L'hypothèse d'invariance par des permutations 4.62 peut être contraignante en pratique ; on peut la relâcher légèrement si le noyau possède de plus une propriété de symétrie. Supposons que pour tout i , il existe une permutation σ_i de $\{1, \dots, p\} \setminus \{i\}$ telle que :

$$q(z'_1, \dots, z'_p|z) = q(z'_{\sigma_i(1)}, \dots, z'_{\sigma_i(i-1)}, z, z'_{\sigma_i(i+1)}, \dots, z'_{\sigma_i(p)}|z'_i) \quad (4.67)$$

Dans ce cas, on peut choisir le système de probabilité a suivant :

$$a(z, z'_i) = \frac{\pi(z'_i)}{\sum_{j=0}^p \pi(z'_j)}, \quad i = 0, p \quad (4.68)$$

La condition de réversibilité pour π est bien vérifiée, en effet :

$$\begin{aligned} \pi(z)K(z, dz') &= \int_{z'_1, \dots, z'_p} a(z, z)\pi(z)q(z'_1, \dots, z'_p|z)\mu(dz'_1) \dots \mu(dz'_p)\delta_z(dz') + \\ &\quad \sum_{i=1}^p \int_{z'^i} a(z, z')\pi(z)q(z'_1, \dots, z'_{i-1}, z', z'_{i+1}, \dots, z'_p|z)\mu^i(dz'^i)\mu(dz') \end{aligned}$$

Le premier terme est évidemment symétrique en z et z' . De plus, chaque terme de la somme est égal à :

$$\begin{aligned} &\int_{z'^i} \frac{\pi(z')\pi(z)}{\sum_{j \neq i} \pi(z'_j) + \pi(z') + \pi(z)} q(z'_1, \dots, z'_{i-1}, z', z'_{i+1}, \dots, z'_p|z)\mu^i(dz'^i) \\ &= \int_{z'^i} \frac{\pi(z')\pi(z)}{\sum_{j \neq i} \pi(z'_j) + \pi(z') + \pi(z)} q(z'_{\sigma_i(1)}, \dots, z'_{\sigma_i(i-1)}, z, z'_{\sigma_i(i+1)}, \dots, z'_{\sigma_i(p)}|z')\mu^i(dz'^i) \\ &= \int_{y^i} \frac{\pi(z')\pi(z)}{\sum_{j \neq i} \pi(y_j) + \pi(z') + \pi(z)} q(y_1, \dots, y_{i-1}, z, y_{i+1}, \dots, y_p|z')\mu^i(dy^i) \end{aligned}$$

Chaque terme de la somme est donc bien symétrique en z et z' , ce qui assure la réversibilité.

Remarquons que si le noyau vérifie la propriété de symétrie 4.67 pour la permutation identité, soit :

$$q(z'_1, \dots, z'_p|z) = q(z'_1, \dots, z'_{i-1}, z, z'_{i+1}, \dots, z'_p|z'_i)$$

alors le noyau vérifie également l'invariance par les permutations 4.62. En effet, en appliquant deux fois la relation précédente pour i et $i+1$, on montre immédiatement que $q(z'_1, \dots, z'_i, z'_{i+1}, \dots, z'_p|z) = q(z'_1, \dots, z'_{i+1}, z'_i, \dots, z'_p|z)$. Le résultat s'obtient en remarquant que les transpositions génèrent les permutations.

Réciproquement, si le noyau vérifie l'invariance par les permutations 4.62, une condition suffisante pour vérifier la relation de symétrie 4.67 est :

$$q_i(z'_i|z) = q_i(z|z'_i) \quad (4.69)$$

q admettant la décomposition :

$$q(z'_1, \dots, z'_p|z) = q_i(z'_i|z)q^i(z'^i|z'_i, z), \quad i = 1, p. \quad (4.70)$$

Utilisation d'une loi auxiliaire

Considérons enfin le cas où la loi cible π se décompose selon :

$$\pi(z) = r(z)\tilde{\pi}(z) \quad (4.71)$$

On peut alors construire un noyau vérifiant la condition de symétrie 4.67 en choisissant un noyau réversible pour $\tilde{\pi}$. La condition est une condition de multi-réversibilité associée à une famille de permutations. Supposons que pour tout i , il existe une permutation σ_i telle que :

$$\tilde{\pi}(z)q(z'_1, \dots, z'_p|z) = \tilde{\pi}(z'_i)q(z'_{\sigma_i(1)}, \dots, z'_{\sigma_i(i-1)}, z, z'_{\sigma_i(i+1)}, \dots, z'_{\sigma_i(p)}|z'_i) \quad (4.72)$$

Dans ce cas, le système de probabilité a devient :

$$a(z, z'_i) = \frac{r(z'_i)}{\sum_{j=0}^p r(z'_j)}, \quad i = 0, p \quad (4.73)$$

Cette expression est notamment intéressante dans le cas de lois conditionnelles. Il suffit de trouver un noyau multi-réversible pour la loi non conditionnelle $\tilde{\pi}$, et l'on dispose immédiatement d'un noyau multiple pour la loi conditionnelle π .

Reprenons l'exemple précédent du noyau conditionnellement indépendant. Si de plus le noyau simple $q(z'_i|z)$ est indépendant de z , on obtient une généralisation de l'échantillonneur indépendant :

$$q(z'_1, \dots, z'_p|z) = \prod_{i=1}^p \tilde{\pi}(z'_i) \quad (4.74)$$

Ce noyau vérifie bien la condition de multi-réversibilité pour $\tilde{\pi}$, en effet :

$$\begin{aligned} \tilde{\pi}(z)q(z'_1, \dots, z'_p|z) &= \tilde{\pi}(z)\tilde{\pi}(z'_1) \dots \tilde{\pi}(z'_p) \\ &= \tilde{\pi}(z'_i)q(z'_1, \dots, z'_{i-1}, z, z'_{i+1}, \dots, z'_p|z'_i), \quad \forall 1 \leq i \leq p \end{aligned}$$

Le cas gaussien

Nous allons considérer un cas particulier où la génération de l'ensemble $\Omega_p(z)$ est peu coûteuse. Prenons pour loi auxiliaire la loi gaussienne : $\tilde{\pi}(z) = \phi(z)$, et considérons la construction suivante :

1. on commence par simuler un état w selon ϕ , indépendamment de l'état courant z ,
2. les états z'_i sont ensuite générés en calculant pour $\theta_1, \dots, \theta_p$ la combinaison linéaire :

$$z'_i = z \cos \theta_i + w \sin \theta_i \quad (4.75)$$

On pose de plus $\theta_0 = 0$ et $z'_0 = z$.

Il s'agit d'une forme particulière de noyau multiple, puisqu'une fois que l'état indépendant w a été généré, tous les autres peuvent être déduits par relation fonctionnelle. Le noyau proposé admet en effet pour expression :

$$q(z'_1, \dots, z'_p|z) = \int_E \phi(w) \prod_{i=1}^p \delta_{z \cos \theta_i + w \sin \theta_i}(z'_i) \mu(dw) \quad (4.76)$$

Ce noyau n'est pas une densité au sens stricte, du fait du produit de fonctions de Dirac, mais cela ne pose pas de problème, car dans l'équation 4.65 on intègre le noyau multiple q pour générer le noyau de transition K .

Propriété 1 *Le noyau multiple donné par 4.76 où les θ_i réalisent une discrétisation régulière de $[0, 2\pi[$:*

$$\theta_i = i \frac{2\pi}{p+1}, \forall i = 1, p \quad (4.77)$$

vérifie la relation de multi-réversibilité 4.72 pour la famille de permutations :

$$\sigma_i(j) = i - j \quad \text{si } j < i, \quad \sigma_i(j) = i - j + p + 1 \quad \text{si } j > i \quad (4.78)$$

La preuve de ce résultat est présentée en annexe D. Pour $p = 1$, θ_1 peut être pris quelconque dans $[0, 2\pi[$.

L'algorithme de simulation est le suivant :

1. simuler w selon ϕ , indépendamment de l'état courant z ,
2. construire :

$$z'_i = z \cos\left(i \frac{2\pi}{p+1}\right) + w \sin\left(i \frac{2\pi}{p+1}\right), \quad i = 1, p, \quad (4.79)$$

3. calculer pour tout $i = 0, p$:

$$a(z, z'_i) = \frac{r(z'_i)}{\sum_{j=0}^p r(z'_j)}, \quad (4.80)$$

4. tirer $z'_i \sim a(z, z'_i)$ et poser $z' = z'_i$.

La construction de l'ensemble de proposition $\Omega_p(z)$ n'exige la simulation que d'un seul état w . Le temps d'exécution de l'algorithme est donc raisonnable si l'évaluation de la fonction r est assez rapide.

La figure 4.5 illustre le principe de cet algorithme, dans le cas de la simulation d'une loi à deux dimensions. Tout d'abord, un état indépendant est simulé selon la loi gaussienne ϕ . Puis, l'ensemble de proposition $\Omega_p(z)$ est construit : il prend la forme d'une ellipse dans l'espace d'états. On comprend qu'un nombre beaucoup plus grand d'états peut être atteint à chaque itération, alors qu'avec un algorithme du type Metropolis-Hastings, un seul état est proposé pour la transition. L'intérêt est que ce grand nombre d'états peut être généré à moindre coût, à partir d'une seule simulation indépendante. De plus, la structure même de la trajectoire construite permet de traverser des régions de forte probabilité selon π .

Cet algorithme est particulièrement adapté à la simulation de lois gaussiennes conditionnelles. Il peut être également utilisé dans le cas où la loi gaussienne est une bonne approximation de la loi *a priori*.

L'utilisation de combinaisons linéaires de fonctions aléatoires gaussiennes de la forme 4.75 n'est pas anodine. En effet, la loi gaussienne est invariante par ce type de combinaisons : si Z_1 et Z_2 sont deux fonctions aléatoires gaussiennes

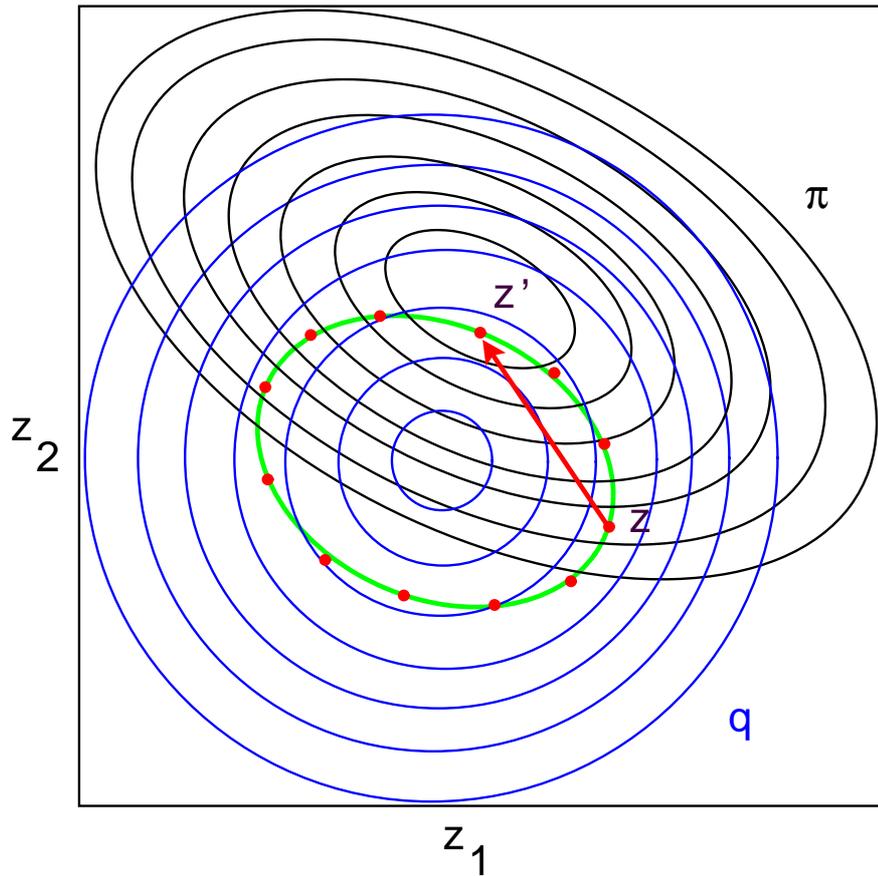


FIG. 4.5 – Transition dans l'espace d'états générée par l'algorithme Metropolis-Hastings multiple (cas gaussien). La loi auxiliaire (gaussienne) est représentée en bleu, la loi cible π en noir. L'ensemble de proposition $\Omega_p(z)$ généré à partir de l'état courant z et d'une simulation indépendante w (points rouges) est situé sur une ellipse (en vert). A chaque état de proposition est associée une probabilité de transition. Dans cet exemple, la transition (en rouge) se fait vers l'état de plus forte π -probabilité.

indépendantes d'espérance nulle et de même covariance C , et Θ une variable aléatoire quelconque indépendante de Z_1 et Z_2 , $Z_1 \cos \Theta + Z_2 \sin \Theta$ est une fonction aléatoire gaussienne de covariance C . Dans [75], Matheron montre même que si Z_1 et Z_2 sont deux fonctions aléatoires gaussiennes stationnaires et Θ une fonction aléatoire intrinsèque, alors la combinaison linéaire définie par $Z(x) = Z_1(x) \cos \Theta(x) + Z_2(x) \sin \Theta(x)$ est une fonction aléatoire gaussienne stationnaire d'ordre 2.

Plus récemment, Hu [58, 59] a utilisé cette relation dans le but de générer des simulations conditionnelles. Dans son approche, le conditionnement est représenté par une fonction de coût. La méthode qu'il propose consiste à choisir à chaque étape la combinaison linéaire qui minimise la fonction de coût, et à répéter l'opération jusqu'à obtenir un minimum local. Cet algorithme, qui a reçu un intérêt certain dans la communauté géostatistique, ne permet pas de simuler rigoureusement la loi cible, comme ses auteurs l'ont eux-mêmes reconnu [89]. L'approche que nous proposons, qui s'appuie sur la théorie des chaînes de Markov, garantit la convergence vers la loi cible.

Version Gibbs-Metropolis dans le cas gaussien

Il est également possible de combiner l'algorithme Metropolis-Hastings multiple avec l'échantillonneur de Gibbs, dans le but de ne relaxer à chaque balayage qu'un sous-vecteur du vecteur z . L'algorithme est le même que pour Metropolis-Hastings simple (paragraphe 4.2.2), mais nous donnons ici les détails dans le cas où l'on utilise une loi auxiliaire $\tilde{\pi}$ gaussienne (algorithme précédent).

Considérons tout d'abord deux ensembles T_1 et T_2 , non nécessairement disjoints. Le problème est de simuler un vecteur Z_1 , restriction de la fonction aléatoire Z au sous-ensemble T_1 conditionnellement au vecteur $Z_2 = z_2$, restriction de Z au sous-ensemble T_2 . On propose pour cela d'utiliser comme loi auxiliaire une loi gaussienne conditionnelle. Notons C_{11} et C_{22} les matrices de covariance sur T_1 et T_2 , et C_{12} la matrice de covariance croisée. La loi à simuler se met sous la forme :

$$\pi(z_1|z_2) \propto r(z_1, z_2)\phi(z_1|z_2) \quad (4.81)$$

Or, $\phi(z_1|z_2)$ est la densité d'une loi gaussienne de moyenne $z_1^* = C_{12}C_{22}^{-1}z_2$ et de matrice de covariance $C_{11} - C_{12}C_{22}^{-1}C_{12}^T$ (cf. annexe A). On peut donc appliquer l'algorithme Metropolis-Hastings multiple dans le cas gaussien ; celui-ci prend la forme suivante :

1. simuler $w = (w_1, w_2)$ selon ϕ (loi gaussienne non conditionnelle), indépendamment de l'état courant z_1 ,
2. construire :

$$z_1^i = z_1^* + (z_1 - z_1^*) \cos\left(i \frac{2\pi}{p+1}\right) + (w_1 - w_1^*) \sin\left(i \frac{2\pi}{p+1}\right), \quad i = 1, p \quad (4.82)$$

où $w_1^* = C_{12}C_{22}^{-1}w_2$,

3. calculer pour tout $i = 0, p$:

$$a(z_1, z_1^i) = \frac{r(z_1^i, z_2)}{\sum_{j=0}^p r(z_1^j, z_2)} \quad (4.83)$$

4. tirer $z_1^i \sim a(z, z_1^i)$ et poser $z_1' = z_1^i$

Cette version de l'algorithme Metropolis-Hastings multiple est particulièrement adaptée à la simulation d'une loi gaussienne conditionnelle où l'information disponible est de deux types : la donnée en certains points de la variable (vecteur z_2) et une information Y supplémentaire.

En appliquant le résultat précédent, il est immédiat de déduire la version Gibbs-Metropolis de notre algorithme dans le cas gaussien. Soit (T_1, \dots, T_p) une partition de S . Au balayage i , il faut générer un vecteur de loi conditionnelle :

$$\pi_{T_i}(z_{T_i} | z^{T_i}) = r(z_{T_i}, z^{T_i}) \phi(z_{T_i} | z^{T_i}) \quad (4.84)$$

qui est de la forme 4.81. On peut donc appliquer l'algorithme précédent, avec $z_1 = z_{T_i}$ et $z_2 = z^{T_i}$. Pour un choix judicieux des groupements par blocs, se reporter au paragraphe 4.2.1.

Il convient d'établir un rapprochement entre l'algorithme Metropolis-Hastings multiple gaussien que nous proposons, notamment dans sa version Gibbs, et l'algorithme de sur-relaxation permettant d'optimiser l'échantillonneur de Gibbs dans le cas gaussien. Cet algorithme a été introduit par Barone et Frigessi [2], et étudié par Green [51] dans une optique d'intégrale de Monte Carlo. Il est également énoncé dans [41]. Supposons qu'une itération de l'échantillonneur de Gibbs implique la relaxation de z_i selon la loi gaussienne $\mathcal{N}(m_i, \sigma_i^2)$. Cependant, il est immédiat de montrer que si l'on relaxe z_i selon la loi $\mathcal{N}((1 + \omega)m_i - \omega z_i, (1 - \omega^2)\sigma_i^2)$ avec $\omega \in [-1, 1]$, la transition est toujours acceptée : c'est le principe de la sur-relaxation de paramètre ω . On peut alors choisir le paramètre ω de manière à augmenter la vitesse de convergence du noyau (par minimisation du rayon spectral), ou à diminuer la variance d'estimation dans l'intégrale de Monte Carlo. Il est notamment intéressant de choisir $\omega > 0$ afin d'obtenir des états successifs anticorrélés. Cet algorithme est étendu naturellement aux cas non gaussiens, dans une approche Gibbs-Metropolis où la loi gaussienne est utilisée comme noyau de proposition [51, 4]. Un résultat intéressant et immédiat est que la probabilité d'acceptation ne dépend pas de ω .

Dans cet algorithme, le paramètre ω joue exactement le même rôle que le coefficient $-\cos \theta$ de l'équation 4.75. La différence essentielle avec l'algorithme Metropolis-Hastings multiple que nous avons proposé est que dans l'approche de Barone le paramètre ω est fixé : une seule transition est donc proposée à chaque itération. En réalité, l'algorithme de sur-relaxation correspond au cas particulier $p = 1$ de notre algorithme.

4.3 Illustration

Nous proposons de comparer les différents algorithmes de simulation exposés au paragraphe précédent. Pour cela, nous utilisons un cas d'étude inspiré de la vision stéréoscopique : la mise en correspondance d'une ligne d'un couple. Considérons deux signaux I_1 et I_2 , dont on suppose que l'un peut se déduire de l'autre par une translation aléatoire τ et un bruit d'acquisition η . Le système est décrit par la relation :

$$I_1(x) = I_2(x + \tau(x)) + \eta(x) \quad (4.85)$$

Les variables I_1 , I_2 , τ et η sont discrétisées selon une maille de longueur n . Dans cette application, τ est un vecteur gaussien de covariance cubique, notée C_p , et η est un bruit blanc gaussien de covariance $\sigma_l^2 \mathbf{I}_n$:

$$\tau \sim \mathcal{N}(0, C_p) \quad (4.86)$$

$$\eta \sim \mathcal{N}(0, \sigma_l^2 \mathbf{I}_n) \quad (4.87)$$

avec $C_p(i, j) = \sigma_p^2 \text{cubic} \left(\frac{|i-j|}{b_p} \right)$ (b_p est la portée, σ_p^2 le pallier).

Le tableau 4.1 indique les valeurs choisies pour les différents paramètres.

n	σ_p	b_p	σ_l
64	1	8	0.1

TAB. 4.1 – Paramètres du modèle

Nous commençons par construire le système de la manière suivante. Tout d'abord, une translation τ est simulée selon le modèle 4.86, et nous procédons de même pour obtenir une réalisation du bruit η . Une fois τ et η fixés, on peut construire à partir d'un signal donné I_2 le signal I_1 . Dans cette application, I_2 est un signal périodique. On note Y le couple (I_1, I_2) . La figure 4.6 représente les deux signaux I_1 et I_2 .

Le problème que l'on se pose est de reconstituer τ à partir des signaux I_1 et I_2 seulement (problème équivalent à celui de la mise en correspondance en vision stéréoscopique). Notamment, on s'intéresse à une fonctionnelle particulière de τ , la longueur totale du déplacement, qui est approchée par :

$$h_\tau \approx \sum_{i=1}^{n-1} \sqrt{(\tau_{i+1} - \tau_i)^2 + 1} \quad (4.88)$$

h_τ est une variable scalaire. Nous allons nous intéresser à l'estimation de cette quantité par Monte Carlo. Il est aisé de représenter l'évolution de cette estimation en fonction du nombre d'itérations de l'algorithme utilisé, et d'étudier les propriétés de convergence des algorithmes de Monte Carlo pour cette fonction particulière de τ .

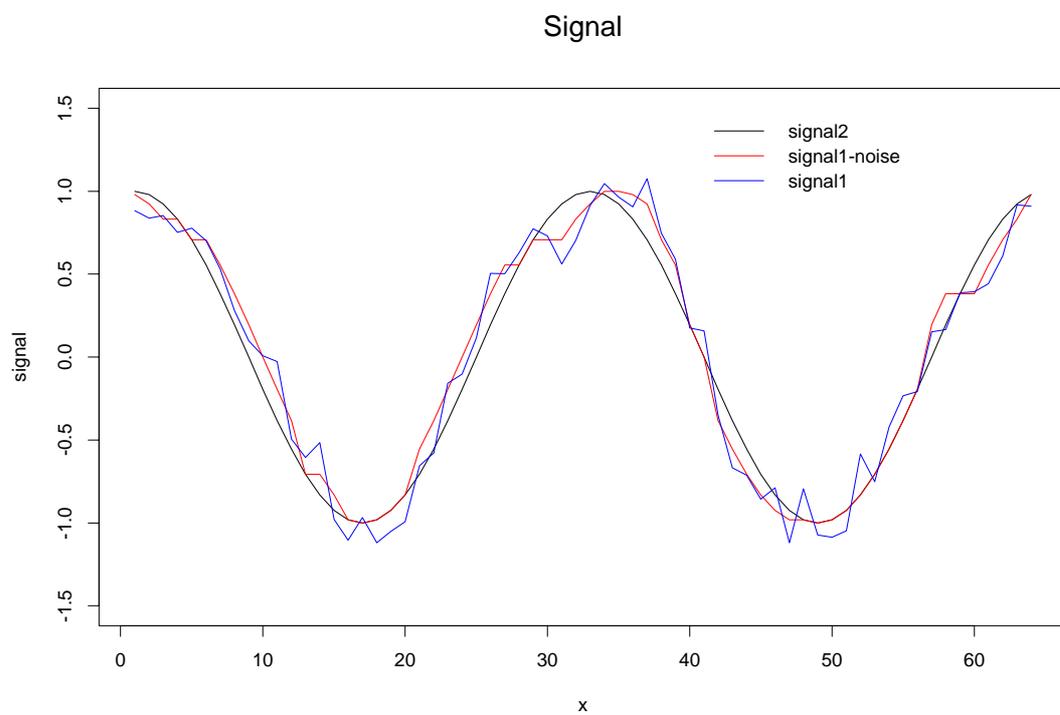


FIG. 4.6 – Signal original I_2 (courbe noire), signal déformé par une translation aléatoire (courbe rouge) et signal bruité I_1 (courbe bleue)

Nous adoptons une approche bayésienne. Dans cet exemple, le modèle stochastique est parfaitement connu et les valeurs des paramètres du tableau 4.1 sont utilisées, la seule inconnue est la variable τ . Cette situation est bien sûr une simplification par rapport à un cas d'étude réel, dans lequel on fait le choix d'un modèle stochastique dont les paramètres doivent être estimés.

A translation fixée, la différence des deux signaux suit une loi gaussienne (de matrice de covariance réduite à la diagonale). La loi *a priori* (ou non-conditionnelle) de la translation τ est une loi gaussienne de covariance C_p . La loi *a posteriori* π_τ de τ connaissant $Y = y$ prend l'expressions suivante :

$$\pi_\tau(\tau|y) \propto \exp\left(-\sum_{i=1}^n \frac{(I_1(x_i) - I_2(x_i + \tau_i))^2}{2\sigma_l^2}\right) \exp\left(-\frac{1}{2}\tau^\top C_p^{-1}\tau\right) \quad (4.89)$$

Dans la suite nous comparons les résultats obtenus avec les algorithmes suivants :

- un échantillonneur indépendant (**IndSampler**), qui utilise comme loi de transition la loi *a priori* 4.86,
- une marche aléatoire (**RandWalk**) de loi de transition gaussienne,
- une diffusion de Langevin (**Langevin**),
- un échantillonneur multiple (**MMH**) gaussien, où la loi gaussienne ϕ utilisée est la loi *a priori* 4.86.

Notons que la variance utilisée pour les transitions de la marche aléatoire et la diffusion de Langevin sont identiques, et que l'échantillonneur multiple est de dimension $p = 24$. Ces valeurs ont été choisies de telle sorte à obtenir des taux d'acceptation de l'ordre de 30%, mais ceci n'est qu'un critère empirique.

Le nombre total d'itérations pour chaque algorithme a été fixé à $25 \cdot 10^6$ itérations, et on a retenu une simulation toutes les 100 itérations, soit un ensemble de $l = 25 \cdot 10^4$ échantillons dans chaque cas.

Nous avons tout d'abord calculé la moyenne des simulations, le maximum *a posteriori* (MAP), ainsi que l'écart type. La moyenne des simulations est une estimation de l'espérance conditionnelle $E(\tau|Y = y)$, qui est l'estimateur optimal au sens de la minimisation de la variance. Le MAP est l'état simulé qui maximise la probabilité *a posteriori* π_τ . Moyenne et MAP pour chaque algorithme sont représentés à la figure 4.7, l'écart type à la figure 4.8. On constate que **RandWalk**, **Langevin** et **MMH** donnent la même estimation pour la moyenne et l'écart type, ce qui suggère que ces 3 algorithmes ont bien convergé et que l'estimation fournie est bonne. En revanche, dans le cas de **IndSampler**, on remarque certaines différences, notamment dans le calcul de l'écart type. Nous allons voir par la suite que cet algorithme n'a pas convergé. Notons enfin que les MAP calculés dans les 4 cas sont assez différents. Ce n'est pas une surprise puisque chaque algorithme génère des états distincts. Un calcul rigoureux du MAP impliquerait en réalité une étape de recherche d'optimum, par recuit simulé par exemple, qui n'a pas été menée ici. D'un point de vue du problème de la reconstruction, il est intéressant de com-

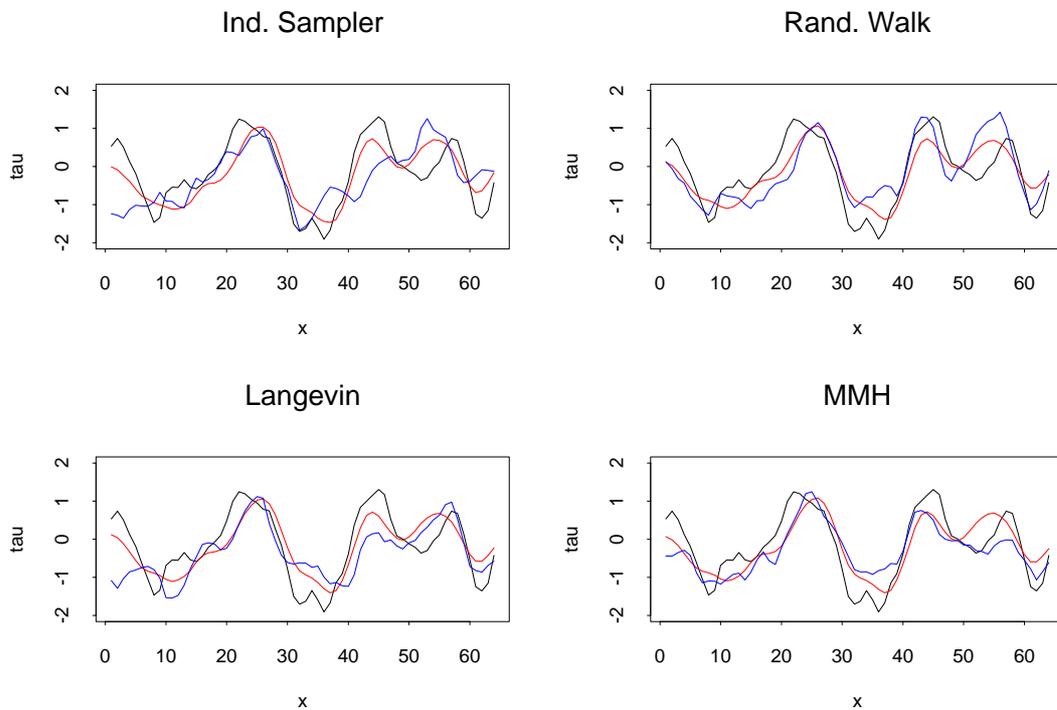


FIG. 4.7 – Translation réelle (courbe noire), espérance conditionnelle (courbe rouge) et MAP (courbe bleue) estimés par différents algorithmes

parer l'estimateur fourni par la moyenne (considérons par exemple l'estimation MMH) et la valeur réelle de τ , connue dans cet exemple. La première remarque est que l'approche bayésienne adoptée permet d'estimer convenablement τ , dont les variations spatiales sont notamment bien reproduites. Les différences qui subsistent sont bien entendu dues à la présence du bruit η . Plus intéressante encore est la donnée de l'écart type, car elle permet de juger de la qualité de l'estimation dans le cadre du modèle 4.89. On constate en particulier que la valeur de l'écart type varie fortement spatialement, ce qui montre bien l'influence locale à la fois du bruit et du signal. On retrouve notamment que les points de fort écart type correspondent bien aux zones où l'appariement entre les deux signaux est le plus ambiguë. Il s'agit des zones où le signal varie peu, au voisinage des minima et maxima locaux (cf. figure 4.8). Cet exemple montre donc bien à quel point le problème de la correspondance en stéréoscopie est sensible au contenu de l'information : plus encore que le bruit du signal, c'est l'ambiguïté de l'information qui est source d'incertitude dans l'estimation.

