



HAL
open science

Modélisation statistique et segmentation d'images TEP : application à l'hétérogénéité et au suivi de tumeurs

Zacharie Irace

► To cite this version:

Zacharie Irace. Modélisation statistique et segmentation d'images TEP : application à l'hétérogénéité et au suivi de tumeurs. Autre [cs.OH]. Institut National Polytechnique de Toulouse - INPT, 2014. Français. NNT : 2014INPT0086 . tel-04261719

HAL Id: tel-04261719

<https://theses.hal.science/tel-04261719v1>

Submitted on 27 Oct 2023

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.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Image, Information et Hypermédia

Présentée et soutenue par :

M. ZACHARIE IRACE

le mercredi 8 octobre 2014

Titre :

MODELISATION STATISTIQUE ET SEGMENTATION D'IMAGES TEP.
APPLICATION A L'HETEROGENEITE ET AU SUIVI DE TUMEURS.

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. HADJ BATATIA

M. ALAIN AYACHE

Rapporteurs :

M. HICHEM SNOUSSI, UNIVERSITE DE TECHNOLOGIE DE TROYES

Mme FLORENCE FORBES, INRIA GRENOBLE - RHONE ALPES

Membre(s) du jury :

M. JEAN YVES TOURNERET, INP TOULOUSE, Président

M. ALAIN AYACHE, INP TOULOUSE, Membre

M. HADJ BATATIA, INP TOULOUSE, Membre

M. STEVE MC LAUGHLIN, HERIOT WATT UNIVERSITY, Membre

Résumé

Cette thèse étudie le traitement statistique des images TEP. Plus particulièrement, la distribution binomiale négative est proposée pour modéliser l'activité d'une région mono-tissulaire. Cette représentation a l'avantage de pouvoir prendre en compte les variations d'activité biologique (ou hétérogénéité) d'un même tissu. A partir de ces résultats, il est proposé de modéliser la distribution de l'image TEP entière comme un mélange spatialement cohérent de lois binomiales négatives. Des méthodes Bayésiennes sont considérées pour la segmentation d'images TEP et l'estimation conjointe des paramètres du modèle. La cohérence spatiale inhérente aux tissus biologiques est modélisée par un champ aléatoire de Potts-Markov pour représenter la dépendance locale entre les composantes du mélange. Un algorithme original de Monte Carlo par Chaîne de Markov (MCMC) est utilisé, faisant appel aux notions d'échantillonnage dans un espace Riemannien et d'opérateurs proximaux. L'approche proposée est appliquée avec succès à la segmentation de tumeurs en imagerie TEP. Cette méthode est ensuite étendue d'une part en intégrant au processus de segmentation des informations anatomiques acquises par tomodensitométrie (TDM), et d'autre part en traitant une série temporelle d'images correspondant aux différentes phases de respiration. Un modèle de mélange de distributions bivariées binomiale négative - normale est proposé pour représenter les images dynamiques TEP et TDM fusionnées. Un modèle Bayésien hiérarchique a été élaboré comprenant un champ de Potts-Markov à quatre dimensions pour respecter la cohérence spatiale et temporelle des images PET-TDM dynamiques. Le modèle proposé montre une bonne qualité d'ajustement aux données et les résultats de segmentation obtenus sont visuellement en concordance avec les structures anatomiques et permettent la délimitation et le suivi de la tumeur.

Mots-clés : Tomographie par Emission de Positons (TEP), Hétérogénéité, Fusion multimodale, Suivi de tumeurs, Modèle Bayésien, Distribution binomiale négative.

Abstract

This thesis studies statistical image processing of PET images. More specifically, the negative binomial distribution is proposed to model the activity of a single tissue. This representation has the advantage to take into account the variations of biological activity (or heterogeneity) within a single tissue. Based on this, it is proposed to model the data of the entire PET image as a spatially coherent finite mixture of negative binomial distributions. Bayesian methods are considered to jointly perform the segmentation and estimate the model parameters. The inherent spatial coherence of the biological tissue is modeled by a Potts-Markov random field to represent the local dependence between the components of the mixture. An original Markov Chain Monte Carlo (MCMC) algorithm is proposed, based on sampling in a Riemannian space and proximal operators. The proposed approach is successfully applied to the segmentation of tumors in PET imaging. This method is further extended by incorporating anatomical information acquired by computed tomography (CT) and processing a time series of images corresponding to the phases of respiration. A mixture model of bivariate negative binomial - normal distributions is proposed to represent the dynamic PET and CT fused images. A hierarchical Bayesian model was developed including a four dimensional Potts-Markov field to enforce the spatio-temporal coherence of dynamic PET-CT images. The proposed model shows a good fit to the data and the segmentation results obtained are visually consistent with the anatomical structures and allow accurate tumor delineation and tracking.

Keywords : Positron Emission Tomography (PET), Heterogeneity, Multimodal fusion, Tumor tracking, Bayesian model, Negative binomial distribution.

Table des matières

Introduction	xxi
1 Contexte médical	1
1.1 Introduction	1
1.2 Le cancer du thorax et de l'abdomen	1
1.2.1 Statistiques	2
1.2.2 Causes	3
1.2.3 Diagnostic	3
1.2.4 Catégories de cancer pulmonaires	4
1.2.5 Traitements	4
1.3 Facteurs de pronostic/Indicateurs d'avancement du cancer	6
1.3.1 Volume	6
1.3.2 Activité	7
1.3.3 Cinétique de fixation du traceur	7
1.3.4 Hétérogénéité	8
1.4 Modalités d'imagerie non invasive du thorax et de l'abdomen	9
1.4.1 Importance de l'imagerie médicale	9
1.4.2 Imagerie anatomique vs imagerie fonctionnelle	9
1.4.3 Fusion des modalités	11
1.5 La tomographie par émission de positons	11
1.5.1 Principes physiques	13
1.5.2 Reconstruction	18
1.6 La tomodensitométrie	25
1.6.1 Principes physiques	25
1.6.2 Reconstruction	29
1.7 Conclusion	30

2	Modélisation des tumeurs hétérogènes et estimation Bayésienne de paramètres	39
2.1	Introduction	39
2.2	Position du problème	40
2.2.1	Modélisation des données TEP	40
2.2.2	Estimation des paramètres de la DBN	42
2.3	Cadre Bayésien	43
2.3.1	Vraisemblance	44
2.3.2	Loi a priori sur les paramètres du modèle	44
2.3.3	Loi a posteriori	45
2.4	Méthode	47
2.4.1	MALA sur variété	47
2.4.2	MALA Proximal sur variété	52
2.4.3	Algorithme d'échantillonnage	55
2.5	Résultats et discussions	56
2.5.1	Estimation des paramètres sur données synthétiques	56
2.5.2	Modélisation des données TEP	82
3	Segmentation Bayésienne de tumeurs en TEP par mélange de lois binomiales négatives.	97
3.1	Introduction	97
3.2	Position du problème	99
3.3	Modèle Bayésien	100
3.3.1	Vraisemblance	101
3.3.2	A priori sur les paramètres	101
3.3.3	A priori sur les étiquettes	102
3.3.4	Distribution a posteriori de $(\boldsymbol{\theta}, \mathbf{z})$	104
3.4	Echantillonneur de Gibbs hybride	105
3.4.1	Distribution conditionnelle de $p(\boldsymbol{\theta} \mathbf{z}, \mathbf{x})$	105
3.4.2	Distribution conditionnelle de $P[\mathbf{z} \boldsymbol{\theta}, \mathbf{x}]$	106
3.4.3	Algorithme d'échantillonnage	107
3.5	Résultats et discussions	108
3.5.1	Données synthétiques	108
3.5.2	Données TEP réelles	114
3.5.3	Comparaison avec les méthodes de seuillage	123

4	Fusion spatio-temporelle de données TEP-TDM pour le suivi de tumeur	131
4.1	Introduction	131
4.2	Position du problème	133
4.3	Modèle Bayésien hiérarchique	135
4.3.1	Vraisemblance (modèle des données)	135
4.3.2	A priori sur les paramètres du modèle θ	136
4.3.3	A priori sur les étiquettes \mathbf{z}	136
4.3.4	Distribution a posteriori $p(\theta, \mathbf{z} \mathbf{x})$	137
4.4	Échantillonneur de Gibbs Hybride	137
4.4.1	Distribution conditionnelle $p(\theta \mathbf{z}, \mathbf{x})$	137
4.4.2	Distribution conditionnelle $p(\mathbf{z} \theta, \mathbf{x})$	138
4.4.3	Résumé de l'algorithme	139
4.5	Résultats expérimentaux	140
4.5.1	Résultats de la la segmentation	142
4.5.2	Influence du nombre de classes	144
4.5.3	Influence de la granularité	146
4.5.4	Suivi	148
A	Détails des calculs du MALA sur variété pour la loi binomiale négative	165
A.1	MALA	166
A.2	MALA sur variété	167
A.3	Estimation de $E_x [\Psi^2(x + k)]$	168
B	Détails des calculs du MALA sur variété pour la loi bivariée binomiale négative - normale	169
B.1	MALA	170
B.2	MALA sur variété	171
C	Modèle de mélange de distributions Poisson bivariées	175
C.1	Modèle des données / Vraisemblance	175
C.2	Distribution conditionnelle $p(\theta \mathbf{z}, \mathbf{x})$ dans un échantillonneur de Gibbs hybride.	177

Table des figures

1.1	Anatomie des poumons et de l'abdomen [108, 88]	2
1.2	Différentes modalités d'imagerie médicale [101].	10
1.3	Acquisition combinée TEP/TDM [138].	12
1.4	Système combiné TEP/TDM de General Electric [47].	12
1.5	Principes de la TEP [142].	13
1.6	Glucose et 18F-FDG [138]	15
1.7	Désintégration et annihilation [138].	16
1.8	Détecteurs et leur arrangement [36].(a) : Bloc de cristaux couplés à un photomultiplicateur, (b) Schéma de l'organisation des cristaux.	16
1.9	Coïncidences indésirables [98].	17
1.10	Construction du sinogramme [98]	18
1.11	Principe de projection : la mesure $y_\theta(u)$ correspond à l'intégrale de la fonction à reconstruire le long d'une droite orthogonale à l'axe u [101].	19
1.12	Principe des opérations de projection et de rétroprojection à partir de deux angles de vue (0° et 90°) [98].	20
1.13	Représentation schématique d'un tomodensitomètre [138].	26
1.14	Les deux modes d'acquisition : séquentiel à gauche, hélicoïdal à droite [101].	29
2.1	Graphe orienté acyclique pour le modèle Bayésien proposé. Les hyperparamètres fixes sont représentés dans des cases en pointillé.	45
2.2	Forme générale de la log-posterior de la DBN ($\mu = 1 ; k = 1$).	46
2.3	Echantillonnage à réflexion	53
2.4	Schéma progressif-rétrogressif	55
2.5	Trois types de marche aléatoire initialisés au même point ($\mu_0 = 10 ; k_0 = 15$) : RWMH (bleu), MALA (vert) et mMALA (rouge). (Valeur théorique : $\mu = 1 ; k = 1 ; 1000$ itérations)	57

2.6	Evolution du MPSRF le long des premières 11000 iterations pour $N = 50$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).	58
2.7	Evolution du MPSRF le long des premières 11000 iterations pour $N = 100$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).	59
2.8	Evolution du MPSRF le long des premières 11000 iterations pour $N = 500$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).	60
2.9	Evolution du MPSRF le long des premières 11000 iterations pour $N = 1000$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).	61
2.10	Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 50$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.	68
2.11	Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 100$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.	69
2.12	Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 500$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.	70
2.13	Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 1000$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.	71
2.14	Image originale (a-c), partition (d-f) et indice des différentes classes (g-l).	84
2.15	Adéquation aux données pour la loi normale (en vert), la loi Poisson (en jaune) et la loi binomiale négative (en rouge)	87
3.1	Structure de voisinage 2-D du premier ordre (à gauche) et du second ordre (à droite)	103

3.2	Structure de voisinage 3-D du premier ordre (à gauche) et du second ordre (à droite)	103
3.3	Graphe orienté acyclique pour le modèle Bayésien proposé. Les hyperparamètres fixes sont représentés dans des cases en pointillé.	104
3.4	Étiquettes théoriques (a), Valeurs générées (b) et estimations a posteriori des labels pour différentes valeurs de granularité (c-f) pour un mélange de 3 classes.	109
3.5	Histogramme des échantillons générés par les chaînes MCMC après la période de chauffe. Valeurs théoriques $\boldsymbol{\mu} = [1500, 3500, 7000]$ et $\boldsymbol{k} = [7, 11, 6]$	111
3.6	Evolution du PSRF lors du burn-in pour chaque paramètre.	113
3.7	3 coupes consécutives de l'image TEP réelle.	114
3.8	Labels estimés sur les trois coupes ($Z = 6, \beta = 0.9$)	115
3.9	Segmentation obtenue en modélisant les données par différents mélanges ($Z = 6, \beta = 0.9$) pour la coupe 2.	115
3.10	Ajustement aux données des modèles de mélange testés sur une échelle linéaire. En bleu : histogramme des données; en rouge : mélange de DBN; en vert : mélange de distributions Gaussiennes; en jaune : mélange de distributions Poisson.	117
3.11	Résultats de la segmentation selon le nombre de classes recherchées.	120
3.12	Influence de la granularité	122
3.13	3 coupes TEP successives de différents patients et délimitations obtenues par la méthode proposée (en rouge), par une technique de seuillage adaptatif (en vert) et manuellement (en bleu).	126
4.1	4-D neighborhood	136
4.2	(a) Image TEP originale; (b) Image TDM originale; (c) Segmentation sur le canal TEP seul, basée sur un modèle de mélange de distributions binomiales négatives; (d) Segmentation sur le canal TDM seul, basée sur un modèle de mélange de loi normales.	141
4.3	Résultats de la segmentation pour les différents modèles de mélange testés.	143
4.4	Influence du nombre de classes Z recherchées	145
4.5	Influence de la granularité β	147
4.6	Segmentation des 6 bins temporels, correspondant aux différents stades du mouvement respiratoire.	149
4.10	Suivi temporel : Patient A	152
4.11	Suivi temporel : Patient B	153

4.12	Suivi temporel : Patient C	153
4.7	Patient A : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée; vert : segmentation par seuillage adaptatif; bleu : segmentation manuelle. .	154
4.8	Patient B : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée; vert : segmentation par seuillage adaptatif; bleu : segmentation manuelle. .	155
4.9	Patient C : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée; vert : segmentation par seuillage adaptatif; bleu : segmentation manuelle. .	156

Liste des tableaux

2.1	Nombre d'itérations nécessaire pour que $MPSRF < 1.2$ ($N = 50$). Le caractère "-" signifie que le nombre nécessaire d'itérations dépasse 101000.	63
2.2	Nombre d'itérations nécessaire pour que $MPSRF < 1.2$ ($N = 100$). Le caractère "-" signifie que le nombre nécessaire d'itérations dépasse 101000.	64
2.3	Nombre d'itérations nécessaire pour que $MPSRF < 1.2$ ($N = 500$). Le caractère "-" signifie que le nombre nécessaire d'itérations dépasse 101000.	65
2.4	Nombre d'itérations nécessaire pour que $MPSRF < 1.2$ ($N = 1000$). Le caractère "-" signifie que le nombre nécessaire d'itérations dépasse 101000.	66
2.5	Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 50$)	73
2.6	Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 100$)	74
2.7	Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 500$)	75
2.8	Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 1000$)	76
2.9	Nombre moyen d'échantillons indépendants tirés par seconde ($N = 50$)	78
2.10	Nombre moyen d'échantillons indépendants tirés par seconde ($N = 100$)	79
2.11	Nombre moyen d'échantillons indépendants tirés par seconde ($N = 500$)	80
2.12	Nombre moyen d'échantillons indépendants tirés par seconde ($N = 1000$)	81
2.13	Paramètres estimés pour chacun des modèles de distribution : Normale (\mathcal{N}), Poisson (\mathcal{P}) et binomiale negative (\mathcal{BN})	85
2.14	PSRF	86

2.15	Test de Kolmogorov-Smirnov (KS), de Cramer von Mises (CvM) et du χ^2	88
2.16	Critère d'Information d'Akaike (AIC) et Critère d'Information Bayésien(BIC)	89
3.1	Matrice de confusion pour différentes valeurs de β	110
3.2	Estimation des paramètres calculée sur 50 chaînes indépendantes. . .	112
3.3	Tests d'adéquation aux données, obtenus pour différents modèles de mélange	116
3.4	Critère d'Information d'Akaike (AIC) et Critère d'Information Bayésien(BIC)	116
3.5	Estimations MMSE pour chaque paramètre, calculés sur 50 chaînes indépendantes de 5000 itérations.	118
3.6	PSRF atteint à la fin du burn-in, pour chaque paramètre, calculés sur 50 chaînes indépendantes.	119
3.7	Résultats des tests de Kolmogorov-Smirnov et de Cramer von Mises obtenues pour différents nombre de classes Z	121
3.8	Résultats des tests de Kolmogorov-Smirnov et de Cramer von Mises obtenues pour différentes granularité β	123
3.9	Coefficients DICE obtenus en comparant la segmentation manuelle à la segmentation estimée par notre méthode, ainsi qu'à la segmentation obtenue par différents seuillages	124
3.10	Seuil nécessaire pour considérer tous les voxels segmentés par la méthode proposée et seuil adaptatif estimé	125
4.1	Coefficients DICE obtenus en comparant la segmentation manuelle à la segmentation estimée par notre méthode, ainsi qu'à la segmentation obtenue par différents seuillages	151
4.2	Seuil nécessaire pour considérer tous les voxels délimités par la méthode proposée et seuil adaptatif mesuré, et seuil adaptatif estimé.	152

Remerciements

J'aimerais en premier lieu remercier mon directeur de thèse Hadj Batatia pour son encadrement consciencieux. Je le remercie en particulier pour sa disponibilité, son soutien constant et pour les nombreuses discussions que nous avons pu échanger, autant sur les notions scientifiques que sur le moyen de les formuler. J'ai beaucoup apprécié par ailleurs ses conseils avisés en pédagogie et nos discussions passionnantes relatives à l'enseignement. Je remercie également Alain Ayache pour le soutien qu'il a pu m'apporter lorsque cela a été nécessaire. Je tiens à remercier les rapporteurs Florence Forbes et Hichem Snoussi, ainsi que les examinateurs Stephen McLaughlin et Jean-Yves Tourneret pour l'intérêt qu'ils ont porté à mes travaux de thèse et pour leurs remarques avisées. J'aimerais aussi exprimer ma profonde gratitude envers Marcelo Pereyra pour son accompagnement durant ma thèse, pour sa sympathie et pour tout le savoir qu'il m'a transmis. Je remercie grandement le service de médecine nucléaire de l'Institut Claudius Regaud pour m'avoir transmis les données cliniques, et plus spécialement Frédéric Courbon pour ses remarques pertinentes liées à l'aspect médical, et les perspectives intéressantes qu'il propose, ainsi que David Didierlaurent pour son aide généreuse.

Ces remerciements ne seraient pas complets sans l'expression de toute ma sympathie envers l'ensemble de mes collègues de travail, avec une pensée spéciale pour mes compagnons d'aventure Andrea, Mohanad, Facundo et Abdelghafour. J'adresse également un grand merci à Nicolas Dobigeon et une nouvelle fois à Jean-Yves Tourneret, pour leurs bons conseils et le temps qu'ils m'ont consacré. Je tiens par ailleurs à témoigner ma reconnaissance aux équipes enseignantes de la Prépa des INP et de l'IPST-CNAM, pour m'avoir permis de faire mes premiers pas en enseignement dans des conditions optimales. Je souhaite par ailleurs remercier chaleureusement les secrétaires Sylvie Eichen et Sylvie Armengaud dont l'efficacité et la gentillesse ont grandement facilité mes démarches administratives.

Enfin, j'aimerais exprimer ma gratitude envers tous mes amis et envers ma famille pour leurs encouragements et pour leur bonne humeur communicative. Et j'adresse bien évidemment mes derniers remerciements à ma merveilleuse compagne Blanca, qui a partagé avec moi ces années et a su rendre mes journées heureuses en toute occasion. Merci.

Zacharie

Introduction

Positron Emission Tomography (PET) is an imaging modality commonly used in oncology. Its applications include diagnostics, tissue characterization and therapeutic monitoring. The diagnosis consists primarily in assessing the extent and the severity of the tumor. This requires accurate segmentation to delineate and quantify tumors. However, lesions are soft tissues without geometry. Their borders are often difficult to identify. In addition, tissues, and especially tumors, may show variations of activities that can significantly alter the delineation process. It is therefore important to base the segmentation on the statistics of tissues rather than the intensity of activity. This requires a data model appropriate to the imaging modality.

The motivation of this thesis is to propose a new automatic method for segmenting PET images to help clinicians to better delineate and quantify the lesions. A new statistical model is proposed to characterize PET data. This model is based on the image formation process and takes into account the heterogeneity of the tissues. The segmentation requires the estimation of the model parameters, and vice versa. This thesis presents a Bayesian method to solve these two tasks jointly. This work also proposes to incorporate anatomical information during the segmentation in order to obtain a more precise delimitation using Computed Tomography (CT) images. The fusion of PET and CT modalities is only possible if the images are spatially and temporally coherent. To overcome the problems associated with moving tissue due to respiratory movements, the literature suggests partitioning PET and CT images into temporal bins and reconstructing images for every bin. This thesis proposes a segmentation method of bimodal images in a 4-D environment while ensuring both spatial and temporal coherence.

Three main methodological contributions are proposed :

1. A statistical model based on the negative binomial distribution is proposed for single tissues in PET, and a novel method for estimating the parameters is developed.

2. A Bayesian method based on a mixture of negative binomial distributions and a 3-D Potts-Markov random field is developed for segmenting PET images [69, 70].
3. A Bayesian method for segmenting bimodal PET-CT data based on a bivariate mixture of negative binomial - normal distributions and a 4-D Potts field [68].

The thesis is organized into four chapters :

Chapter 1 introduces the medical context that motivates this work, with general notions about chest cancer, PET and CT imaging. The focus is placed on the image formation process to justify the proposed data model.

Chapter 2 examines the statistical distribution of a single, isolated tissue in a PET image. The negative binomial distribution (NBD) is proposed. This model has the advantage of considering the variations of activity within the tissue. However, the estimation of the NBD's parameters is a well known difficult problem. This paper proposes a Bayesian method for the estimation of the parameters of the NBD based on a Markov Chain Monte Carlo (MCMC) technique . The corresponding posterior density presents a correlation between the parameters and a strong anisotropy. Moreover, the restriction of parameters to positive values may hinder the sampling process. To overcome these difficulties, a new sampling technique is proposed combining manifold-MALA and proximal MCMC, leading to a very efficient sampling. The algorithm has been validated on synthetic data. It gives good accuracy and reproducibility when the conventional method of Metropolis-within-Gibbs fails. In addition, the NBD model has been tested on pre-segmented real PET data and shows a good fit supported by excellent KS-tests.

Chapter 3 addresses the problem of segmenting PET images. The delineation performed by this algorithm discriminates tissues with different statistics while taking into account the heterogeneity within the tissues. Data from a PET image with various tissues are described by spatially coherent mixture of negative binomial distributions. The spatial coherence, inherent to biological tissues, is enforced using a Potts-Markov random field, representing the local dependence between the components of the mixture. A hierarchical Bayesian model is developed to account for the dependency between the parameters of the model. The resulting posterior distribution is estimated using an MCMC algorithm that jointly estimates the model parameters and labels of different classes. The efficiency of the algorithm is assessed on synthetic data. Results on PET images produced in real clinical conditions are also used to illustrate the flexibility of the proposed algorithm. We show the different

tissues can be discriminated, even if their biological activities are heterogeneous.

Chapter 4 presents an unsupervised technique for the segmentation of 4-D bimodal PET-CT images. A hierarchical Bayesian model is proposed where the bimodal data are represented by a mixture of bivariate negative binomial and normal distributions. An MCMC algorithm is developed to jointly estimate the labels of each voxel and the parameters of the mixture model. Furthermore, a 4-dimensional Potts-Markov field was integrated to reflect the temporal and spatial coherence of the mixture components. The method has been successfully applied to 4-D PET-CT data from actual patients with lung tumors. The results show that the conducted segmentation allow to segment and track the tumors.

Introduction

La Tomographie par Emission de Positons (TEP) est une modalité d'imagerie couramment utilisée en oncologie. Ses applications incluent le diagnostic, la caractérisation tissulaire et le suivi thérapeutique. Le diagnostic consiste principalement à évaluer la portée et la sévérité de la tumeur. Cela exigerait une segmentation précise permettant d'estimer le volume et les frontières des tumeurs. Cependant, les frontières de la lésion sont souvent difficiles à localiser avec précision. De plus, les tissus, et particulièrement les tumeurs, peuvent présenter des variations d'activités qui peuvent fortement altérer le processus de délimitation. Pour cela, il semble important de baser la segmentation sur les statistiques des tissus plutôt que leur activité mesurée. Le choix du modèle des données doit alors être adapté à la modalité d'imagerie utilisée.

La motivation de cette thèse est de proposer une nouvelle méthode automatique de segmentation d'image TEP afin de permettre aux cliniciens de mieux délimiter les lésions et de mieux les quantifier. Pour cela, un nouveau modèle statistique est proposé pour modéliser les données TEP. Ce modèle est basé sur le processus de formation de l'image et permet de prendre en considération l'hétérogénéité des tissus. Il existe cependant une forte corrélation entre la segmentation et l'estimation des paramètres du modèle, pouvant rendre la résolution du problème difficile. Cette thèse présente un moyen de contourner ce problème en résolvant ces deux tâches de manière conjointe. Ce travail propose également d'intégrer des informations anatomiques lors de la segmentation afin d'obtenir une délimitation plus précise. On considère pour cela les images de tomodensitométrie (TDM ou CT) acquises lors du même examen. La fusion des modalités TEP et TDM n'est possible qu'à condition d'assurer la concordance spatiale et temporelle des images. Pour pallier les problèmes liés aux déplacements des tissus dus aux mouvements respiratoires, la littérature propose de découper les images TEP et TDM en bins temporels. Un autre objectif de cette thèse est alors de permettre de segmenter l'image bimodale dans un environnement 4-D tout en assurant à la fois la cohérence spatiale et la cohérence temporelle du résultat.

Trois contributions méthodologiques principales sont proposées :

1. Un modèle statistique de distribution binomiale négative pour la représentation d'un tissu en TEP, et une méthode robuste d'estimation des paramètres.
2. Une méthode Bayésienne de segmentation d'images TEP fondée sur un modèle de mélange de distributions binomiales négatives et un champ aléatoire de Potts-Markov 3-D [69, 70].
3. Une méthode Bayésienne de segmentation d'images bimodales PET-CT fondé sur un mélange de distributions bivariées binomiales négatives - normales et un champ de Potts 4-D [68].

La thèse est organisée en quatre chapitres :

Le chapitre 1 introduit le contexte médical qui motive ce travail en présentant des notions générales sur le cancer du thorax et les modalités d'imagerie TEP et CT. Le focus est porté sur la formation de l'image afin de justifier le modèle de données proposé.

Le chapitre 2 étudie la distribution statistique d'un tissu unique et isolé dans une image TEP. Le modèle de distribution binomiale négative (DBN) est proposé. Ce modèle a l'avantage de respecter essentiellement les variations de l'activité à l'intérieur des tissus. Hélas, l'estimation des paramètres de la DBN est un problème reconnu difficile. Ce document propose une méthode d'estimation Bayésienne des paramètres de la DBN en utilisant une technique de Monte-Carlo par Chaîne de Markov (MCMC). Cependant, la densité a posteriori présente une corrélation entre ses paramètres et une forte anisotropie. De plus, la restriction des paramètres à des valeurs positives peuvent entraver le processus d'échantillonnage. Afin de surmonter les difficultés de l'échantillonnage, la recherche du maximum a posteriori fait appel à une technique combinant MALA-sur-variété et MCMC proximal. Utilisés conjointement, ces deux méthodes permettent un échantillonnage très performant. L'algorithme a été validé sur des données synthétiques et donne de bons résultats de précision et de reproductibilité lorsque la méthode classique de Metropolis-dans-Gibbs échoue. En outre, le modèle de DBN a été testé sur des données réelles issues d'images TEP pré-segmentées et des résultats à des tests d'adéquation aux données confirment le modèle proposé.

Le chapitre 3 traite le problème de la délimitation des différents tissus des images TEP. La segmentation réalisée par cet algorithme discrimine les tissus présentant des statistiques différentes et prend en considération l'hétérogénéité dans l'activité des

tissus. Pour cela, le processus de segmentation repose sur le modèle des données de DBN présenté dans le chapitre précédent. Les données d'une image TEP présentant divers tissus seront ainsi décrites par mélange spatialement cohérent de distributions binomiales négatives. La cohérence spatiale, inhérente aux tissus biologiques, est modélisée par un champ aléatoire de Potts-Markov représentant la dépendance locale entre les composantes du mélange. Un mécanisme Monte-Carlo par Chaîne de Markov (MCMC) dans un cadre Bayésien est mis en oeuvre pour réaliser conjointement l'estimation des paramètres du modèle et le processus de segmentation. L'efficacité de l'algorithme a été mesurée sur des données synthétiques. De plus, les résultats sur des images TEP réalisées dans des conditions cliniques réelles illustrent la flexibilité de l'algorithme proposé pour la segmentation de tissus même si leur activité biologique est hétérogène.

Le chapitre 4 présente une technique non-supervisée pour la segmentation d'images 4-D bimodales PET-TDM. Un modèle Bayésien hiérarchique est proposé où les données bimodales sont représentées par un mélange de distributions bivariées binomiales négatives - normales. Un algorithme MCMC est développé pour estimer les étiquettes de chaque voxel ainsi que les paramètres du modèle de mélange de manière conjointe. De plus, un champ de Potts-Markov à 4 dimensions a été intégré pour tenir compte de la cohérence spatio-temporelle des composants du mélange. La méthode a été appliquée avec succès sur des images 4-D PET-TDM réelles de plusieurs patients ayant une tumeur au niveau du poumon. Les résultats montrent que la segmentation réalisée permet de segmenter et de suivre la tumeur.

Chapitre 1

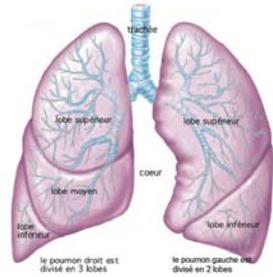
Contexte médical

1.1 Introduction

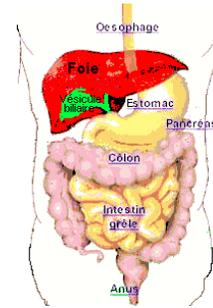
Nous commençons cette thèse par introduire le contexte médical qui motive ce travail. Dans un premier temps, des notions générales sur le cancer du thorax et de l'abdomen sont présentées (section 1.2). Le diagnostic du cancer repose en grande partie sur des indicateurs reflétant le stade du cancer, ils seront présentés dans la section 1.3. Ensuite, nous présentons en section 1.4.2 les différentes modalités d'imagerie médicale utilisées dans les examens du thorax en nous focalisant sur les principes et la terminologie de la Tomographie par Emission de Positons (section 1.5) et de la tomодensitométrie (section 1.6).

1.2 Le cancer du thorax et de l'abdomen

Le corps humain est un organisme vivant complexe. Il est constitué d'un ensemble d'organes faits de cellules assurant des fonctions spécialisées et pouvant interagir de manière à assurer les fonctions vitales. Il arrive cependant que l'ADN de certaines cellules soit perturbé et que ces cellules se mettent à fonctionner de manière anarchique. Il y a alors situation de cancer. Les cellules de l'organisme sont contrôlées en permanence par le système immunitaire, chargé de détecter si des cellules sont infectées par un organisme pathogène ou si elles sont cancéreuses. Lorsque la mutation touche les gènes régissant la croissance et la division des cellules, et si les mécanismes de défense de l'organisme ne réussissent pas à détruire ces cellules cancéreuses, le phénomène peut se propager vers les tissus environnants et former une tumeur. Les cancers représentent la deuxième cause de mortalité en France après les maladies



Anatomie des poumons



Anatomie de l'abdomen

FIGURE 1.1 – Anatomie des poumons et de l'abdomen [108, 88]

cardio-vasculaires.

La forme de cancer la plus incidente est le cancer du poumon [50], qui peut se situer dans la trachée, dans les voies aériennes ou dans le tissu pulmonaire (voir fig. 1.1).

Même si le focus de cette étude est principalement porté sur le cancer du poumon, les moyens et résultats mis en jeu dans cette thèse peuvent s'appliquer plus généralement aux cancers du thorax et de l'abdomen, voire tout autre cancer visible en imagerie TEP.

Le cancer du thorax ne se limite pas au cancer du poumon, des tumeurs peuvent se situer au niveau du médiastin, de la trachée, de la plèvre, de la paroi thoracique ou des voies nerveuses traversant la poitrine (voir fig. 1.1). Ces cancers sont toutefois plus rares et généralement moins graves.

L'abdomen, situé entre le thorax et l'aîne, est également le siège de nombreux cancers. Ils touchent principalement le tube digestif (le côlon ou gros intestin, l'estomac, et beaucoup moins l'intestin grêle), les glandes qui lui sont associées (le pancréas et le foie) et la vésicule biliaire. Ils peuvent aussi toucher les reins ou les ganglions lymphatiques. Chez la femme, une tumeur peut également se développer dans les ovaires ou dans l'utérus. Les cancers de l'abdomen ne sont souvent diagnostiqués qu'à un stade avancé de la maladie d'où un pronostic souvent pessimiste.