Afin de comparer les performances des différents algorithmes de simulation, nous nous intéressons maintenant plus particulièrement à l'estimation de h_τ . A la figure 4.9 est représentée l'évolution de $S_l(h_\tau)$, estimateur par Monte Carlo

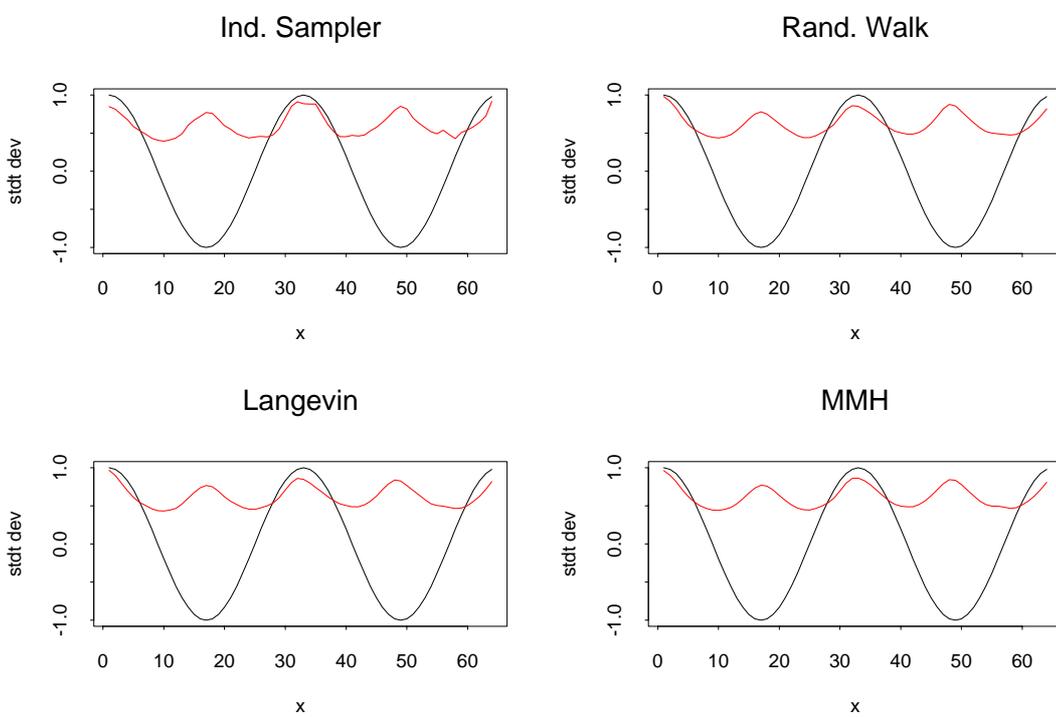


FIG. 4.8 – Ecart type (courbe rouge) estimé par chaque algorithme. En comparaison, le signal I_2 (courbe noire)

MCMC estimation

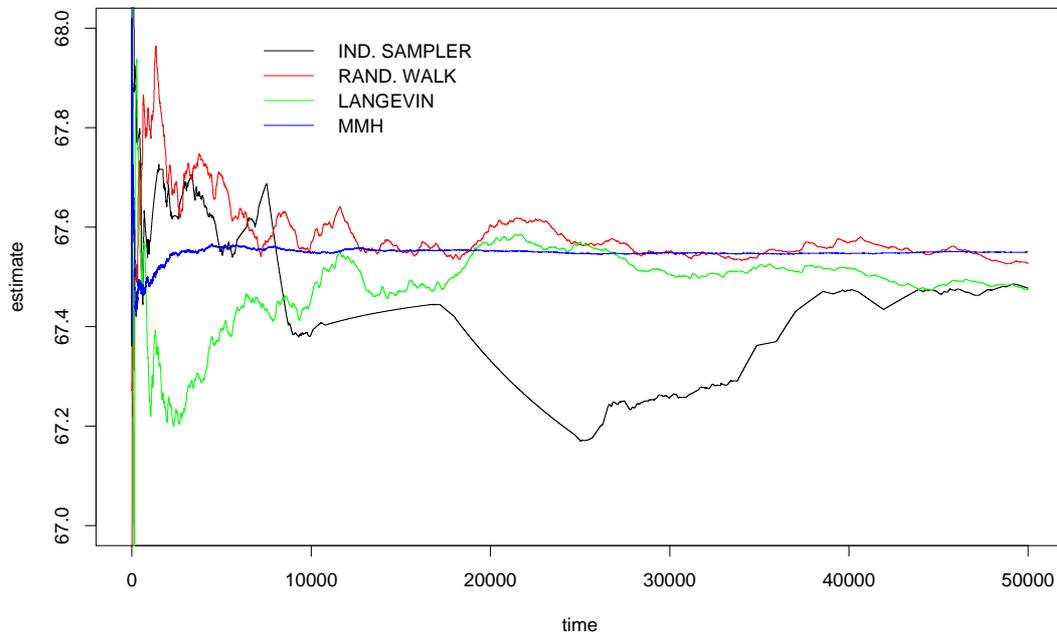


FIG. 4.9 – Evolution au cours du temps d’intégration de l’estimation de h_τ (les 50 000 premières itérations sont représentées)

de h_τ , en fonction du nombre l de simulations consécutives utilisées (on peut assimiler le paramètre l au temps, et considérer les états successifs issus de la chaîne de Markov comme une série temporelle). La situation est très contrastée. A l’issue des 50 000 premières simulations, les 4 algorithmes ont convergé vers la même valeur (et l’on accepte qu’il s’agit de la bonne valeur). Cependant, les vitesses de convergence sont très différentes. L’échantillonneur indépendant est le plus lent, la valeur limite étant pendant un grand nombre d’itération sous estimée. En fait, nous allons voir que la chaîne générée reste coincée à plusieurs reprises dans certains états, ce qui explique la lenteur de convergence de l’intégrale de Monte Carlo. **RandWalk** et **Langevin** ont des comportements semblables, avec des temps de coalescence (le moment où l’estimateur atteint la valeur limite et ne varie presque plus) d’environ 15 000 et 20 000 itérations respectivement. La convergence est de loin la plus rapide dans le cas de l’échantillonneur multiple (moins de 1 000 simulations sont nécessaires), l’estimateur qu’il donne est de plus le plus stable des 4 au cours du temps.

Afin de mieux comprendre ces différents comportements, on peut tracer les valeurs de h_τ obtenues au cours du temps (figure 4.10). La comparaison de ces 4 trajectoires montre à quel point le mélange à l’intérieur de la chaîne varie d’un

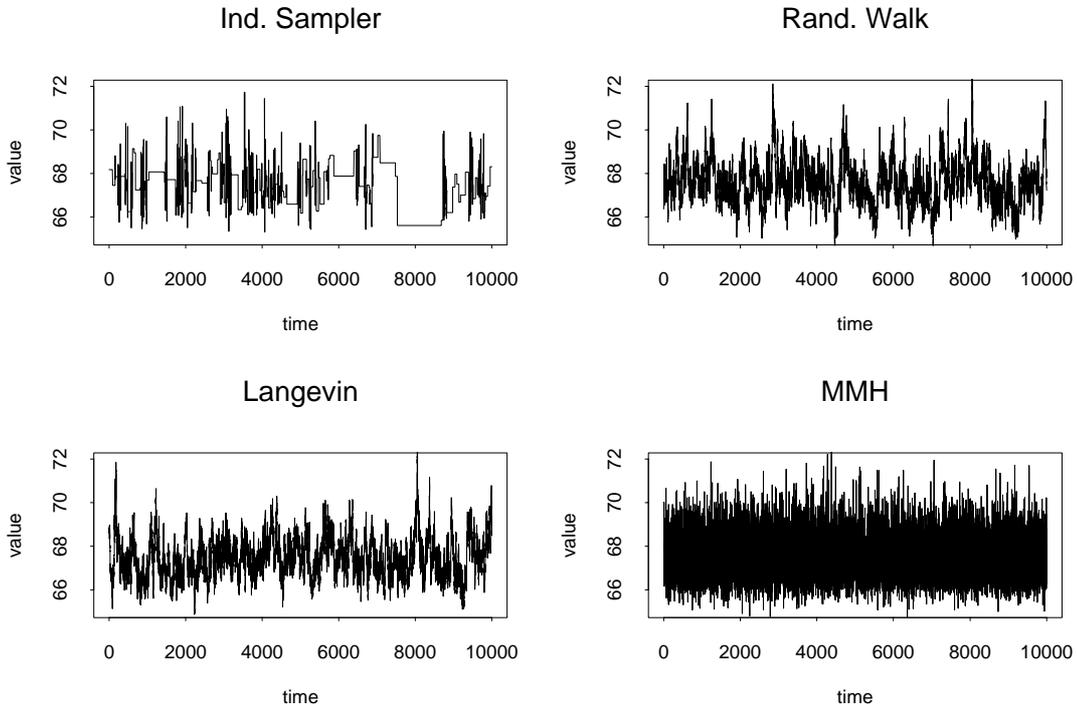


FIG. 4.10 – Trajectoire de h_τ pour les différents algorithmes

algorithme à l'autre. On constate tout d'abord que dans le cas `IndSampler`, la chaîne reste bloquée assez fréquemment. Le mélange est donc faible, et cela explique les problèmes de convergence mentionnés précédemment. Les séries obtenues avec `RandWalk` et `Langevin` sont assez proches, toutes les deux exhibent des corrélations temporelles assez fortes, ce qui montre que le mélange est relativement faible. En revanche, la série de l'algorithme `MMH` s'apparente plutôt à un bruit blanc (absence de corrélation). Le mélange y est très important.

A titre indicatif, les histogrammes de h_τ , qui sont une estimation de la loi *a posteriori* de h_τ , sont représentés à la figure 4.11. On constate notamment la similitude entre les 3 histogrammes obtenus avec `RandWak`, `Langevin` et `MHH`. En revanche, l'histogramme obtenu avec `IndSampler` est fortement perturbé par les problèmes de mélange à l'intérieur de la chaîne.

Les précédentes observations peuvent être précisées par le calcul de la fonction d'autocorrélation $A(h_\tau)$ des différentes séries (figure 4.12). Rappelons que la portée intégrale relative à une fonction g correspond à l'intégrale de $A(g)$. Logiquement, le graphe obtenu pour `IndSampler` met en évidence de fortes corrélations. Celles-ci sont dues aux périodes où la chaîne reste coincée. Dans le cas de `RandWalk` et `Langevin`, les fonctions d'autocorrélation empiriques sont similaires et présentent des portées assez importantes, ce qui explique la relative lenteur de la convergence

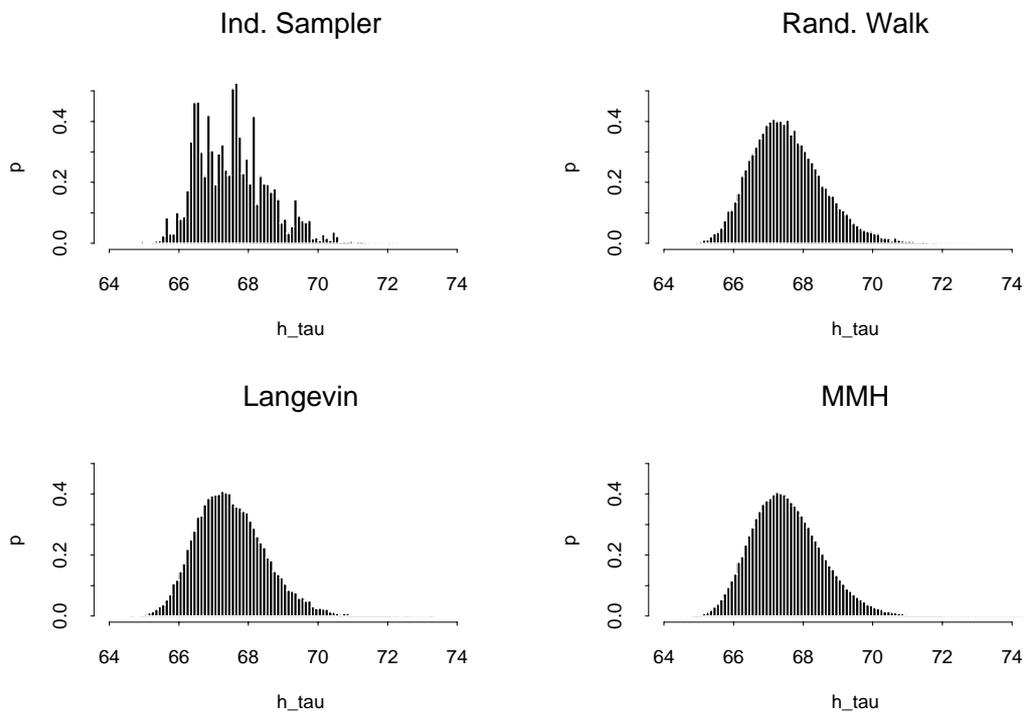


FIG. 4.11 – Histogramme de h_{τ}

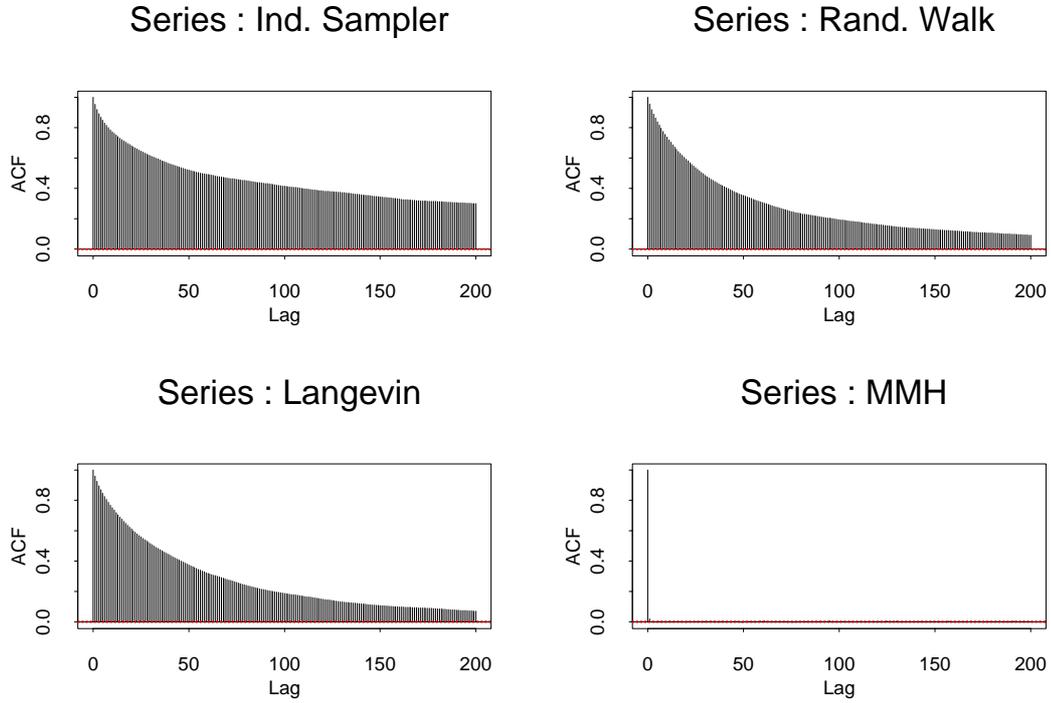


FIG. 4.12 – Fonctions d'autocorrélation des séries h_τ obtenues avec les différents algorithmes

de l'intégrale de Monte Carlo de la figure 4.9. En revanche, pour l'algorithme MMH, la portée intégrale est pratiquement égale à 1, on est donc proche de la situation d'échantillons indépendants. Dans ce cas, on pourrait certainement diminuer avantageusement le nombre d'itérations entre deux états retenus.

On peut calculer à l'aide de moyennes par paquets la variance d'estimation de h_τ . On a en effet montré que sous l'hypothèse d'ergodicité de la chaîne, et pour k assez grand, on doit avoir (cf. paragraphe 4.1.3) :

$$\frac{1}{s-1} \sum_{i=1}^s (S_k^i(h_\tau) - S_l(h_\tau))^2 \approx \frac{1}{k} A(h_\tau) \text{Var}_\pi(h_\tau)$$

soit en échelle log :

$$\log \left(\frac{1}{s-1} \sum_{i=1}^s (S_k^i(h_\tau) - S_l(h_\tau))^2 \right) \approx \log(A(h_\tau) \text{Var}_\pi(h_\tau)) - \log k$$

On peut donc tester cette hypothèse en traçant en fonction de $\log k$ la variance expérimentale $\log \left(\frac{1}{s-1} \sum_{i=1}^s (S_k^i(h_\tau) - S_l(h_\tau))^2 \right)$ et vérifier que la courbe admet bien une asymptote de pente -1 . La valeur en 0 de cette asymptote permet ainsi de calculer la valeur du produit $A(h_\tau) \text{Var}_\pi(h_\tau)$. Les courbes $\log - \log$ ainsi que

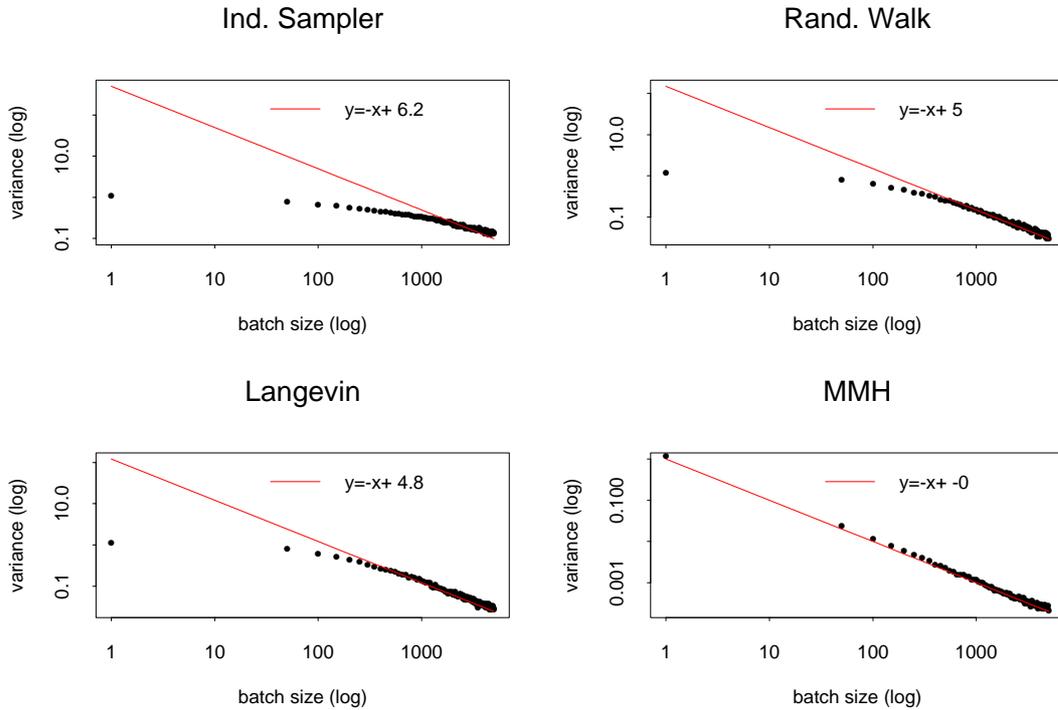


FIG. 4.13 – Estimation de la variance $A(h_\tau)\text{Var}_\pi(h_\tau)/k$ à l'aide de moyennes par paquets en fonction de la taille k des paquets (échelle $\log - \log$). En rouge, l'asymptote correspondante.

les asymptotes estimées sont représentées à la figure 4.13. Seul le graphe obtenu pour l'algorithme `IndSampler` n'exhibe pas le comportement attendu. L'estimation de l'asymptote dans ce cas n'est donnée qu'à titre de comparaison. Dans le cas de `RandWalk` et `Langevin`, c'est pour des valeurs de k supérieures à 1 000 que le comportement linéaire peut être mis évidence. L'algorithme `MMH` offre les meilleurs résultats, la courbe $\log - \log$ étant parfaitement linéaire.

Le tableau 4.2 résume les performances de chaque algorithme : taux d'acceptation, temps de calcul Δt entre 2 états simulés et valeurs estimées de $A(h_\tau)\text{Var}_\pi(h_\tau)$. Nous proposons de plus comme mesure de l'efficacité de l'algorithme le produit $A(h_\tau)\text{Var}_\pi(h_\tau) \times \Delta t$. En effet, le temps de calcul nécessaire pour réduire la variance d'estimation d'un facteur k est proportionnel à $kA(h_\tau)\text{Var}_\pi(h_\tau)\Delta t$. En choisissant ce critère et pour ce cas d'étude, on constate que l'algorithme Metropolis-Hastings multiple est plus de 60 fois plus performant que le plus performant des 3 autres, en l'occurrence `RandWalk`.

Ce cas d'étude permet de tirer certaines conclusions sur l'efficacité relative des algorithmes de simulation présentés au paragraphe 4.2. La première remarque importante est le succès des méthodes de Monte Carlo pour traiter le problème de

	IndSampler	RandWalk	Langevin	MMH
taux d'acceptation (%)	0.04	30.11	26.73	35
Δt (ms)	29.8	29.6	81.2	71.8
$A(h_\tau)\text{Var}_\pi(h_\tau)$	493	148	122	1
$A(h_\tau)\text{Var}_\pi(h_\tau) \times \Delta t$	14.69	4.38	9.90	0.07

TAB. 4.2 – Performances des algorithmes

la correspondance. L'exemple étudié est certes une simplification à une dimension du problème réel, mais il est instructif. En effet, généralement, il est difficile de valider les résultats obtenus par les méthodes de Monte Carlo puisque le calcul numérique ne peut être mis en œuvre. Ici, en comparant quatre algorithmes différents de simulation, dont trois ont donné des résultats semblables, nous avons pu nous assurer de la convergence des algorithmes. De plus, comme la solution du problème (la translation τ) était d'avance connue, il était possible de juger la qualité de l'estimation.

La seconde remarque concerne les propriétés de convergence des algorithmes proposés. On peut être surpris par le comportement de l'échantillonneur indépendant. En fait, plus la dimension du système augmente, plus la probabilité de proposer un état dans une région proche du mode de la loi visée diminue. En particulier, plus on se rapproche du mode, plus la probabilité d'accepter la transition diminue, ce qui explique qu'à plusieurs moments la chaîne reste coincée. Ce phénomène est évité dans le cas de la marche aléatoire, puisque la loi de transition varie en fonction de l'état courant, ce qui explique la plus grande robustesse de l'algorithme. On aurait également pu s'attendre à ce que l'algorithme construit sur la diffusion de Langevin soit plus performant que la marche aléatoire. Le résultat principal est l'excellent comportement de l'échantillonneur multiple. Le coût supplémentaire de calcul qu'il engendre est largement compensé par la rapidité de convergence dans les calculs d'intégrale par Monte Carlo (tableau 4.2). L'avantage principal du noyau de transition utilisé est qu'à chaque itération un continuum de transitions est proposée, continuum qui comprend l'état courant et un état indépendant. Un équilibre se met en place entre des transitions proches, qui permettent d'explorer localement l'espace, et des transitions éloignées, qui permettent de sauter d'éventuelles barrières locales. La dimension p de l'échantillonneur multiple est un paramètre important qui n'a cependant pas été étudié en détails. Les résultats empiriques montrent cependant que si on augmente p , on augmente le taux d'acceptation mais on favorise des transitions locales. Le coup de calcul supplémentaire ne se traduit donc pas nécessairement par un mélange accru. En pratique, on peut donc choisir une valeur qui permet d'obtenir un taux d'acceptation entre 30 et 70% (c'est l'ordre de grandeur conseillé dans [10]).

Chapitre 5

Cas d'étude

Nous proposons d'appliquer les modèles du chapitre 3 et l'algorithme de simulation MMH gaussien du paragraphe 4.2.3 à l'étude de plusieurs extraits du couple stéréoscopique SPOT de la région d'Aix-Marseille. Il s'agit notamment de déterminer les avantages respectifs des modèles GMF-0, GMF-1 et MIXT, et de valider la méthodologie proposée.

5.1 Présentation des outils statistiques

Dans l'ensemble des applications étudiées dans ce chapitre, nous avons utilisé les statistiques suivantes :

- moyenne et écart type ponctuels de la disparité,
- probabilité de dépassement de seuil : ponctuelle, sur un domaine, à l'échelle de la carte,
- minimum et maximum ponctuels de la disparité.

Moyenne et écart type sont des statistiques générales, que nous ne calculons ici qu'à titre d'illustration, et éventuellement pour tester la pertinence d'un modèle particulier. La carte moyenne représente l'espérance conditionnelle de la disparité connaissant le couple stéréoscopique. En ce sens, elle est l'estimateur qui minimise la variance conditionnelle. Bien que nous n'envisageons pas d'utiliser les méthodes de Monte Carlo pour calculer un estimateur de la carte de disparité, il est intéressant de comparer cet estimateur à la carte calculée par corrélation. Quant aux cartes d'écart type, elles sont avant tout une indication de la variabilité de la disparité dans le cadre du modèle choisi. Elles permettent de mettre en évidence les zones du couple stéréoscopique discriminantes du point de vue de l'information radiométrique, ou au contraire les zones plus ambiguës (cf. exemple du paragraphe 4.3).

Cependant, en dehors du cas gaussien, l'écart type ne donne pas d'information sur les queues de la distribution. Or, lorsque l'on s'intéresse à la fiabilité d'un système, ce sont justement les valeurs extrêmes, qui apparaissent avec des probabilités faibles, qui sont importantes. Dans ces applications, il convient alors de calculer la fonction de répartition, ou la fonction de répartition complémentaire, de la loi étudiée. Nous utilisons par la suite le terme de probabilité de dépassement de seuil.

Notons \hat{d} la carte de disparité calculée par corrélation et supposons que l'on dispose d'un échantillon $\{z_i, i = 1, l\}$ de la loi $\pi_D(\cdot|y)$. Nous définissons la probabilité ponctuelle de dépassement du seuil s au point $(u, v)^\top$ comme suit :

$$\text{pour } s \geq 0, P\left(D(u, v) - \hat{d}(u, v) \geq s\right) \approx \frac{1}{l} \sum_{i=1, l} 1_{z_i(u, v) - \hat{d}(u, v) \geq s} \quad (5.1)$$

$$\text{pour } s \leq 0, P\left(D(u, v) - \hat{d}(u, v) \leq s\right) \approx \frac{1}{l} \sum_{i=1, l} 1_{z_i(u, v) - \hat{d}(u, v) \leq s} \quad (5.2)$$

La probabilité ponctuelle de dépassement de seuil au point (u, v) dépend uniquement de la loi marginale de $\pi_D(\cdot|y)$ au point (u, v) . En ce sens, elle est une mesure locale de la précision de \hat{d} .

Dans d'autres applications, cette mesure locale n'est pas suffisante : considérons par exemple un objet en déplacement au-dessus de la zone observée, dont la trajectoire est définie à partir du modèle numérique de terrain. Dans ce cas, il faut pouvoir garantir que la probabilité de collision au cours de son déplacement est en deçà d'un certain risque fixé. Il est donc nécessaire de calculer la probabilité de dépassement de seuil relative à un domaine T du domaine S étudié. Cette probabilité peut être calculée à l'aide d'une statistique du maximum (ou du minimum) sur T de la disparité, soit :

$$\text{pour } s \geq 0, \quad P\left((D - \hat{d})_T \geq s\right) \approx \frac{1}{l} \sum_{i=1, l} 1_{\max_{(u, v) \in T} (z_i(u, v) - \hat{d}(u, v)) \geq s} \quad (5.3)$$

$$\text{pour } s \leq 0, \quad P\left((D - \hat{d})_T \leq s\right) \approx \frac{1}{l} \sum_{i=1, l} 1_{\min_{(u, v) \in T} (z_i(u, v) - \hat{d}(u, v)) \leq s} \quad (5.4)$$

Enfin, on peut également calculer une statistique relative à l'ensemble de la carte pour un seuil donné : la proportion de points pour lesquels la différence entre la disparité simulée et la disparité calculée par corrélation est supérieure (ou inférieure pour un seuil négatif) au seuil s fixé. Ceci permet de chiffrer la précision de la carte pour le seuil choisi.

On peut également s'intéresser à la situation inverse, dans laquelle un risque α est fixé, et il s'agit de calculer au point $(u, v)^\top$ les valeurs $d_{\text{inf}}(u, v)$ et $d_{\text{sup}}(u, v)$ telles que la probabilité pour que l'intervalle $[d_{\text{inf}}(u, v), d_{\text{sup}}(u, v)]$ contienne la valeur réalisée soit supérieure à $1 - \alpha$, i.e. :

$$P(D(u, v) \in [d_{\text{inf}}(u, v), d_{\text{sup}}(u, v)]) \geq 1 - \alpha \quad (5.5)$$

Un tel intervalle est un intervalle de confiance pour le risque α . Nous proposons de construire cet intervalle en utilisant le min et le max des simulations. En effet, soit $\{Z_1, \dots, Z_l\}$ l variables aléatoires indépendantes et de même loi que Z , alors on a les résultats suivants :

$$P(Z \leq \min_{i \leq l} Z_i) = \frac{1}{l+1} \quad (5.6)$$

$$P(Z \geq \max_{i \leq l} Z_i) = \frac{1}{l+1} \quad (5.7)$$

Soit donc un échantillon $\{z_1, \dots, z_l\}$ indépendant de loi $\pi_D(\cdot|y)$, et notons z_{inf} et z_{sup} les deux vecteurs tels que :

$$z_{\text{inf}}(u, v) = \min_{i \leq l} z_i(u, v) \quad (5.8)$$

$$z_{\text{sup}}(u, v) = \max_{i \leq l} z_i(u, v) \quad (5.9)$$

z_{inf} et z_{sup} permettent de construire un intervalle de confiance pour le risque $\frac{2}{l+1}$. Ce résultat, qui suppose l'indépendance entre les z_i , ne dépend que de la taille l de l'échantillon, et non de la loi $\pi_D(\cdot|y)$. En pratique, il suffit de choisir $l \geq \frac{2}{\alpha} - 1$. Dans la suite, on note cet intervalle de confiance **minmax**. Remarquons que la valeur du risque associé à l'intervalle **minmax** n'est qu'une approximation lorsque l'échantillon $\{z_i, i = 1, l\}$ n'est plus indépendant.