1.2.1 Statistiques

Depuis plusieurs décennies, le cancer du poumon est le cancer le plus fréquent dans le monde tant en terme d'incidence que de mortalité. L'OMS estime à 1.8 million le nombre d'incidences en 2012 (soit 12.9% des cancers) et à 1.6 million le nombre de

cas mortels (soit 19.4% des cancers ou presque 1 cancer sur 5) [50]. Chez les hommes, le cancer du poumon représente le cancer à la fois le plus incident et le plus mortel (1.2 million de cas et 1.1 millions de décès)[50]. Chez les femmes, le taux d'incidence est moindre mais reste la deuxième cause la plus fréquente de décès par cancer après le cancer du sein (0.6 millions de cas et 0.5 millions de décès)[50].

En France, l'Institut National du Cancer (INC) chiffre à 39 500 le nombre de nouveaux cas de cancer du poumon diagnostiqués en 2012 dont 28 200 chez les hommes et 11 300 chez les femmes. D'après la même source, environ 21320 hommes et 8 600 femmes en décèderaient, faisant de ce cancer la première cause de décès par cancer tous sexes confondus.

Le pronostic du cancer du poumon est parmi les plus pessimistes des cancers. En 2011, d'après l'Institut de Veille Sanitaire (InVS), le taux de survie des personnes atteintes de ce cancer 5 ans après le diagnostic est de 14% (de 18 % chez les femmes et de 13 % chez les hommes). Ces chiffres n'ont pas connu d'amélioration franche au cours du temps, malgré l'amélioration récente des prises en charge diagnostique et thérapeutique.

1.2.2 Causes

Plusieurs facteurs ont été reconnus comme pouvant être la cause du cancer du poumon. Toujours d'après InVC, le tabagisme (y compris le tabagisme passif) est de loin le principal facteur, responsable de 81 % des décès par cancers du poumon en France. L'exposition au radon, à l'arsenic, à l'amiante ou aux particules émises par les automobiles peut aussi augmenter le risque de cancer du poumon. Des études récentes ont également montré que des facteurs génétiques peuvent également être en cause [52], ou encore une exposition insuffisante à la lumière du soleil et aux UVB [90].

1.2.3 Diagnostic

Le cancer du poumon est d'autant plus alarmant qu'il est souvent découvert tardivement. Plusieurs symptômes, bien que peu expressifs, peuvent toutefois suggérer un cancer du poumon. Parmi ces signes cliniques on retrouve une difficulté à respirer, une toux chronique, des crachements de sang, des douleurs thoraciques ou un amaigrissement du patient.

Lorsque le médecin soupçonne un cancer du poumon, des procédés d'imagerie médicale tels que la bronchoscopie, la tomодensitométrie (TDM) ou la tomographie par émission de positons (TEP) peuvent confirmer son diagnostic. Une biopsie de

la lésion peut également être réalisée au moyen d'une longue aiguille insérée dans la cavité thoracique pour l'examiner.

1.2.4 Catégories de cancer pulmonaires

La grande majorité des cancers du poumon apparaissent au niveau des bronches (Cancer Bronchique Pulmonaire ou CBP) et est due à des états de prolifération maligne de cellules. Ces cancers sont appelés carcinomes et peuvent être divisés en deux catégories, selon la taille et l'aspect des cellules malignes :

les cancers dits *non à petites cellules* (CBNP) sont les plus communs (environ 80 % des cas). Ils se développent lentement et localement et sont ainsi plus facile à détecter et à traiter. Ils peuvent eux-mêmes se distinguer en trois sous-groupes, partageant des pronostics et des traitements semblables : le carcinome épidermoïde (dont l'origine est généralement près d'une bronche majeure), l'adénocarcinome (dont l'origine est généralement dans le tissu pulmonaire périphérique) et le carcinome à grandes cellules indifférenciées.

les cancers dits *à petites cellules* (CBPC) sont moins courants mais représentent la forme la plus dangereuse du cancer du poumon. Contrairement au premier type, le cancer CBPC se développe rapidement et peut se propager très précocement en métastases à d'autres parties du corps. En effet, le poumon est une zone où le contact avec les vaisseaux sanguins et lymphatiques est étroit et le risque que des métastases se soient déjà propagées au moment du diagnostic est élevé.

Cette distinction, basée sur des critères histologiques, a des implications importantes sur le traitement à suivre.

1.2.5 Traitements

La médecine dispose actuellement de trois types de traitement principaux contre le cancer : la chirurgie, la chimiothérapie et la radiothérapie, ces stratégies pouvant être associées. Le traitement du cancer dépend du type exact de cellule de la tumeur, du stade du cancer, et de l'état général de santé du patient.

La chirurgie

Le traitement chirurgical est le traitement curatif principal mais n'est efficace que si les tumeurs peuvent être enlevées entièrement. La chirurgie est donc parti-

culièrement intéressante sur les carcinomes non à petites cellules, circonscrite à un seul poumon et dont le volume n'est pas trop important.

On a souvent recours à l'imagerie médicale, plus précisément au scanner ou à la tomographie par émission de positons (TEP), pour déterminer si le cancer est localisé et possible à traiter par chirurgie.

Lorsque c'est le cas, c'est durant l'intervention chirurgicale elle-même que le médecin décide de l'étendue de l'exérèse pulmonaire. Si la réserve respiratoire du patient est insuffisante, la résection cunéiforme (résection d'une partie d'un lobe) sera souvent préférée. Si la réserve respiratoire le permet, la lobectomie (résection d'un lobe) ou la pneumonectomie (poumon entier) pourront être envisagés car le risque d'une récurrence locale est minimisé.

L'opération chirurgicale en elle-même est particulièrement risquée. C'est la raison pour laquelle certaines conditions doivent être respectées. Le patient ne doit présenter en particulier aucun trouble cardiaque ni pulmonaire grave. Afin d'éviter une chirurgie inutile, il est également important qu'aucun signe de métastases ne soit présenté.

Toutefois, même si le traitement chirurgical permet d'éliminer 10 % à 35 % des tumeurs cancéreuses du poumon, il ne guérit pas tous les cancer. Ainsi, le carcinome à petites cellules répond généralement mieux à la chimiothérapie ou à la radiothérapie.

La chimiothérapie

La chimiothérapie est un traitement fondé sur la prise de médicaments anticancéreux comme le cisplatine ou l'étoposide. Elle est particulièrement utilisée pour traiter le carcinome du poumon à petites cellules ou pour un cancer du poumon métastatique.

La radiothérapie

La radiothérapie est un traitement du cancer utilisant des radiations pour détruire les cellules cancéreuses tout en épargnant les tissus sains périphériques.

Elle est souvent utilisée dans le cas de carcinomes non à petites cellules qui ne sont pas susceptibles d'être opérés chirurgicalement (car siégeant trop près de la trachée par exemple). L'irradiation de la poitrine peut également être recommandée pour le carcinome à petites cellules en addition avec la chimiothérapie. Dans le cas d'un cancer non potentiellement curable, la radiothérapie permet de ralentir voir d'arrêter la croissance de la tumeur.

On distingue plusieurs techniques de radiothérapie, la plus utilisée étant la radiothérapie externe où la source de rayonnement est à l'extérieur du patient. La

curiethérapie quant à elle consiste en une source radioactive placée au contact direct avec la tumeur.

Dans tous les cas, une information précise sur la localisation et la taille des tumeurs est primordiale. On fait appel pour cela à des techniques d'imagerie médicale telles que la tomодensitométrie ou le tomographie par émission de positons.

1.3 Facteurs de pronostic/Indicateurs d'avancement du cancer

Afin d'assurer une prise en charge optimale du patient, il est important de définir des indicateurs permettant d'extraire des caractéristiques de la tumeur et d'en déduire la meilleure stratégie à adopter pour les traiter. Ces indicateurs sont qualitatifs ou quantitatifs et sont souvent estimés à partir des images acquises par tomодensitométrie (TDM) ou par tomographie par émission de positons (TEP).

Bien que le diagnostic repose en premier abord sur l'inspection visuelle des images (qualitative et souvent subjective), il est important de considérer également des indicateurs quantitatifs. Ceux-ci rendent possible les comparaisons entre patients différents et le suivi thérapeutique du même patient au cours du temps.

Certains de ces paramètres quantitatifs sont présentés dans la suite de cette section.

1.3.1 Volume

Un des principaux facteurs relevé pour quantifier l'évolution d'un cancer est la taille des tumeurs. La tomодensitométrie permet une définition précise du volume cible lorsque la tumeur est visible. On parle alors de *gross tumor volume* (GTV), très important en balistique de radiothérapie

On peut aussi considérer l'*internal tumor volume* (ITV) qui inclut également les mouvements physiologiques de la tumeur.

On s'intéresse également au volume mesuré sur les images TEP, appelé *volume biologique tumoral* ou *biological tumor volume* (BTV). Ce dernier indicateur se montre pertinent pour améliorer le pronostic, pour préciser la balistique en radiothérapie ou pour évaluer la réponse à un traitement

La délimitation des tumeurs est un problème difficile, particulièrement en TEP, où de nombreux facteurs détériorent l'image comme l'effet de volume partiel, le faible rapport signal sur bruit, ou même l'hétérogénéité des tumeurs. En clinique,

la délimitation de ces volumes repose en général sur un seuillage adaptatif en pourcentage de l'activité maximale mesurée, quand elle n'est pas manuelle. De nombreuses méthodes de segmentation automatique ont toutefois été proposées et seront détaillées en 3.

1.3.2 Activité

Avec le volume des tumeurs, l'indicateur le plus utilisé est l'activité, ou l'intensité de fixation des tumeurs. Celle-ci est mesurée sur les images TEP où la valeur de chaque voxel correspond à l'intensité radioactivité du traceur par unité de volume (en kBq.mL^{-1}). L'activité mesurée directement dépend toutefois de nombreux paramètres, tels que la quantité de radiotracer injectée, la voie d'administration, la corpulence du patient, etc.

Afin de permettre la comparaison d'exams entre plusieurs patients, l'activité volumique est normalisée en fonction de l'activité injectée et du poids du patient. Le nouveau paramètre est appelé *standardized uptake value* (SUV), et s'exprime en g.mL^{-1} . Il est obtenu selon la formule suivante :

$$SUV = \text{poids du patient} \times \frac{\text{Activité mesurée}}{\text{Activité injectée}} \quad (1.1)$$

Des études ont alors montré que l'activité pic SUV_{\max} était corrélée au pronostic de survie [80, 10], et pouvait être un indicateur pertinent dans le cadre de l'évaluation d'un traitement [63].

L'indicateur d'activité maximum a l'avantage de ne pas être altéré par l'effet de volume partiel, mais diverses études ont mis en avant les limites du SUV comme mesure quantitative pour le diagnostic [75, 17].

1.3.3 Cinétique de fixation du traceur

Des méthodes plus récentes s'intéressent également à la cinétique de fixation du traceur en TEP dynamique (4-D) [97, 81]. Ces indicateurs sont déduits des courbes de temps-activité (ou *Time Activity Curves* (TAC)) et fournissent des informations quant à l'évolution temporelle de l'activité tumorale durant le temps d'acquisition. Ces paramètres sont surtout utilisés pour l'évaluation de la réponse à un traitement [63]. Ces indices sont très prometteurs, mais difficile à mettre en œuvre car le seul problème de la reconstruction 4D est toujours une tâche difficile et n'est généralement pas utilisé dans la pratique clinique.

1.3.4 Hétérogénéité

Il a été rapporté que les images TEP présentent souvent des variations dans l'absorption du traceur au sein d'une même tumeur [61].

Cette hétérogénéité peut être due à une nécrose, à la prolifération cellulaire, au flux sanguin, à la densité des microvaisseaux ou à l'hypoxie [135]. L'hétérogénéité dans l'activité de la tumeur est liée à l'hétérogénéité dans son métabolisme et peut être considérée comme un indicateur à part entière.

Cet indicateur peut être intéressant dans diverses applications de l'oncologie. Premièrement, l'hétérogénéité est très utile pour l'évaluation du diagnostic et du pronostic [39, 76]. Au moment du pronostic, la pertinence d'un indicateur d'hétérogénéité a été soulignée pour plusieurs affections malignes, parmi lesquelles des systèmes de gradation se basent directement sur cette caractéristique [8].

L'indice d'hétérogénéité est très important dans l'évaluation de réponse aux traitements [99, 39]. Selon Basu et al. [8], l'utilité majeure de l'étude de l'hétérogénéité des tumeurs serait son implication sur les résultats d'un traitement thérapeutiques car elle permettrait d'identifier une réponse multiple pour une même tumeur. En d'autres termes ils proposent que l'hétérogénéité des tumeurs est invariante à l'échelle, c'est à dire que l'hétérogénéité observée en imagerie soit le reflet de la variabilité génétique et cellulaire de la tumeur qui est elle-même en lien direct avec son comportement face à la thérapie. Cette invariance d'échelle expliquerait pourquoi un cancer où les tumeurs sont hétérogènes ne peut être guéri par une unique modalité de traitement et devrait faire l'objet d'une combinaison de régimes de chimiothérapie.

Enfin, l'hétérogénéité des tumeurs est devenue un indicateur intéressant dans le domaine de la radiothérapie guidée par l'image, où une carte de l'activité précise de la tumeur est nécessaire pour concentrer les radiations sur les sous-volumes où la fonction biologique est la plus accrue, et permet donc de mieux épargner les tissus sains voisins [83, 132, 21].

Malgré son importance prometteuse, seules quelques études récentes ont présenté des méthodes de quantification de l'hétérogénéité intra-tumorale.

Une méthode basique a été proposée où la distribution intra-tumorale de l'absorption du FDG est comparée à un objet solide elliptique avec une densité homogène [99, 100]. Récemment, une méthode basée sur le calcul de l'aire sous la courbe des histogrammes cumulés SUV-volume (AUC -CSH) ou intensité-volume (ASC-HVA) a été proposée pour quantifier l'hétérogénéité [41, 135].

1.4 Modalités d'imagerie non invasive du thorax et de l'abdomen

1.4.1 Importance de l'imagerie médicale

L'imagerie médicale joue un rôle primordial durant toute la durée de la prise en charge du patient.

Pour le diagnostic : Les cancers du thorax et de l'abdomen sont souvent détectés par radiographie thoracique ou par scanner. Lorsqu'une biopsie est réalisée, elle est généralement réalisée par bronchoscopie ou guidée par scanner.

Pour le bilan d'extension et le choix du traitement : La place de l'imagerie TEP/TDM pour la caractérisation des nodules isolés, pour le bilan d'extension ganglionnaire et métastatique est important afin de choisir le meilleur traitement.

Lors du traitement : L'imagerie médicale se montre utile quelque soit le traitement choisi. En chirurgie et en radiothérapie, on utilise généralement le scanner TDM ou TEP pour localiser avec précision la tumeur à réséquer ou la zone à irradier. L'imagerie médicale a également un intérêt majeur pour évaluer la réponse aux traitements lors d'une chimiothérapie.

Pour le suivi : Même si le traitement a été efficace dans un premier temps, une récurrence du cancer est fréquente dans les mois ou les années qui suivent. Le patient guéri devra subir des examens de suivi réguliers durant lesquels il est nécessaire que des examens d'imagerie non invasifs soient réalisés.

1.4.2 Imagerie anatomique vs imagerie fonctionnelle

Pour étudier le thorax et l'abdomen, deux principaux types d'imagerie peuvent être utilisés selon l'information prospectée : l'imagerie anatomique et l'imagerie fonctionnelle.

L'imagerie anatomique

Les modalités anatomiques (ou structurelles) fournissent une information géométrique détaillée des tissus et structures des organes. Pour le thorax, on a souvent recours à l'imagerie par résonance magnétique (IRM), à l'imagerie ultrasons ou à la tomodesitométrie (TDM).

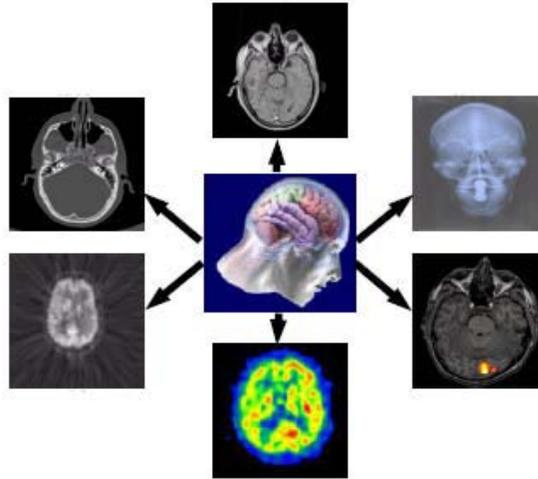


FIGURE 1.2 – Différentes modalités d'imagerie médicale [101].

L'imagerie par résonance magnétique (IRM) est une modalité permettant la visualisation des tissus mous avec un très bon contraste. Elle a l'avantage d'être non invasive et de ne soumettre le patient à aucune radiation.

L'imagerie ultrasonore (US) a l'avantage de pouvoir fournir des images en temps réel et en fait donc un outil de choix pour imager des organes en mouvement comme le coeur. Elle est souvent préférée également pour son faible coût et sa facilité d'utilisation, mais les images qu'elle fournit sont fortement caractérisées par le bruit de speckle.

La tomodensitométrie (TDM) offre une représentation visuelle très confortable du volume exploré et permet de déceler les infections, les hémorragies et les tumeurs. Quand le volume des tumeurs le permet, la TDM peut fournir des informations précises quant aux caractéristiques géométriques des tumeurs telles que leur emplacement, leur taille, leur volume ou leur forme.

L'imagerie fonctionnelle

L'imagerie fonctionnelle permet, quant à elle, de fournir des informations relatives au fonctionnement des organes. En oncologie, la tomographie par émission de positons (TEP) est une modalité courante. Elle consiste à mesurer la répartition

spatiale d'un traceur radioactif préalablement injecté dans le corps du patient. Les imageurs TEP ne procurent pas des images de très bonne qualité visuelle mais leur avantage est de visualiser les tumeurs à un stade précoce, bien avant qu'elles soient visibles en imagerie structurale [78]. Le diagnostic et le suivi de cancers pulmonaires utilisent souvent cette modalité.

1.4.3 Fusion des modalités

Il est à noter que la prise en charge du patient nécessite souvent l'utilisation combinée de ces deux types de modalité complémentaires. Depuis l'essor des technologies mixtes, de nombreuses études se sont intéressées au couplage de ces deux types d'imagerie qui, une fois fusionnés correctement, permettrait au praticien d'établir un diagnostic plus précis (voir fig. 1.3 et 1.4). Plus de détails seront donnés à cet égard dans la suite de cette thèse, lorsqu'une technique de segmentation d'images bimodales TEP/TDM sera présentée en chapitre 4.

En oncologie thoracique et abdominale, les modalités d'imagerie les plus répandues sont la tomographie par émission de positons et la tomodensitométrie. Ces deux modalités reposent sur le principe de la tomographie où le volume à étudier est projeté selon de multiples directions à l'aide d'un phénomène physique de radiation. Les informations mesurées représentent alors les intégrales-ligne du volume observé. L'image est ensuite reconstruite à partir de ces projections en résolvant un problème inverse.

Les sections suivantes présentent en détail ces deux modalités d'imagerie, en se focalisant sur les processus physiques sous-jacent à chacune des modalités et la formation de l'image.

1.5 La tomographie par émission de positons

La tomographie par émission de positons (TEP) est une modalité d'imagerie fonctionnelle de médecine nucléaire permettant de visualiser le métabolisme de l'organisme. Un traceur radioactif est injecté dans l'organisme et va se désintégrer en émettant des particules. Ces particules seront recueillies et exploitées afin de reconstruire une image tridimensionnelle correspondant à la distribution spatiale de la fonction examinée. La fonction examinée (par exemple la consommation de glucose, ou d'oxygène) dépend du produit radiopharmaceutique injecté. Ces informations s'avèrent très utiles pour le diagnostic de pathologies car des changements fonctionnels de l'organisme précèdent souvent les changements anatomiques.

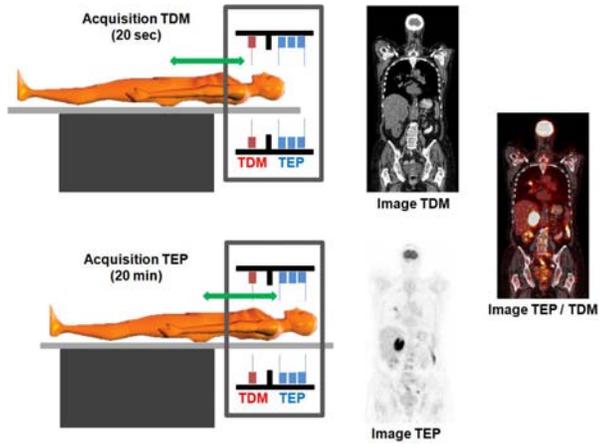


FIGURE 1.3 – Acquisition combinée TEP/TDM [138].

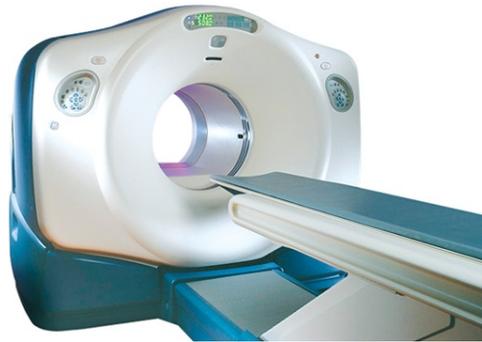


FIGURE 1.4 – Système combiné TEP/TDM de General Electric [47].

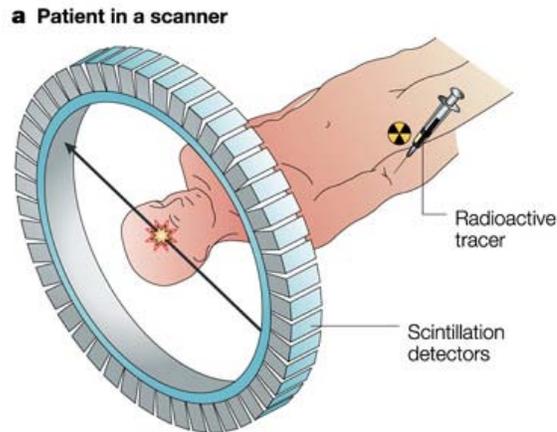


FIGURE 1.5 – Principes de la TEP [142].

Les applications de la TEP sont vastes. La TEP est avant tout un outil courant en oncologie, pour le diagnostic, la caractérisation tumorale et le suivi thérapeutique. Elle est aussi utilisée par des industries pharmaceutiques ou des instituts de recherche biomédicale pour étudier de nouvelles stratégies thérapeutiques et évaluer de nouveaux traitements. Des travaux récents se sont intéressés également à l'imagerie TEP dans le domaine de la radiothérapie guidée par imagerie fonctionnelle. Les applications de la TEP s'étendent également en cardiologie ou en encore en neurologie.

Nous commençons par présenter les principes physiques de la TEP, puis les brièvement les différentes méthodes de reconstruction. Nous expliquons enfin comment les images TEP peuvent être modélisées statistiquement.

1.5.1 Principes physiques

Un traceur contenant un isotope radioactif est injecté dans le corps du patient et se répartit spatialement selon la fonction ciblée. Cette substance radioactive émet des positons qui s'annihilent avec des électrons en libérant à chaque fois deux photons γ de manière coaxiale mais dans des directions opposées. Ces photons sont ensuite détectés par des capteurs et enregistrés pour pouvoir être utilisés dans un algorithme de reconstruction d'image. Cette section détaille chacune de ces étapes.

Isotopes émetteurs de positons et choix du radiotracteur

Comme pour les autres examens de médecine nucléaire, le protocole d'acquisition des images TEP nécessite l'injection d'un traceur radioactif, ou radiotracteur. Un radiotracteur résulte d'un couplage chimique entre un vecteur moléculaire ciblant le phénomène physiologique d'intérêt et d'un isotope radioactif émetteur de positons, permettant de localiser la distribution de la molécule au sein de l'organisme.

L'isotope radioactif est un élément instable caractérisé par sa demi-vie, ou période radioactive. Le choix de l'isotope radioactif doit être compatible avec les contraintes cliniques. Cette durée doit donc être suffisamment courte pour limiter la dose d'irradiation chez le patient, mais aussi suffisamment longue pour permettre un diagnostic complet. Pour palier à ces problèmes, la production des isotopes radioactifs est souvent réalisée de manière artificielle à proximité des hôpitaux, par bombardement de particules grâce à un cyclotron. Le fluor 18 (^{18}F) est un isotope radioactif particulièrement adapté aux contraintes liées à un usage en routine clinique. Sa demi-vie est égale à 110 minutes. Notez cependant que d'autres éléments radioactifs peuvent être utilisés en TEP, tels que le carbone 11 (^{11}C), le gallium 68 (^{68}Ga) ou le cuivre 64 (^{64}Cu).

Lorsque l'on intègre le fluor 18 dans une molécule de glucose, en remplacement d'un couple hydroxyle (OH), on obtient du fluorodeoxyglucose (^{18}F -FDG) (voir fig. 1.6). En clinique, ce radiotracteur est le plus utilisé en imagerie TEP. Il permet de mesurer le taux de consommation de glucose. Or, pour certains types de cancers, il a été établi que la prolifération cellulaire nécessite une consommation anormalement élevée de glucose. Cette corrélation rend donc les tissus cancéreux susceptibles de fixer le ^{18}F -FDG [140]. Il est important de noter que le ^{18}F -FDG ne se fixe pas uniquement sur les cellules cancéreuses. Certaines fixations sont d'origine physiologiques et ne reflètent pas de profil pathologique. La concentration de glucose est en effet très importante dans le cerveau et le cœur, également dans le foie de manière moins intense et parfois dans les muscles et le tube digestif. Le ^{18}F -FDG est également très concentré au niveau des reins et de la vessie. Enfin, il peut être fixé dans des tissus inflammatoires, qu'ils soient d'origine cancéreuse ou pas. À l'inverse, il arrive toutefois que certaines tumeurs ne fixent pas le ^{18}F -FDG.

Désintégration et annihilation

Désintégration L'isotope radioactif est un élément instable. Le nombre de noyaux diminue avec le temps selon une décroissance exponentielle :

$$N(t) = N_0 e^{-\lambda t} \quad (1.2)$$

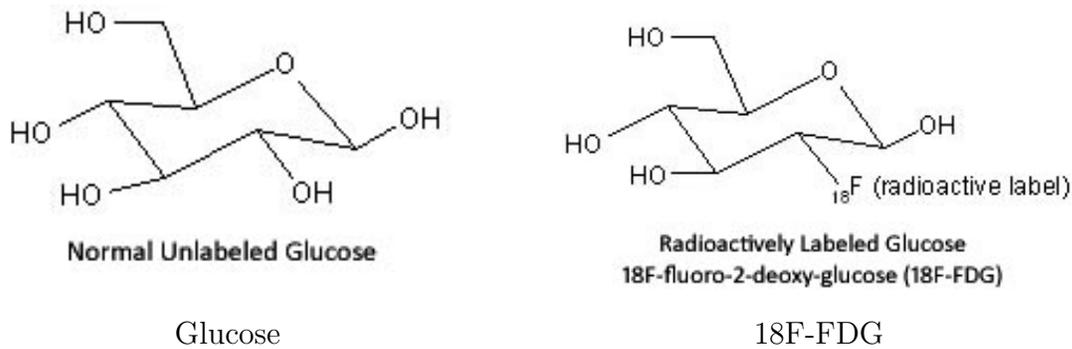


FIGURE 1.6 – Glucose et 18F-FDG [138]

où $N(t)$ représente le nombre de radionucléides restant au temps t , N_0 est le nombre initial de noyaux non-désintégrés et λ est la constante radioactive de l'élément.

Dans le cas du fluor 18, la désintégration est de type β^+ . Un proton est converti en neutron et à l'issue de cette transformation sont émis un positon e^+ ainsi qu'un neutrino ν_e



Ainsi le fluor 18 possédant 9 protons et 9 neutrons se désintègre en l'isotope stable de l'oxygène ${}^{18}_8\text{O}$ possédant 8 protons et 10 neutrons. Un positon est libéré de cette transformation.

Annihilation A l'issue de la désintégration, le positon parcourt une distance de 1 à 3 mm avant de perdre son énergie cinétique et d'interagir avec un électron. Le positon étant de l'anti-matière, l'interaction entre ces deux anti-particules provoque une annihilation et conduit à l'émission de deux photons γ :



Ces deux photons γ de 511 keV sont émis colinéairement dans des sens opposés (voir fig. 1.7). Leur direction est isotrope.

Il est intéressant de noter à ce point que le nombre d'annihilations apparues dans une période de temps fixée peut être représenté par une loi de Poisson. En effet, les processus de Poisson et exponentiels sont liés par le fait que dans un processus de Poisson, le temps d'attente entre deux événements indépendants suit une loi exponentielle. Ainsi, en considérant l'équation (1.2), on peut en déduire que la probabilité que le nombre d'annihilations vaille n sachant que le nombre moyen d'annihilation vaut

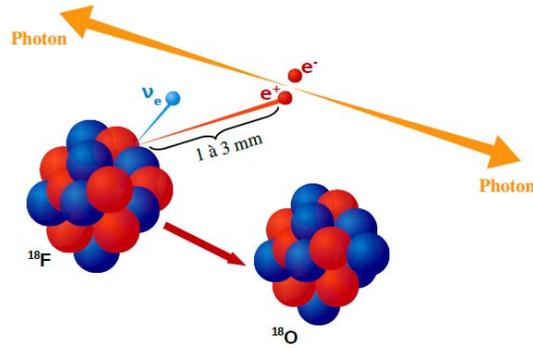


FIGURE 1.7 – Désintégration et annihilation [138].

\bar{n} peut être exprimée ainsi :

$$P[n] = \frac{\bar{n}^n \exp(-\bar{n})}{n!}. \quad (1.5)$$

Détection et organisation des données

Détection de coïncidences Les deux rayons γ atteignent ensuite chacun un détecteur du dispositif, placés sur des anneaux concentriques entourant le patient. Ces détecteurs sont composés de cristaux scintillateurs couplés à des photo-multiplicateurs émettant un signal électrique à chaque détection (voir fig. 1.8).

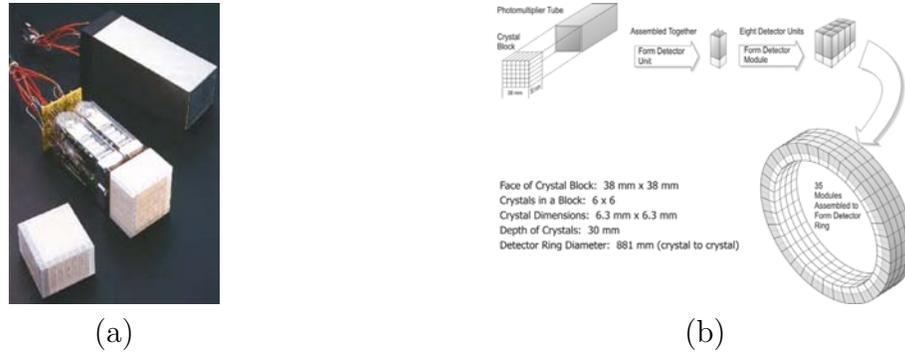


FIGURE 1.8 – Détecteurs et leur arrangement [36]. (a) : Bloc de cristaux couplés à un photomultiplicateur, (b) Schéma de l'organisation des cristaux.

On appelle coïncidence un couple de détections respectant des contraintes temporelles et énergétiques. Plus précisément, la détection des deux photons provenant de la même annihilation doit être faite dans un délai temporel relativement court (de l'ordre de 10ns). De plus, ne sont considérés que les photons dont l'énergie est proche des 511 keV initiaux. En effet, les photons dont l'énergie est faible au moment de leur détection sont ceux qui ont beaucoup interagit avec les tissus et ont plus de risque d'avoir été déviés. Ces coïncidences définissent ainsi une ligne de réponse (*Line Of Response*, LOR) reliant les deux détecteurs atteints par chacun des photons. Chacune des LOR obtenues a la caractéristique de passer par le lieu où s'est produit une annihilation. Ce sont par la suite ces informations qui seront utilisées pour reconstruire l'image.

Il arrive toutefois que certaines coïncidences détectées ne correspondent pas à une annihilation sur la LOR en question (voir fig. 1.9).

Malgré la considération d'une fenêtre énergétique, il arrive que des coïncidences, dites *diffusées* soient prises en compte alors qu'au moins un des deux photons a subi une déviation avant sa détection. La modification de sa trajectoire peut être due à une interaction avec un électron du milieu appelée *effet Compton* [29]. D'autre part, il arrive que deux annihilations se produisent dans un laps de temps très courts et qu'elles soient détectées dans la même fenêtre temporelle. Ces coïncidences sont dites *fortuites*. Ces deux types de coïncidences indésirables auront pour effet de considérer des LOR ne passant pas par le lieu de l'annihilation.