Si le calcul de la moyenne, de l'écart type, du min, du max ou des probabilités ponctuelles de dépassement de seuil ne requièrent que la connaissance de la loi marginale au point $(u, v)^T$, les probabilités de dépassement de seuil relatives à un domaine ou à la carte nécessitent la connaissance de la loi spatiale tout entière. Cet aspect motive pleinement l'emploi des méthodes de Monte Carlo, parfois jugées trop exigeantes en capacité de calcul, mais qui elles seules permettent d'estimer ces quantités. Notons de plus l'extrême souplesse de cette méthode : une fois l'échantillon généré, toute propriété du système peut être estimée.

Les statistiques de moyenne, d'écart type, de probabilité ponctuelle de dépassement de seuil, et du min et du max se présentent sous la forme de carte de même taille que la disparité. Les probabilités de dépassement de seuil sur un domaine peuvent également être représentées par des cartes superposables à la carte de disparité en affectant à tous les points d'un domaine T la même valeur. La statistique de dépassement de seuil relative à la carte entière est un scalaire, dont on peut aisément suivre les variations au cours de la simulation.

5.2 Extrait de zone urbaine

Nous considérons tout d'abord un extrait du couple stéréoscopique d'Aix-Marseille en zone urbaine (figure 5.1). La carte de disparité correspondante, calculée à l'aide de l'algorithme de corrélation d'Istar, est représentée à la figure 5.2.

Pour cet exemple, nous disposons également d'un couple stéréoscopique haute-résolution, issu de l'imagerie aérienne. A partir de ce couple, il est possible de calculer une carte de disparité, qui a été projetée dans le système de coordonnées du couple SPOT et dégradée à la résolution de ce dernier (figure 5.2). Nous proposons d'utiliser cette carte comme référence afin de valider nos résultats. La difficulté liée à la validation des résultats est double. Tout d'abord, nous devons admettre que la carte de disparité calculée à partir du couple haute-résolution ne comporte elle-même pas d'erreurs et que le processus de dégradation de la résolution est valide. Puis, comme nous ne disposons que d'une seule carte de référence à partir de laquelle il s'agit de valider des probabilités, nous sommes dans l'obligation de remplacer des statistiques d'ensemble par des statistiques spatiales, procédé qui rigoureusement nécessite la stationnarité spatiale de la grandeur considérée et l'indépendance spatiale. Il est évident que ces deux conditions ne sont pas vérifiées. Ces deux remarques limitent donc fortement la portée de la procédure de validation que nous présentons dans la suite; cette dernière n'en reste pas moins essentielle.

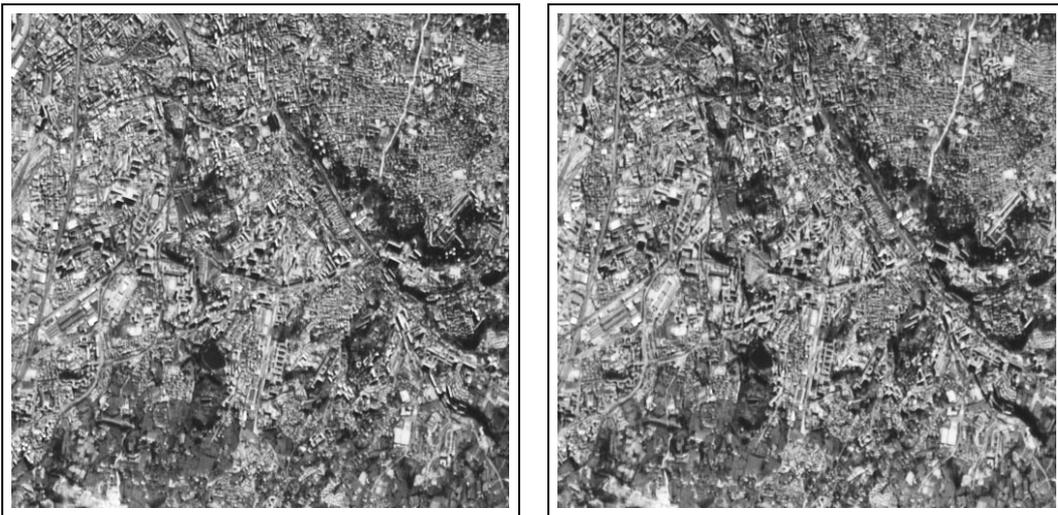


FIG. 5.1 – Extrait du couple stéréoscopique d'Aix-Marseille.

La taille des images du couple est 512×512 (taille totale : $n = 2^{18}$). La carte de disparité obtenue par corrélation est calculée avec un échantillonnage de facteur 4, c'est-à-dire en ne retenant qu'un pixel sur deux dans chaque direction (taille de la carte de disparité : $256 \times 256 = 2^{16}$). Les simulations ont été calculées avec le même taux d'échantillonnage. L'algorithme MMH a été utilisé, avec les deux modifications suivantes : groupement par blocs selon des bandes horizontales (de largeur 8 pixels), simulation sur une sous-grille par la méthode d'échantillonnage d'importance (facteur de réduction 4 dans chaque direction, soit des grilles de taille $64 \times 64 = 2^{12}$). Les simulations ont été générées après une période de "chauf-

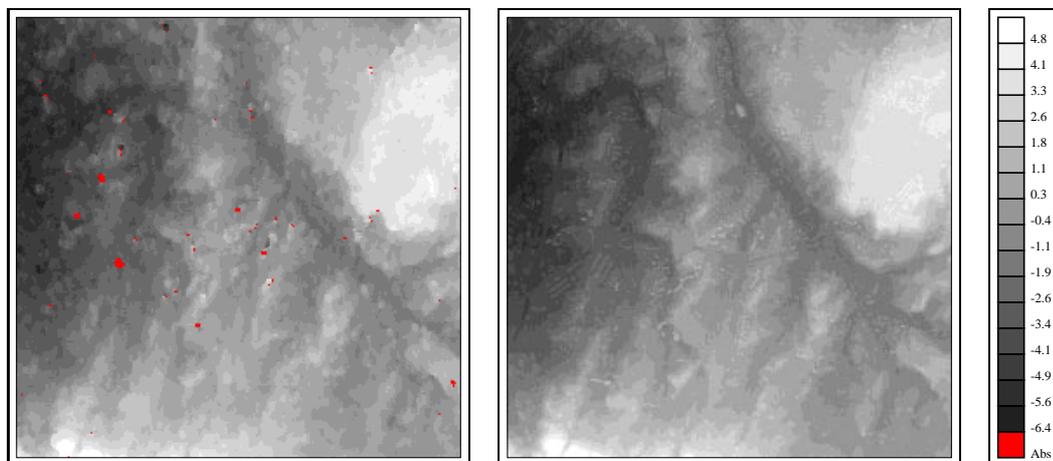


FIG. 5.2 – Figure de gauche : carte de disparité du couple de la figure 5.1 calculée par corrélation (en rouge : valeurs manquantes). Figure de droite : carte de disparité de référence obtenue à partir d’un couple stéréoscopique haute-résolution.

fe” de 10 000 itérations, en ne retenant qu’une simulation toutes les 100 itérations, afin d’approcher l’indépendance entre les états. Un total de $l = 2\,000$ simulations ont été ainsi calculées, ce qui représente $2\,10^5$ itérations.

Le modèle *a priori* est un modèle de fonctions aléatoires gaussiennes, de covariance sphérique (portée : 360 pixels, pallier 6). Celui-ci a été inféré à partir de la carte de disparité de la figure 5.2 (cf. paragraphe 3.4). Dans la suite, nous comparons les résultats obtenus avec les trois modèles de vraisemblance GMF-0, GMF-1 et MIXT présentés au chapitre 3.

Nous avons calculé les différentes statistiques introduites au paragraphe 5.1. Pour les probabilités de dépassement de seuil, nous avons utilisé :

- les seuils 1, 2, -1 et -2 pixels pour les probabilités ponctuelles,
- les seuils 2, 3, -2 et -3 pixels pour les probabilités sur des domaines de taille 64×64 pixels,
- le seuil 1, 2, -1 , et -2 pixels pour les statistiques de la carte entière.

Un seuil de 1 à 2 pixels correspond à des erreurs significatives de la disparité (de l’ordre de quelques dizaines de mètres en erreurs altimétriques), supérieures à la précision que l’on peut attendre d’un algorithme de corrélation subpixelique, mais qui ne peuvent pas être aisément détectées par des procédures automatiques (à la différence de valeurs aberrantes).

5.2.1 Estimation de la disparité

Moyenne et écart type pour chaque modèle sont représentés aux figures 5.3 à 5.5. Comme attendu, les 3 cartes de moyenne calculées ont une structure spa-

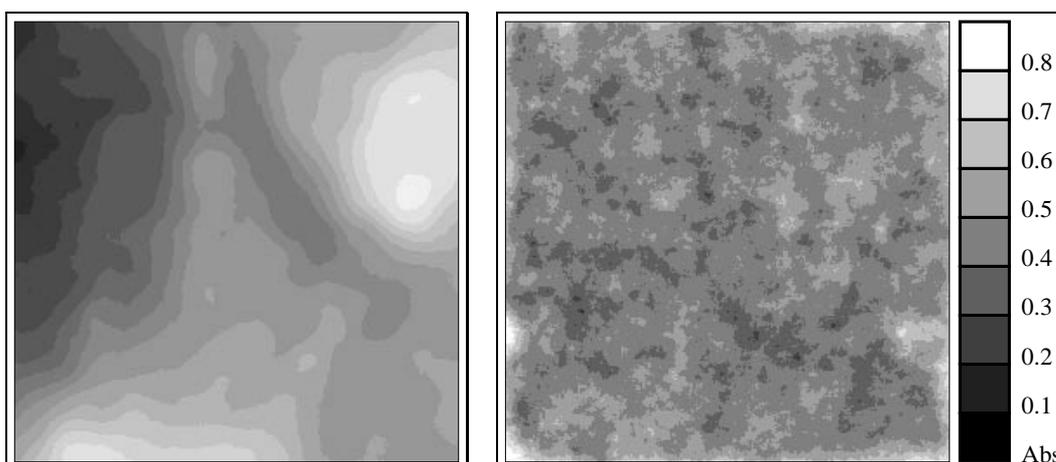


FIG. 5.3 – Moyenne et écart type obtenus pour le modèle GMF-0.

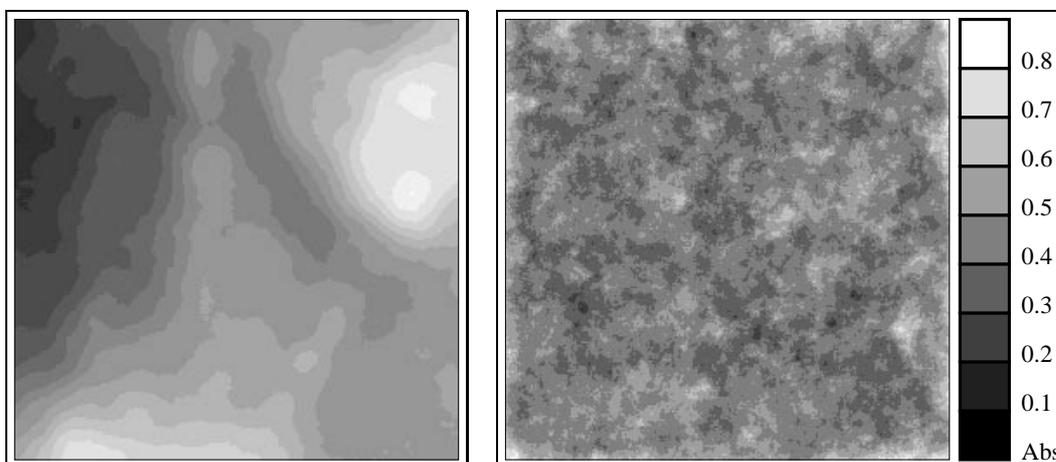


FIG. 5.4 – Moyenne et écart type obtenus pour le modèle GMF-1.

tiale plus “lisse” que la carte calculée par corrélation. Il est bien connu en effet que la carte moyenne a une variance plus faible que chaque réalisation. Mais ce phénomène est également lié au modèle utilisé, caractérisé par une portée relativement importante. En particulier, il est évident que les microstructures ne sont pas prises en compte dans ce modèle. Si on s’intéressait également à ces dernières, on pourrait les modéliser par le biais d’une structure de covariance gigue [24].

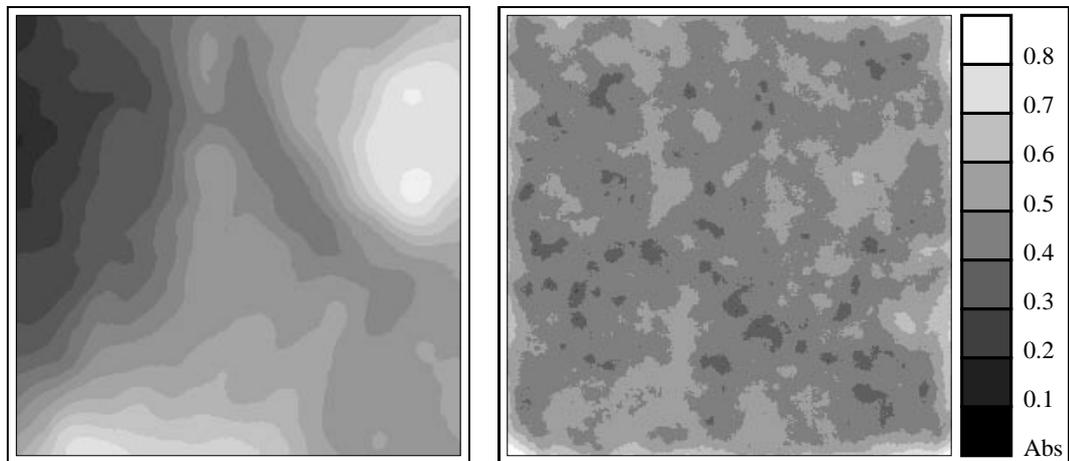


FIG. 5.5 – Moyenne et écart type obtenus pour le modèle MIXT.

Les cartes d'écart type exhibent une structure non stationnaire, mettant en évidence des zones de plus forte variabilité (écart type supérieur à 0.5) et des zones de plus faible variabilité (écart type de l'ordre de 0.2). On remarque également un effet de bord non négligeable aux extrémités gauche et droite. Cette plus forte variabilité s'explique par le fait que le point homologue n'est pas nécessairement dans l'image. De manière générale, on peut vérifier que les zones de forte variabilité correspondent à des régions du couple d'intensité relativement uniforme (zones d'ombre ou de l'eau par exemple), donc peu discriminantes, et qu'au contraire les zones de faible variabilité traduisent la présence d'un élément fortement discriminant (comme un immeuble blanc sur fond noir). Enfin, le modèle de vraisemblance utilisé a une influence non négligeable sur la structure spatiale de la carte d'écart type. La carte d'écart type obtenue avec le modèle MIXT apparaît notamment plus lisse que les cartes correspondant aux deux autres modèles. Nous revenons dans la suite sur ce point.

Nous avons également calculé le biais et l'écart quadratique moyen centré de la carte moyenne par rapport à la carte de référence (tableau 5.1). Compte tenu de la précision du calcul de la disparité, les biais, de l'ordre de 0.15 pixel, sont faibles. On peut d'ailleurs penser que ce biais est plutôt dû à la carte de référence, la précision de la projection dans le système de coordonnées du couple utilisé étant justement de cet ordre de grandeur. Les écarts quadratiques moyens sont compatibles avec les valeurs d'écart type précédentes.

A titre de validation, nous avons comparé l'estimation de la moyenne et la valeur référence aux points où l'algorithme de corrélation n'avait pas pu estimer la disparité (en rouge sur la carte de la figure 5.2). Il s'agit donc de points particulièrement délicats du point de vue de la mise en correspondance. Les courbes correspondantes sont tracées aux figures 5.6 à 5.8. De manière générale et quel

	GMF-0	GMF-1	mélange
biais	0.140	0.141	0.130
écart quadratique moyen	0.273	0.265	0.273

TAB. 5.1 – Biais et écart quadratique moyen entre la moyenne et la carte de disparité de référence pour les différents modèles

que soit le modèle utilisé, ces valeurs sont correctement estimées, seuls deux points étant en dehors de l'intervalle de confiance.

Estimation de la disparité aux points manquants

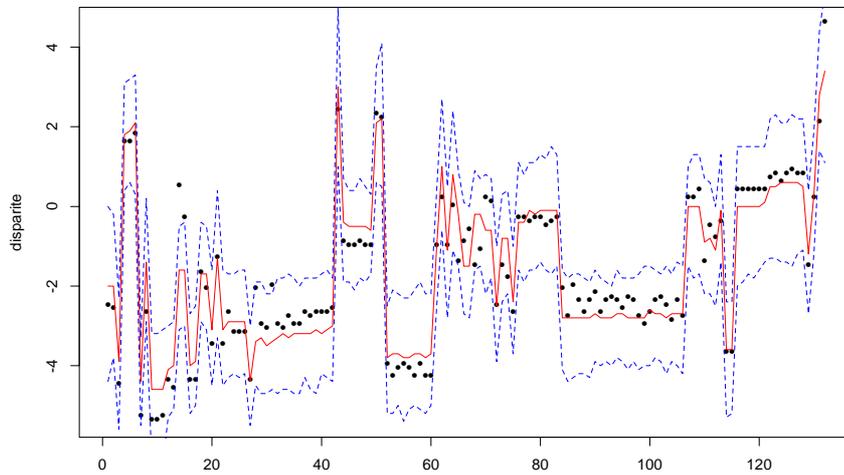


FIG. 5.6 – Estimation de la disparité aux points manquants : valeur de référence (points noirs), moyenne conditionnelle (courbe rouge) et intervalle de confiance minmax (pointillés bleus). Modèle GMF-0.

Estimation de la disparité aux points manquants

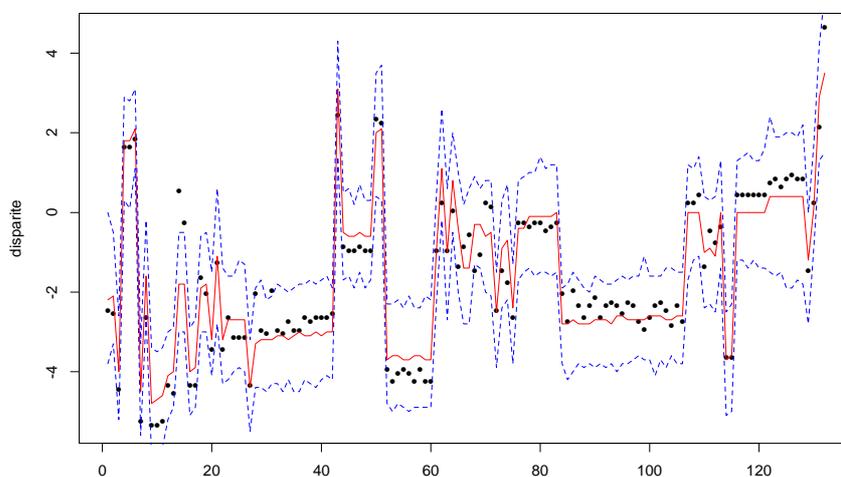


FIG. 5.7 – Estimation de la disparité aux points manquants : valeur de référence (points noirs), moyenne conditionnelle (courbe rouge) et intervalle de confiance minmax (pointillés bleus). Modèle GMF-1.

Estimation de la disparité aux points manquants

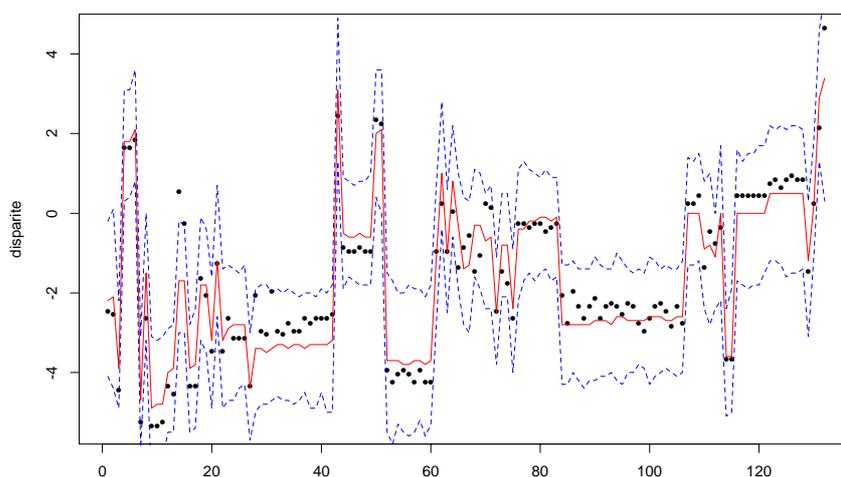


FIG. 5.8 – Estimation de la disparité aux points manquants : valeur de référence (points noirs), moyenne conditionnelle (courbe rouge) et intervalle de confiance minmax (pointillés bleus). Modèle MIXT.

5.2.2 Probabilité de dépassement de seuil

Probabilité ponctuelle

Intéressons-nous maintenant aux cartes de probabilité ponctuelle de dépassement de seuil. Les cartes correspondant aux seuils 2 et -2 pixels sont représentées aux figures 5.9 à 5.11.

Les trois cartes ont des structures très semblables : elles mettent clairement en

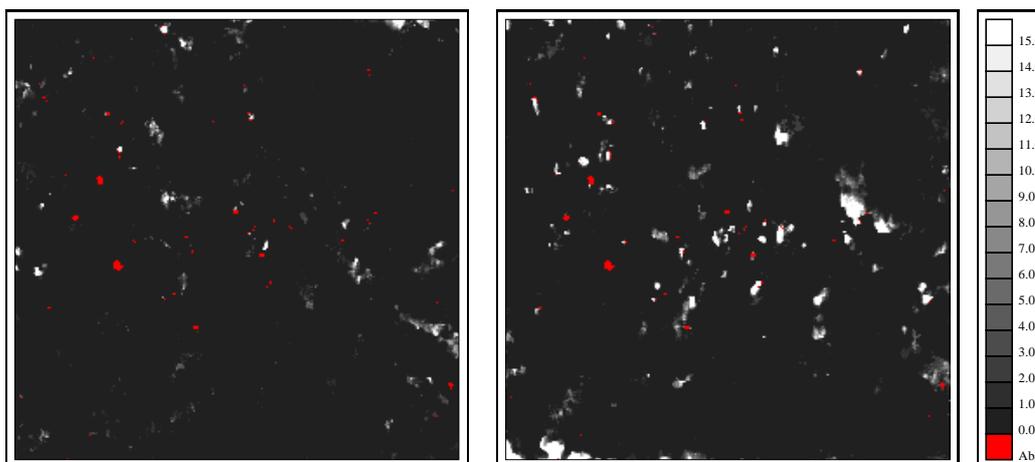


FIG. 5.9 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes). Modèle GMF-0.

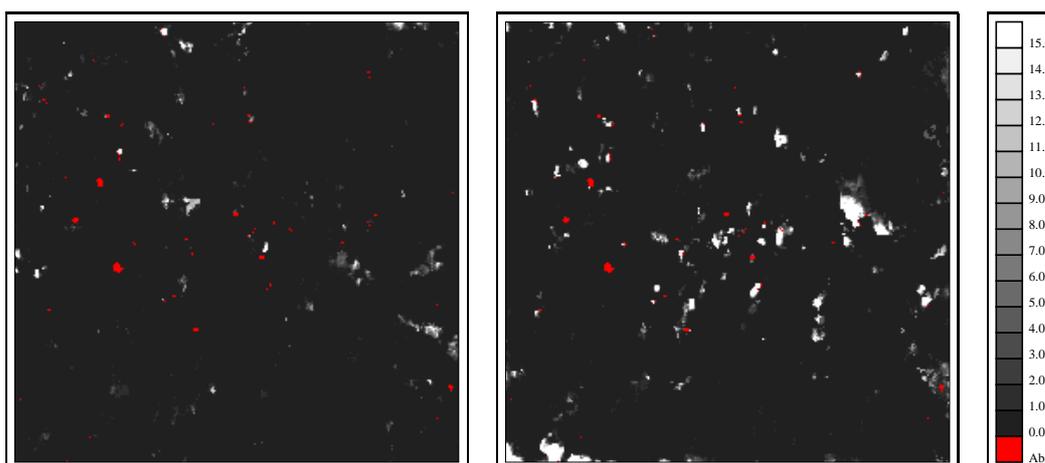


FIG. 5.10 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes). Modèle GMF-1.

évidence des groupes de points pour lesquels la probabilité d'une erreur supérieure

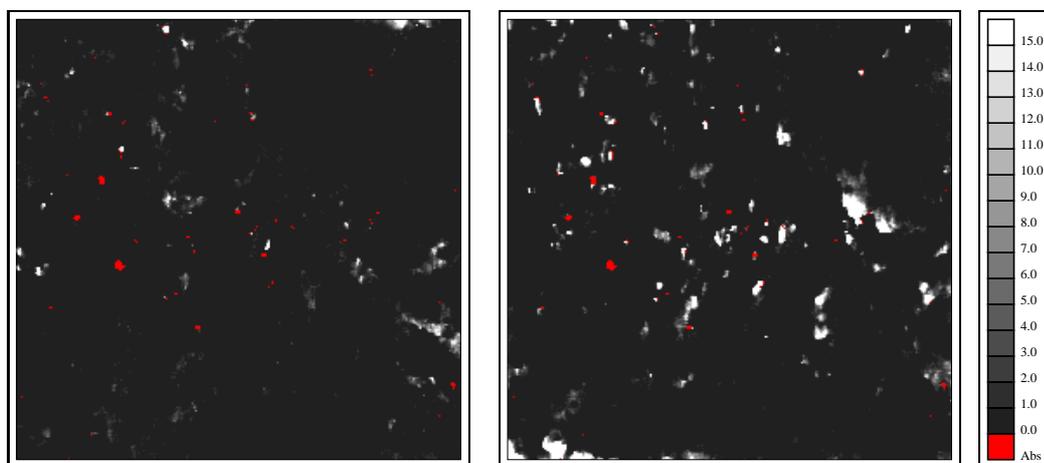


FIG. 5.11 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes). Modèle MIXT.

à 2 pixels est significative. L'information apportée par ces cartes de probabilité est plus fine que la simple donnée de l'écart type. Elles permettent en effet de localiser précisément les valeurs de la disparité susceptibles d'être erronées, et de chiffrer le risque d'erreur. A titre d'indication, environ 5% de l'image a une probabilité significative (par exemple supérieure à 1%) de dépassement de seuil, ce qui montre que l'information est bien localisée. Le seuil de 2 pixels choisi correspond à des erreurs importantes, bien au-delà de la précision que l'on peut attendre d'un algorithme de corrélation subpixelique, sans pour autant être des valeurs aberrantes. Ce type d'erreurs est donc particulièrement délicat à détecter par les tests habituels. Notons que la grande majorité des points dont la probabilité de dépassement de seuil est forte (supérieure à 90% par exemple) correspond bien à des erreurs supérieures à 2 pixels.

La figure 5.12, qui représente un détail de la carte de disparité étudiée, permet de juger de la pertinence de l'information apportée par ces cartes. Est représentée une zone en bordure de relief (au centre droit de la carte de la figure 5.2) mal rendue par l'algorithme de corrélation, la disparité étant surestimée. Cette zone apparaît en effet très sombre sur le couple stéréoscopique, ce qui peut expliquer l'erreur. La carte de probabilité de dépassement de seuil (ici le seuil -2 pixels) exhibe exactement sur cette zone des valeurs fortes, et permet donc de localiser précisément ces erreurs.

Afin de valider les résultats de chaque modèle, nous avons calculé des statistiques de dépassement de seuil à partir de la carte de disparité de référence. Nous avons procédé comme suit : à partir d'une carte de probabilité p relative au seuil s et d'un risque α , on construit l'ensemble $T_{p \geq \alpha}$ de points pour lesquels la probabilité p est supérieure à α , et l'ensemble complémentaire $T_{p < \alpha}$. A partir de la

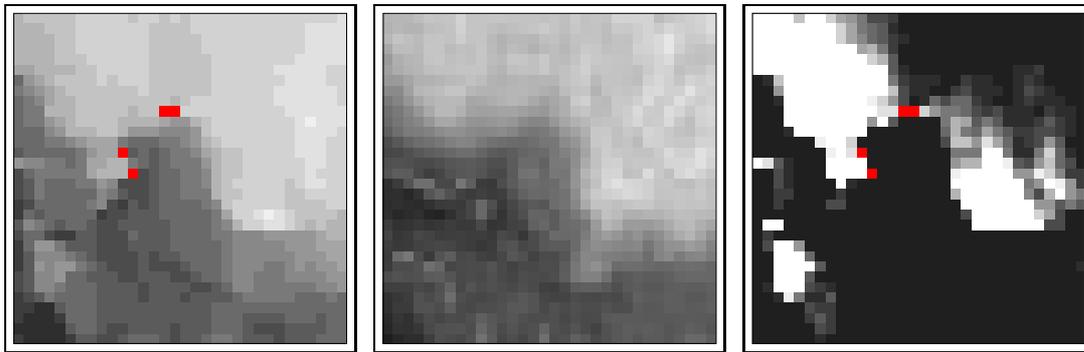


FIG. 5.12 – Extrait de la carte de disparité. De gauche à droite : disparité calculée par corrélation, disparité de référence, carte de probabilité de dépassement de seuil.

carte de disparité de référence, on peut alors calculer deux fonctions de répartition expérimentales : une relative à $T_{p \geq \alpha}$, une autre relative à $T_{p < \alpha}$. De manière intuitive, ces deux courbes doivent exhiber un comportement radicalement différent. Notamment, dans le cas de $T_{p \geq \alpha}$, la proportion de points pour laquelle le seuil s est atteint doit être supérieure à α , et inférieure à α dans le cas de $T_{p < \alpha}$. Il s'agit ici d'une comparaison purement heuristique, dont le but est simplement de déterminer si les cartes de probabilité de dépassement de seuil apportent une information pertinente.

Les fonctions de répartition expérimentales correspondant aux seuils 2 et -2 pixels et à un risque $\alpha = 5\%$ sont tracées aux figures 5.13 à 5.15. On vérifie bien que les fonction de répartition expérimentales des sélections $T_{p \geq 0.05}$ ont des queues de distribution significativement plus longues que celles des sélections $T_{p < 0.05}$. Dans les exemples des figures 5.13 à 5.15, le quantile correspondant à la valeur 5% est légèrement supérieur à 2 (respectivement inférieur à -2) pour la sélection $T_{p \geq 0.05}$ au seuil 2 (respectivement -2), alors qu'il est inférieur à 1 (respectivement supérieur à -1) pour la sélection $T_{p < 0.05}$.

Les tableaux 5.2 et 5.3 résument l'ensemble des résultats obtenus avec les seuils 1, -1 , 2 et -2 et les risques 10% et 5%. Dans l'ensemble, les proportions de points pour lesquels le seuil s est réellement dépassé sont supérieures au risque α correspondant, ce qui montre donc une bonne adéquation entre les prédictions des modèles et les statistiques expérimentales. On constate cependant une dissymétrie importante pour les seuils 1 et -1 , la proportion correspondant au seuil 1 étant anormalement basse. Cet effet est certainement à rapprocher du léger biais mis en évidence pour la moyenne conditionnelle. Les 3 modèles offrent des résultats très similaires.