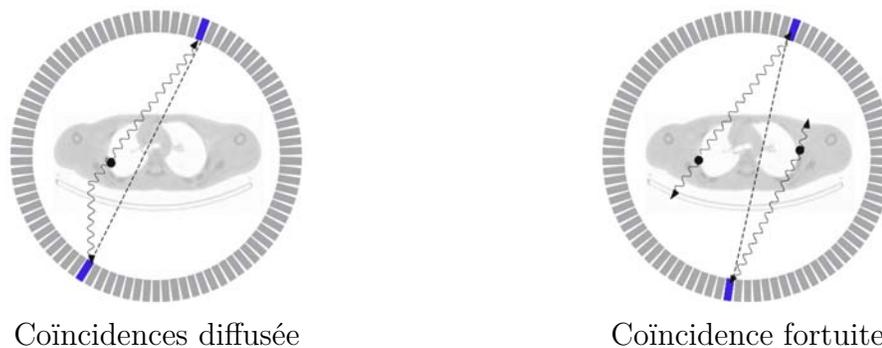


FIGURE 1.9 – Coïncidences indésirables [98].

Organisation des données Les données brutes collectées par le scanner TEP sont le nombre de coïncidences comptées pour chaque couple de capteurs. Ces données peuvent être triées chronologiquement dans un format appelé *list mode*. Toutefois, il

est parfois intéressant de représenter ces informations sous la forme d'un *sinogramme*. Un sinogramme est une matrice où chaque ligne correspond à la projection d'un plan 2-D de l'image originale selon différentes incidences (voir fig. 1.10). Ainsi, à un point du sinogramme correspond une LOR.

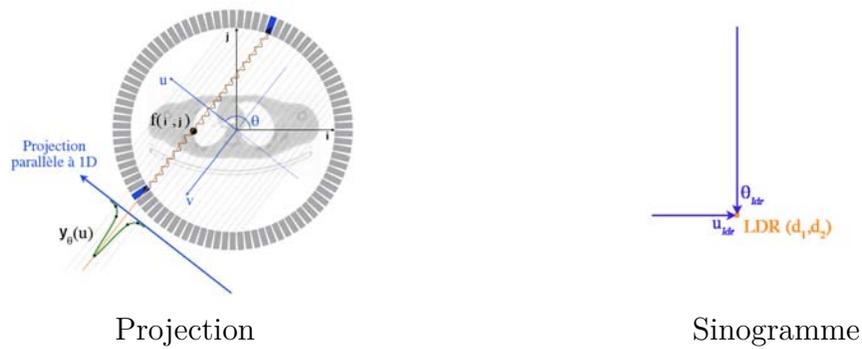


FIGURE 1.10 – Construction du sinogramme [98]

En 3-D, le sinogramme se généralise en une structure appelée michelogramme, composé d'un sinogramme pour chaque couple d'anneaux coaxiaux du scanner.

1.5.2 Reconstruction

Comme beaucoup de modalités d'imagerie médicale, le modèle direct d'observation TEP se formule comme :

$$\mathbf{y} = \mathcal{S}\mathbf{x} \quad (1.6)$$

où \mathbf{y} est le vecteur d'observation (des paires de coïncidences de photons γ en TEP). \mathbf{x} est le vecteur représentant tous les pixels de l'image (inconnue). \mathcal{S} est l'opérateur d'observation, ou matrice système.

Il existe deux grands types de méthodes de reconstruction :

Les méthodes analytiques qui fournissent une solution "exacte" en résolvant un système d'équations.

Les méthodes itératives dans lesquelles une image intermédiaire converge itérativement vers la "vraie" image.

Méthodes analytiques

En 1917, Radon a formulé les bases mathématiques de la reconstruction d'un objet à partir de ses projections [112]. A cet égard, il est intéressant de remarquer

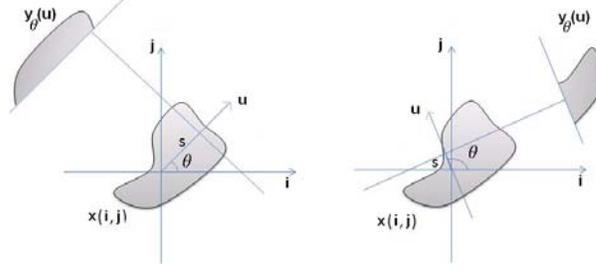


FIGURE 1.11 – Principe de projection : la mesure $y_\theta(u)$ correspond à l'intégrale de la fonction à reconstruire le long d'une droite orthogonale à l'axe u [101].

qu'en 2-D, pour les protocoles d'imagerie basés sur une tomographie, comme c'est le cas en TEP, \mathcal{S} correspond à la matrice de projection de Radon. Les sinogrammes \mathbf{y} peuvent alors être vus comme la transformée de Radon (discrète) de l'image originale. De manière générale, si \mathbf{x} est une fonction définie de \mathbb{R}^2 dans \mathbb{R} , la transformée de Radon \mathbf{y} de \mathbf{x} dans la direction θ à une distance u de l'axe est égale à :

$$y_\theta(u) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(i, j) \delta [i \cos(\theta) + j \sin(\theta) - u] didj \quad (1.7)$$

où $\delta(\cdot)$ est l'impulsion de Dirac. La transformée de Radon correspond donc à l'ensemble des intégrales lignes selon chacune des directions θ , et chaque direction correspond à une ligne du sinogramme.

Les solutions analytiques se basent sur cette propriété et consistent à appliquer la transformation inverse de Radon sur le sinogramme y pour retrouver l'image x .

Reconstruction par transformée de Fourier Une solution analytique de l'inverse de Radon est basée sur le théorème de la coupe centrale. Ce théorème stipule que la transformée de Fourier unidimensionnelle d'une projection $\mathcal{F}_{1D}\{y_\theta(u)\}$ est égale à une ligne passant par l'origine et faisant un angle θ avec l'axe i , de la transformée de Fourier bidimensionnelle de l'image $\mathcal{F}_{2D}\{x(u \cos(\theta), u \sin(\theta))\}$. La résolution du problème revient donc à calculer la transformée de Fourier de chaque projection $\mathcal{F}_{1D}\{y_\theta(u)\}$, de reporter les lignes obtenues dans le plan fréquentiel, puis de réaliser une transformée de Fourier bidimensionnelle inverse.

En pratique, la reconstruction par l'inversion directe de la transformée de Fourier est simple et rapide mais s'avère très peu robuste au bruit car la transformée de Radon mesurée est dégradée.

Rétroprojection filtrée (Filtered Back-Projection : FBP) La résolution par la méthode de la rétroprojection filtrée consiste en deux étapes. Dans un premier temps, on applique un filtre rampe à chaque ligne du sinogramme, puis, dans un deuxième temps, une rétroprojection simple est faite dans le plan des projections filtrées pour chaque angle. La résolution analytique par la méthode de rétroprojection filtrée peut se formuler ainsi :

$$x(i, j) = \int_0^\pi y_\theta(u) \otimes h(u) d\theta \quad (1.8)$$

où \otimes est l'opérateur de convolution et $h(u)$ est le noyau d'un filtre rampe souvent combiné à un filtre passe-bas pour éviter l'amplification du bruit (voir fig. 1.12).

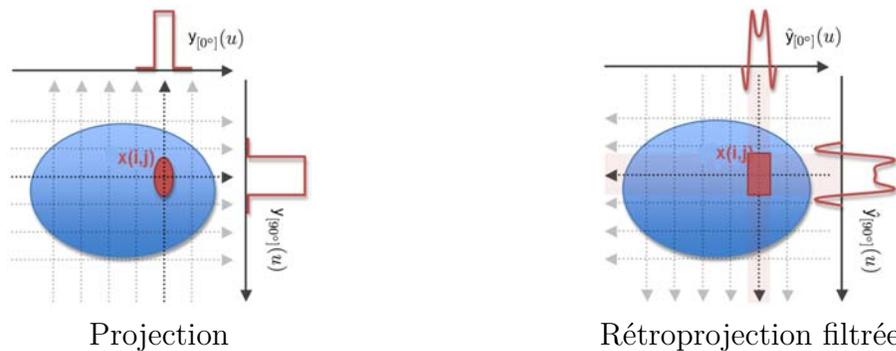


FIGURE 1.12 – Principe des opérations de projection et de rétroprojection à partir de deux angles de vue (0° et 90°) [98].

Les méthodes analytiques déterministes de reconstruction ont l'avantage d'être rapides et faciles à mettre en oeuvre. Mais elles sont peu robustes lorsqu'elles sont employées sur des données discrètes et se basent sur un modèle physique très simpliste. Pour pallier ces problèmes, des méthodes itératives ont été développées.

Méthodes itératives

Le problème de reconstruction tomographique peut être vu comme la résolution d'un système discret linéaire $\mathbf{y} = \mathcal{S}\mathbf{x}$ où le nombre d'équations correspond au nombre d'éléments dans le sinogramme, et les inconnues sont les valeurs des voxels de l'image

à reconstruire. L'équation (1.6) s'écrit plus précisément :

$$\begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^I \end{pmatrix} = \begin{pmatrix} s_1^1 & s_2^1 & \cdots & s_J^1 \\ s_1^2 & s_2^2 & \cdots & s_J^2 \\ \vdots & \vdots & \ddots & \vdots \\ s_1^I & s_2^I & \cdots & s_J^I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{pmatrix} \quad (1.9)$$

où s_j^i est la probabilité qu'une annihilation ayant eu lieu dans le voxel x_j soit détectée sur la LOR y^i . Notez que \mathcal{S} correspond ici à la forme matricielle de la projection associée à la transformée de Radon discrète. La matrice \mathcal{S} est souvent appelée *matrice système* et peut inclure le modèle géométrique d'acquisition aussi bien que la prise en compte des phénomènes physiques influençant les mesures (comme les phénomènes d'atténuation).

La résolution du système (1.9) consiste à déterminer \mathbf{x} connaissant \mathcal{S} et \mathbf{y} . Or, les valeurs singulières de \mathcal{S} sont petites. Le problème est donc mal conditionné et l'inversion directe du projecteur \mathcal{S} est impossible en pratique. On a donc souvent recours à des méthodes itératives. Leur principe est de répéter en boucle un cycle de projection et de retroprojection jusqu'à la convergence de l'algorithme, c'est à dire quand la différence entre le sinogramme mesuré et le sinogramme associé à l'image estimée est négligeable.

Il existe principalement deux types d'approches discrètes itératives :

Les méthodes algébriques ont été les premières méthodes de reconstruction itératives à être utilisées, notamment avec l'algorithme ART (Algebraic Reconstruction Techniques) [51] ou son amélioration SIRT [48] et autres dérivés. Ces méthodes considèrent la solution du système (1.9) comme l'intersection d'hyperplans et consistent en une série de projections sur chacun de ces hyperplans. Mais ces méthodes ne considèrent pas le caractère stochastique des données et sont donc sensibles au bruit.

Les méthodes statistiques qui se fondent sur une expression probabiliste du problème de reconstruction et permettant de prendre en considération le bruit des données. La résolution est formulée comme un problème d'optimisation visant à maximiser une vraisemblance basée sur le modèle statistique choisi (généralement Poissonien). Parmi les méthodes statistiques proposées dans la littérature, MLEM est l'algorithme le plus utilisé en TEP et sera présenté dans le paragraphe suivant.

Une description générale des différentes méthodes de reconstruction itératives est présentée dans [111].

Maximum de vraisemblance par Espérance-Maximization (MLEM) La méthode Maximum-Likelihood Expectation-Maximization (MLEM) [126] repose sur la représentation du bruit des données selon une statistique de Poisson maximisant une fonction de croyance, plus précisément le logarithme de la vraisemblance.

On discrétise l'espace en voxels $\{j\}$. En ignorant le déplacement des positons entre leur apparition et leur annihilation (de l'ordre de quelques millimètres), le nombre d'émissions x_j provenant d'un voxel j donné dépendra de la concentration du traceur dans ce voxel et de la période de l'isotope radioactif. D'après (1.5), on suppose ici que les données mesurées sont des réalisations de variables aléatoires distribuées selon une loi de Poisson. Pour chaque voxel j , on associe donc une variable aléatoire X_j suivant une loi de Poisson de paramètre \bar{x}_j .

$$P[X_j = x_j] = e^{-\bar{x}_j} \frac{\bar{x}_j^{x_j}}{x_j!}. \quad (1.10)$$

Certaines des ces émissions seront détectées par le système. On peut alors considérer :

$$p(i, j) \triangleq p(\text{événement détecté par la LOR } i | \text{événement émis par le voxel } j) \quad (1.11)$$

Comme les photons sont émis selon une direction aléatoire uniformément distribuée, la plupart d'entre eux ne seront pas détectés et

$$p(j) = \sum_i p(i, j) \leq 1. \quad (1.12)$$

On peut toutefois se ramener à l'égalité en ne considérant que les photons détectés. En posant

$$s(i, j) \triangleq p(i, j)/p(j) \quad (1.13)$$

on a

$$s(j) = \sum_i s(i, j) = 1 \quad (1.14)$$

$s(i, j)$ représente alors la probabilité que parmi les événements détectés, un événement émis par le voxel j soit détecté par la LOR i . Il est intéressant de noter que l'on retrouve ici la matrice \mathcal{S} de l'équation (1.9).

En définissant y_i le nombre de détections faites par la LOR i on peut écrire :

$$y_i = \sum_j x_i^j \quad (1.15)$$

où x_i^j est le nombre d'événements provenant du voxel j détectés par la LOR i . Les x_i^j sont des variables indépendantes suivant une loi de Poisson, une somme finie de ces variables suit donc également une loi de Poisson.

Selon cette modélisation, le nombre de coïncidences y_i détectées par la LOR i peut donc être vu comme une réalisation d'une variable aléatoire Y_i suivant une loi de Poisson :

$$P[Y_i = y_i] = e^{-\bar{y}_i} \frac{\bar{y}_i^{y_i}}{y_i!} \quad (1.16)$$

dont le paramètre \bar{y}_i est lié aux paramètres des X_j par la formule :

$$\bar{y}_i = E[Y_i] = \sum_j s(i, j) \bar{x}_j. \quad (1.17)$$

Autrement dit, le nombre moyen de photons détectés dans une LOR est égal à la somme sur tous les voxels, du produit du nombre moyen d'émissions issues de ce voxel par la fraction de photons émis depuis ce voxel selon la LOR considérée [128].

Bien que le modèle de Poisson soit approprié en théorie, il arrive que des corrections effectuées avant, durant ou après le processus de reconstruction viennent dénaturer le caractère poissonnien et indépendant des données. En effet, des tâches comme la correction d'atténuation, la correction des coïncidences aléatoires, ou une simple renormalisation de l'image peuvent modifier la densité de probabilité de la variable aléatoire et un simple lissage peut corrélérer spatialement les mesures. D'autres modèles de données tels que des modèles gaussiens [3, 91] ont été proposés. D'autres travaux ont également été réalisés pour essayer de préserver les statistiques Poissonniennes [89]. Le modèle de Poisson reste tout de même le modèle le plus utilisé pour décrire les données TEP [120, 126, 130].

Une fois le modèle des données fixé, le problème de reconstruction d'image consiste à estimer \bar{x}_j connaissant les données observées y_i . En supposant que les détections sont Poissonniennes, la vraisemblance des données peut donc s'écrire :

$$L(\bar{\mathbf{x}}) = P[y|\bar{\mathbf{x}}] = \prod_i \exp^{-\bar{y}_i} \frac{\bar{y}_i^{y_i}}{y_i!} \quad (1.18)$$

où \bar{y}_i a été défini dans (1.17). En reprenant l'expression (1.17) et en passant au log on obtient la log-vraisemblance suivante :

$$\mathcal{L}(\bar{\mathbf{x}}) = \log [L(\bar{\mathbf{x}})] = - \sum_j \sum_i \bar{x}_j s(i, j) + \sum_i y_i \log \left[\sum_j \bar{x}_j s(i, j) \right] - \sum_i \log(y_i!). \quad (1.19)$$

En considérant maintenant les première et deuxième dérivées de la log-vraisemblance, il peut être montré que la matrice des deuxièmes dérivées est semi-définie négative et que $\mathcal{L}(\bar{\mathbf{x}})$ est concave [126]. Par conséquent, une condition suffisante pour

qu'un vecteur \bar{x} minimise \mathcal{L} est donnée par la condition de Kuhn-Tucker :

$$0 = \left[\bar{x}_j \frac{\partial \mathcal{L}(x)}{\partial \bar{x}_j} \right]_{\bar{x}_j} = \bar{x}_j - \sum_i \frac{y_i s(i, j) \bar{x}_j}{\sum_{j'} s(j', i) \bar{x}_{j'}}. \quad (1.20)$$

En interprétant la formule précédente comme un problème de point fixe, on obtient enfin la formule de mise à jour de \bar{x} voxel par voxel :

$$\bar{x}_j^{(k+1)} = \bar{x}_j^{(k)} \sum_i s(i, j) \frac{y_i}{\sum_{j'} s(j', i) \bar{x}_{j'}}. \quad (1.21)$$

Dans l'équation (1.21), on reconnaît la projection directe de l'image estimée \bar{x}^k sur l'espace de Radon à l'itération k :

$$\tilde{y}_j \triangleq \sum_{j'} s(j', i) \bar{x}_{j'} \quad (1.22)$$

et la rétroprojection :

$$\bar{x}_j^{(k+1)} \triangleq \bar{x}_j^{(k)} \sum_i s(i, j) \frac{y_i}{\tilde{y}_j}. \quad (1.23)$$

L'algorithme MLEM converge alors vers l'unique maximum de vraisemblance \bar{x} . Le nombre d'itérations nécessaires avant d'obtenir une image pertinente est toutefois important et rend l'algorithme MLEM lent.

Ordered-Subsets Expectation Maximization (OSEM) De nombreux travaux ont été développés afin de trouver des alternatives convergeant plus rapidement. L'algorithme le plus utilisé en routine clinique est l'algorithme des sous-ensembles ordonnés (Ordered Subsets EM, OSEM [66]). Les projections sont regroupées en blocs disjoints et chaque itération est composée d'une sous-itération pour chacun des blocs. Durant chacune de ces sous-itérations, l'algorithme MLEM est appliqué au sous-ensemble correspondant. A chaque itération, l'image est donc mise à jour autant de fois qu'il y a de blocs, et la convergence est accélérée proportionnellement. Le schéma itératif de l'algorithme de reconstruction OSEM s'écrit

$$x_j^{(k*B+b+1)} = x_j^{(k')} \sum_{i \in \mathcal{B}_b} s(i, j) \frac{y_i}{\tilde{y}_i^{(k*B+b)}} \quad (1.24)$$

où B est le nombre de bloc et \mathcal{B}_b représente les indices des LOR appartenant au bloc b .

Notons par ailleurs qu'une version nommée OPL-EM (One-Pass List-Mode Expectation Maximization), adaptée au format list-mode a également été développée [115].

Ces algorithmes sont largement utilisés dans les protocoles cliniques du fait de leur rapidité d'exécution. Ils conduisent cependant à des images très bruitées. Il est alors fréquent d'ajouter un terme de régularisation à la fonction à minimiser ou d'appliquer un filtre passe-bas à l'image reconstruite.

1.6 La tomodensitométrie

La tomodensitométrie (TDM), ou computerized tomography (CT), est une modalité d'imagerie anatomique non invasive mettant en évidence la différence d'atténuation d'un faisceau de rayons X à travers les différents tissus de l'organisme. Elle fut étudiée initialement par Cormack [30, 31] dans les années 1960 et développée par Hounsfield [65] au début des années 1970. Considérée comme une des plus importantes en radiologie, cette découverte valut à Cormack et à Hounsfield le Prix Nobel de médecine en 1979. La TDM est utilisée aujourd'hui dans de vastes domaines d'applications, que ce soit pour l'exploration du thorax et de l'abdomen ou pour l'évaluation des fractures osseuses. La TDM est également d'intérêt majeur pour la définition du volume cible en radiothérapie et l'image obtenue est également utilisée lors de la reconstruction des images TEP quant à la correction des phénomènes d'atténuation. Des études plus récentes s'intéressent aussi à la TDM dans le domaine de la coloscopie virtuelle.

1.6.1 Principes physiques

Le système d'acquisition du TDM est composé d'un tube à rayons X couplé à des détecteurs situés en vis-à-vis autour d'un lit sur lequel le patient est allongé. Les rayons X émis traversent le corps à imager et sont plus ou moins atténués selon la densité des tissus rencontrés. Les récepteurs mesurent alors la quantité de rayons X reçus après la traversée du corps. Ces informations sont converties en signaux électriques et correspondent aux coefficients d'atténuation linéique des tissus traversés selon l'angle considéré. La source et les récepteurs sont animés d'un mouvement de rotation synchrone autour du patient et les projections mesurées selon les différents angles peuvent alors être utilisées pour reconstruire mathématiquement une coupe axiale transverse de l'objet. L'image obtenue fait alors ressortir les différences d'atténuation des rayons X dans les différents tissus et correspond donc à une carte de densité du corps examiné.

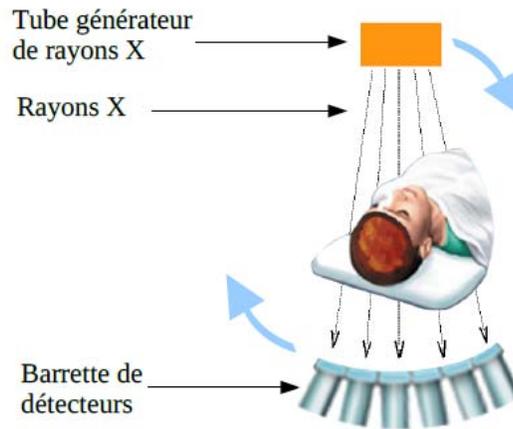


FIGURE 1.13 – Représentation schématique d'un tomographe [138].

Génération des rayons X

La génération des rayons X est assurée par un tube radiogène. Il est composé d'une cathode émettant des électrons par effet thermoionique, qui sont eux mêmes accélérés vers une anode, et un faisceau de rayons X est émis par rayonnement de freinage. A la sortie du tube, la source est colimatée de manière à ce que le faisceau soit en forme d'éventail et couvre toute l'étendue du corps à examiner. L'intégralité du faisceau est ensuite intercepté par des centaines de détecteurs diamétralement opposés à la source et disposés en arc de cercle.

Afin de limiter l'exposition du patient aux rayonnements ionisants, l'intensité du faisceau peut être adaptée par le système en fonction de la corpulence du patient.

Interaction avec la matière

Entre leur émission et leur détection, les rayons X vont interagir avec la matière qu'ils traversent. Ces interactions peuvent être résumées aux trois principes physiques suivants :

- l'effet photoélectrique
- l'effet de diffusion Compton
- l'effet de diffusion Rayleigh

et résultent en une atténuation de l'énergie des rayons X.

Le flux du faisceau détecté après la traversée d'un tissu de densité uniforme répond à la loi d'atténuation de Beer-Lambert, valable pour un faisceau monochromatique :

$$I = I_0 e^{-\mu(E)x} \quad (1.25)$$

où I_0 est l'intensité initiale du rayon, I représente l'intensité du rayonnement après la traversée, x l'épaisseur du matériau traversé et μ est le coefficient d'atténuation linéique du milieu, dépendant de l'énergie E du photon X et définissant les propriétés de la matière à atténuer les rayons X.

Dans le cas de la traversée de tissus hétérogènes, cette même loi se généralise en :

$$I = I_0 e^{-\int_L \mu(x,y,E) ds} \quad (1.26)$$

où I_0 et I représentent respectivement les flux de photons incidents et transmis et $\mu(x, y, E)$ est le coefficient d'atténuation de la matière à la position (x, y) . L'intégrale est calculée le long de la ligne L partant de la source vers le détecteur.

La fraction transmise peut alors être exprimée comme le rapport entre le flux sortant et le flux entrant :

$$TF = \frac{I}{I_0} = e^{-\int_L \mu(x,y,E) ds} \quad (1.27)$$

En se basant sur l'équation (1.27), on peut alors définir la projection P à l'angle θ et à la distance t depuis l'origine par

$$P(\theta, t) \triangleq -\ln\left(\frac{I}{I_0}\right) = \int_{L(\theta,t)} \mu(x, y, E) ds \quad (1.28)$$

Il est alors intéressant de constater que l'équation (1.28) représente la somme des coefficients d'atténuation linéique le long du chemin de projection et qu'il peut être directement calculé à partir des flux incidents I_0 et transmis I .

Détection des rayons X et modes d'acquisition

Méthode de conversion des rayons X La détection des rayons X peut se faire par conversion directe ou indirecte.

Dans le cas d'une conversion directe, l'énergie déposée par les particules dans le détecteur est transformée directement en charges électriques dans le substrat du semi-conducteur. Toutefois, l'efficacité de détection dépend de l'énergie reçue et ce mode de conversation est souvent touché par ailleurs par la faible efficacité de détection des capteurs de type semi-conducteurs.

En détection indirecte, les rayons X sont d'abord convertis en lumière visible, dans un scintillateur via les mécanismes de luminescence. Ces photons lumineux vont ensuite créer des charges dans un photodétecteur.

Mode de détection Une fois les détections réalisées et converties en signaux électriques, il existe deux manières de quantifier ces détections : par intégration de charge ou par comptage de photons.

Les détecteurs de rayons X utilisés en TDM les plus utilisés en routine clinique fonctionnent en mode intégration de charges. Dans ce mode, les charges reçues par chaque détecteur sont accumulées pendant une période de temps donnée et ne sont lues qu'à la fin de cette durée d'intégration. Or, la charge créée par un photon est fonction de son énergie, le bruit Poissonien des photons de haute énergie aura tendance à altérer les mesures.

Dans le cas du mode comptage de photons, chaque détection est traitée individuellement. L'information mesurée correspond alors au nombre de photons détectés, indépendamment de leur énergie.

Modes d'acquisition Il existe de plus deux modes d'acquisition selon le protocole utilisé : le mode séquentiel et le mode hélicoïdal (voir fig. 1.14).

Le mode séquentiel, aussi appelé incrémental, était le premier à être utilisé. Il consiste à acquérir les coupes successives de manière isolée. Selon ce mode, le lit est fixe et l'ensemble émetteur/récepteurs tournent autour du patient avec une période de révolution allant de 0.5 à 4 secondes. Les différentes projections acquises pendant la rotation, sont utilisées pour reconstruire une coupe à une position axiale précise. Puis, le lit est déplacé et une nouvelle coupe est acquise.

Depuis les années 1990, les technologies des systèmes TDM ont connues de nombreuses évolutions, et le mode d'acquisition séquentiel a laissé place au mode hélicoïdal, au point de faire de ce dernier le mode d'acquisition le plus utilisé de nos jours. Dans le mode hélicoïdal, ou continu, la table d'examen effectue une translation linéaire à vitesse constante pendant la durée d'exposition. On introduit alors un paramètre appelé pitch qui correspond au rapport entre la vitesse de la table et la période de rotation du tube. Ce mode d'acquisition permet d'effectuer des acquisitions rapides du volume. Des interpolations linéaires sont toutefois nécessaires pour compenser le manque de données à chaque position axiale.

Plus récemment, des systèmes multicoupes ont été développés. A l'inverse des systèmes monocoupes acquérant une coupe à chaque rotation du tube, les systèmes multicoupes sont équipés de plusieurs rangées de détecteurs permettant d'obtenir et de reconstruire sans artefacts une partie assez large du corps à chaque rotation.

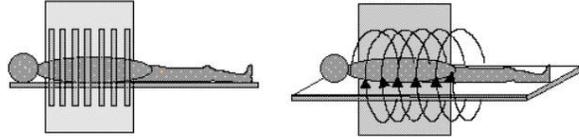


FIGURE 1.14 – Les deux modes d’acquisition : séquentiel à gauche, hélicoïdal à droite [101].

1.6.2 Reconstruction

Quelques soient les techniques utilisées, les données acquises sont mesurées dans l’espace des projections et non dans le domaine spatial directement. En tomodensitométrie, ces projections diffèrent de la TEP du fait qu’à une position du tube donnée, tous les rayons sont émis depuis la même source et les projections sont donc à géométrie conique et non plus parallèle. Ces projections peuvent toutefois être représentées par un sinogramme comme en TEP et peuvent être interprétées comme la transformée de Radon de la répartition spatiale des coefficients d’atténuation. L’image peut donc être reconstruite en appliquant un algorithme de reconstruction tomographique usuel. En imagerie TDM, c’est la méthode analytique de rétroprojection filtrée qui est la plus utilisée (voir 1.5.2).

L’image obtenue correspondra alors aux valeurs du coefficient d’atténuation linéique μ en tout point de l’espace. Ces valeurs sont toutefois renormalisées selon une échelle standardisée, dite de Hounsfield, qui a l’avantage d’être relativement stable et peu dépendante de l’énergie d’émission des rayons X. Cette nouvelle échelle rapporte chaque coefficient μ à celui de l’eau. Ainsi, la valeur finale N_i (en unités HU) du voxel i , dont le coefficient d’atténuation linéique vaut μ_i est fixée par la relation

$$N_i = \frac{\mu_i - \mu_{\text{eau}}}{\mu_{\text{eau}}} * 1000 \quad (1.29)$$

Selon cette nouvelle échelle, les valeurs obtenues varient de -1000 (pour l’air) à $+3000$ (pour le métal), avec $N_{\text{eau}} = 0$ et $N_{\text{os}} = 1000$.

La résolution fine de la TDM (de l’ordre de 1mm dans le sens transaxial et de 2 à 5mm dans le sens axial) permet une visualisation précise des structures anatomiques du corps étudié.

1.7 Conclusion

La prise en charge du cancer reste de nos jours un enjeu de santé publique. Dans ce domaine, l'imagerie médicale joue un rôle très important. Elle permet par exemple d'extraire des indicateurs, utiles au diagnostic et au suivi des patients. Souvent, ces indicateurs sont mesurés sur une région d'intérêt préalablement délimitée par un processus de segmentation habituellement semi-manuelle. La qualité de la segmentation est donc un facteur critique pour améliorer la précision du diagnostic.

Cette thèse propose une méthode de segmentation pour la tomographie par émission de positons (TEP). Cette modalité d'imagerie constitue un outil prometteur permettant souvent de détecter un cancer bien avant sa visibilité par tomodensitométrie. La segmentation en TEP est toutefois un problème difficile. Une approche orientée "contour" n'est pas possible car les organes ne sont pas délimités par des contours géométriques. Une approche orientée "région" est difficile car les données (même associée à une même classe) présentent une forte variabilité à cause du bruit ajouté durant la détection, ou durant la formation de l'image, ou même à cause de l'objet exploré lui-même comme c'est le cas des tumeurs hétérogènes. Les méthodes de segmentation orientées "région" appliquées directement sur les données mesurées, comme les techniques de seuillage, sont peu robustes à cette variabilité.

Une modélisation statistique des données est donc nécessaire. L'algorithme de segmentation proposé aura alors pour objectif de discriminer les classes présentant des statistiques différentes. Le cadre statistique permettra ainsi de trouver le modèle, les délimitations ou directement les indicateurs *les plus probables*.

Afin de modéliser au mieux les données, le modèle statistique utilisé doit répondre à certains critères. En premier lieu, il doit être robuste aux variations d'activité pouvant avoir lieu dans une même classe. Ensuite, il doit être adapté aux indicateurs recherchés. Enfin, il doit se baser sur les processus physiques de formation de l'image. Ce document présente un modèle statistique répondant à ces critères. Il sera présenté chapitre 2.

Comme la modélisation des données et la segmentation sont étroitement liées le chapitre 3 présentera un algorithme résolvant ces deux problèmes de manière conjointe.

Avec l'essor des technologies mixtes, permettant d'acquérir une image anatomique et une image fonctionnelle durant le même examen, il semble intéressant d'ajouter des informations géométriques pour affiner le processus de segmentation. Toujours dans un cadre statistique, une technique de segmentation d'images bimodales TEP/TDM sera présentée en chapitre 4.

Modeling heterogeneous tumors and Bayesian parameters estimation

Introduction

Pathophysiological studies report that the biological activity is heterogeneous in tissues, and especially in tumors. This chapter addresses the statistical modeling of PET data, so that intra-tissue variations of activity are considered.

This chapter also provides a robust and efficient method to estimate the parameters of this distribution in a general context of the Monte Carlo approach.

Problem statement

Modeling PET data

It is proposed to model the activity of a single tissue by a negative binomial distribution (NBD).

Its analytical expression can be formulated using the parameterization in 2.2.

This model has the advantage of taking into consideration the variations of activity within the same tissue.

Parameter estimation

Despite its popularity, the estimation of the NBD parameters is a well known difficult problem.

This paper proposes a Bayesian method to estimate the parameters of the NBD using an MCMC technique.

Bayesian model

A hierarchical Bayesian model is established. The parameter estimation problem is formulated as a *Maximum a Posteriori* (MAP) problem (see (2.7)).

Likelihood

As the observations are mutually independent, the likelihood $P[\mathbf{x}|\boldsymbol{\theta}]$ (see eq.(2.9)) can be defined directly from the formulation of the density of the NBD (2.2).

Prior on the model parameters

Gamma distributions were chosen as prior for both the mean μ and inverse dispersion k (eq. (2.11)).

The choice of these priors are motivated by the fact that the gamma distributions have their values limited to \mathbb{R}_+^* , corresponding to the positivity constraint of μ and k .

Assuming that the parameters are independent, the joint prior distribution of the vector θ is expressed in (2.14).