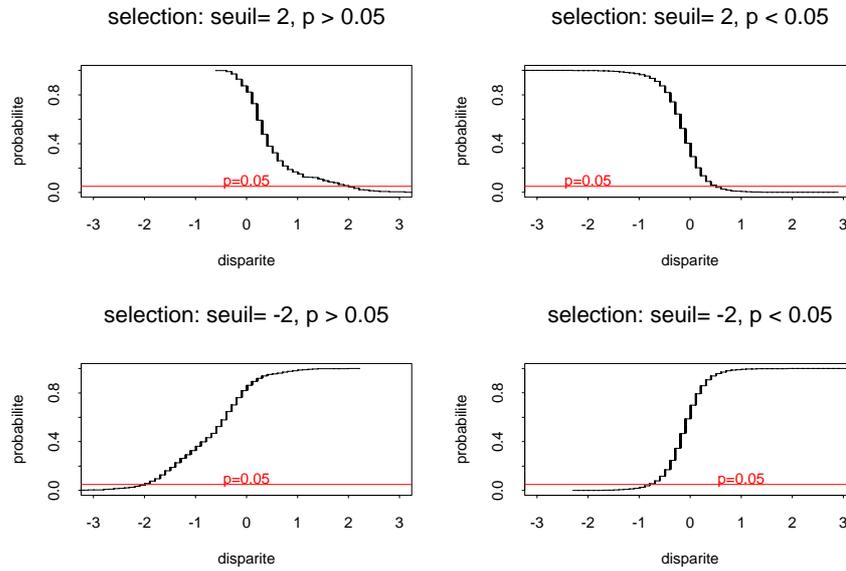


FIG. 5.13 – Fonctions de répartition expérimentales pour les seuils 2 et -2 pixels et le risque 5%. Modèle GMF-0.

	$s = 1$	$s = -1$	$s = 2$	$s = -2$
GMF-0	2.32 (14478)	13.02 (13689)	13.10 (229)	8.87 (981)
GMF-1	2.44 (13469)	13.72 (12917)	13.24 (204)	8.47 (1004)
mélange	2.35 (14818)	13.03 (14277)	13.89 (216)	7.84 (1110)

TAB. 5.2 – Statistiques de la sélection $T_{p \geq \alpha}$: proportion (en %) de points pour lesquels le seuil s est atteint, pour différents seuils s et un risque $\alpha = 10\%$. Entre parenthèse, la taille de l'échantillon $T_{p \geq \alpha}$.

Probabilité sur un domaine

Les cartes de probabilité de dépassement de seuil sur un domaine (figures 5.16 à 5.18) sont données ici à titre indicatif, le nombre de domaines utilisés (64) étant trop faible pour autoriser une validation statistique à partir de la carte de référence.

Il serait éventuellement intéressant de comparer ces cartes aux cartes de probabilité ponctuelle de dépassement de seuil. Il existe en effet une relation simple entre probabilités ponctuelles et probabilités sur un domaine de non dépassement

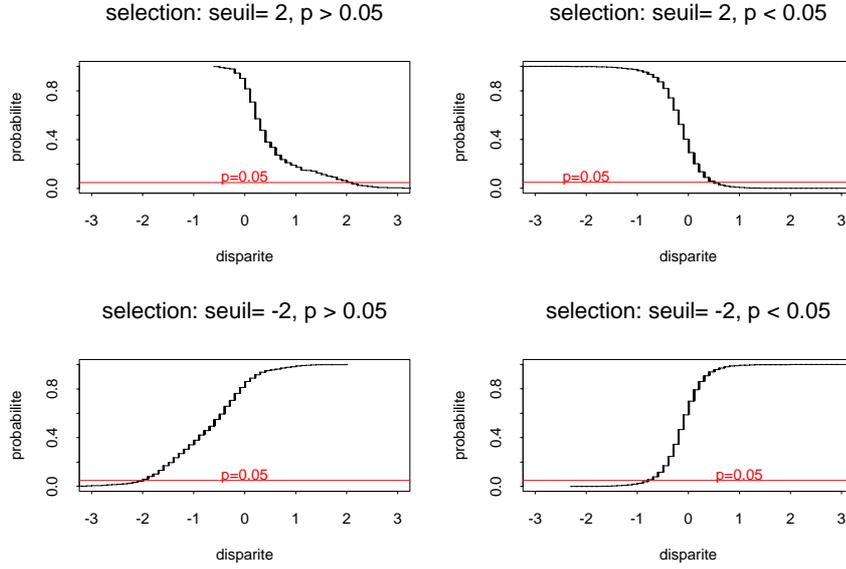


FIG. 5.14 – Fonctions de répartition expérimentales pour les seuils 2 et -2 pixels et le risque 5%. Modèle GMF-1.

	$s = 1$	$s = -1$	$s = 2$	$s = -2$
GMF-0	1.80 (22736)	9.73 (19910)	5.81 (585)	5.73 (1657)
GMF-1	1.87 (20745)	10.45 (18685)	6.68 (464)	5.99 (1537)
mélange	1.74 (23296)	9.53 (21120)	6.23 (546)	5.10 (1823)

TAB. 5.3 – Statistiques de la sélection $T_{p \geq \alpha}$: proportion (en %) de points pour lesquels le seuil s est atteint, pour différents seuils s et un risque $\alpha = 5\%$. Entre parenthèse, la taille de l'échantillon $T_{p \geq \alpha}$.

de seuil sous l'hypothèse d'indépendance spatiale :

$$s \geq 0, \quad P\left((D - \hat{d})_T < s\right) = \prod_{(u,v) \in T} P\left(D(u,v) - \hat{d}(u,v) < s\right) \quad (5.10)$$

$$s \leq 0, \quad P\left((D - \hat{d})_T > s\right) = \prod_{(u,v) \in T} P\left(D(u,v) - \hat{d}(u,v) > s\right) \quad (5.11)$$

Dans le cas de dépendance spatiale, les probabilités sont modifiées, et on peut notamment penser que si les corrélations spatiales sont positives, la probabilité de non dépassement de seuil sur T augmente. Cette intuition est rigoureusement vérifiée dans le cas particulier d'une fonction aléatoire gaussienne. En effet, un résultat dû à Slepian énonce que si $Z_1(t)$ et $Z_2(t)$ sont deux processus gaussiens centrés tels que $\forall u, v \in T, T$ compact,

$$\mathbb{E}(Z_1^2(u)) = \mathbb{E}(Z_2^2(u)), \quad \mathbb{E}(Z_1(u)Z_1(v)) \geq \mathbb{E}(Z_2(u)Z_2(v)), \quad (5.12)$$

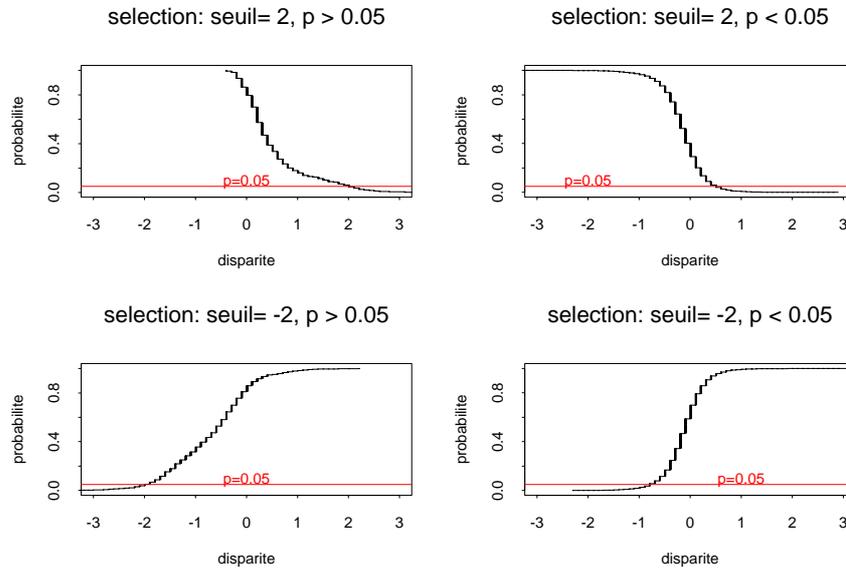


FIG. 5.15 – Fonctions de répartition expérimentales pour les seuils 2 et -2 pixels et le risque 5%. Modèle MIXT.

alors, pour tout seuil s , l'inégalité suivante est vérifiée :

$$P\left(\sup_T Z_1(u) \geq s\right) \leq P\left(\sup_T Z_2(u) \geq s\right) \quad (5.13)$$

Autrement dit, un processus qui est plus corrélé qu'un autre a une probabilité plus faible de dépasser un seuil fixé. Pour d'autres résultats sur la distribution du maximum d'une fonction aléatoire, on peut se rapporter à [28].

Nombre moyen de dépassements de seuil

Les statistiques du nombre moyen de dépassements de seuil par carte peuvent être directement, et en toute rigueur, comparées avec la valeur calculée à partir de la carte de référence. Le tableau 5.4 résume les résultats. Par rapport à la valeur expérimentale, le nombre de dépassements de seuil est systématiquement surestimé, d'un facteur qui varie fortement d'un seuil à l'autre. Cependant, l'ordre de grandeur que donnent les trois modèles est raisonnable, et il est certainement préférable de surestimer le nombre de cas défavorables que de le sous-estimer. Cette surestimation est légèrement plus forte dans le cas du modèle GMF-0 et du modèle MIXT.

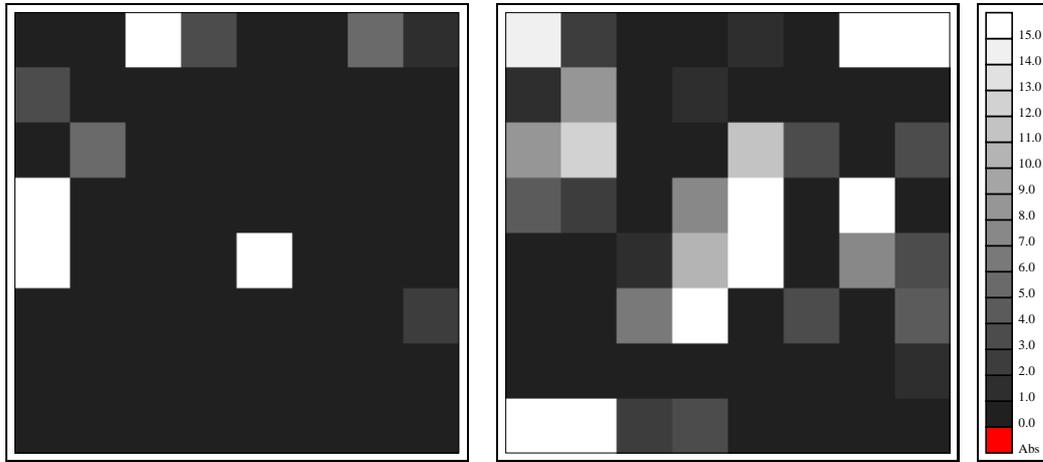


FIG. 5.16 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels. Modèle GMF-0

	$s = 1$	$s = -1$	$s = 2$	$s = -2$
exp.	0.98	3.43	0.10	0.16
GMF-0	7.12	8.18	0.22	0.77
GMF-1	6.61	7.80	0.18	0.74
mélange	7.20	8.65	0.24	0.84

TAB. 5.4 – Nombre de dépassements par carte du seuil s , exprimé en pourcentage du nombre de points total de la carte : valeur expérimentale obtenue avec la carte de référence et valeurs calculées par Monte Carlo pour différents modèles.

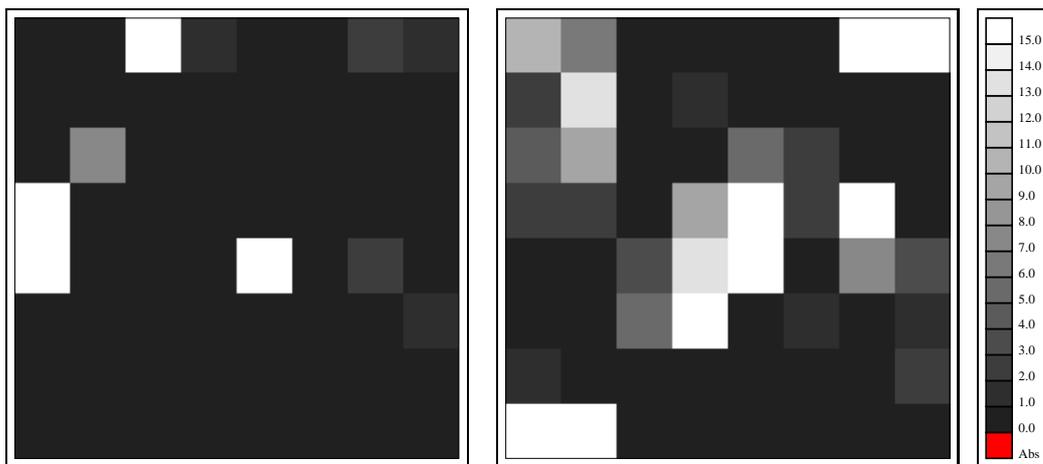


FIG. 5.17 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels. Modèle GMF-1

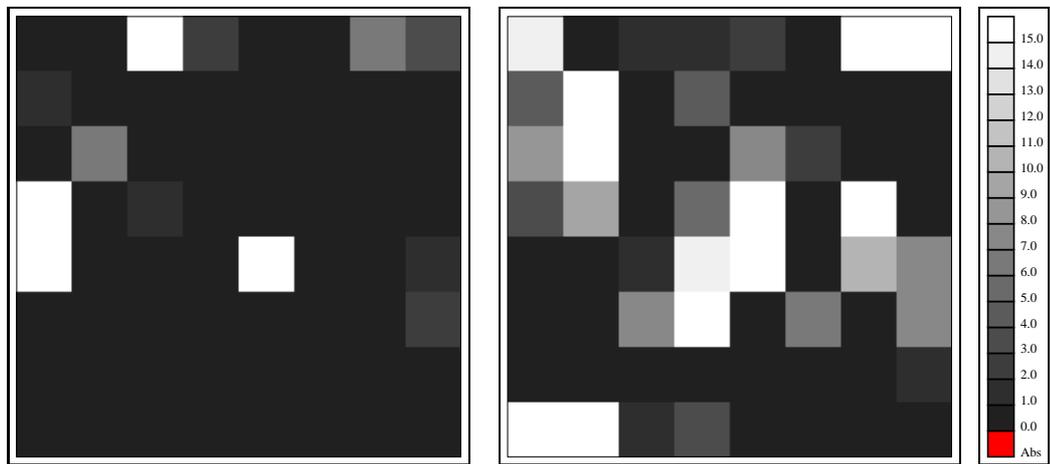


FIG. 5.18 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels. Modèle MIXT

5.2.3 Intervalle de confiance

L'intervalle de confiance minmax peut être vu comme une enveloppe de la disparité. Aux figures 5.19 à 5.21, nous avons représenté des coupes de cette enveloppe réalisées le long de certaines lignes de la carte de disparité. L'intérêt de cet

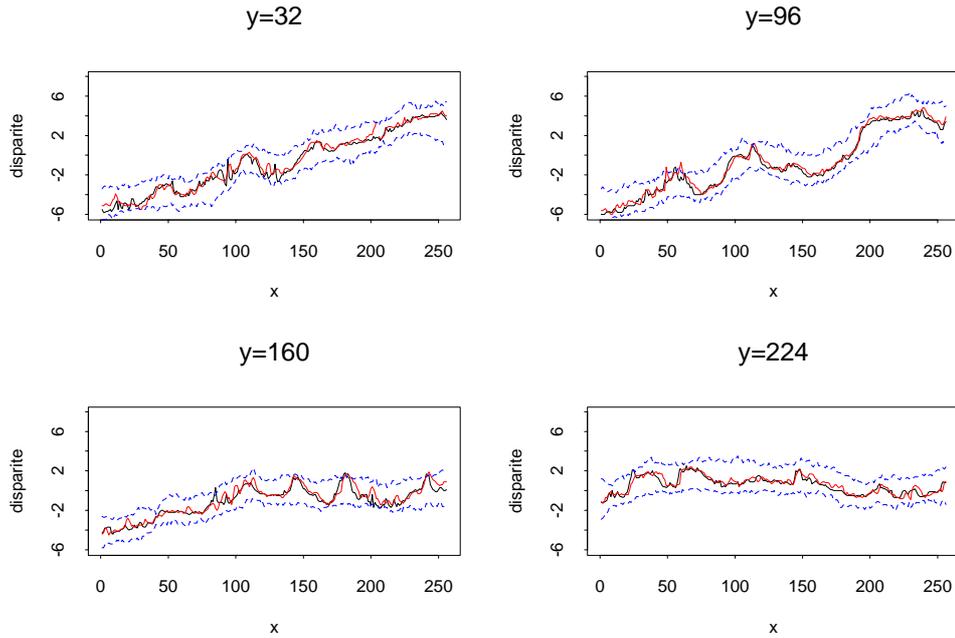


FIG. 5.19 – Intervalle de confiance : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) pour différentes lignes de la carte de disparité. Modèle GMF-0.

intervalle de confiance est double. Tout d'abord, il permet de donner des bornes inférieure et supérieure correspondant à un risque α choisi. De plus, l'intervalle de confiance peut être utilisé pour détecter des erreurs éventuelles de la carte de disparité, là encore avec une fiabilité égale à $1 - \alpha$.

Il est intéressant d'analyser la qualité de l'intervalle de confiance calculé aux points d'erreur forte. Pour cela, nous avons retenu les points pour lesquels la différence entre la carte de référence et la carte calculée par corrélation est supérieure à 2 pixels en valeur absolue. Le résultat est présenté aux figures 5.22 à 5.24. Ces courbes mettent clairement en évidence l'intérêt de l'intervalle de confiance : seuls en quelques points la valeur de référence est en dehors de l'intervalle calculé, et encore les dépassements sont faibles (inférieurs à 1 pixel).

Les remarques précédentes peuvent être complétées par des statistiques de la disparité de référence. Avec 2000 simulations, le risque associé aux intervalles $[d_{\text{inf}}, +\infty[$ et $] - \infty, d_{\text{sup}}]$ est $\frac{1}{2001} \approx 0.05\%$, et $\frac{2}{2001} \approx 0.1\%$ pour l'intervalle

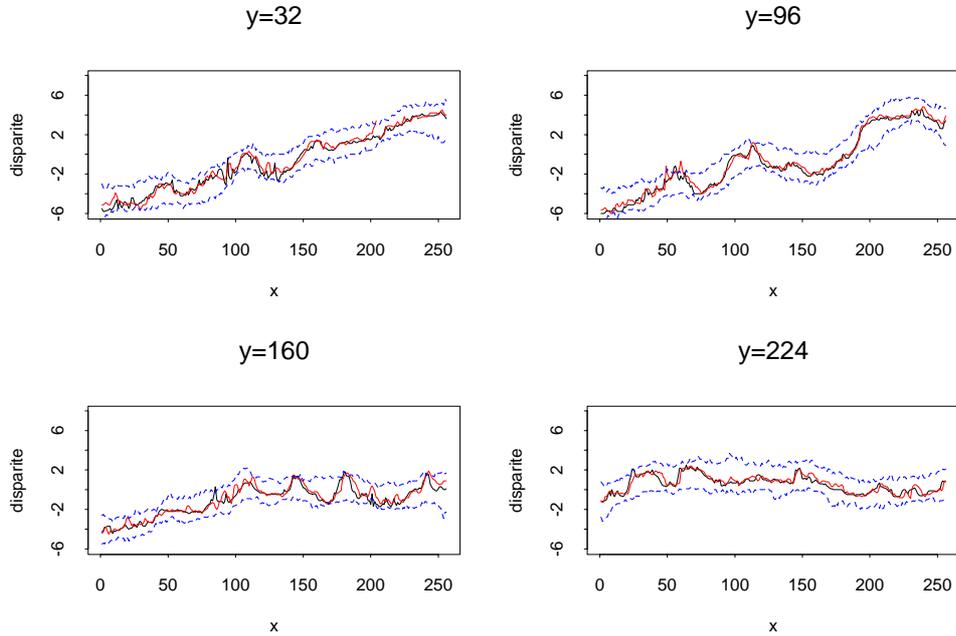


FIG. 5.20 – Intervalle de confiance : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) pour différentes lignes de la carte de disparité. Modèle GMF-1.

$[d_{\text{inf}}, d_{\text{sup}}]$. Il est donc possible de comparer ces probabilités avec la proportion de valeurs en dehors des intervalles considérés (tableau 5.5). Cette comparaison est

	$[d_{\text{inf}}, +\infty[$	$] - \infty, d_{\text{sup}}]$	$[d_{\text{inf}}, d_{\text{sup}}]$	$ d_{\text{sup}} - d_{\text{inf}} $
th.	0.05	0.05	0.10	-
GMF-0	0.23	0.69	0.92	2.92
GMF-1	0.38	0.83	1.22	2.73
mélange	0.08	0.57	0.66	3.12

TAB. 5.5 – Proportion de points en dehors de l'intervalle de confiance minmax pour les différents modèles, et valeurs théoriques associées (en %). La dernière colonne donne la largeur moyenne de l'intervalle de confiance pour chaque modèle.

valide, même dans le cas de valeurs non indépendantes. En effet, soit Z_1, \dots, Z_n un échantillon de n variables non nécessairement indépendantes, et des seuils s_1, \dots, s_n tels que $P(Z_i > s_i) = \alpha$. La statistique que nous considérons est la proportion de dépassement de seuils, soit $\frac{1}{n} \sum_{i=1}^n 1_{z_i > s_i}$. Cette statistique a pour espérance $\frac{1}{n} \sum_{i=1}^n P(Z_i > s_i) = \alpha$, et est donc une estimation non biaisée de α . Dans le cas indépendant, cette statistique a pour variance $\frac{1}{n} \alpha(1 - \alpha)$, mais elle est

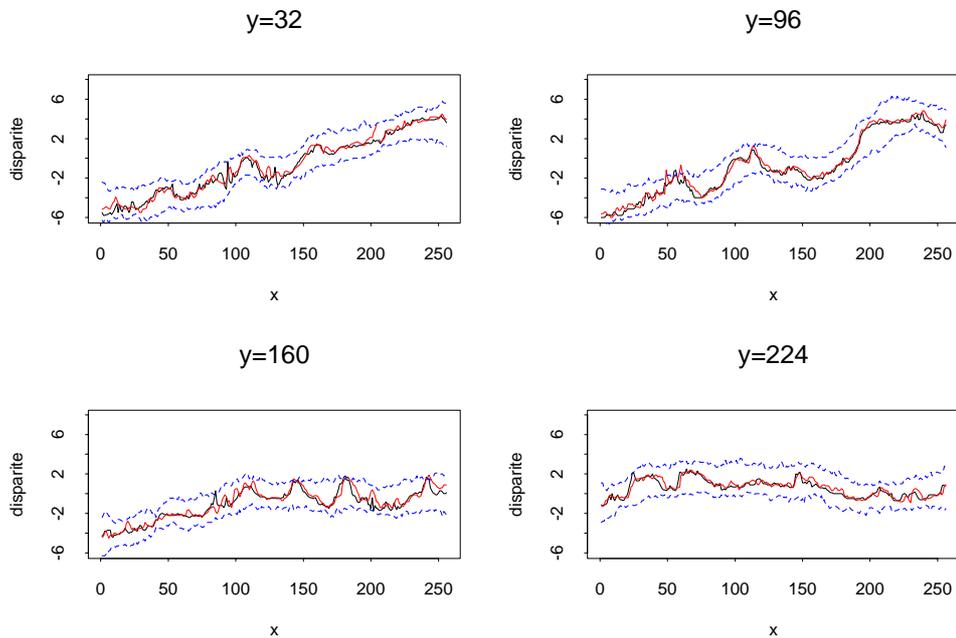


FIG. 5.21 – Intervalle de confiance : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) pour différentes lignes de la carte de disparité. Modèle MIXT.

bien plus importante dans le cas contraire.

Les résultats mettent en évidence une inadéquation entre le risque théorique associé à l'intervalle de confiance minmax et les valeurs expérimentales, plus élevées d'un facteur 6 à 12. L'intervalle de confiance le plus étroit (en moyenne) est obtenu pour le modèle GMF-1, mais le risque expérimental associé à cet intervalle est bien supérieur à la valeur théorique. La plus grande variabilité du modèle MIXT se traduit par des intervalles de confiance plus larges en moyenne, mais les différences totales restent assez faibles (0.4 pixel entre le modèle GMF-0 et le modèle MIXT). Comparativement aux deux autres modèles, le modèle MIXT offre les résultats expérimentaux les plus proches des statistiques annoncées, notamment en ce qui concerne la borne inférieure.

5.2.4 Bilan

Les résultats présentés au paragraphe précédent sont assez contrastés. Si la capacité des différents modèles à décrire l'incertitude associée à la carte de disparité est vérifiée, on constate en revanche une certaine inadéquation entre statistiques expérimentales et prévisions des modèles. Les statistiques de dépassement de seuil apparaissent comme une information pertinente permettant de repérer les points

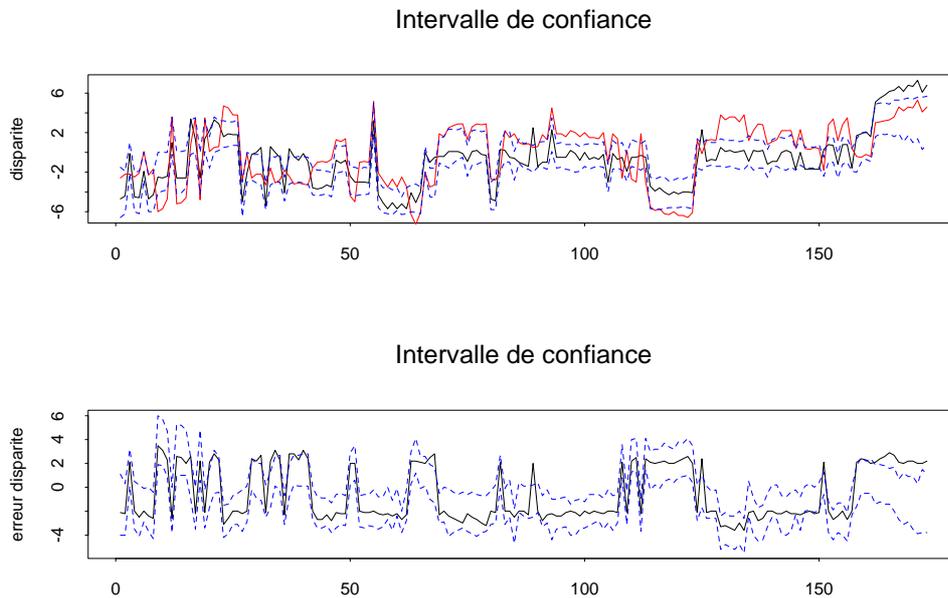


FIG. 5.22 – Haut : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) aux points de fortes erreurs (supérieures à 2 pixels). Bas : erreur de disparité (disparité de référence moins disparité calculée par corrélation) et intervalle de confiance associé (intervalle précédent moins disparité calculée par corrélation). Modèle GMF-0.

critiques avec une bonne estimation du risque associé. En revanche, les statistiques relatives à l'intervalle de confiance montrent que certaines erreurs fortes ne sont pas décrites par les modèles utilisés. Deux explications peuvent être avancées : le modèle *a priori* est mis en défaut localement, notamment en des points de discontinuité, fortement pénalisés par la forme de la covariance choisie (continue à l'origine), ou certaines ambiguïtés du couple ne sont pas prises en compte par le terme de vraisemblance. Cependant, compte tenu des hypothèses simplificatrices que nous avons émises, notamment des hypothèses gaussiennes et de stationnarité, et des remarques restrictives sur le processus de validation utilisé, l'écart entre statistiques expérimentales et prévisions des modèles reste acceptable.

Cette étude nous permet également de porter un jugement comparatif sur les différents modèles utilisés. La première conclusion est la robustesse des résultats relativement au choix du modèle de vraisemblance. La principale différence est une plus grande variabilité du modèle MIXT par rapport aux modèles GMF-0 et GMF-1, ce qui est conforme avec les hypothèses sous-tendant ces modèles. Mais moyenne, écart type et probabilités de dépassement de seuil sont relativement semblables. Si un choix doit donc s'opérer, c'est sur un critère de sécurité. Si l'on

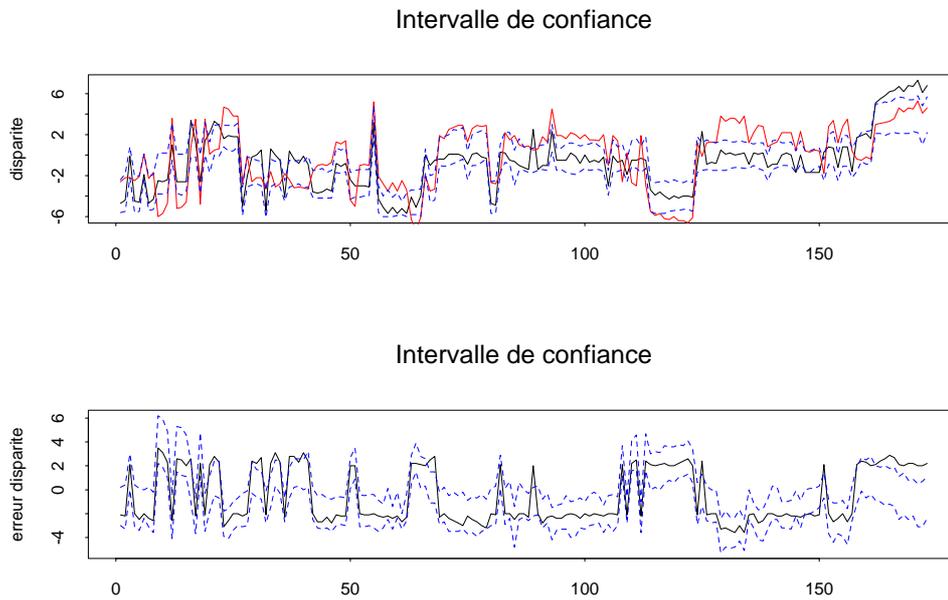


FIG. 5.23 – Haut : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) aux points de fortes erreurs (supérieures à 2 pixels). Bas : erreur de disparité (disparité de référence moins disparité calculée par corrélation) et intervalle de confiance associé (intervalle précédent moins disparité calculée par corrélation). Modèle GMF-1.

désire sélectionner avec une forte probabilité de réussite les points erronés, alors les modèles GMF-0 et GMF-1 semblent légèrement mieux adaptés. Si au contraire il s'agit de fournir des bornes fiables pour la disparité, alors le modèle MIXT apparaît le plus approprié, le modèle GMF-1 étant à rejeter. Notons également que d'un point de vue algorithmique et pour des choix identiques de la loi de transition, la convergence pour la simulation du modèle MIXT semble plus rapide, en termes de nombre d'itérations (corrélations entre états successifs plus faible dues notamment à un taux de transition plus élevé).