To summarize the relationships between the various parameters and hyper-parameters, figure 2.1 presents the proposed hierarchical Bayesian model as a directed acyclic graph.

Posterior distribution

The shape of the obtained posterior density is shown in figure 2.2.

Its complex form has the following characteristics :

- First, the target density exhibits a strong curve, indicating a correlation between its parameters.
- The target density is also anisotropic, as illustrated by the stretch along the k dimension.
- Finally, the parameter values are constrained to be positive.

These three factors make the sampling process difficult. In this work, we develop a sampling method to address these problems.

Method

This section proposes a method to efficiently sample densities with correlated parameters under non-negativity constraints.

To address the problem of the correlation between parameters, a random walk based on *Manifold MALA* allows an efficient exploration of the target density, by reducing the difficulties due to anisotropy.

In addition, the use of a *proximal MCMC* algorithm effectively ensures non-negativity of the parameters.

The two methods used together provide efficient mixing properties even if vague priors are used.

Manifold MALA

Random Walk Metropolis-Hastings

Using a standard RWMH algorithm is not advisable for exploring an anisotropic parameter space because the different dimensions behave in very different ways and it is difficult, if not impossible, to choose a scale factor ϵ that fits all dimensions. This can lead to excessively slow mixing.

Hybrid Gibbs Sampler

Gibbs hybrid methods fail if there is a correlation between the elements to be sampled. For this reason, a sampler proposing a pair of parameters (μ, k) is preferred to Gibbs sampling, which works separately in each dimension.

Metropolis Adjusted Langevin Algorithm (MALA)

In the Metropolis Adjusted Langevin Algorithm (MALA), the gradient favors the movement of the chain according to the steepest slope. Thus, the chain will converge faster to the ergodic distribution and fewer steps will be required during burn-in. A new value θ^* is proposed as in (2.16).

However, it is clear that the isotropic diffusion will be ineffective for highly correlated θ with very different variances, because the step size ϵ will always fit the random variable with the smallest variance.

Covariance Matrix

This problem can be circumvented by using a preconditioning matrix \mathbf{M} , as in (2.17). But again, a global level of preconditioning may be inappropriate when the transient and stationary regimes of the Markov process are different.

MALA on Riemannian manifold

Manifold MALA (mMALA) is a generalization of MALA. Instead of using the Euclidean gradient as in the MALA, the idea is to consider the gradient with a more appropriate metric, exploiting the Riemannian geometry of the parameter space.

In this regard, the Fisher information matrix (see (2.18)) defines a local metric of a Riemannian manifold. From a Bayesian perspective, the joint probability of the data and the parameters defines the metric tensor; this tensor is the expected value

of the Fisher information matrix added to the opposite of the Hessian of the log-prior distributions (see (2.23)).

The new proposal mechanism is described in (2.24).

The parameter θ^* is accepted with probability (2.29).

The proposed sampling method is superior to the conventional RWMH method in several ways.

- The chain converges faster to the ergodic distribution therefore fewer iterations are required during burn-in.
- The step size ϵ is compensated by the deviation term that takes into account the shape of the target distribution. In other words, the algorithm is more flexible to the different step sizes required for transient and stationary phases and for the different behavior in each direction of the parameter space.
- The proposal distribution $q(\theta^*|\theta^i)$ is closer to the posterior distribution. This results in a better simulator.

Proximal manifold MALA

The manifold MALA method requires differentiability at each point of the parameter space. However, the two parameters of the NBD are limited to positive values. In practice, accurately simulating near the borders of this space will often lead to candidates who are out of bounds.

A simple solution would be to reject these proposed values, but in this case the step size can become extremely low and could significantly slow down the algorithm.

A better solution would be to consider a reflecting sampling scheme to keep successive candidates within the bounds (see fig. 2.3). The main drawback of this method is that the candidates may not be in the direction of the gradient, especially if the step size is too large.

In this chapter, we address this problem by using proximal operators.

We consider f , the extension of \mathcal{P} on \mathbb{R}^2 . f is not everywhere differentiable but MALA can be generalized to a wide range of non-differentiable distributions by exploiting the proximal applications.

In this case, a candidate θ^* is proposed according to the transition (2.34).

Because f is convex, its proximal application can be effectively approximated by using a forward-backward proximal splitting method (see (2.33)).

The above results can be easily extended to the consideration of the geodesic gradient (see (2.35)).

The parameter θ^* is then accepted with probability (2.36).

The forward-backward algorithm can be seen as a progressive step, guided by the gradient of f , followed by a retrogressive step where a projection is made according to h , on the positive area (see fig. 2.4).

The generalization of the gradient by a sub-gradient allows to restrict the generated candidates in the differentiable domain and thus allows to explore the parameter space efficiently.

Sampling algorithm

The final algorithm is detailed in the listing 1.

Results and discussions

The proposed method was validated on synthetic data and performance of the algorithm was evaluated and compared to other methods. Then, experiments were carried out on a real PET pre-segmented image to validate the proposed model.

Parameter estimation on synthetic data

Six sampling algorithms were compared :

GIBBS Metropolis-within-Gibbs

RWMH Random Walk Metropolis-Hastings

MALA Metropolis Adjusted Langevin Algorithm

prox-MALA proximal MALA

mMALA Manifold MALA

prox-mMALA Proposed proximal manifold-MALA

Convergence to the ergodic distribution

Figure 2.5 shows the first steps of the three different types of random walks, starting from the same starting point. We can see that the chain associated to the mMALA method reaches the top of the ergodic distribution faster than other chains. We also note that the MALA chain rapidly reaches the ergodic distribution, but as

the step size is adapted to the component with the smallest variance, the chain "*struggles*" to explore the entire density.

The simulations were repeated 20 times for each method, and the Multivariate Potential Scale Reduction Factor (MPSRF) was measured.

Figures 2.6, 2.7, 2.8 and 2.9 relate the evolution of MPSRF throughout the iterations for the six methods. Unlike the other methods, MPSRF associated with mMALA and prox-mMALA methods become low in the first iterations.

Tables 2.1, 2.2, 2.3 and 2.4 indicate the number of iterations required to achieve a MPSRF below 1.2. We see that only a few iterations are required for the algorithms mMALA and prox-mMALA for convergence, while the convergence is much slower for the other algorithms.

Samples correlation

The autocorrelation factor was calculated for each of the algorithms over the 100,000 iterations after burn-in (see figures 2.10, 2.11, 2.12 and 2.13).

The mMALA methods and prox-mMALA offer uncorrelated samples after fewer iterations than needed for other methods.

Effective sample size

The effective sample size has been deducted from above autocorrelation measurements.

Tables 2.5, 2.6, 2.7 and 2.8 show the number of effective samples within the 100000 iterations after burn-in.

One can see that the mMALA and prox-mMALA algorithms allow to provide much more uncorrelated samples than the other algorithms.

Effective sample size by second

From the previous results of the ESS and the computation time required to perform the simulations, the average number of independent samples drawn per second was calculated for each method. They are presented in tables 2.9, 2.10, 2.11 and 2.12.

The method using the proximal operators is faster than that based on a reflective pattern, and therefore provides more uncorrelated samples per time interval.

Modeling PET data

To confirm that the NBD correctly models the PET data, the proposed method was applied to real PET images. Figure 2.14 shows three successive slices of an image centered on the abdomen.

Goodness of fit tests have been made for the proposed model and compared to other statistical models used in the literature. Specifically, the distribution of the activity of each tissue has been modeled by a normal distribution, a Poisson distribution and the proposed negative binomial distribution.

Parameter estimation

The parameters associated to the different distributions have been estimated for each tissue. The MMSE estimates of each model are presented in table 2.13 for each of the 6 classes. The parameter estimation algorithms give a high reproducibility.

Moreover, the PSRF values obtained on the 50 chains for each parameter of each model shows that all chains have converged (see table 2.14).

Both results show that the estimated parameters for each model were estimated by accurate and robust algorithms.

Goodness of fit

The estimated parameters have been used to model the data by different distributions. Figure 2.15 shows the histogram of the data and the curves associated to the different models.

Figure 2.15 shows that the probability density function estimated using the proposed negative binomial model correctly fits the histogram data. The normal distribution model gives the worst results while the Poisson model does not correspond to the data.

The quantitative results of the goodness of fit are shown in tables 2.15 and 2.16. Table 2.15 reports the results of Kolmogorov-Smirnov (KS), Cramer von Mises (CVM) and χ^2 tests for each model and each class. Table 2.16 presents the Akaike (AIC) and Bayesian (BIC) information criteria for each model and each class. The negative binomial model outperforms the two other models, whatever the criterion used.

Discussions

The results of the Kolmogorov-Smirnov, Cramer von Mises and χ^2 tests show that, whatever the model, the goodness of fit is better for tissues with high activity,

particularly the tumor.

Although in theory the Poisson model is more legitimate than the Gaussian model, it does not fit the data. This can be due to reconstruction method that does not preserve the statistics of the data (artificial zeros, smoothing, scaling, contrast adjustment).

It is interesting to note that the negative binomial distribution is robust to artifacts.

In conclusion, we note that the proposed model correctly fits the activity of tumors, but also other biological tissues (liver, skin, ...) and outperforms Gaussian and Poisson models in terms of goodness of fit.

Chapitre 2

Modélisation des tumeurs hétérogènes et estimation Bayésienne de paramètres

2.1 Introduction

La tomographie par émission de positons (TEP) est une modalité d'imagerie fonctionnelle qui offre la possibilité de localiser les tumeurs dès leurs premiers stades en mesurant la répartition spatiale d'un traceur radioactif préalablement injecté dans le patient. Elle est souvent utilisée en oncologie pour porter un diagnostic, évaluer une thérapie ou faire un suivi de traitement. En diagnostic, une bonne connaissance des caractéristiques des tumeurs est essentielle pour permettre une évaluation précise et choisir le meilleur traitement. Cela nécessite une segmentation précise des tumeurs. Les méthodes existantes en clinique se basent principalement sur des techniques de seuillage [43, 95, 96, 32, 72, 38, 11] qui considèrent uniquement les valeurs des pixels séparément. Une segmentation précise ne peut être obtenue qu'en s'appuyant sur une modélisation appropriée des données.

Le mélange de Gaussiennes est le modèle souvent proposé pour représenter les données TEP [3, 91]. Or, en TEP, la valeur d'un pixel représente le nombre d'annihilations qui se produisent dans le tissu sous-jacent. Il s'agit donc d'un phénomène de comptage d'événements se produisant dans un interval de temps fixe. En supposant que la probabilité d'observer un événement est indépendant du temps écoulé depuis l'événement précédent, de telles données sont généralement modélisées par un processus de Poisson. Notons que dans l'acquisition TEP, les observations représentent le

nombre de photons arrivant sur chacun des capteurs durant la durée de l'examen. Ces données brutes suivent une loi de Poisson [113, 89, 110]. La reconstruction de l'image par des algorithmes de type MLEM se fait en considérant que les données de l'image sont aussi Poisson (1.10). Pour ces raisons, il a été établi que les voxels peuvent être modélisés par une variable aléatoire suivant une loi de Poisson [120, 126, 130]. Il s'ensuit facilement que la vraisemblance peut être exprimée comme un mélange fini de distributions de Poisson si λ_n est uniforme dans chaque tissu biologique.

Or, plusieurs études physiopathologiques rapportent que l'activité biologique est très hétérogène dans les tissus, et en particulier dans les tumeurs [62, 61]. Certaines tumeurs ont en effet des sous-régions ayant une activité biologique plus élevée que le volume restant. L'hétérogénéité dans le métabolisme des tissus conduit évidemment à l'hétérogénéité de leurs activités. Ceci explique le fait que les images TEP présentent souvent des variations d'absorption du traceur à l'intérieur d'un même tissu. Ceci conduit généralement à une *sur-variation* des données que le modèle de mélange de lois de Poisson ne parvient pas à modéliser.

Dans ce travail, on se propose de prendre en considération ce phénomène d'hétérogénéité et de le représenter par un paramètre qui estime l'éloignement de la distribution des données TEP à une loi Poisson.

Le chapitre est organisé comme suit. La section 2.2 pose le problème de la modélisation des images TEP de manière à tenir compte des variations d'activité intra-tissulaires. La section 2.3 présente le contexte Bayésien dans lequel le problème d'estimation des paramètres du modèle proposé est formulé. Ensuite, la section 2.4 détaille la méthode proposée pour construire un échantillonneur MCMC performant. Enfin, la section 2.5 rapporte les résultats expérimentaux validant le modèle proposé et la méthode d'estimation de ses paramètres.

2.2 Position du problème

2.2.1 Modélisation des données TEP

Soit $\mathbf{x} = \{x_1, \dots, x_N\} \in \mathbb{N}^N$, une image TEP contenant N voxels. On considère que l'image \mathbf{x} est composée de Z différents tissus biologiques \mathcal{Z}_z ($z = 1, \dots, Z$). Ces différents tissus forment une partition de l'image ($\bigcup_z \mathcal{Z}_z = \mathbf{x}$ et $\bigcap_z \mathcal{Z}_z = \emptyset$). Su et al. [130] modélisent l'image comme un mélange de distributions de Poisson, où l'activité des voxels d'un tissu donné suit une distribution de Poisson de paramètre λ_z :

$$x_n | (x_n \in \mathcal{Z}_z) \sim \mathcal{P}(\lambda_z) \quad (2.1)$$

où \mathcal{P} désigne la distribution de Poisson et λ_z est l'activité moyenne à l'intérieur du tissu \mathcal{Z}_z . Comme mentionné en introduction, cette modélisation considère l'activité au sein d'un même tissu homogène, ne permettant pas de considérer l'hétérogénéité intra-tissulaire. Ce chapitre propose un nouveau modèle de données TEP qui tient compte des variations de l'activité à l'intérieur d'un même tissu.

Lorsque l'on souhaite modéliser des données de comptage qui ont été générées à partir d'une population hétérogène trop dispersée pour être modélisée par des distributions de Poisson, on a souvent recours à la distribution binomiale négative (DBN). La DBN est utilisée dans de nombreuses applications telles que la sécurité routière [147, 87, 104], écologie [82, 18], zoologie [44], biologie [12, 143], épidémiologie [84], et plus récemment dans le domaine de la santé [148]. Elle est aussi utilisée pour modéliser des phénomènes aussi divers que le comportement d'achat des consommateurs [40], la productivité d'un vendeur [19], ou la circulation de livres dans les bibliothèques [16]. Son expression analytique peut être formulée en utilisant la paramétrisation suivante :

$$P[X = x|\mu, k] = \binom{x+k-1}{x} \left(\frac{\mu}{\mu+k}\right)^x \left(\frac{k}{\mu+k}\right)^k \quad (2.2)$$

où $\binom{a}{b} = \frac{a!}{b!(a-b)!}$, μ est le paramètre associé à la moyenne des échantillons et k est le paramètre d'inverse dispersion. Notons que :

$$\begin{cases} E(X|\mu, k) = \mu \\ Var(X|\mu, k) = \mu + \frac{1}{k}\mu^2. \end{cases} \quad (2.3)$$

On remarque que lorsque $k \rightarrow \infty$ alors $Var(X) \rightarrow \mu$, donnant une loi Poisson. La distribution binomiale négative peut donc être vue comme une généralisation de la loi Poisson permettant de prendre en compte la sur-dispersion des données et le paramètre k peut être interprété comme un indicateur de proximité de la distribution des observations à la loi de Poisson.

Il est intéressant par ailleurs de noter que modéliser l'activité des voxels d'un même tissu par une DBN revient à considérer que chaque voxel suit une distribution de Poisson propre de paramètre λ_n :

$$x_n|\lambda_n \sim \mathcal{P}(\lambda_n) \quad (2.4)$$

et que λ_n est lui-même la réalisation d'une variable aléatoire suivant une loi gamma, propre au tissu \mathcal{Z}_z auquel appartient x_n :

$$\lambda_n|(x_n \in \mathcal{Z}_z) \sim \Gamma(\alpha_z, \beta_z) \quad (2.5)$$

où $\alpha_z \in \mathbb{R}^+$ and $\beta_z \in \mathbb{R}^+$ sont respectivement les paramètres de forme et d'échelle associés au tissu \mathcal{Z}_z . En effet, la distribution marginale $\int_0^\infty \mathcal{P}(x_n|\lambda_n)\Gamma(\lambda_n|\alpha_z, \beta_z)d\lambda_n$ correspond alors à la loi Poisson-Gamma qui, après reparamétrisation, n'est autre que la loi binomiale négative :

$$\begin{aligned} P[x_n|\alpha_z, \beta_z, x_n \in \mathcal{Z}_z] &= \int_0^\infty \mathcal{P}(x_n|\lambda)\Gamma(\lambda|\alpha_z, \beta_z)d\lambda \\ &= \binom{x_n + \alpha_z - 1}{x_n} \left(\frac{1}{1 + \beta_z}\right)^{\alpha_z} \left(\frac{\beta_z}{1 + \beta_z}\right)^{x_n} \\ &= \mathcal{BN}(\mu_z, k_z) \end{aligned} \quad (2.6)$$

avec $\mu_z = \alpha_z$ et $k_z = \alpha_z\beta_z$.