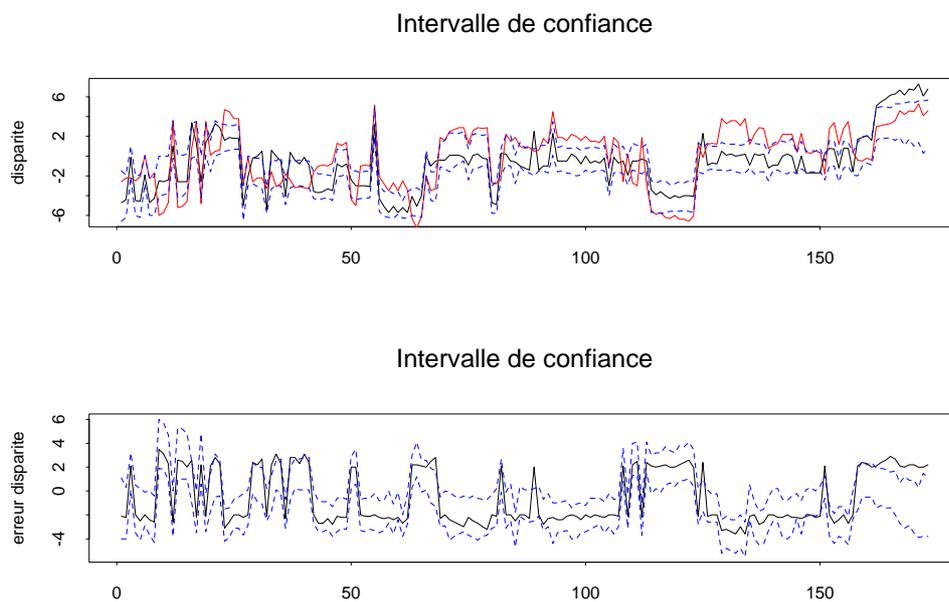


FIG. 5.24 – Haut : disparité de référence (courbe noire), disparité calculée par corrélation (courbe rouge) et intervalle de confiance (pointillés bleus) aux points de fortes erreurs (supérieures à 2 pixels). Bas : erreur de disparité (disparité de référence moins disparité calculée par corrélation) et intervalle de confiance associé (intervalle précédent moins disparité calculée par corrélation). Modèle MIXT.

5.3 Extrait de zone montagneuse

Nous considérons maintenant un extrait du couple d'Aix-Marseille comprenant une zone montagneuse (figure 5.25). Pour cet extrait, nous ne disposons pas de carte de référence. La taille de chaque image est 512×512 , mais carte de disparité et simulations ont été calculées sur une grille de taille 256×256 . La carte de disparité obtenue par corrélation (figure 5.26) met en évidence un relief important, avec des valeurs de la disparité supérieures à 30 pixels en valeur absolue. On constate également qu'en un nombre non négligeable de points, l'algorithme de corrélation n'a pu réaliser la mise en correspondance (critère de symétrie de l'optimum non vérifié). Ceci est dû aux déformations géométriques importantes induites par le relief, ainsi qu'à l'existence de zones d'ombre. La situation est donc très différente de l'exemple de zone urbaine du paragraphe 5.2.

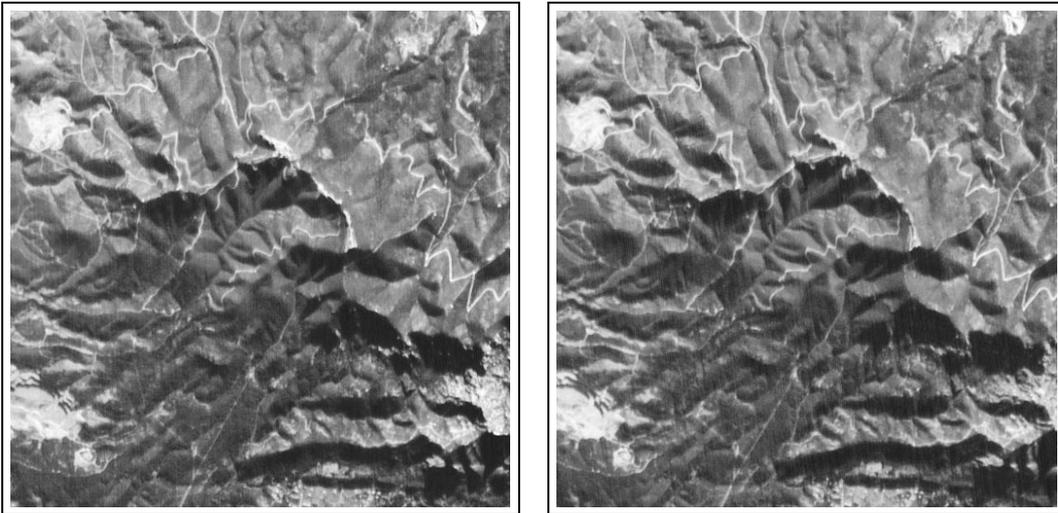


FIG. 5.25 – Extrait du couple stéréoscopique d'Aix-Marseille.

Comme précédemment, nous avons échantillonné la loi *a posteriori* à l'aide de l'algorithme MMH, avec les deux modifications suivantes : groupement par blocs selon des bandes horizontales (de largeur 8 pixels), simulation sur une sous-grille par la méthode d'échantillonnage d'importance (facteur de réduction 2 dans chaque direction, soit des grilles de taille $128 \times 128 = 2^{14}$). Les simulations ont été générées après une période de "chauffe" de 10 000 itérations, en ne retenant qu'une simulation toutes les 100 itérations. Un total de $l = 2\,000$ simulations ont été ainsi calculées, ce qui représente $2 \cdot 10^5$ itérations.

L'étude variographique de la carte de disparité 5.26 a permis d'inférer le modèle *a priori*. Le modèle de covariance utilisé est un modèle gigogne : la première structure, de portée égale à 240 pixels et de pallier égal à 20, est un modèle sphérique, tandis que la seconde structure, de portée égale à 420 pixels et de pallier

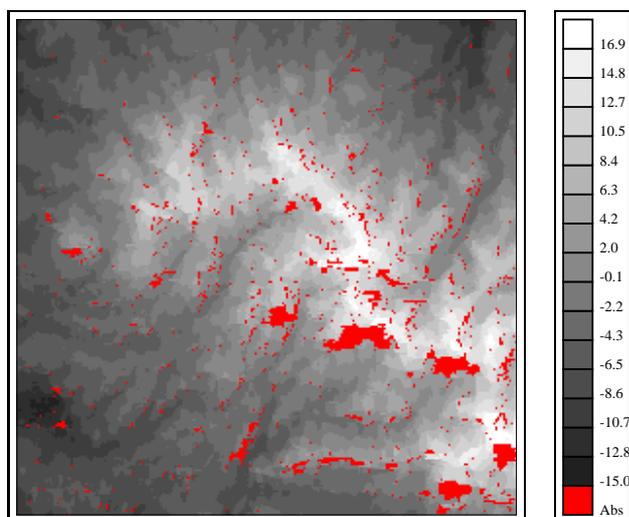


FIG. 5.26 – Carte de disparité du couple de la figure 5.25 calculée par corrélation (en rouge : valeurs manquantes).

égal à 35, est un modèle cubique. Compte tenu des remarques du paragraphe 5.2 concernant la pertinence des différents modèles de vraisemblance, nous avons utilisé dans ce cas uniquement le modèle de mélange MIXT.

A titre indicatif, nous avons représenté à la figure 5.27 la moyenne des simulations. Là encore, on observe une structure plus lisse que la carte de corrélation, mais la structure générale du relief y est clairement visible.

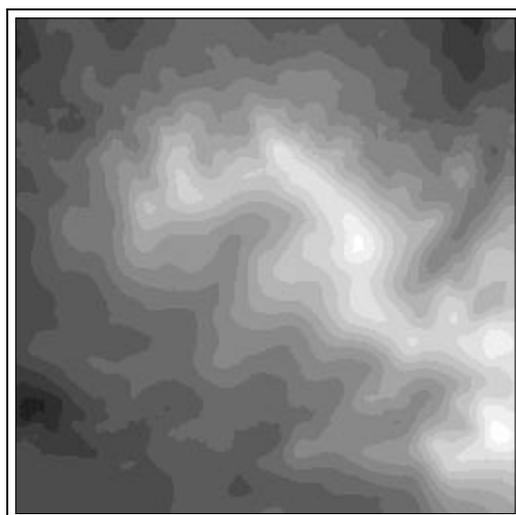


FIG. 5.27 – Moyenne

5.3.1 Probabilité de dépassement de seuil

Probabilité ponctuelle

Nous avons calculé les cartes de probabilité ponctuelle de dépassement de seuil correspondant aux valeurs 2, -2, 3 et -3 pixels. Elles sont représentées aux figures 5.28 et 5.29.

Par rapport au couple stéréoscopique précédent, la situation est assez différente.

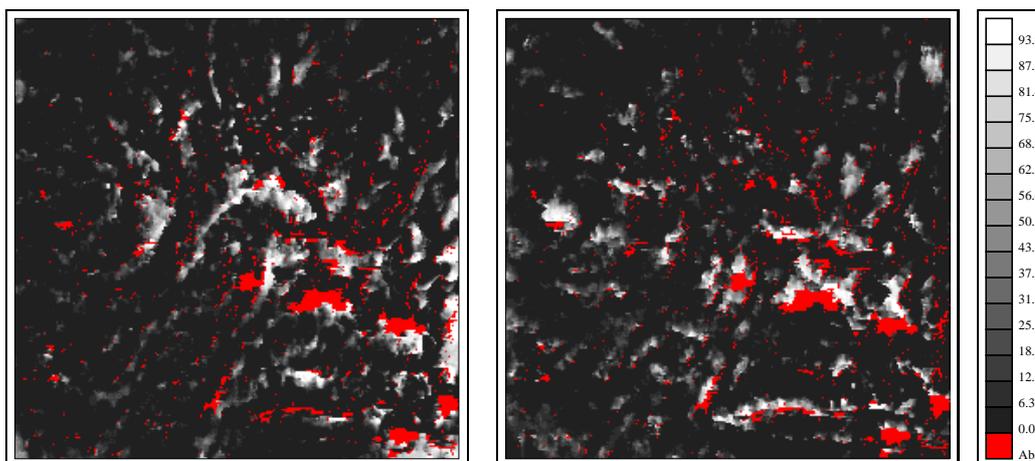


FIG. 5.28 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes).

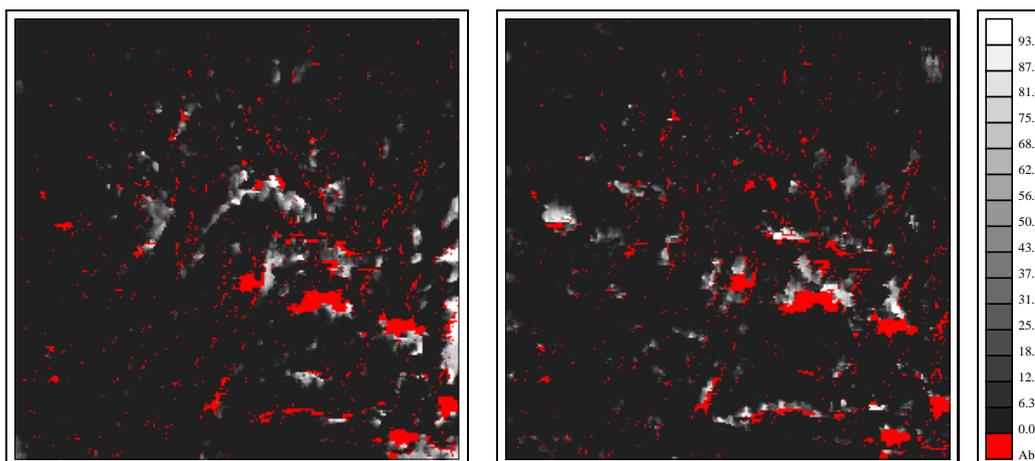


FIG. 5.29 – Probabilité (en %) ponctuelle de dépassement des seuils 3 (gauche) et -3 (droite) pixels (en rouge : valeurs manquantes).

L'information semble moins bien localisée : de larges zones exhibent des proba-

bilités d'erreur fortes, et cela même pour les seuils élevés de 3 et -3 pixels. Les queues de distribution semblent donc beaucoup plus longues que dans l'exemple précédent.

Il n'est pas surprenant d'obtenir à la fois une plus forte variabilité de la disparité et de plus grandes zones de forte probabilité d'erreur que dans l'exemple urbain. Ceci est dû tout d'abord à la nature du couple stéréoscopique, et notamment à la présence de déformations géométriques importantes et de larges zones d'ombre. On trouve également moins d'éléments fortement contrastés qui permettent "d'accrocher" la disparité, comme c'était le cas avec les bâtiments dans l'exemple précédent, si ce n'est les routes en lacets, aisément reconnaissables. Mais ces cartes reflètent également le choix de la covariance du modèle *a priori*. Celle-ci est en effet caractérisée par une forte variance (55) et de longues portées (240 et 420 pixels).

Il est intéressant d'analyser la nature des zones jugées critiques au regard de ces cartes de probabilité. Dans l'ensemble, les points de forte probabilité pour les seuils négatifs (-2 et -3 pixels) sont localisés en bordure de crêtes de la carte de disparité. Or, les lignes de crête sont aisément identifiables sur le couple stéréoscopique, car elles correspondent à des transitions sombre / clair. Comme le contraste est important le long de ces lignes, la corrélation produit un artefact assez typique : la ligne de crête est élargie, d'une longueur qui correspond grosso modo à la moitié de la taille de la fenêtre de corrélation utilisée. Ce phénomène est assez visible au centre de la carte de disparité. Il est intéressant de voir que ces erreurs assez importantes (puisque le gradient de disparité dans ces zones est fort) sont parfaitement repérables sur les cartes de probabilité de dépassement de seuil. Inversement, la plupart des points de forte probabilité pour les seuils positifs sont localisés dans les fonds de vallon. Dans ce cas, l'artefact dû à l'algorithme de corrélation est moins évident, car le contraste correspondant au fond de vallon est moindre que pour les crêtes.

Ces cartes d'erreur mettent en évidence un autre phénomène très intéressant, à savoir l'anisotropie de l'incertitude. La contrainte d'épipolarité implique que deux points homologues sont situés sur la même ligne des images rectifiées. Considérons donc une ligne de l'image : si la radiométrie est fortement contrastée le long de la ligne, il est aisé d'établir la mise en correspondance. Dans le cas contraire, des erreurs d'appariement risquent de se produire. Ainsi, dans une zone montagneuse, lorsque les crêtes sont perpendiculaires aux lignes de l'image et donc le contraste le long de la ligne fort, la mise en correspondance est précise, et est au contraire délicate si elles sont presque parallèles aux lignes de l'image. Ceci est tout à fait visible sur les cartes 5.28 et 5.29, notamment dans la partie inférieure droite de l'image où le relief est orienté selon les lignes du couple : tout le long de la crête, la probabilité de dépassement de seuil est forte.

Probabilité sur un domaine

Nous avons également calculé des cartes de probabilité de dépassement de seuil sur un domaine. La taille des domaines est de 64×64 pixels, et les seuils choisis sont 3, -3 , 4 et -4 pixels. Les résultats sont représentés aux figures 5.30 et 5.31.

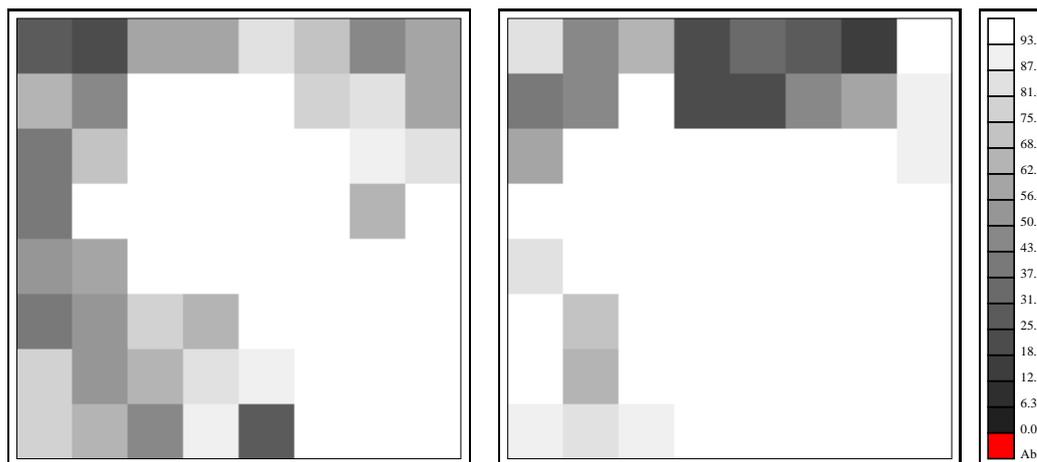


FIG. 5.30 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels.

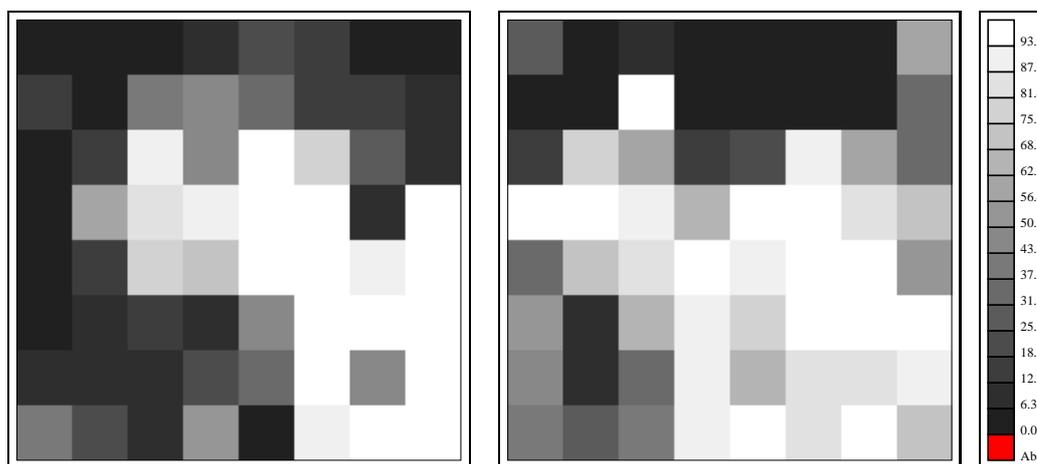


FIG. 5.31 – Probabilité (en %) sur un domaine de dépassement des seuils 4 (gauche) et -4 (droite) pixels.

Ces cartes mettent bien en évidence l'existence de deux zones sur la carte de disparité : une première, correspondant à peu près à la zone de fort relief, pour

laquelle la probabilité d'erreur, positive ou négative, est élevée, et une seconde, caractérisée par des valeurs plutôt faibles de la disparité, pour laquelle la probabilité d'erreur est bien inférieure.

5.3.2 Quelques remarques

Le deuxième exemple que nous venons de présenter est radicalement différent du premier : de part le type de relief que l'on cherche à reconstituer, marqué par des variations importantes, et de part la nature de l'information radiométrique, plus ambiguë (déformations géométriques, texture pauvre et larges zones d'ombre). Il s'agit d'un couple délicat à traiter pour le calcul de la disparité par corrélation, comme en témoigne la présence sur la carte de disparité d'importantes zones non estimées.

Comme noté au paragraphe précédent, les cartes de probabilité de dépassement de seuil permettent de détecter certaines erreurs grossières dues à l'algorithme de corrélation (généralement localisées en bordure de zones non estimées et alignées). Mais, et cela est bien mis en évidence par cet exemple, elles mettent également en évidence un artefact typique des cartes de disparité calculées par corrélation, sur lesquelles les éléments de relief fortement contrastés (comme les lignes de crête) ont tendance à "baver".

Cependant, il convient également de montrer les limites de la méthode utilisée. Nous avons en effet souligné que le choix du modèle de covariance dans le modèle *a priori* a une forte incidence sur les résultats. Il faut notamment être bien conscient du piège de l'hypothèse stationnaire. En effet, lorsque l'on doit modéliser un phénomène qui présente des variations importantes mais fortement structurées, on est confronté à un double choix : soit on décide que ces variations sont dues à une variation de la moyenne, soit à une variance forte couplée à une longue portée. Ainsi, sous l'hypothèse stationnaire, comme moyenne et variance sont supposées constantes, on est contraint de choisir un modèle à forte variance et longue portée pour décrire des zones de fort relief. La loi *a posteriori*, qui n'est certes plus stationnaire, est alors caractérisée également par de fortes variances et des corrélations élevées. Le choix du modèle *a priori* biaise donc en un certain sens les résultats.

Chapitre 6

Une alternative

Les résultats du chapitre précédent sont contrastés : s'ils mettent en évidence un intérêt certain de la méthode d'échantillonnage proposée, qui permet entre autres d'obtenir une information localisée précise, ils montrent également qu'un problème majeur subsiste : le choix du modèle *a priori*, dont la structure est particulièrement délicate à inférer. Il semble notamment que l'hypothèse de stationnarité utilisée jusqu'ici présente certaines limites en cas de fort relief. Or, le choix du modèle *a priori* a une incidence importante sur la forme de la distribution *a posteriori*. De plus, si le modèle stationnaire permet de traiter des extraits de couple stéréoscopique, l'application à l'analyse de couples stéréoscopiques entiers semble délicate. Une approche plus "pragmatique" s'impose. Nous examinons dans ce chapitre quelques idées pour proposer une méthode pouvant être mise en œuvre pour l'analyse automatique de MNTs.

6.1 Retour sur le modèle *a priori*

6.1.1 Quelle hypothèse de stationnarité ?

Le problème du choix du modèle *a priori* est réel. Nous avons souligné au chapitre 3 que plusieurs hypothèses de stationnarité étaient envisageables : modèle stationnaire d'ordre 2, modèle intrinsèque ou modèle avec dérive. Or, il apparaît que nous avons principalement travaillé en hypothèse de stationnarité d'ordre 2 (cas où la covariance est définie et ne dépend que du vecteur déplacement entre deux points). Ce choix n'est par fortuit. Tout d'abord, il convient de tester en premier lieu les hypothèses les plus simples. De plus, il existe un grand nombre d'algorithmes de simulation pour de tels modèles ; certains d'entre eux (ceux reposant sur la transformée de Fourier) permettent de générer des champs de grande taille avec un temps de calcul très court.

Or, le modèle stationnaire d'ordre 2 a des limites. En effet, si l'on considère une scène SPOT entière, on distingue des zones accidentées, voire montagneuses, des zones de plaine, rurales ou urbaines, et des zones de transition. On ne peut vouloir décrire une telle scène avec un même modèle stationnaire. Une solution pourrait constituer à scinder la scène considérée en zones homogènes, avec chacune un modèle différent. Intellectuellement, cette solution n'est pas satisfaisante, et pratiquement, elle est peu envisageable : comment effectuer automatiquement cette partition ? Mais même si l'on considère des zones de taille plus réduite, où l'hypothèse de stationnarité peut être envisagée, il subsiste des difficultés : le cas de l'extrait de zone montagneuse du chapitre 5 en est un exemple. En effet, dans de telles configurations, l'hypothèse stationnaire oblige à choisir une structure caractérisée par une portée spatiale importante (supérieure à la moitié du champ considéré), et une variance forte. On a ainsi tendance à surestimer la variabilité, donc l'incertitude liée à la disparité en zone montagneuse. De plus, du fait des fortes portées de corrélation, on fait apparaître des zones critiques (zones où la probabilité d'erreur est significative) certainement trop étendues.

Enfin, et c'est un argument pratique important, la simulation de champs à longues portées est plus lourde numériquement. En effet, avec un algorithme du type FFT, on ne peut simuler que des champs dont la taille est supérieure à la distance pour laquelle la covariance est presque nulle [22, 23]. Ainsi, même lorsque l'on a une approche du type Gibbs par blocs, où simplement une partie du champ est renouvelée à chaque itération, il faut simuler des blocs beaucoup plus grands que le bloc à renouveler. De plus, et même si l'on utilise des approximations, il faut également augmenter la taille du voisinage de krigeage utilisé pour conditionner la simulation.

Lorsque l'hypothèse de stationnarité d'ordre 2 apparaît trop limitée, une solution peut consister à formuler une hypothèse de stationnarité moins contraignante. On pourrait donc penser s'orienter vers l'hypothèse intrinsèque, dans laquelle les accroissements de la variable seulement sont supposés stationnaires, voire les modèles plus complexes que sont les fonctions aléatoires intrinsèques d'ordre k (FAI- k). Une FAI- k est une fonction aléatoire dont les accroissements généralisés d'ordre k sont stationnaires et de moyenne nulle [74, 24]. A une FAI- k est associée une covariance généralisée, définie à un polynôme de degré $2k$ près. Dans cette approche, deux FAI- k qui diffèrent d'un polynôme de degré k sont représentées par le même modèle.

Cependant, malgré le formalisme intéressant qu'introduit la notion de FAI- k , l'inférence du modèle est extrêmement délicate, car l'on ne peut calculer à partir d'une réalisation d'une FAI- k un estimateur expérimental de la covariance généralisée. De plus, le type de non-stationnarité auquel nous sommes confrontés ne correspond pas réellement au comportement polynômial qui est au cœur du formalisme des FAI- k , ne serait-ce que parce que la disparité est bornée. D'ailleurs les résultats préliminaires obtenus avec un modèle intrinsèque de variogramme

linéaire n'ont pas été concluants.

Reste donc le modèle avec dérive. Si l'on cherche à décomposer la fonction D en un terme de dérive et un terme stationnaire, on aboutit aux relations suivantes :

$$\mathbb{E}(D(u, v)) = m_p(u, v) \quad (6.1)$$

$$\text{Cov}(D(u_1, v_1), D(u_2, v_2)) = C_p(|u_1 - u_2|, |v_1 - v_2|) \quad (6.2)$$

La dérive m_p , non constante, correspond à l'espérance de la fonction D , et le résidu $D - m_p$ est une fonction aléatoire stationnaire d'ordre 2 de moyenne nulle. Mais là encore, on se heurte à une difficulté pratique, qui est l'estimation de la dérive m_p . Lorsque la non stationnarité est due à l'influence d'une autre variable (c'est souvent le cas en physique) et qu'il existe une relation fonctionnelle entre cette variable externe et la variable étudiée, m_p traduit cette relation fonctionnelle et l'on parle alors de dérive externe. Ce n'est évidemment pas le cas dans le problème que nous considérons.

6.1.2 Vers une solution

Face à ces difficultés théoriques (par ailleurs récurrentes en statistiques spatiales), on peut avoir une approche pragmatique qui laisse de côté le formalisme rigoureux du chapitre 3 pour proposer une méthode opérationnelle. La validité d'une telle approche peut être discutée, elle a le mérite d'apporter une solution.

Modèle de disparité résiduelle

Considérons de nouveau la dichotomie en terme de dérive m_p et terme résiduel stationnaire d'ordre 2, de covariance C_p . La dérive m_p est inconnue, mais nous disposons d'une estimation de la disparité \hat{d} calculée par corrélation. Notons bien que \hat{d} dépend du couple stéréoscopique. On peut alors prendre \hat{d} comme estimation de m_p , le terme résiduel s'identifiant maintenant à l'erreur de disparité. On aboutit formellement à un modèle *a priori*, dont l'expression est :

$$\pi_D(d) \propto |C_p|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(d - \hat{d})^\top C_p^{-1}(d - \hat{d})\right) \quad (6.3)$$

La simulation de ce modèle se fait comme énoncé au chapitre 3, en utilisant un des trois modèles de vraisemblance GMF-0, GMF-1 ou MIXT.

Structure spatiale

Dans le modèle de dichotomie proposé, la covariance C_p est implicitement à courte portée. La simulation de cette disparité résiduelle est donc plus rapide : on

peut utiliser l'algorithme de simulation par FFT sans avoir à simuler "inutilement" des champs de grande taille.

Si l'on cherche même à optimiser l'algorithme, on peut éventuellement choisir un modèle de Gibbs, dans lequel la matrice inverse C_p^{-1} est fortement creuse (voir l'exemple de l'annexe C). Il faut alors se résoudre à travailler sur une grille fixe et l'algorithme d'échantillonnage d'importance proposé au paragraphe 4.1.4, qui permet de scinder la simulation en deux étapes successives, ne peut plus être utilisé.

Le modèle de dichotomie présente cependant deux difficultés majeures. La première est d'ordre statistique. On sait bien que si l'on estime tout d'abord le terme de dérive par une procédure d'optimisation, puis le terme résiduel, à l'aide du même jeu de données, les résultats sont biaisés, car l'on a tendance à sous-estimer la variabilité du résidu. La seconde critique est d'ordre pratique. Comment inférer la structure spatiale, i.e. la covariance C_p , du modèle *a priori*, puisque l'on ne peut plus utiliser la carte de disparité \hat{d} ? A ces deux questions, on peut répondre en reprenant l'approche proposée au paragraphe 3.3 et traiter la variance du modèle *a priori* comme une inconnue que l'on simule en même temps que la disparité. Certes, les paramètres de forme de la covariance (ainsi que le choix du modèle) restent toujours inaccessibles, mais des études expérimentales peuvent permettre de choisir des valeurs raisonnables.

6.2 Illustration

Nous reprenons l'exemple du couple stéréoscopique en zone montagneuse du chapitre précédent (figure 5.25). La carte de disparité de la figure 5.26 est utilisée comme moyenne du modèle *a priori*, les valeurs aux points non estimés par la corrélation étant obtenues par interpolation. La covariance C_p est un modèle sphérique de portée 12 pixels. La variance est fixée initialement à 0.5, puis est estimée par Monte Carlo au cours de la période de "chauffe" (10 000 itérations). La valeur limite obtenue est de l'ordre de 0.35, soit un écart type proche de 0.6. Cette valeur correspond bien intuitivement à la précision moyenne que l'on peut attendre de l'algorithme de corrélation. Nous utilisons le modèle de vraisemblance MIXT.

L'algorithme d'échantillonnage est l'algorithme MMH gaussien, avec un groupement par blocs selon des bandes horizontales de largeur 8 pixels. Comme précédemment, on utilise une sous-grille de simulation (facteur de réduction 2 dans chaque direction, soit des grilles de taille $128 \times 128 = 2^{14}$), et la méthode d'échantillonnage d'importance décrite au paragraphe 4.1.4. Un total de $l = 2000$ simulations ont été calculées, avec 50 itérations markoviennes entre deux simulations.

Moyenne et écart type calculés par Monte Carlo sont représentés à la figure 6.1. La carte moyenne est bien sûr assez proche de la carte de disparité obtenue par

corrélation (cf. figure 5.26 pour comparaison), mais localement, des différences importantes (supérieures à 5 pixels en valeur absolue) apparaissent. La carte d'écart type, qui ici encore n'est qu'une indication de la variabilité de la disparité dans le cadre du modèle, exhibe une structure assez intéressante. La variance d'estimation varie en effet fortement spatialement : les valeurs fortes (au centre et en bas) correspondent principalement aux versants sombres du relief observé sur le couple stéréoscopique, tandis que les zones de valeurs plus faibles (partie supérieure de la carte) reprennent le tracé des routes en lacets visibles sur le couple.

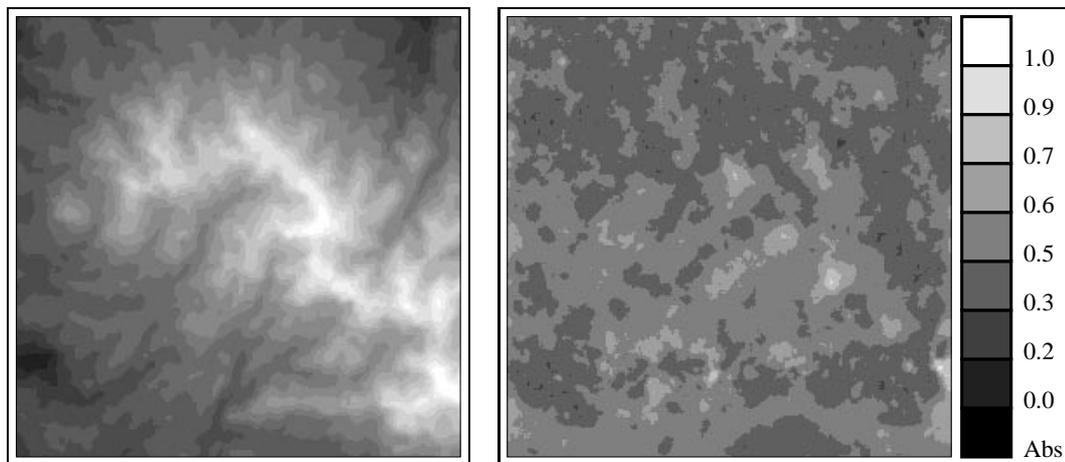


FIG. 6.1 – Moyenne et écart type obtenus avec le modèle de disparité résiduelle

6.2.1 Probabilité de dépassement de seuil

Probabilité ponctuelle

Nous avons calculé la probabilité ponctuelle de dépassement relative aux seuils 2 et -2 pixels. Ces cartes, représentées à la figure 6.2, sont très différentes de celles obtenues au paragraphe 5.3. Les zones critiques (pour lesquelles la probabilité d'erreur est significative) sont beaucoup plus localisées. On retrouve en outre fréquemment des groupements de points alignés le long des lignes de l'image, ce qui correspond à une erreur caractéristique de l'algorithme de corrélation : chaque ligne étant traitée indépendamment l'une de l'autre, et la contrainte d'ordre ne permettant que des solutions monotones, un appariement faux entraîne fréquemment à sa suite plusieurs erreurs.

On peut remarquer que les points pour lesquels la disparité est surestimée avec une forte probabilité (seuil -2) sont principalement situés en bordure de crêtes. Nous avons signalé au chapitre précédent que sur la carte de disparité estimée par corrélation, les crêtes ont tendance à "baver". C'est donc cet artefact qui est

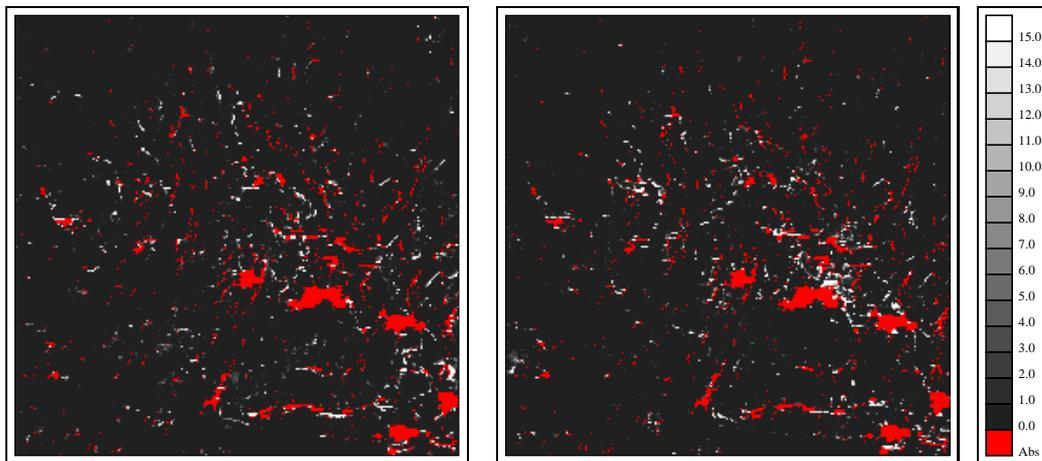


FIG. 6.2 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes).

identifié sur les cartes de probabilité. En revanche, par comparaison aux cartes du paragraphe 5.3, l'étendue de ces zones est plus restreinte. Ceci est bien entendu à mettre en relation avec la structure spatiale du modèle *a priori*. Dans le modèle de disparité résiduelle, la variabilité et la corrélation spatiale sont plus faibles que dans le modèle stationnaire. Enfin, il semble que certains points critiques correspondent à des fortes discontinuités locales de la carte de disparité.

Probabilité sur un domaine

Les cartes de probabilité sur un domaine de dépassement relatives aux seuils 3 et -3 pixels sont représentées à la figure 6.3. Leur structure est assez semblable à celles des cartes obtenues avec le modèle stationnaire (cf. figure 5.30), mais bien entendu les probabilités sont plus faibles. On retrouve notamment, et de manière encore plus nette, l'existence de deux zones, la probabilité d'erreur en zone de fort relief étant plus forte qu'en absence de relief marqué. Ainsi, bien que le modèle *a priori* soit de variance constante et que les fortes variations du relief y soient décrites par la moyenne, non constante, la loi *a posteriori* exhibe bien un comportement radicalement différent selon le type de relief observé.

6.2.2 Sensibilité aux paramètres

Contrairement au modèle stationnaire pour lequel nous avons montré au chapitre 3 que les paramètres peuvent être raisonnablement estimés, le choix de la covariance C_p et des paramètres de forme et de portée est beaucoup plus subjectif dans le cas du modèle de disparité résiduelle. Il est donc important de tester la

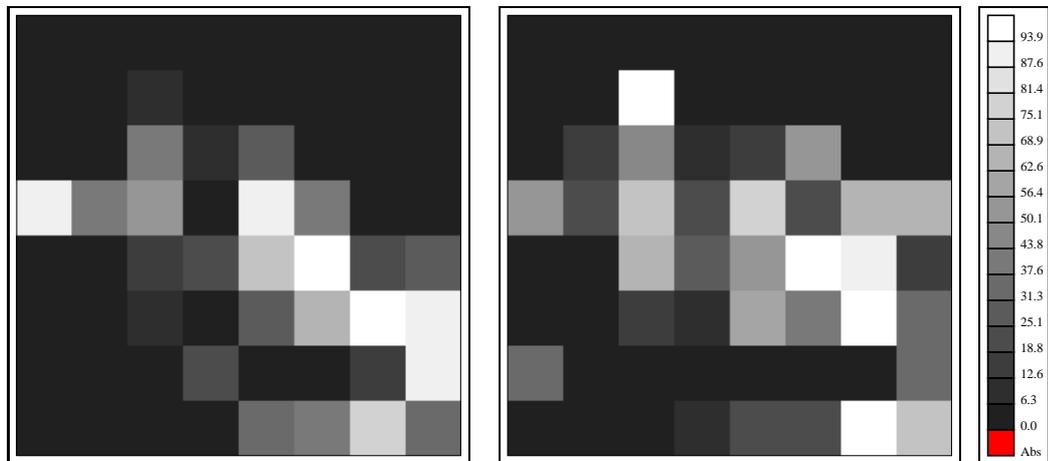


FIG. 6.3 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels.

sensibilité des résultats au choix de ces paramètres.

A titre d'illustration, nous avons effectué les mêmes calculs que précédemment, mais avec une portée double (24 pixels) et une variance initiale (avant estimation par Monte Carlo) double également (1 au lieu de 0.5). L'estimation par Monte Carlo de la variance du modèle *a priori* (de l'ordre de 0.35) est semblable à la valeur obtenue précédemment. Sont représentées pour comparaison les cartes de moyenne et d'écart type (figure 6.4), de probabilité ponctuelle pour les seuils 2 et -2 pixels (figure 6.5) et sur un domaine pour les seuils 3 et -3 pixels (figure 6.6). La principale influence d'une augmentation du paramètre de portée est une diminution de la variabilité locale, visible sur la carte d'écart type. Comme souligné au chapitre 5, plus la corrélation spatiale est forte, plus la probabilité de dépassement d'un seuil fixé est faible. C'est en effet ce que l'on observe aux figures 6.5 et 6.6, où les zones critiques sont moins nombreuses que précédemment. Cependant, on observe une certaine stabilité pour les points de forte probabilité de dépassement, qui restent les mêmes pour les deux choix de paramètres.

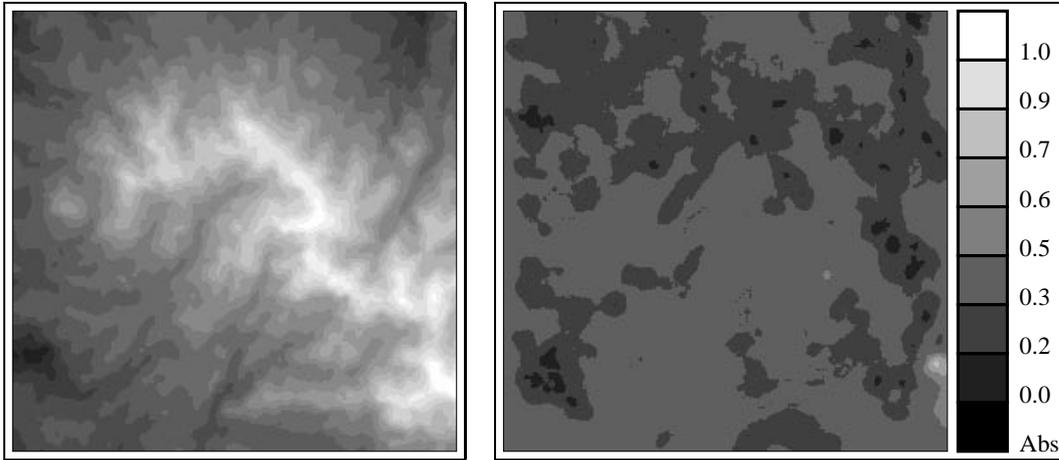


FIG. 6.4 – Moyenne et écart type obtenus

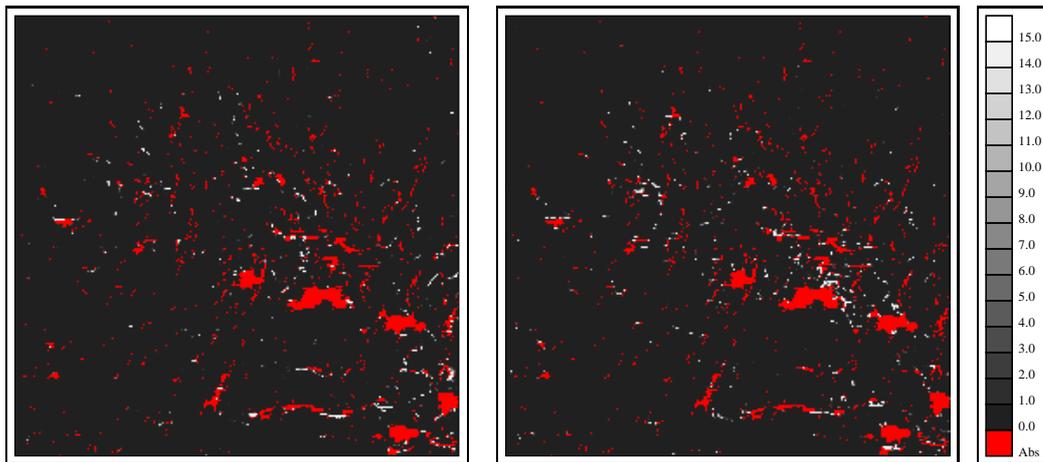


FIG. 6.5 – Probabilité (en %) ponctuelle de dépassement des seuils 2 (gauche) et -2 (droite) pixels (en rouge : valeurs manquantes).

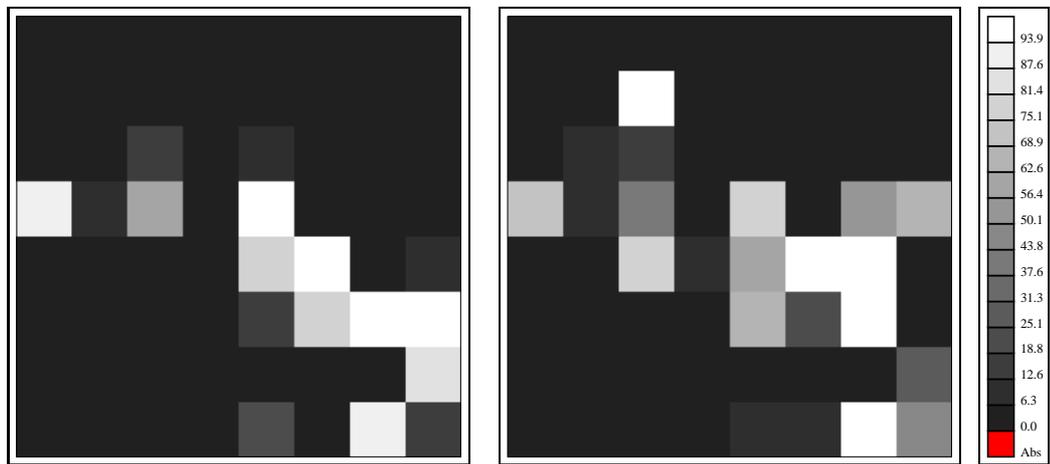


FIG. 6.6 – Probabilité (en %) sur un domaine de dépassement des seuils 3 (gauche) et -3 (droite) pixels.

6.2.3 Performances relatives

A plusieurs reprises, nous avons évoqué les contraintes liées au temps de calcul. Nous revenons ici sur cet aspect du problème, pour l'exemple du couple stéréoscopique en zone montagneuse étudié précédemment. Le tableau 6.1 montre une comparaison des temps de calcul selon le modèle *a priori* et la taille du champ à simuler. Les deux modèles comparés sont le modèle stationnaire utilisé au paragraphe 5.3 et le modèle de disparité résiduelle utilisé précédemment. Pour indication, nous avons également indiqué les temps que l'on obtiendrait si on avait choisi un modèle de disparité résiduelle où le résidu est modélisé par un champ de Markov d'ordre 1 (voir annexe C pour les détails). Dans tous les cas, l'algorithme de simulation utilisé est l'algorithme MMH gaussien avec un renouvellement par blocs selon des bandes horizontales. Les champs gaussiens sont générés à l'aide d'un algorithme par transformée de Fourier [22, 23], sauf dans le cas du modèle de disparité résiduelle avec champ de Markov où l'on utilise une décomposition de la matrice de covariance par Choleski. Les résultats ont été obtenus sur une station Sun Ultra 60 équipée d'un processeur SPARC II à 400 Mhz.

taille du champ	64×64	128×128
modèle stationnaire	40	330
modèle de disparité résiduelle	15	61
modèle de disparité résiduelle et structure de champ de Markov	3	22

TAB. 6.1 – Temps de calcul (en secondes) correspondant à 100 itérations markoviennes (algorithme MMH de dimension 24) en fonction du modèle *a priori* et de la taille de la grille.

Cette comparaison montre clairement l'avantage du modèle de disparité résiduelle : les temps de calcul sont bien inférieurs à ceux correspondant au modèle stationnaire (facteur 2.5 pour un champ de taille 64×64 , et 5 pour un champ de taille 128×128). En effet, dans le modèle stationnaire, la covariance s'annule pour les grandes distances (supérieures à la moitié de la taille du champ), ce qui nécessite de simuler des blocs de taille bien supérieure à la taille du bloc à renouveler. Ce problème disparaît avec le modèle de disparité résiduelle, car la covariance s'annule pour des distances inférieures à la taille du bloc à renouveler. Notons également que l'utilisation d'une structure de champ de Markov permet de gagner encore un facteur 5 dans le cas du champ 64×64 , et 3 dans le cas du champ de taille 128×128 .

6.3 Exemples d'application

Nous présentons pour clore cette étude quelques exemples d'application de notre méthode à l'analyse de l'incertitude liée au MNT lui-même. Jusqu'à présent, tous les résultats que nous avons indiqués concernaient la carte de disparité, or, *in fine*, c'est le MNT qui est la variable d'intérêt. Comme expliqué au paragraphe 3.1, pour obtenir une information relative au MNT, il suffit de calculer pour chaque carte de disparité simulée le MNT correspondant et d'effectuer l'intégrale de Monte Carlo sur cette variable. Cette souplesse est justement l'un des intérêts de l'approche par Monte Carlo.

A partir de la carte de disparité 5.26 obtenue par corrélation et d'une procédure d'interpolation, on déduit le MNT représenté à la figure 6.7 (pour simplifier, nous avons utilisé une approximation linéaire et avons projeté le MNT dans le système de coordonnées de l'image droite du couple rectifié).

6.3.1 Altitude de vol

Une des utilisations possibles des MNTs dans l'industrie aéronautique est la détermination de la trajectoire et de l'altitude de vol. Nous avons ici considéré le cas où la trajectoire est fixée, et il s'agit de calculer une altitude de vol pour laquelle la probabilité de collision est en deçà d'un certain seuil. Les deux trajectoires envisagées ont été choisies parallèles l'une aux lignes de l'image, l'autre aux colonnes (cf. figure 6.7). Les profils correspondants sont représentés à la figure 6.8. Sur ces profils, nous nous sommes tout d'abord intéressés à la probabilité pour que l'altitude réelle soit supérieure de 25 m à l'altitude prédite par le MNT. Les zones pour lesquelles cette probabilité est supérieure à 1% sont représentées en bleu sur les figures. Il s'agit principalement des flans et des fonds de vallée. Nous avons également calculé l'altitude supérieure maximale correspondant au risque 0.05%. C'est l'altitude de référence qui peut par exemple être choisie pour déterminer la trajectoire optimale. La différence entre cette altitude maximale et l'altitude donnée par le MNT, que l'on peut interpréter comme la marge de sécurité, est dans ces exemples de l'ordre de 25 m, avec quelques valeurs dépassant 50 m. Enfin, nous avons calculé la probabilité de collision le long de chacune des deux trajectoires, pour différentes altitudes de vol (tableau 6.2). Les résultats montrent que pour le profil vertical, une altitude de 50 m au-dessus du MNT est suffisante, en revanche, pour le profil horizontal, une altitude de 100 m est nécessaire. Il est intéressant de constater que l'incertitude relative au profil horizontal est plus forte, bien que l'altitude moyenne soit plus faible : en effet, sur ce profil les crêtes sont parallèles aux lignes de l'image, et donc le contraste d'intensité utile pour calculer la disparité est faible. Nous avons déjà signalé ce phénomène précédemment.

altitude de vol au-dessus du MNT	profil vertical	profil horizontal
25 m	32.67	100.
50 m	0.	64.97
100 m	0.	0.

TAB. 6.2 – Probabilité (en %) de collision le long d’une trajectoire pour différentes altitudes de vol.

6.3.2 Zone de visibilité d’un point

Un autre exemple d’application est le calcul de la zone de visibilité d’un point de la scène. Il s’agit de déterminer l’ensemble des points de la scène qui peuvent être vus depuis un point d’observation fixé. Deux points sont visibles l’un par rapport à l’autre si la droite les joignant ne coupe pas la surface topographique. Ce type d’information est par exemple utile en urbanisme pour prévoir l’impact d’une nouvelle construction.

Nous avons choisi comme point d’observation le point de la scène d’altitude la plus élevée. La zone de visibilité calculée à partir du MNT de la figure 6.7 est représentée à la figure 6.9. Seule une assez faible portion de la scène est visible, et notamment pratiquement toute la partie supérieure est dissimulée. Afin de chiffrer l’incertitude relative à cette variable, on peut calculer à partir des simulations générées la probabilité pour qu’un point soit visible. Cette carte de probabilité est représentée à la figure 6.10. La comparaison des deux cartes montre que la zone de visibilité calculée à partir du MNT est certainement quelque peu optimiste : il suffit en effet que le point d’observation soit situé légèrement plus bas pour que cette zone soit fortement réduite. Il est intéressant également de constater que de nouvelles zones éventuellement visibles apparaissent dans la partie supérieure de la scène.

6.4 Bilan

Le modèle de disparité résiduelle que nous avons utilisé dans ce chapitre représente une alternative pratique au modèle stationnaire d’ordre 2 dès qu’il s’agit d’étudier automatiquement des scènes SPOT entières. En effet, dans un contexte d’application industrielle, il convient de développer des méthodes pouvant être aisément mises en œuvre (i.e. de manière quasiment automatique), ne nécessitant pas un coût de calcul exorbitant et permettant d’obtenir des résultats à la fois robustes et précis.

A titre indicatif, la méthodologie suivante peut être proposée :

- calcul de la carte de disparité par un algorithme du type corrélation, et interpolation si nécessaire,

- division de la scène SPOT en portions de taille comprise entre 128×128 et 512×512 pixels,
- sur chaque portion, simulation (en parallèle éventuellement) du modèle stochastique, avec les choix suivants :
 - modèle *a priori* de disparité résiduelle, et structure stationnaire à courte portée,
 - modèle de vraisemblance MIXT,
- estimation des paramètres du modèle par Monte Carlo (ou algorithme SEM) pendant la période de “chauffe” (entre 5 000 et 10 000 itérations),
- utilisation de l’algorithme MMH avec un renouvellement par blocs (la taille des blocs dépend de la portée utilisée, et peut-être ajustée de manière à garantir un taux d’acceptation des transitions comprise entre 20% et 40%),
- calcul pour chaque carte de disparité générée du MNT correspondant,
- calcul par Monte Carlo des probabilités de dépassement de seuil (éventuellement le long de trajectoires particulières) et d’intervalles de confiance.

Si l’on choisit de travailler à grille fixe, le modèle *a priori* peut être un champ de Markov semblable à celui proposé en annexe C. Si au contraire l’approche par des fonctions aléatoires est conservée, l’algorithme d’échantillonnage d’importance peut être utilisé.

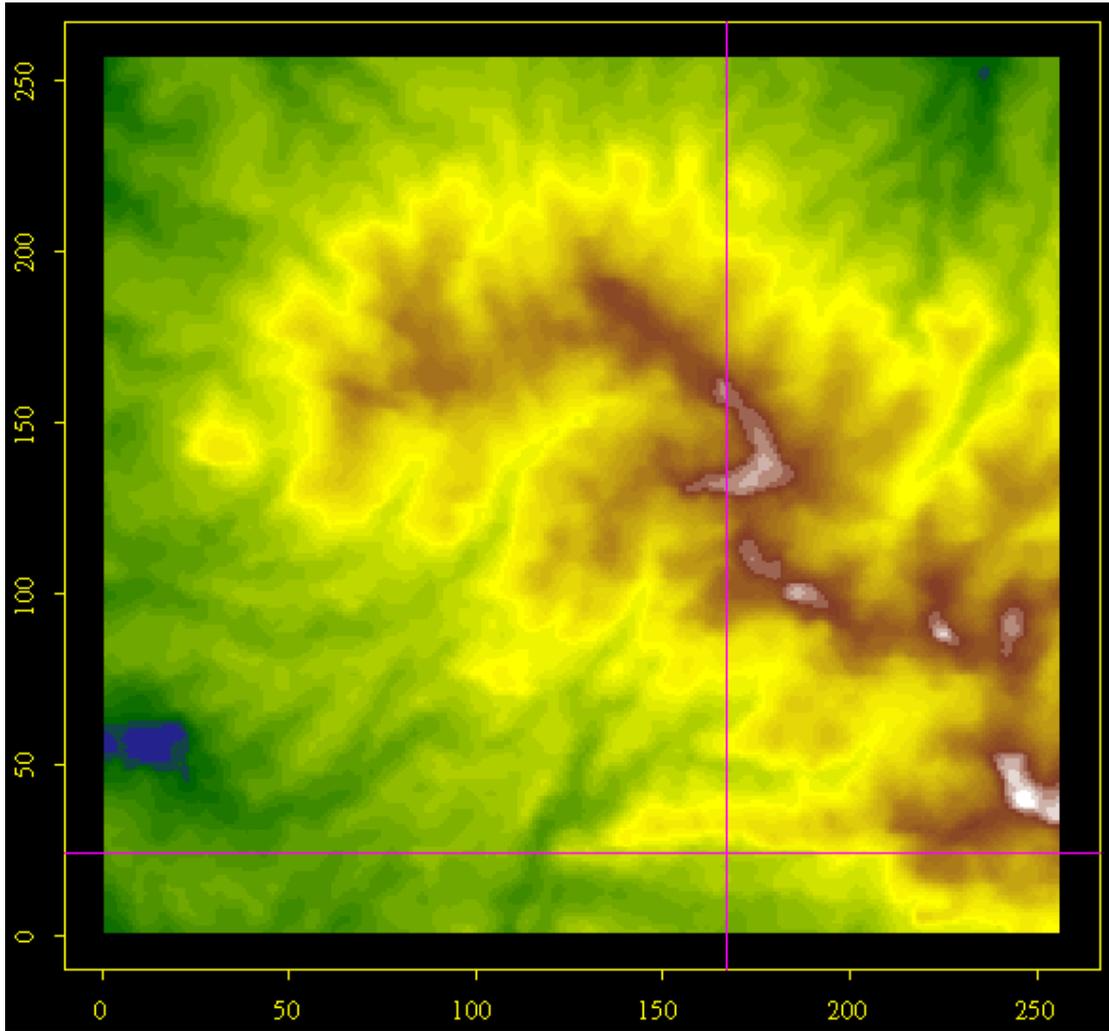


FIG. 6.7 – MNT déduit de la carte de disparité 5.26. Les deux droites représentées symbolisent les deux trajectoires envisagées.

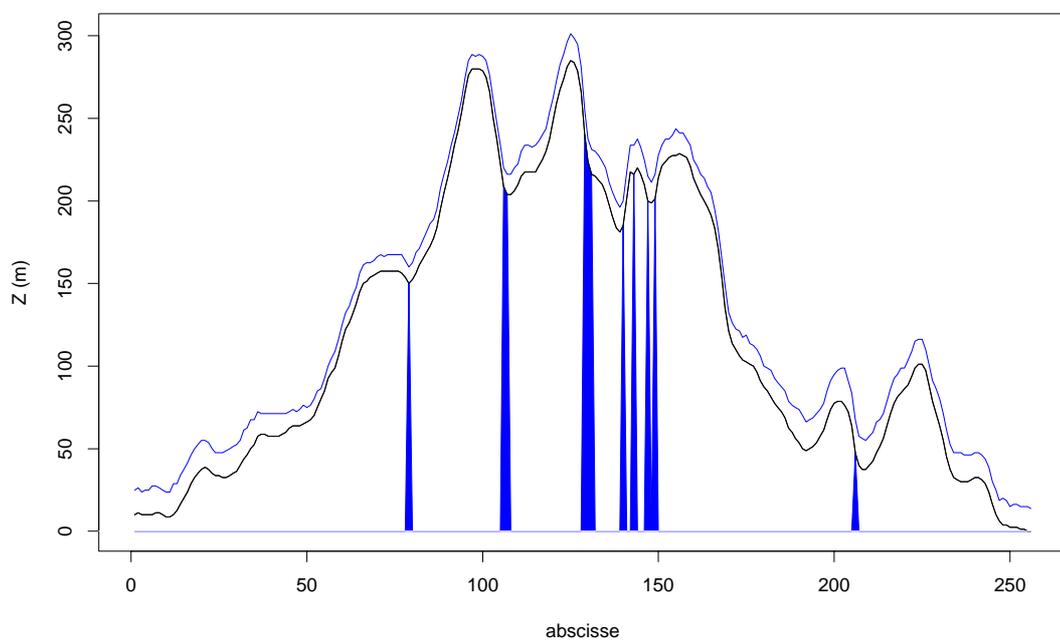
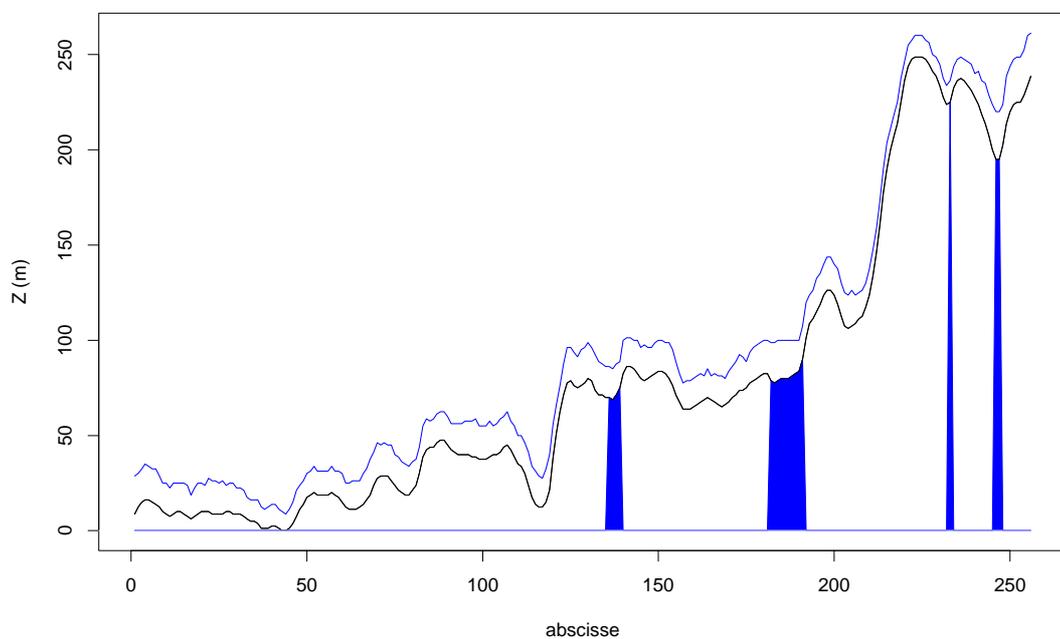


FIG. 6.8 – Profils d'altitude correspondant aux deux trajectoires de la figure 6.7 (courbe noire) : en haut profil horizontal, en bas profil vertical. Les zones en bleu représentent les points de probabilité significative (supérieure à 1%) d'erreur supérieure à 25 m. La courbe bleue est la borne supérieure de l'intervalle de confiance, le risque associé est 0.05%.

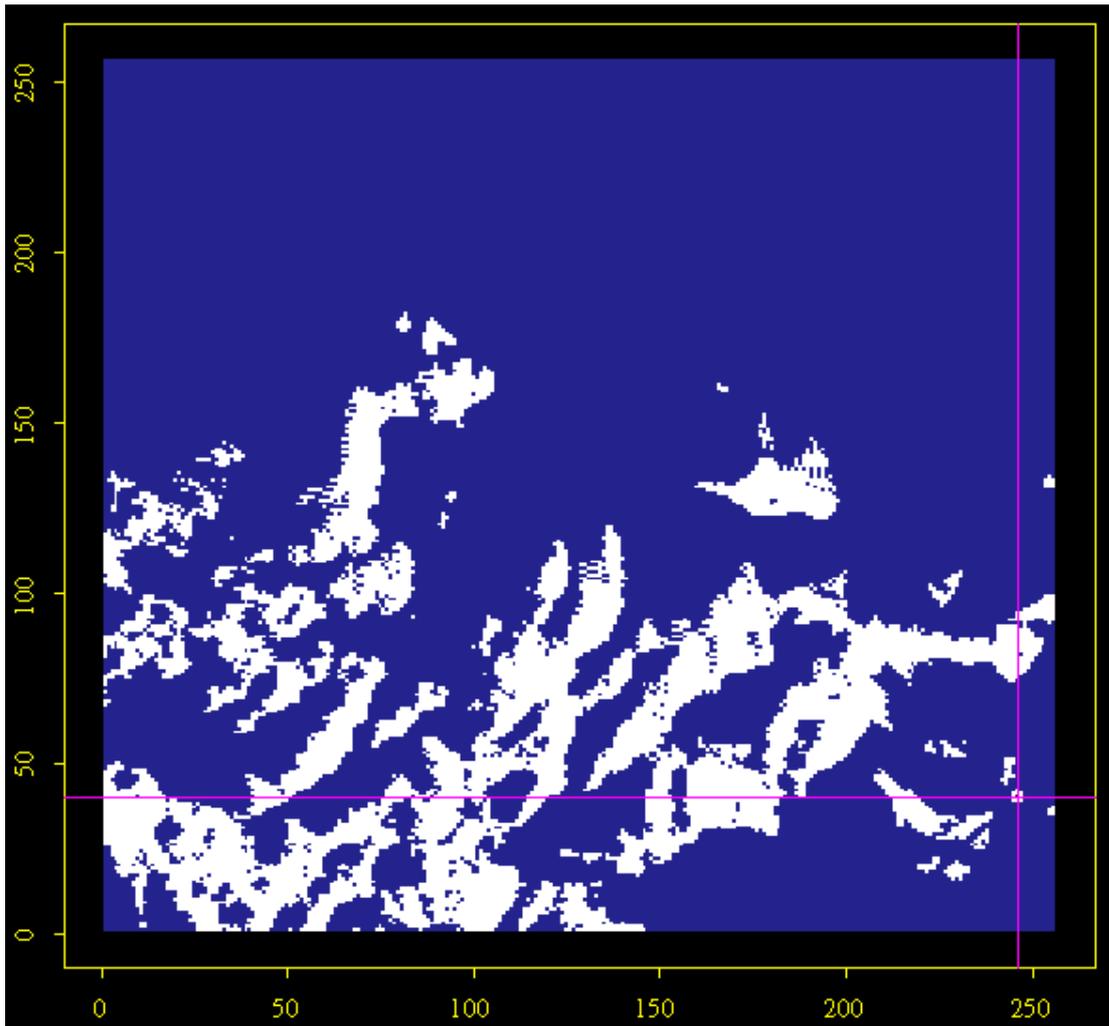


FIG. 6.9 – Zone de visibilité d'un point. Le point d'observation est symbolisé par l'intersection des deux droites. En blanc : points visibles, en bleu : points non visibles.

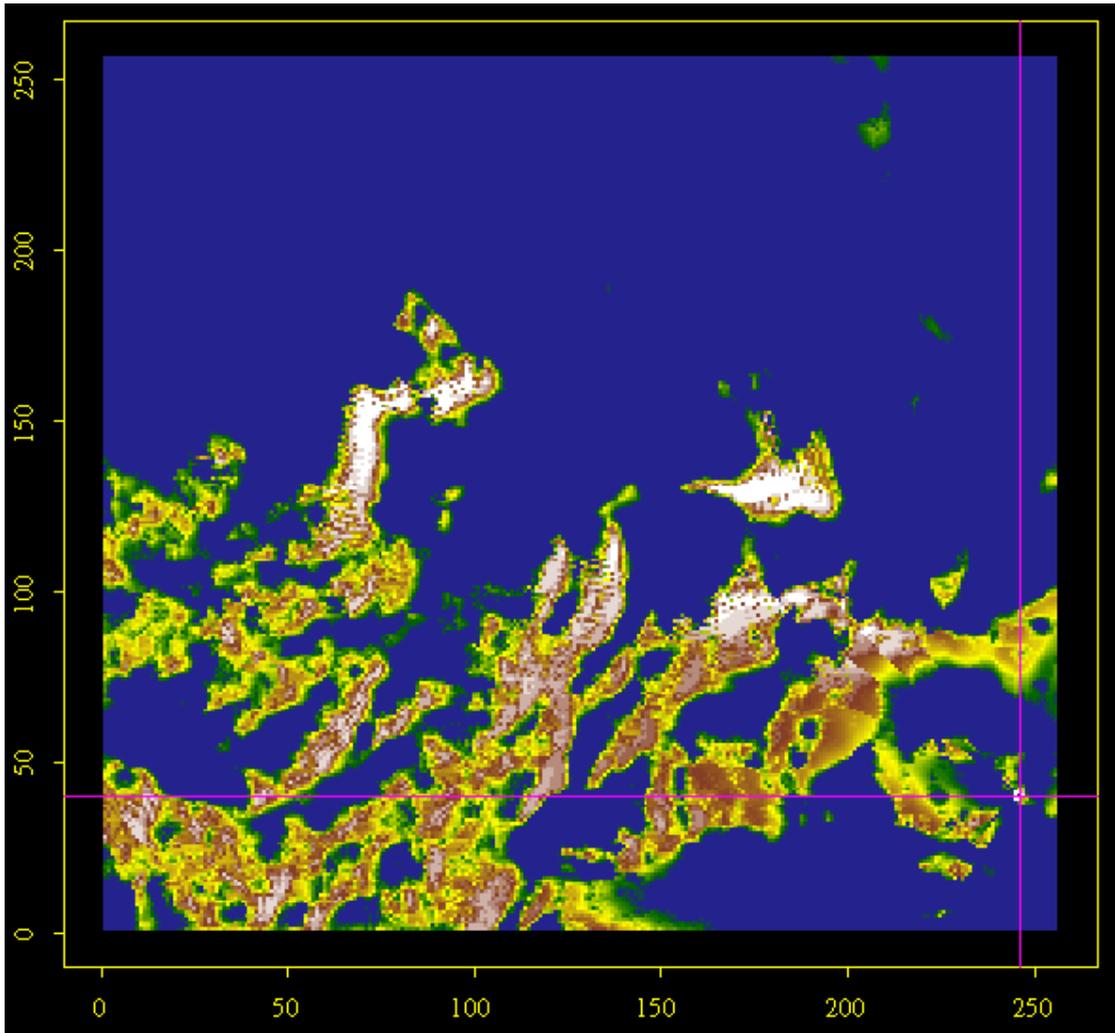


FIG. 6.10 – Probabilité de visibilité d'un point. Le point d'observation est symbolisé par l'intersection des deux droites.

Chapitre 7

Conclusion générale et perspectives

Dans ce travail, nous nous sommes efforcés d'apporter une réponse au problème de l'incertitude en vision stéréoscopique. Une des difficultés majeures de cette problématique réside dans le sens à donner au terme incertitude. La plupart du temps, ce problème est abordé en terme d'efficacité de l'algorithme de calcul, selon la démarche suivante : à partir d'un jeu de données de référence, différents algorithmes de reconstruction 3D sont testés et classés en fonction de plusieurs critères (écart quadratique moyen, proportion d'erreurs supérieures à un certain seuil, proportion de points non estimés...). Ce type d'évaluation fait même l'objet d'un projet récent de recherche [104]. Or, il est assez frappant de constater que malgré le nombre d'algorithmes testés et la variété des approches considérées, une limite semble être atteinte : des erreurs subsistent. C'est que l'incertitude est une part inhérente du paradigme de la vision stéréoscopique, et qu'il convient de traiter le problème en tant que tel. Il faut donc changer radicalement de point de vue, et abandonner la croyance illusoire qui consiste à vouloir reconstituer fidèlement la réalité. En effet, les observations fournies par chaque image du couple sont imparfaites et partielles, et des ambiguïtés existent.

Au problème posé, nous avons cherché à donner une formulation mathématique simple. Pour cela, nous avons proposé un cadre probabiliste et une approche bayésienne, nous fondant sur la remarque suivante : la reconstruction géométrique d'une scène à partir d'un couple stéréoscopique est un problème inverse. Il n'existe notamment pas de bijection entre l'information disponible (le couple stéréoscopique $Y = (I_1, I_2)$) et la variable à estimer (la géométrie Z de la scène, ou ce qui est équivalent, la disparité D). Les valeurs que peut prendre la disparité sont alors décrites par la probabilité *a posteriori* $\pi_D(.|y)$ de D connaissant le couple $Y = y$. Ainsi, la multiplicité des solutions est explicitement considérée. Cette formulation a un double avantage. Tout d'abord, elle donne une signification

à la notion d'incertitude évoquée initialement. Il s'agit de calculer la probabilité pour que la valeur de la disparité soit dans un certain domaine (cas des probabilités de dépassement de seuil), ou, au contraire, de calculer le domaine correspondant à un niveau de risque fixé (cas de l'intervalle de confiance). Mais, on peut également envisager de s'intéresser à des problématiques plus ciblées, comme la probabilité pour qu'un objet évoluant selon une trajectoire prédéfinie dans l'environnement tridimensionnel reconstitué ne percute aucun élément de la scène. De plus, une fois que le problème devient le calcul d'une loi *a posteriori*, il est possible de mettre en œuvre un certain nombre de méthodes dont la validité est assurée. Il s'agit dans ce cas particulier des méthodes MCMC qui permettent de calculer des intégrales à partir d'un échantillon statistique de la loi considérée. Le but est de simuler la loi *a posteriori* en générant une chaîne de Markov qui admet cette loi pour loi stationnaire. Ainsi, si l'estimation de la disparité est un problème d'optimisation, l'étude de l'incertitude liée à la disparité est un problème d'échantillonnage. Remarquons enfin que dans cette approche c'est l'incertitude inhérente au couple stéréoscopique qui est étudiée. On se détache ainsi complètement de l'algorithme de calcul de la disparité utilisé.

Cependant, il faut être également conscient de l'artifice utilisé. Nous sommes en effet tributaires du formalisme qu'impose l'approche proposée. Une première remarque est que la formulation bayésienne n'est pas seulement une méthode de calcul, mais également une méthode de modélisation, de part la décomposition de la loi *a posteriori* en terme de vraisemblance et terme *a priori*. Ces deux termes ont des statuts radicalement différents. Le terme de vraisemblance décrit la relation entre variable observée et variable à estimer. Il est donc possible de proposer un modèle à partir d'une compréhension et d'une analyse du processus d'observation. Même s'il repose sur certaines hypothèses statistiques (comme la distribution des erreurs d'observation), celles-ci peuvent être confrontées aux données disponibles. La subjectivité du choix est donc réduite. En revanche, elle est totale quand il s'agit de choisir le modèle *a priori*, puisque celui-ci n'est qu'une vue de l'esprit. Or, et cela a été mis en évidence dans les chapitres précédents, le choix du modèle *a priori* a une grande influence sur la forme de la distribution *a posteriori*. La seconde remarque concerne la validité des résultats : il est évident que ceux-ci n'ont de signification que dans le cadre du formalisme et du modèle choisis. Pour une discussion originale de cet aspect du cadre probabiliste, nous renvoyons le lecteur à l'essai de G. Matheron [76].

On comprend dès lors qu'une part importante de ce travail ait été consacré au choix et à l'étude du modèle. Nous nous sommes en partie appuyés sur des hypothèses courantes en vision par ordinateur (modèle *a priori* gaussien et bruit d'acquisition gaussien), et avons proposé un certain nombre d'extensions. En ce qui concerne le modèle *a priori*, nous avons considéré un cadre très général qui est celui des fonctions aléatoires gaussiennes, dans lequel la structure spatiale, décrite par la fonction de covariance, ne dépend pas de la grille de discrétisation

choisie (à l'inverse des modèles de champs de Markov). Ce cadre offre une souplesse certaine, notamment en ce qui concerne les hypothèses de stationnarité (modélisation de la variable ou de ses accroissements) et la diversité des structures spatiales, plus ou moins régulières, dont on peut rendre compte avec un choix approprié du modèle de covariance et de son comportement à l'origine. Enfin, n'étant pas associé à une grille de calcul, il permet des approches multi-échelles ou multi-grilles intéressantes. Nous nous sommes en particulier intéressés aux modèles stationnaires d'ordre 2, dont nous avons mis en évidence les possibilités et les limites. Il apparaît que ces modèles sont bien adaptés à l'étude de portions de couple et permettent d'obtenir des résultats pertinents. En revanche, l'estimation des paramètres ne peut être entièrement automatisée (à cause du choix de la covariance et des paramètres de forme et de portée), et le coût de calcul associé reste très important (plusieurs heures pour une image 512×512). Si l'on veut donc traiter de grands couples stéréoscopiques, une autre approche doit être mise en œuvre. Le modèle de disparité résiduelle représente une solution "pratique", qui peut être mise en œuvre de manière entièrement automatique.

Les trois modèles de vraisemblance que nous avons étudiés permettent de décrire des situations assez variées : bruit d'acquisition de type bruit blanc (GMF-0), existence d'une structure spatiale (GMF-1), existence de valeurs aberrantes (MIXT). Les deux derniers modèles constituent notamment un apport original en comparaison des modèles utilisés couramment en traitement bayésien de la vision stéréoscopique. Le choix parmi ces trois modèles dépend principalement de l'application visée : si l'on désire localiser avec une grande marge de sécurité les points critiques et obtenir des intervalles de confiance raisonnables, le modèle MIXT est bien adapté ; en revanche, si l'on souhaite avoir un bon score de localisation des points erronés, les modèles GMF-0 et GMF-1 sont préférables.

Pour chacun des modèles présentés, nous avons proposé des méthodes d'estimation des paramètres. Deux stratégies peuvent être envisagées. La première consiste à utiliser la carte de disparité estimée par un algorithme de stéréovision (dans notre cas de type corrélation), et à mener une étude variographique pour inférer les paramètres du modèle *a priori*, ou à calculer le maximum de vraisemblance (ou éventuellement des approximations du maximum de vraisemblance), lorsque c'est possible. Cette approche est semi-automatique, mais est assez aisée à mettre en œuvre, sauf dans le cas des paramètres du modèle MIXT pour lesquels une méthode purement heuristique a été proposée. Afin d'éviter les artefacts dus à la carte de disparité estimée, une autre stratégie consiste à randomiser les paramètres du modèle et à les simuler conjointement à la disparité. Cette stratégie repose sur l'échantillonneur de Gibbs et l'utilisation des lois conditionnelles des paramètres connaissant la disparité et le couple stéréoscopique. Cette méthode est plus satisfaisante d'un point de vue théorique, mais il faut cependant veiller à ce que la chaîne de Markov générée atteigne bien son régime stationnaire, notamment pour le paramètre de variance du terme de vraisemblance qui a tendance à décroître sans se stabiliser. Cela peut poser d'importants problèmes pour la simulation de

la disparité.

Concernant le choix du modèle, de nombreux points restent à traiter. Si nous avons mis en avant les limites du modèle stationnaire d'ordre 2, nous n'avons pas été en mesure d'apporter une réelle alternative. Les études préliminaires que nous avons menées ont montré que des modèles intrinsèques de variogramme linéaire n'étaient pas adaptés. On pourrait s'intéresser à d'autres modèles intrinsèques, mais bien souvent, les algorithmes de simulation correspondants restent à développer. Dans l'optique d'un modèle stationnaire avec dérive, une véritable méthodologie de définition et d'estimation de cette dérive reste à mettre en place. Mais peut-être faut-il s'orienter vers des modèles radicalement différents, comme les fonctions aléatoires booléennes [111, 88], qui pourraient permettre de prendre naturellement en compte la forme des structures du relief et d'appréhender la non-stationnarité. Cette réflexion sur le choix du modèle *a priori* s'inscrit également dans la perspective d'appliquer la méthodologie proposée à d'autres types d'image. En effet, les derniers satellites d'observation mis en orbite et ceux à venir offrent des résolutions spatiales de l'ordre du mètre, voire en dessous. A cette résolution, la nature de la scène observée change radicalement. Comme nous l'avons déjà indiqué, des modèles de géométrie stochastique, proches dans l'esprit des fonctions aléatoires booléennes, semblent bien adaptés dans ce cas. Cependant, des méthodes de simulation efficaces de ces modèles doivent être développées.

D'un point de vue algorithmique et statistique, ce travail nous a menés à nous intéresser aux méthodes de simulation spatiale et aux calculs d'intégrale par Monte Carlo. Nous avons évoqué dans un premier temps l'intérêt des schémas d'échantillonnage d'importance pour développer des approches multi-grilles pour la simulation. La réflexion est la suivante : dans les méthodes MCMC, c'est l'étape de simulation de la chaîne de Markov qui est la plus délicate et la plus lourde d'un point de vue calcul. En pratique, on est notamment obligé de sauter un certain nombre d'états successifs de la chaîne pour approcher l'indépendance de l'échantillon. Cela signifie qu'un grand nombre d'états générés ne sert pas *in fine*, ce qui entraîne un coût de calcul important. Ce problème devient plus critique à mesure que la taille du vecteur simulé augmente, d'autant plus que le mélange à l'intérieur de la chaîne a tendance à diminuer avec la taille. L'approche que nous proposons consiste donc à simuler la chaîne de Markov sur un sous-espace de taille réduite, puis à générer pour chaque état retenu l'ensemble du vecteur par simulation conditionnelle "classique". L'utilisation de techniques d'échantillonnage d'importance permet de corriger par la suite l'approximation faite en modifiant convenablement les poids de chaque échantillon dans l'intégrale de Monte Carlo. Cet algorithme a été mis en œuvre dans les applications, mais nous n'avons apporté aucun développement précis, nous contentant en pratique de vérifier que les poids d'échantillonnage d'importance n'étaient pas dégénérés (tous nuls sauf quelques-uns). Il serait donc intéressant de s'intéresser de plus près à ce schéma, qui peut avoir des applications nombreuses en statistique spatiale, et notamment d'évaluer

le gain en terme de réduction du temps de calcul que permet cet algorithme, ainsi que de proposer des critères permettant de choisir de manière optimale le facteur de réduction.

L'autre point important a été l'étude d'algorithmes de simulation par chaînes de Markov. L'apport essentiel réside dans l'extension de l'algorithme de Metropolis-Hastings dans une perspective multi-dimensionnelle. Le principe est de proposer à chaque itération un ensemble d'états, et non un seul, pour la transition. Nous avons développé les bases théoriques reposant sur l'utilisation de noyaux de transition multiples, et montré comment dans le cadre gaussien un algorithme qui reprend le schéma de la sur-relaxation peut être développé. L'étude comparative qui a été menée sur un exemple inspiré de la vision stéréoscopique a mis en évidence l'intérêt de cet algorithme.

Cependant, un grand nombre de points restent à aborder. Tout d'abord, le système de probabilité d'acceptation proposé s'appuie sur une généralisation de la dynamique de Barker. Or, il est connu que pour l'algorithme de Metropolis-Hastings c'est la dynamique de Metropolis qui est la plus performante. Reste donc à étudier comment celle-ci peut être appliquée dans le cadre de l'algorithme Metropolis-Hastings multiple. De plus, si nous avons établi un certain nombre de résultats concernant la convergence de l'algorithme vers la loi cible, aucune analyse théorique de la vitesse de convergence n'a été menée. Il serait intéressant d'étendre les résultats obtenus dans ce domaine pour l'algorithme de Metropolis-Hastings et l'échantillonneur de Gibbs, au moins pour certaines lois particulières. De même, nous n'avons pas étudié, si ce n'est de manière purement empirique, l'influence de la taille p de l'ensemble de proposition. Or, il est clair qu'un certain équilibre doit être établi, permettant de garantir un mélange fort pour un coût de calcul raisonnable. Enfin, nous nous sommes particulièrement intéressés au cas de l'algorithme MMH gaussien que nous avons utilisé dans les applications. Compte tenu du rapprochement qui existe avec l'algorithme de sur-relaxation, une étude comparative s'impose. Une autre direction de recherche est le développement de schémas de transition similaires pour des lois autres que gaussiennes, en utilisant par exemple des propriétés d'invariance de la loi sous certaines transformations.

Enfin, ce travail s'est inscrit dans une problématique technologique pratique et il convient de dégager un certain nombre de pistes concernant la mise en application de la méthode proposée. Les résultats présentés ici montrent que l'approche par Monte Carlo est une réponse au problème posé mais nécessite des temps de calcul très importants. Cette limitation implique donc le choix de modèles stochastiques certes plus restrictifs, mais pouvant être simulés plus rapidement. Le modèle de disparité résiduelle représente à cet égard une solution intéressante. D'un point de vue algorithmique, il reste encore à déterminer si l'approche par champs de Markov pour modéliser la disparité résiduelle est plus efficace. Surtout, il nous reste à appliquer les algorithmes proposés à l'étude systématique de scènes SPOT. Le passage de l'étude d'un extrait de couple au traitement automatique

d'un couple entier est en lui-même un problème important, auquel nous nous intéressons actuellement. De plus, suivant l'application considérée, il est nécessaire de développer les outils les plus pertinents : ainsi, dans le cas d'applications aéronautiques, c'est l'aide au calcul de trajectoires optimales qui constitue un axe de recherche intéressant ; en téléphonie mobile, c'est la position des relais et l'influence d'erreurs dans le MNT sur le modèle de propagation d'ondes électromagnétiques qui importe. Enfin, et l'on rejoint la discussion sur le choix du modèle, l'application à d'autres types de données constitue une perspective intéressante : on pense en particulier aux modèles de ville en 3D calculés à partir d'images aériennes, mais il existe toute une série de problèmes similaires où il s'agit de reconstituer une variable à partir d'une information partielle : reconstitution 3D d'organes en imagerie médicale, détection d'objets à partir d'images radar, reconstitution du sous-sol à partir d'information sismique...

Annexe A

Simulation conditionnelle d'un vecteur gaussien

On s'intéresse ici au problème de la simulation d'une fonction aléatoire gaussienne Z conditionnellement à des valeurs de Z en certains points. La méthode décrite ici, dite de conditionnement par krigeage, comprend deux étapes distinctes : une étape de simulation non conditionnelle, et une étape de conditionnement, qui utilise les équations de krigeage. Nous ne présentons pas ici les méthodes de simulation non conditionnelle, nous renvoyons à [74, 39] pour une description de l'algorithme des bandes tournantes, à [82] pour une approche à l'aide de voisinages glissants et à [84, 22, 20, 23] pour l'utilisation de la transformée de Fourier. Des références plus générales sont [24, 65]. Dans le cas de simulation de fonctions aléatoires gaussiennes sur des grilles régulières en 2 ou 3 dimensions, l'utilisation de la transformée de Fourier rapide permet d'obtenir des algorithmes de simulation très rapides.

Soit une fonction aléatoire gaussienne Z définie sur \mathbb{R}^d et à valeurs dans $F = \mathbb{R}$, de moyenne m et de covariance C . Soit T_1 un ensemble fini de cardinal n_1 de points de \mathbb{R}^d . On désigne par Z_1 la restriction de Z à T_1 .

Supposons que l'on connaisse la valeur de Z en un certain nombre n_2 de points, qui forment un sous-ensemble T_2 (non nécessairement disjoint de T_1). On note alors Z_2 la restriction de Z à T_2 . Le problème est de simuler le vecteur Z_1 conditionnellement à Z_2 . On adopte les notations suivantes : m_1 et m_2 désignent respectivement les vecteurs moyenne de Z_1 et Z_2 , $C_{11} \in \mathbb{R}^{n_1 \times n_1}$ est la matrice de covariance de Z_1 , $C_{22} \in \mathbb{R}^{n_2 \times n_2}$ celle de Z_2 , et $C_{12} = C_{21}^\top \in \mathbb{R}^{n_1 \times n_2}$ est la matrice de covariance entre Z_1 et Z_2 . Il est connu que la loi conditionnelle de Z_1 à Z_2 est une loi gaussienne, dont la moyenne est donnée par les équations de krigeage. Cette moyenne s'écrit :

$$Z_1^* = E(Z_1|Z_2) = m_1 + C_{12}C_{22}^{-1}(Z_2 - m_2) \quad (\text{A.1})$$

et le résidu $Z_1 - Z_1^*$ a pour covariance :

$$C_{11} - C_{12}C_{22}^{-1}C_{12}^T \quad (\text{A.2})$$

De plus, dans le cas gaussien, l'espérance conditionnelle Z_1^* et le résidu $Z_1 - Z_1^*$ sont indépendants. Cette propriété suggère la décomposition suivante :

$$Z_1 = Z_1^* + Z_1 - Z_1^* \quad (\text{A.3})$$

où Z_1 apparaît comme la somme de deux variable indépendantes. Une méthode de simulation consiste ainsi à simuler un résidu $Z_1 - Z_1^*$ et à l'ajouter à l'espérance conditionnelle Z_1^* . L'algorithme, dit de conditionnement par krigeage, est le suivant :

- simuler W de moyenne nulle et de covariance C (simulation sur $T_1 \cup T_2$),
- calculer le krigeage W_1^* à partir de W_2 ,
- ajouter le résidu simulé $W_1 - W_1^*$ au krigeage Z_1^* .

La simulation conditionnelle s'écrit alors :

$$Z_1 = Z_1^* + W_1 - W_1^* \quad (\text{A.4})$$

ou encore :

$$Z_1 = W_1 + C_{12}C_{22}^{-1}(Z_2 - W_2) \quad (\text{A.5})$$

On peut trouver une présentation plus complète de cet algorithme dans [24].

Annexe B

Gibbs par blocs dans le cas gaussien

Nous nous intéressons au schéma de l'échantillonneur de Gibbs dans le cas gaussien. Les notations sont identiques à celles de l'annexe A.

Considérons pour simplifier une partition (T_1, T_2) de S . On envisage donc de relaxer dans un premier temps le bloc T_1 , puis le bloc T_2 . Pour alléger les notations, on indexe par i toute quantité relative au bloc T_i . Le vecteur gaussien à simuler se décompose en $Z = (Z_1, Z_2)$. On le suppose, sans perte de généralité, de moyenne nulle, et on note C sa matrice de covariance, qui se met sous la forme :

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \quad (\text{B.1})$$

avec $C_{21} = C_{12}^\top$. Il est immédiat de montrer que l'inverse Q de C s'écrit :

$$Q = \begin{pmatrix} R_{11}^{-1} & -R_{11}^{-1}\Lambda_{12} \\ -\Lambda_{12}^\top R_{11}^{-1} & C_{22}^{-1} + \Lambda_{12}^\top R_{11}^{-1}\Lambda_{12} \end{pmatrix} \quad (\text{B.2})$$

où les matrices R_{11} et Λ_{12} sont définies comme suit :

$$\Lambda_{12} = C_{12}C_{22}^{-1} \quad (\text{B.3})$$

$$R_{11} = C_{11} - C_{12}C_{22}^{-1}C_{21} \quad (\text{B.4})$$

Λ_{12} est la matrice des poids de krigeage de Z_1 par Z_2 , et R_{11} la matrice de covariance correspondante.

Une itération de l'échantillonneur de Gibbs construit selon les blocs T_1 et T_2 se décompose en deux étapes : la simulation de Z'_1 conditionnellement à Z_2 , puis celle de Z'_2 conditionnellement à Z'_1 . Comme rappelé en annexe A, ces simulations conditionnelles s'écrivent :

$$Z'_1 = \Lambda_{12}Z_2 + W_1 \quad (\text{B.5})$$

$$Z'_2 = \Lambda_{21}Z'_1 + W_2 \quad (\text{B.6})$$

où W_1 et W_2 sont des vecteurs gaussiens indépendants de Z_2 et Z'_1 , de matrices de covariance R_{11} et R_{22} respectivement.

On voit alors qu'une itération complète peut s'écrire sous la forme :

$$Z' = BZ + W \quad (\text{B.7})$$

où B est la matrice de transition suivante :

$$B = \begin{pmatrix} 0 & C_{12}C_{22}^{-1} \\ 0 & C_{21}C_{11}^{-1}C_{12}C_{22}^{-1} \end{pmatrix} \quad (\text{B.8})$$

$$= \begin{pmatrix} 0 & -Q_{11}^{-1}Q_{12} \\ 0 & Q_{22}^{-1}Q_{21}Q_{11}^{-1}Q_{12} \end{pmatrix} \quad (\text{B.9})$$

La décomposition précédente se généralise sans aucune difficulté à un nombre $p \geq 2$ de blocs [97]. Pour cela, introduisons la matrice A telle que :

$$A = I - \begin{pmatrix} Q_{11}^{-1} & 0 & \dots & 0 \\ 0 & Q_{22}^{-1} & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & Q_{pp}^{-1} \end{pmatrix} Q \quad (\text{B.10})$$

On décompose A selon ses matrices triangulaires inférieure et supérieure :

$$A = L + U \quad (\text{B.11})$$

La matrice de transition B s'écrit alors :

$$B = (I - L)^{-1}U \quad (\text{B.12})$$

Finalement, l'échantillonneur de Gibbs peut être vu comme une itération stochastique de matrice de transition B :

$$Z' = BZ + W \quad (\text{B.13})$$

Annexe C

Modèles de Gibbs et champs de Markov

Soit $S = \{1, 2, \dots, n\}$ un ensemble fini de sites, (F, \mathcal{F}) un espace d'états mesurable, muni d'une mesure de référence ν . On note alors $E = F^n$, $\mathcal{E} = \mathcal{F}^{\otimes n}$ et $\mu = \nu^{\otimes n}$. En pratique, F est l'ensemble des réels \mathbb{R} , et \mathcal{F} la tribu des boréliens \mathcal{B} . La valeur d'un vecteur aléatoire définie sur S est notée $z = (z_1, \dots, z_n)$. L'exposé qui suit est largement inspiré de [54].

C.1 Modèles de Gibbs

Soit U une application mesurable de \mathcal{E} dans \mathbb{R} . Un *modèle de Gibbs* sur (E, \mathcal{E}) d'énergie U est une loi π absolument continue par rapport à μ , de densité :

$$\pi(z) = R^{-1} \exp U(z) \tag{C.1}$$

avec la condition

$$R = \int_E \exp U(z) \mu(dz) < +\infty \tag{C.2}$$

Une telle densité $\pi(z)$ est partout strictement positive. A l'inverse, toute densité strictement positive peut s'écrire de cette façon en prenant pour énergie $U(z) = \ln \pi(z)$.

Un modèle de Gibbs est caractérisé par des formes particulières d'énergie U . Soit \mathcal{C} une famille de parties de S , contenant l'ensemble vide et tous les singletons. Une partie T de \mathcal{C} est appelée une *clique*. On note z_T la restriction de z à T , et z^T son complémentaire, de telle sorte que $z = (z_T, z^T)$. On dit que T dérive d'un potentiel $\Phi = \{\Phi_T, T \in \mathcal{C}\}$ si

$$U = \sum_{T \in \mathcal{C}} \Phi_T \tag{C.3}$$

où $\Phi_T : F^T \rightarrow \mathbb{R}$, le potentiel attaché à la clique T , ne dépend que des coordonnées de z dans T : $\Phi_T(z) = \Phi_T(z_T)$. Les modèles intéressants sont ceux pour lesquels \mathcal{C} est différent de l'ensemble des parties de S .

C.2 Lois conditionnelles

A partir d'une famille de cliques \mathcal{C} , on peut définir une relation de voisinage $\langle \cdot, \cdot \rangle$:

$$\text{Pour } i \neq j, (i, j) \in S^2 : \quad \langle i, j \rangle \Leftrightarrow \exists T \in \mathcal{C} \text{ telle que } \{i, j\} \subset T \quad (\text{C.4})$$

Ainsi, i et j sont voisins s'ils appartiennent à une même clique.

On note $\partial T = \{j \in \mathcal{S}, \exists i \in T \text{ t.q. } \langle i, j \rangle\}$ la frontière de voisinage d'une partie T de S . En particulier, ∂i désigne l'ensemble des voisins de i au sens de la relation C.4.

Considérons les lois conditionnelles associées à la loi π d'énergie U définie par la relation C.3 :

$$\pi(z_i | z^i) = \exp \left(\sum_{T \in \mathcal{C}} \Phi_T(z) - \sum_{T \in \mathcal{C}} \Phi_T(z^i) \right) \quad (\text{C.5})$$

$$= \exp \left(\sum_{T \in \mathcal{C}, i \in T} \Phi_T(z) \right) \quad (\text{C.6})$$

$$= \pi(z_i | z_{\partial i}) \quad (\text{C.7})$$

Le modèle de Gibbs vérifie ainsi une propriété markovienne.

C.3 Champs de Markov

Soit une relation de voisinage $\langle \cdot, \cdot \rangle$ définie sur S . Un *champ de Markov* sur E associé à $\langle i, j \rangle$ est une loi telle que pour toute partie T de S , la loi conditionnelle s'écrive :

$$\pi_T(z_T | z^T) = \pi_T(z_T | z_{\partial T}) \quad (\text{C.8})$$

A partir de la relation de voisinage $\langle \cdot, \cdot \rangle$, il est possible de définir une famille de cliques \mathcal{C} : \mathcal{C} contient l'ensemble vide, les singletons et toute partie T d'au moins deux éléments vérifiant :

$$T \in \mathcal{C} \Leftrightarrow \forall i, j \in T^2, \langle i, j \rangle \quad (\text{C.9})$$

La proposition suivante établit l'équivalence entre la définition d'un champ de Markov à partir d'une relation de voisinage et des distributions conditionnelles et la définition d'un modèle de Gibbs à partir d'une famille de cliques et de potentiels.

Proposition 1 *Hammersley-Clifford [7]*

(i) *Tout champ de Gibbs π de famille de cliques \mathcal{C} est un champ de Markov pour la relation de voisinage C.4.*

(ii) *Réciproquement, tout champ de Markov π de relation de voisinage $\langle \cdot, \cdot \rangle$ vérifiant la condition de positivité : $\forall x \in T, \pi(x) > 0$ est un champ de Gibbs pour la famille de cliques \mathcal{C} définie par la relation C.9.*

C.4 Champs de Markov gaussiens

Considérons une loi gaussienne π définie sur S , de moyenne m et de matrice de covariance C . Notons $Q = C^{-1}$ la matrice inverse de C . On peut identifier l'énergie U liée à π :

$$U(z) = -\frac{1}{2}(z - m)^\top Q(z - m) \quad (\text{C.10})$$

La loi gaussienne est un modèle de Gibbs où les cliques ne font intervenir que les singletons et les paires et où les potentiels sont des polynômes :

$$\Phi_i(z) = -\frac{1}{2}q_{ii}z_i^2 + z_i \sum_{j \in S} q_{ij}m_j \quad (\text{C.11})$$

$$\Phi_{ij}(z) = -\frac{1}{2}q_{ij}z_i z_j, \quad i \neq j \quad (\text{C.12})$$

En particulier, les lois conditionnelles π_i s'écrivent :

$$\pi_i(z_i | z^i) \propto \exp \left(-\frac{1}{2}q_{ii}(z_i - m_i + \sum_{j \neq i} \frac{q_{ij}}{q_{ii}}(z_j - m_j))^2 \right) \quad (\text{C.13})$$

Considérons par exemple un champ de Markov gaussien de moyenne nulle sur une grille régulière à deux dimensions $(z_{u,v})$, $1 \leq u \leq n_x$, $1 \leq v \leq n_y$, avec la relation de voisinage $\partial_{uv} = \{(u-1, v), (u+1, v), (u, v-1), (u, v+1)\}$, et la matrice

Annexe D

Le cas gaussien dans Metropolis-Hastings multiple

On reprend l'algorithme exposé dans 4.2.3. Le noyau multiple proposé a pour expression :

$$q(z'_1, \dots, z'_p | z) = \int_w \phi(w) \prod_{j=1}^p \delta_{z \cos \theta_j + w \sin \theta_j}(z'_j) \mu(dw) \quad (\text{D.1})$$

Cherchons une condition suffisante sur les θ_i pour que ce noyau vérifie la relation de multi-réversibilité pour ϕ . Remarquons tout d'abord que l'on a :

$$\begin{aligned} \phi(z)q(z'_1, \dots, z'_p | z) &= \int_w \phi(z)\phi(w) \prod_{j=1}^p \delta_{z \cos \theta_j + w \sin \theta_j}(z'_j) \mu(dw) \\ &= \int_w \int_y \phi(y)\phi(w) \prod_{j=0}^p \delta_{y \cos \theta_j + w \sin \theta_j}(z'_j) \mu(dw) \mu(dy) \end{aligned}$$

Pour i fixé, on choisit alors le changement de variable :

$$\begin{aligned} y' &= y \cos \theta_i + w \sin \theta_i \\ w' &= y \sin \theta_i - w \cos \theta_i \end{aligned}$$

Le jacobien correspondant vaut 1. De plus, il est immédiat de vérifier que :

$$\phi(y)\phi(w) = \phi(y' \cos \theta_i + w' \sin \theta_i)\phi(y' \sin \theta_i - w' \cos \theta_i) = \phi(y')\phi(w')$$

et

$$y \cos \theta_j + w \sin \theta_j = y' \cos(\theta_i - \theta_j) + w' \sin(\theta_i - \theta_j)$$

Après ce changement de variable on obtient :

$$\phi(z)q(z'_1, \dots, z'_p | z) = \int_{w'} \int_{y'} \phi(y')\phi(w') \prod_{j=0}^p \delta_{y' \cos(\theta_i - \theta_j) + w' \sin(\theta_i - \theta_j)}(z'_j) \mu(dw') \mu(dy')$$

Soit :

$$\phi(z)q(z'_1, \dots, z'_p|z) = \int_{w'} \phi(z'_i)\phi(w') \prod_{0 \leq j \leq p, j \neq i} \delta_{z'_i \cos(\theta_i - \theta_j) + w' \sin(\theta_i - \theta_j)}(z'_j) \mu(dw')$$

Pour que la relation de symétrie soit vérifiée, il faut qu'il existe une permutation σ_i telle que :

$$\int_{w'} \phi(z'_i)\phi(w') \prod_{0 \leq j \leq p, j \neq i} \delta_{z'_i \cos(\theta_i - \theta_j) + w' \sin(\theta_i - \theta_j)}(z'_j) \mu(dw') = \quad (D.2)$$

$$\int_{w'} \phi(z'_i)\phi(w') \prod_{j=1}^p \delta_{z'_i \cos(\theta_i) + w' \sin(\theta_i)}(z'_{\sigma_i(j)}) \mu(dw') \quad (D.3)$$

Pour $p = 1$, l'égalité est vérifiée quel que soit θ_1 . Pour $p \geq 2$, une condition suffisante est de prendre

$$\theta_i = i \frac{2\pi}{p+1} \pmod{2\pi} \quad (D.4)$$

Alors, $\theta_i - \theta_j = (i - j) \frac{2\pi}{p+1} \pmod{2\pi}$. La permutation associée est :

$$\sigma_i(j) = i - j \quad \text{si } j < i \quad (D.5)$$

$$\sigma_i(j) = i - j + p + 1 \quad \text{si } j > i \quad (D.6)$$

Pour cette permutation, on a bien :

$$\phi(z)q(z'_1, \dots, z'_p|z) = \phi(z'_i)q(z'_{\sigma_i(1)}, \dots, z'_{\sigma_i(i-1)}, z, z'_{\sigma_i(i+1)}, \dots, z'_{\sigma_i(p)}|z'_i)$$

Bibliographie

- [1] E.P. Baltsavias and D. Stallmann. SPOT stereo matching for digital terrain model generation. In *Proc. of 2nd Swiss Symposium on Pattern Recognition and Computer Vision, Zurich*, pages 61–72, 1993.
- [2] P. Barone and A. Frigessi. Improving stochastic relaxation for Gaussian random fields. *Probability in the Engineering and Informational Sciences*, 4 :369–389, 1990.
- [3] P. Barone, A. Frigessi, and M. Piccioni. *Stochastic Models, Statistical Methods and Algorithms in Image Analysis. Lecture Notes in Statist.* 74. Springer-Verlag Berlin, 1992.
- [4] P. Barone, G. Sebastiani, and J. Stander. Over-relaxation methods and coupled Markov chains for Monte Carlo simulation. *Statistics and Computing*, 12 :17–26, 2002.
- [5] P.N. Belhumeur. *A Bayesian Approach to the Stereo Correspondence Problem*. PhD thesis, Harvard University, 1993.
- [6] P.N. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3) :237–262, 1996.
- [7] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B*, 36 :192–326, 1974.
- [8] J. Besag. On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B*, 48 :259–302, 1986.
- [9] J. Besag. Comments on "Representations of knowledge in complex systems" by U. Grenander and M.I. Miller. *J. Roy. Statist. Soc. Ser. B*, 56, 1994.
- [10] J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems. *Statistical Science*, 10(1) :3–66, 1995.
- [11] J. Besag and P.A.P. Moran. On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika*, 62(3) :555–562, 1975.
- [12] J. Besag, J. York, and A. Mollié. Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43(1) :1–59, 1991.

- [13] S.D. Blostein and T.S. Huang. Error analysis in stereo determination of 3-D point positions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(6) :752–765, 1987.
- [14] S.P. Brooks and G.O. Roberts. Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing*, 8 :319–335, 1998.
- [15] D. Brown and T. Bara. Recognition and reduction of systematic errors in elevation and derivative surfaces from $7\frac{1}{2}$ -minute DEMs. *Photogrammetric Engineering and Remote Sensing*, 60(2) :189–194, February 1994.
- [16] G. Celeux and J. Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Statist. Quater.*, 2 :73–82, 1985.
- [17] B. Chalmond. An iterative Gibbsian technique for the reconstruction of m-ary images. *Pattern Recognition*, 22 :747–761, 1989.
- [18] C. Chang and S. Chatterjee. Multiresolution stereo - A Bayesian approach. In *IEEE International Conference on Pattern Recognition*, volume 1, pages 908–912, 1990.
- [19] J.P. Chilès. How to adapt kriging to non-classical problems : three case-studies. In M. Guarascio, M. David, and C. Huijbregts, editors, *Advanced Geostatistics in the Mining Industry*, pages 69–89. Reidel, Dordrecht, Holland, 1976.
- [20] J.P. Chilès. Quelques méthodes de simulation de fonctions aléatoires intrinsèques. In *Cahiers de Géostatistique*, volume 5, pages 97–112. Ecole des Mines de Paris, 1995.
- [21] J.P. Chilès and P. Chauvet. Le krigeage : méthode de cartographie des fonds marins. *Revue hydrographique internationale*, 52(1) :29–45, Janvier 1975.
- [22] J.P. Chilès and P. Delfiner. Using FFT techniques for simulating Gaussian random fields. In *Proceedings of the Conference on Mining Geostatistics, Berg-en-Dal, South Africa, September 1994*, pages 131–140, 1994.
- [23] J.P. Chilès and P. Delfiner. Discrete exact simulation by the Fourier method. In E.Y. Baafi and N.A. Schofield, editors, *Geostatistics Wallagong '96*, pages 258–269. Kluwer Academic Publishers, 1997.
- [24] J.P. Chilès and P. Delfiner. *Geostatistics : Modeling Spatial Uncertainty*. Wiley and Sons, 1999.
- [25] M.K. Cowles and J.S. Rosenthal. A simulation approach to convergence rates for Markov chain Monte Carlo algorithms. *Statistics and Computing*, 8 :115–124, 1998.
- [26] I.J. Cox. A maximum likelihood n -camera stereo algorithm. In *International Conference on Computer Vision and Pattern Recognition*, pages 733–739, 1994.

- [27] F. Dellaert, S. Seitz, S. Thrun, and C. Thorpe. Feature correspondence : A Markov chain Monte Carlo approach. In *Advances in Neural Information Processing Systems 13*, pages 852–858, 2000.
- [28] C. Delmas. *Distribution du maximum d'un champ aléatoire et applications statistiques*. PhD thesis, Université Paul Sabatier, 2001.
- [29] F. Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, Ecole Polytechnique, 1997.
- [30] U.R. Dhond and J.K. Aggarwal. Structure from stereo - A review. *IEEE Transactions on systems, man and cybernetics*, 19(6) :1489–1510, 1989.
- [31] A. Dick, P. Torr, and R. Cipolla. A Bayesian estimation of building shape using MCMC. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision*, volume 2, pages 852–866. Springer-Verlag Berlin-Heidelberg, 2002.
- [32] J. Diebolt and G. Celeux. Asymptotic properties of a stochastic EM algorithm for estimating mixture proportions. *Stochastic Models*, 9 :599–613, 1993.
- [33] J. Diebolt and E. Ip. Stochastic EM : method and application. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 259–274. Chapman and Hall, 1996.
- [34] O. Faugeras. *Three-Dimensional Computer Vision*. Massachussets Institute of Technology Press, 1993.
- [35] P.F. Fisher. First experiments in viewshed uncertainty : The accuracy of the viewshed area. *Photogrammetric Engineering and Remote Sensing*, 57(10) :1321–1327, October 1991.
- [36] P.F. Fisher. Improved modeling of elevation error with geostatistics. *Geoinformatica*, 2(3) :215–233, 1998.
- [37] G. Foody and P. Atkinson. *Uncertainty in Remote Sensing and GIS*. Wiley and Sons, 2002. to appear.
- [38] D. Forsyth, S. Ioffe, and J. Haddon. Bayesian structure from motion. In *International Conference on Computer Vision*, volume 1, pages 660–665, 1999.
- [39] X. Freulon and C. de Fouquet. Pratique des bandes tournantes à 3D. In C. de Fouquet, editor, *Compte rendu des Journées de Géostatistique*, pages 101–117. Ecole des Mines de Paris, 1991.
- [40] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6 :35–49, 1993.
- [41] A. Galli and H. Gao. Convergence of the Gibbs sampler algorithm in the Gaussian case. *Mathematical Geology*, 32(6) :653–677, 2001.

- [42] L. Garcin, X. Descombes, H. Le Men, and J. Zerubia. Building detection by Markov object processes. In *Proceedings of ICIP'01, Thessaloniki, Greece, 2001*.
- [43] A. Gelman. Inference and monitoring convergence. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 115–130. Chapman and Hall, 1996.
- [44] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.
- [45] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.
- [46] W.R. Gilks and G.O. Roberts. Strategies for improving MCMC. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 89–114. Chapman and Hall, 1996.
- [47] G. Gimel'Farb. Stereo terrain reconstruction by dynamic programming. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications - Volume 2 : Signal Processing and Pattern Recognition*, pages 505–530. Academic Press, 1999.
- [48] M.C. Girard and C. Girard. *Traitement des données de télédétection*. Dunod, 1999.
- [49] P.J. Green. MCMC in image analysis. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 381–400. Chapman and Hall, 1996.
- [50] P.J. Green. A primer on Markov chain Monte Carlo. In O.E. Barndorff-Nielsen, D.R. Cox, and C. Klüppelberg, editors, *Complex Stochastic Systems*, pages 1–62. Chappman and Hall, 2001.
- [51] P.J. Green and X.-L. Han. Metropolis methods, Gaussian proposals, and anthitetic variables. In P. Barone, A. Frigessi, and M. Piccioni, editors, *Stochastic Models, Statistical Methods and Algorithms in Image Analysis. Lecture Notes in Statist.* 74, pages 142–164. Springer-Verlag Berlin, 1992.
- [52] R. Gupta and R.I. Hartley. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9), September 1997.
- [53] X. Guyon. *Random Fields on a Network : Modeling, Statistics and Applications*. Springer, 1995.
- [54] X. Guyon. Méthodes numériques par chaînes de Markov, Université Paris 1, France, Cours, 1999.
- [55] M.J. Hannah. Error detection and correction in Digital Terrain Models. *Photogrammetric Engineering and Remote Sensing*, 47(1) :63–69, january 1981.

- [56] R.I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2) :115–127, november 1999.
- [57] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 :97–109, 1970.
- [58] L.Y. Hu. Gradual deformation and iterative calibration of Gaussian-related stochastic models. *Mathematical Geology*, 32(1) :87–108, 2000.
- [59] L.Y. Hu. Combining dependent realizations within the gradual deformation method. In *Proceedings of the 2001 Annual Conference of the International Association for Mathematical Geology*, 2001.
- [60] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window : Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9), september 1994.
- [61] V. Kolmogorov and R. Zabih. Visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, 2001.
- [62] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision*, volume 3, pages 82–96. Springer-Verlag Berlin-Heidelberg, 2002.
- [63] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision*, volume 3, pages 65–81. Springer-Verlag Berlin-Heidelberg, 2002.
- [64] V. Kratky. Rigorous photogrammetric processing of SPOT images at CCM Canada. *ISPRS Journal of Photogrammetry and Remote Sensing*, 44(2) :53–71, 1989.
- [65] C. Lantuéjoul. *Geostatistical simulation : models and algorithms*. Springer Verlag, Berlin, 2002.
- [66] B. Lapeyre, E. Pardoux, and R. Sentis. *Méthodes de Monte Carlo pour les équations de transport et de diffusion*. Springer-Verlag Berlin, 1998.
- [67] J. Lee, P.K. Snyder, and P.F. Fisher. Modeling the effect of data errors on feature extraction from digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 58(10) :1461–1467, February 1992.
- [68] S.Z. Li. *Markov Random Field Modeling in Computer Visions*. Springer-Verlag New York, 1995.
- [69] Z. Li. On the measure of digital terrain model accuracy. *Photogrammetric Record*, 12(72) :873–877, 1988.
- [70] Z. Li. A comparative study of the accuracy of digital terrain models (DTMs) based on various data models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(1) :2–11, 1994.

- [71] J.L. Lotti and G. Giraudon. Adaptive window algorithm for aerial image stereo. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 701–703. October 1994.
- [72] H. A. Mallot. Stereopsis : Geometrical and global aspects. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications - Volume 2 : Signal Processing and Pattern Recognition*, pages 485–504. Academic Press, 1999.
- [73] G. Matheron. The theory of regionalized variables, and its applications. In *Cahiers du centre de Morphologie Mathématique, Fontainebleau*, volume 5. Ecole des Mines de Paris, 1971.
- [74] G. Matheron. The intrinsic random functions and their applications. *Advances in Applied Probability*, 5 :439–468, 1973.
- [75] G. Matheron. La destructure des hautes teneurs et le krigeage des indicatrices. Technical report, CGMM N-761, Centre de Géostatistique, Ecole des Mines de Paris, 1982.
- [76] G. Matheron. *Estimating and Choosing - An Essay on Probability in Practice*. Springer-Verlag, Berlin, 1989.
- [77] K.L. Mengersen, C.P. Robert, and C. Guhenneuc-Jouyaux. MCMC convergence diagnostics : a "review". In J. Berger, J. Bernardo, A.P. Dawid, and A.F.M. Smiths, editors, *Bayesian Statistics 6*, pages 415–440. Oxford Sciences Publications, 1999.
- [78] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087–1091, 1953.
- [79] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag London, 1993.
- [80] R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11 :125–139, 2001.
- [81] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2) :139–154, 1985.
- [82] D.S. Oliver. Moving averages for Gaussian simulation in two and three dimensions. *Mathematical Geology*, 27(8) :939–960, 1995.
- [83] OS. 1 :50,000 Scale Height Data User Manual. Ordnance Survey, Southampton, 1992.
- [84] E. Pardo-Igúzquiza and M. Chica-Olmo. The Fourier integral method : an efficient spectral method for simulation of random fields. *Mathematical Geology*, 25(2) :177–217, 1993.
- [85] P. Peskun. Optimum Monte Carlo sampling using Markov chains. *Biometrika*, 60 :607–612, 1973.

- [86] L. Polidori. Fractal-based evaluation of relief mapping techniques. In *Joint EC-EARSeL Expert Meeting on Fractals in Geosciences and Remote Sensing, Ispra*, April 1994.
- [87] L. Polidori, J. Chorowicz, and R. Guillaude. Description of terrain as a fractal surface, and application to digital elevation model quality assessment. *Photogrammetric Engineering and Remote Sensing*, 57(10) :1329–1332, October 1991.
- [88] F. Préteux and M. Schmitt. Boolean texture analysis and synthesis. In J. Serra, editor, *Image analysis and Mathematical Morphology : Theoretical Advances*, volume 2, pages 377–400. Academic Press London, 1988.
- [89] M. Le Ravalec-Dupin, L.Y. Hu, and B. Noetinger. Sampling the conditional realization space using the gradual deformation method. In W.J. Kleingeld and D.G. Krige, editors, *Geostats 2000, Cape Town*, 2000.
- [90] L. Renouard. *Restitution automatique du relief à partir de couples d'images SPOT*. PhD thesis, Ecole Polytechnique, 1991.
- [91] B.D. Ripley. *Stochastic Simulation*. John Wileys & Sons, 1987.
- [92] C.P. Robert. Mixtures of distributions : inference and estimation. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 441–464. Chapman and Hall, 1996.
- [93] C.P. Robert. *Discretization and MCMC Convergence Assessment*. Springer-Verlag New York, 1998.
- [94] G.O. Roberts. Markov chain concepts related to sampling algorithms. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 45–57. Chapman and Hall, 1996.
- [95] G.O. Roberts and N.P. Polson. On the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B*, 56(2) :377–384, 1994.
- [96] G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Statist. Soc. Ser. B*, 60 :255–268, 1998.
- [97] G.O. Roberts and S.K. Sahu. Updating schemes, correlation structure, blocking and parametrization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B*, 59(2) :291–317, 1997.
- [98] G.O. Roberts and A.F.M. Smith. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications*, 49 :207–216, 1994.
- [99] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, 2 :341–364, 1996.
- [100] G.O. Roberts and R.L. Tweedie. Geometric convergence and central limit theorems for mutlidimensional Hastings and Metropolis algorithms. *Biometrika*, 83 :95–110, 1996.

- [101] J.J. Rodriguez and J.K. Aggarwal. Stochastic analysis of stereo quantization error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :467–470, 1990.
- [102] D. Scharstein and R. Szeliski. Stereo matching with non-linear diffusion. In *International Conference on Computer Vision and Pattern Recognition*, pages 343–350, 1996.
- [103] D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *International Journal on Computer Vision*, 28(2) :155–174, 1998.
- [104] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithm. *International Journal on Computer Vision*, 47 :7–42, 2002.
- [105] J. S en egas. Geostatistical analysis of stereoscopic pairs of satellite images. Technical report, Centre de g eostatistique, Ecole des Mines de Paris, rapport CG N-14/01/G, February 2001.
- [106] J. S en egas. A Markov chain Monte Carlo approach to stereovision. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *European Conference on Computer Vision*, volume 3, pages 97–111. Springer-Verlag Berlin-Heidelberg, 2002.
- [107] J. S en egas. Spatial simulations with Markov chains. In *8th Annual Conference of the International Association for Mathematical Geology, Berlin, Germany*, volume 1, pages 191–196, 2002.
- [108] J. S en egas, M. Schmitt, and P. Nonin. Geostatistical analysis of stereoscopic pairs of satellite images. In P. Monestiez, D. Allard, and R. Froidevaux, editors, *Proceedings of the Third European Conference on Geostatistics for Environmental Applications*, pages 527–528, 2000.
- [109] J. S en egas, M. Schmitt, and P. Nonin. Conditional simulations applied to uncertainty assessment in DTMs. In G. Foody and P. Atkinson, editors, *Uncertainty in Remote Sensing and GIS*. Wiley and Sons, 2002. to appear.
- [110] J. S en egas, M. Schmitt, and P. Nonin. Monte Carlo methods in stereovision. In H.J. Caulfield, S.H. Cheng, H.D. Cheng, R. Duro, V. Honovar, E.E. Kerre, M. Liu, M.G. Romay, T.K. Smith, D. Ventura, P.P. Wang, and Y. Yang, editors, *Proceedings of the 6th Joint Conference on Information Sciences*, pages 885–889, 2002.
- [111] J. Serra. Boolean random functions. In J. Serra, editor, *Image analysis and Mathematical Morphology : Theoretical Advances*, volume 2, pages 317–342. Academic Press London, 1988.
- [112] R. Stoica, X. Descombes, and J. Zerubia. Road extraction in remote sensed images using a stochastic geometry framework. In *Proceedings of MaxEnt’00, Gif Sur Yvette, July 8-13, 2000*.

- [113] O. Stramer and R.L. Tweedie. Geometric and subgeometric convergence of diffusions with given stationary distributions, and their discretizations. Technical report, University of Iowa and Colorado State University, november 1997.
- [114] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, 1989.
- [115] K. Thapa and J. Bossler. Accuracy of spatial data used in geographic information systems. *Photogrammetric Engineering and Remote Sensing*, 58(6) :835–841, June 1992.
- [116] L. Tierney. Introduction to general state-space Markov chain theory. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 59–74. Chapman and Hall, 1996.
- [117] R.L. Tweedie. Markov chains : Structure and applications. In B. N. Shanbhag and C.R. Rao, editors, *Handbook of Statistics 19*, pages 817–851. Elsevier Amsterdam, 1998.
- [118] U.S.G.S. *Digital Elevation Models, Data Users Guide 5*. U.S. Geological Survey, Reston, Virginia, 1987.
- [119] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [120] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Applications of Mathematics, 27. Springer-Verlag Berlin, 1995.
- [121] A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions I*. Springer-Verlag New York, 1987.
- [122] A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions II*. Springer-Verlag New York, 1987.
- [123] N. Yokoya, K. Yamamoto, and N. Funakubo. Fractal-based analysis and interpolation of 3D natural surface shapes and their application to terrain modeling. *Computer Vision, Graphics, and Image Processing*, 46 :284 – 302, 1989.

This work is concerned with the assessment of uncertainty of depth estimates computed from a stereo system. Computing visual correspondence between homologous points of the stereo pair relies on the assumption of local similarity of the two images and requires the radiometric information to be discriminant. In many situations however, these assumptions (geometric deformations, sensor noise, lack of texture. ...) are violated and the resulting matching errors strongly depend on the nature of the radiometric information embedded in the stereo pair and cannot be entirely controlled by the design of the stereo system.

To tackle the uncertainty problem, we propose a Bayesian framework and apply Markov chain Monte Carlo methods. They consist of sampling from the posterior distribution of the disparity field given the stereoscopic pair and allow the prediction of large errors which occur with possibly low probability while accounting for spatial correlations. Several stochastic models are compared and tested on real data sets (SPOT images), and we give insight on how these models can be extended to other types of images. We also address the problem of the parameter estimation of these models, and propose some algorithms for automatic estimation. Lastly, an important part of the work is concerned with the study of sampling algorithms relying on Markov chain theory. The main achievement is the extension of the Metropolis-Hastings algorithm in a multidimensional way. We give an interesting application based on the Gaussian distribution. We show moreover how importance sampling techniques can be used to efficiently speed up the computation.

Key words : Markov chain Monte Carlo, sampling algorithms, Bayesian computation, inverse problems, stereoscopic vision, digital terrain models.

Cette thèse a pour objet l'étude de l'incertitude attachée à l'estimation de la géométrie d'une scène à partir d'un couple stéréoscopique d'images. La mise en correspondance des points homologues d'un couple suppose la similarité locale des deux images et nécessite une information radiométrique discriminante. Dans de nombreuses situations cependant (déformations géométriques, bruit d'acquisition, manque de contraste, ...), ces hypothèses sont mises en défaut et les erreurs d'appariement qui en résultent dépendent fortement de l'information contenue dans le couple et non du système stéréoscopique lui-même.

Afin d'aborder ce problème, nous proposons un cadre bayésien et l'application de méthodes de Monte Carlo par chaînes de Markov. Celles-ci consistent à simuler la distribution conditionnelle du champ de disparité connaissant le couple stéréoscopique et permettent de déterminer les zones où des erreurs importantes peuvent apparaître avec une probabilité éventuellement faible. Différents modèles stochastiques sont comparés et testés à partir de scènes stéréoscopiques SPOT, et nous donnons quelques pistes pour étendre ces modèles à d'autres types d'images. Nous nous intéressons également au problème de l'estimation des paramètres de ces modèles et proposons un certain nombre d'algorithmes permettant une estimation automatique. Enfin, une part importante du travail est consacrée à l'étude d'algorithmes de simulation reposant sur la théorie des chaînes de Markov. L'apport essentiel réside dans l'extension de l'algorithme de Metropolis-Hastings dans une perspective multi-dimensionnelle. Une application performante reposant sur l'utilisation de la loi gaussienne est donnée. De plus, nous montrons comment le recours à des techniques d'échantillonnage d'importance permet de diminuer efficacement le temps de calcul.

Mots clés : Markov chain Monte Carlo, algorithmes de simulation, calcul bayésien, problèmes inverses, vision stéréoscopique, modèles numériques de terrain