2.2.2 Estimation des paramètres de la DBN

Malgré sa popularité, l'estimation des paramètres de la DBN soulève de nombreux problèmes [137]. Certaines méthodes d'estimation ont été étudiées en profondeur, comme la méthode des moments [79, 27, 124, 18] et de maximum-vraisemblance [79, 27, 107]. Une étude complète de l'état de l'art sur l'estimation des paramètres de la DBN, selon les méthodes des moments ou de maximum vraisemblance a été récemment proposée par Cadigan et Tobin [18]. Cependant, les estimateurs du paramètre de dispersion inverse k à l'aide des méthodes des moments ou du maximum vraisemblance sont biaisés et faiblement efficaces [27, 139, 121, 137]. Le problème avec les approches des méthodes des moments est le besoin de connaître la moyenne et la variance des données afin d'avoir une bonne estimation de k , car même un petit écart sur la variance peut causer une grande variation sur k [147]. Par ailleurs, les approches fondées sur le maximum de vraisemblance peinent à trouver un maximum global : l'algorithme EM a tendance à atteindre un maximum local ce qui le rend très sensible aux conditions d'initialisation [45]. De plus, la taille des échantillons doit être très large car cette méthode émane d'une théorie asymptotique. Les conditions de régularité sont souvent violées dans le cas d'un faible nombre de données [45]. Lloyd-Smith [84] met en évidence l'importance du biais selon les valeurs des paramètres de la DBN et selon le nombre d'observations.

Des méthodes Bayésiennes ont été considérées dans quelques travaux sur le problème de la DBN [34, 77, 125]. Cependant, comme il n'existe pas d'a priori conjugué pour la DBN, la loi a posteriori n'appartient à aucune famille de loi simple [104] et la difficulté d'explorer *entièrement* la densité rend l'échantillonnage inefficace. Notons cependant, que Bradlow et al. [13] proposent une approximation de la loi a priori conjuguée pour la DBN en utilisant une approximation polynomiale. Ils

présentent une solution analytique pour le problème d'inférence Bayésienne pour la DBN en exprimant la loi a posteriori comme une somme de termes polynomiaux. Mis à part cette approximation, l'inférence Bayésienne est souvent résolue au moyen de méthodes d'échantillonnage. Par exemple, Malyshkina et al. [87] présentent un échantillonneur de Gibbs hybride dans un modèle de permutation Markovien à deux états. Cependant, ces techniques d'inférence peuvent être entravées par la difficulté d'explorer la densité, surtout lorsque des lois a priori vagues sont employées.

De plus, de nombreuses applications nécessitent de résoudre le problème d'estimation des paramètres de la DBN dans un contexte plus général, comme par exemple pour l'estimation des paramètres d'un mélange de lois binomiales négatives. Récemment, Park et Lord [104] ont proposé un estimateur pour les paramètres d'un mélange fini de DBN utilisant une méthode d'échantillonnage de Gibbs hybride, avec une application à des données modélisant des accidents de véhicules. Un autre échantillonneur de Gibbs hybride a été proposé par Irace et al. [69] pour estimer les paramètres d'un mélange de DBN en tomographie par émission de positons où un champ aléatoire de Markov de type Potts a été utilisé comme a priori spatial dans l'image.

Il a été rapporté toutefois que la méthode d'échantillonnage de Gibbs hybride débouche sur des estimateurs biaisés [85]. Les auteurs ont remarqué en effet que les estimations MCMC aboutissent souvent à des écarts de simulation trop grands. Ils ont aussi mis en évidence que ces inexactitudes n'apparaissent pas seulement quand le nombre d'échantillons est faible ou que leur valeur moyenne est petite, mais peuvent également apparaître lorsque la taille des échantillons est relativement correcte. De plus, des études ont montré plus généralement que l'échantillonnage de type Gibbs peut conduire à des estimations peu précises lorsque les lois a priori sont vagues [93].

Ce chapitre répond à la problématique de la modélisation statistiques des données TEP, de manière à ce que les variations d'activité intra-tissulaire soient considérées. Ainsi, étant donné un tissu, son activité sera modélisée par une distribution binomiale négative. Ce chapitre propose également un moyen robuste et efficace d'estimer les paramètres de cette distribution dans un contexte général de Monte-Carlo, de manière à pouvoir être étendu efficacement à des problèmes de plus grande dimension (cas de l'estimation des paramètres d'un mélange par exemple).

2.3 Cadre Bayésien

Soit $\mathbf{x} = \{x_1, \dots, x_N\}$, N observations indépendantes et identiquement distribuées selon la loi binomiale négative dont l'expression a été définie en (2.2).

En posant le problème dans un cadre Bayésien, l'estimation des paramètres peut être vu comme une recherche du *Maximum a Posteriori* (MAP) :

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\boldsymbol{\theta}|\mathbf{x}) \quad (2.7)$$

où $\boldsymbol{\theta} = (\mu, k)$ est le vecteur de paramètres inconnus à trouver.

En utilisant le théorème de Bayes, la distribution a posteriori du vecteur de paramètres $\boldsymbol{\theta}$ peut être exprimée comme :

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto P[\mathbf{x}|\boldsymbol{\theta}]\pi(\boldsymbol{\theta}) \quad (2.8)$$

où $P[\mathbf{x}|\boldsymbol{\theta}]$ est la vraisemblance et $\pi(\boldsymbol{\theta})$ la loi a priori. Leur expression est définie dans les sous-sections suivantes.

2.3.1 Vraisemblance

Comme les observations sont mutuellement indépendantes, la vraisemblance $P[\mathbf{x}|\boldsymbol{\theta}]$ peut être définie directement à partir de la formulation de la densité de probabilité de la DBN (2.2) ainsi :

$$P[\mathbf{x}|\boldsymbol{\theta}] = \prod_{i=1}^N P[x_i|\boldsymbol{\theta}] \quad (2.9)$$

$$= \prod_{i=1}^N \binom{x_i + k - 1}{x_i} \left(\frac{\mu}{\mu + k}\right)^{x_i} \left(\frac{k}{\mu + k}\right)^k \quad (2.10)$$

2.3.2 Loi a priori sur les paramètres du modèle

Des distributions Gamma ont été choisies comme distribution a priori à la fois pour le paramètre de moyenne μ que pour le paramètre d'inverse dispersion k :

$$\begin{aligned} \mu &\sim \Gamma\left(1 + a_\mu, \frac{-1}{b_\mu}\right) \\ k &\sim \Gamma\left(1 + a_k, \frac{-1}{b_k}\right) \end{aligned} \quad (2.11)$$

où a_μ , a_k , b_μ et b_k sont les hyper-paramètres du modèle, pouvant être adaptés en fonction du problème, en se basant sur des informations qui peuvent avoir été obtenues à partir d'études préliminaires ou à partir d'informations subjectives.

La loi log-a priori s'écrira donc :

$$\begin{aligned}\log [\pi(\mu)] &= a_\mu \log(\mu) + b_\mu \mu + c \\ \log [\pi(k)] &= a_k \log(k) + b_k k + c\end{aligned}\tag{2.12}$$

où c est une constante. Le choix de ces distributions a priori a été motivé par le fait que les distributions gamma ont leurs valeurs limitées à \mathbb{R}_+^* , ce qui correspond aux contraintes de positivité des paramètres μ et k , et que la matrice Hessienne des distributions log-a priori (dont l'utilisation sera expliquée dans la partie 2.4.1) prend alors la forme diagonale simple suivante :

$$H(\mu, k) = - \begin{pmatrix} \frac{a_\mu}{\mu^2} & 0 \\ 0 & \frac{a_k}{k^2} \end{pmatrix}\tag{2.13}$$

Enfin, en supposant que les paramètres sont indépendants, la distribution a priori jointe du vecteur $\boldsymbol{\theta}$ est :

$$\pi(\boldsymbol{\theta}) = \pi(\mu)\pi(k)\tag{2.14}$$

où les logarithmes de $\pi(\mu)$ et de $\pi(k)$ ont été définis dans (2.12).

Pour résumer les relations entre les différents paramètres et hyper-paramètres, la figure 2.1 représente le modèle Bayésien hiérarchique proposé sous la forme d'un graphe orienté acyclique.

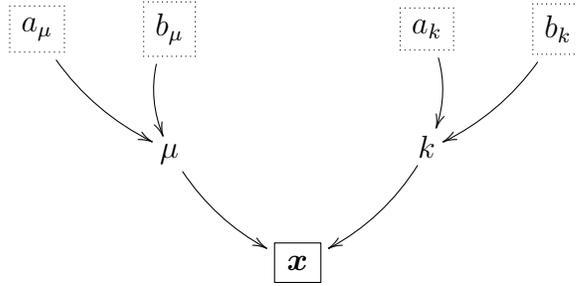


FIGURE 2.1 – Graphe orienté acyclique pour le modèle Bayésien proposé. Les hyperparamètres fixes sont représentés dans des cases en pointillé.

2.3.3 Loi a posteriori

La forme de la densité a posteriori obtenue est représentée figure 2.2.

Sa forme complexe présente les caractéristiques suivantes :

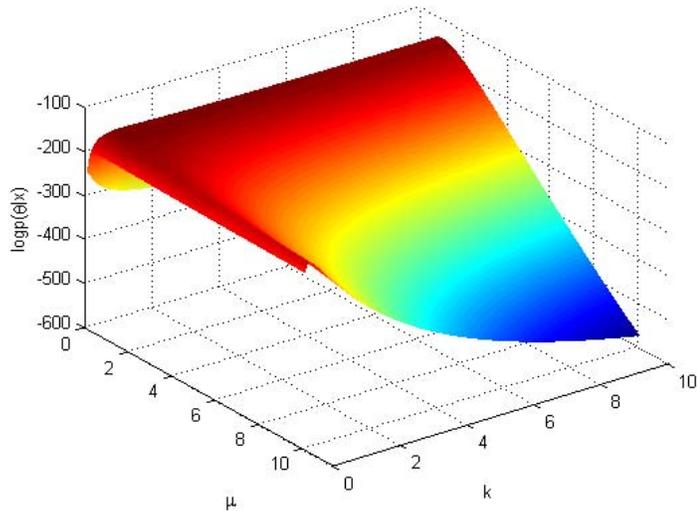


FIGURE 2.2 – Forme générale de la log-posterior de la DBN ($\mu = 1 ; k = 1$).

- En premier lieu, on remarque une courbure de la densité cible, témoignant de l’existence d’une corrélation entre les paramètres.
- La densité cible est également anisotrope, ce qui signifie que la distribution est étirée le long d’une dimension spécifique (ici le long de k).
- Enfin, les valeurs des paramètres sont contraintes à être positives.

Ces trois facteurs rendent le processus échantillonnage difficile. En effet, l’exploration de l’espace complet de la densité est tout d’abord impossible en pratique en raison de la corrélation entre ses paramètres. Le fait que les variables soient corrélées explique en particulier pourquoi les méthodes de Gibbs hybrides ne mélangent pas bien en deux dimensions et sont inefficaces dans les dimensions supérieures [85]. En second lieu, l’anisotropie de la densité cible rend le choix du facteur d’échelle difficile, voire impossible, et entrave le processus de génération d’échantillons indépendants. Enfin, le fait que les valeurs des paramètres soient contraintes à être positives peut entraver le processus d’échantillonnage lorsque un échantillon négatif est proposé.

Dans ce travail, nous développons une méthode d’échantillonnage répondant à ces difficultés. Ce chapitre examine le problème visant à simuler des échantillons dans un contexte général de Monte-Carlo, pouvant conduire à l’échantillonnage dans un espace de grande dimension. Simuler des échantillons selon une distribution de grande dimension est généralement difficile, en particulier dans les modèles où

l'échantillonnage de Gibbs standard n'est pas possible ou est inefficace [116, p. 339]. L'algorithme proposé combine une technique MALA-sur-variété pour atténuer le problème de paramètres corrélés et MCMC proximal pour remédier à la contrainte de positivité. Les deux techniques sont présentées dans la section suivante.

2.4 Méthode

Cette section propose une méthode pour échantillonner efficacement des densités cibles présentant une corrélation entre leurs paramètres sous une contrainte de non-négativité. Pour répondre au problème de la corrélation entre les paramètres, une marche aléatoire basée sur *MALA-sur-variété* permettra d'explorer la densité cible plus efficacement, atténuant également les difficultés causées par l'anisotropie de l'espace des paramètres. En outre, l'utilisation d'un algorithme *MCMC proximal* permettra d'assurer efficacement la non-négativité des paramètres. Les deux méthodes utilisées conjointement offrent des propriétés de mélange efficaces même si des lois a priori vagues sont utilisées. Les deux techniques seront plus détaillées dans cette même section.

Bien que cette section propose de résoudre le problème de l'estimation des paramètres de la DBN dans un contexte de Monte-Carlo général, nous n'illustrerons dans cette section que l'estimation des paramètres dans le cas bi-dimensionnel pour des raisons de clarté. Les résultats seront ensuite généralisés dans le cas des espaces de grande dimension dans le chapitre 3, où les paramètres d'un mélange de DBN sont estimés.

2.4.1 MALA sur variété

En raison de la forme complexe de la distribution de la loi cible, le processus d'échantillonnage doit être adaptée à sa forme. L'objectif est de créer des échantillons en fonction de la densité a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$ définie dans (2.8). Les lecteurs intéressés peuvent trouver plus de détails sur cette méthode dans le travail original de Girolami et Calderhead [49], où un cadre général est présenté.

Marche aléatoire de Metropolis-Hastings

Un des algorithmes de type Monte Carlo par Chaîne de Markov (MCMC) couramment utilisés est l'algorithme de Metropolis-Hastings à marche aléatoire ou *Random Walk Metropolis-Hastings*(RWMH). Dans sa forme classique, la règle du mouvement

pour la loi de proposition $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ guidant la chaîne de Markov propose des candidats $\boldsymbol{\theta}^*$ tels que :

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^i + \epsilon \mathbf{z}^i \quad (2.15)$$

où \mathbf{z}^i indique un mouvement Brownien et ϵ est un facteur d'échelle. Cependant, l'utilisation d'un algorithme RWMH classique est déconseillée pour explorer un espace de paramètre anisotrope car les différentes dimensions se comportent de façons très différentes, et il est difficile voire impossible de choisir un facteur d'échelle ϵ qui convienne à toutes les dimensions à la fois et peut conduire à un mélange excessivement lent. Le facteur d'échelle ϵ ne doit pas être trop grand, car les valeurs proposées peuvent être rejetées trop souvent, ni trop court, car la séquence d'échantillons pourrait présenter une grande auto-corrélation et la chaîne de Markov serait plus susceptible de se retrouver piégée dans un mode local et ne parviendrait pas à explorer la totalité de la distribution postérieure en un temps fini [94]. De plus, la taille du pas ϵ optimale peut ne pas être la même selon la position de la chaîne.

Echantillonneur de Gibbs hybride

L'échantillonnage de Gibbs, consiste à considérer la distribution multivariée comme composée d'un ensemble de variables aléatoires individuelles et de choisir un nouvel échantillon pour chaque dimension alternativement, plutôt que de choisir un échantillon dans toutes les dimensions à la fois. Un algorithme de Metropolis-Hastings unidimensionnel peut ensuite être utilisé pour choisir ces échantillons individuels comme expliqué précédemment. Cette technique est appelée *Metropolis-within-Gibbs* ou *échantillonneur de Gibbs hybride*.

Toutefois, les méthodes de Gibbs hybride échoueront s'il existe une corrélation entre les éléments du vecteur à échantillonner. Pour cette raison, un échantillonnage acceptant ou rejetant un couple de paramètre (μ, k) est généralement préféré à un échantillonnage de Gibbs, travaillant dans chaque dimension individuellement.

Metropolis Adjusted Langevin Algorithm (MALA)

L'algorithme de Metropolis ajusté Langevin ou *Metropolis Adjusted Langevin Algorithm* (MALA) est un algorithme récent de marche aléatoire qui propose une alternative pour atténuer ce problème. Le principe de l'algorithme MALA est de guider la marche aléatoire selon un pas de la méthode de Newton. Autrement dit, le gradient de la densité cible est incorporé dans la loi de proposition [119]. Considérons la log-densité $\mathcal{P}(\boldsymbol{\theta}^i) \triangleq \log\{p(\boldsymbol{\theta}^i|\mathbf{x})\}$, une nouvelle valeur $\boldsymbol{\theta}^*$ sera proposée telle que :

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^i + \frac{\epsilon^2}{2} \nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) + \epsilon \mathbf{z}^i \quad (2.16)$$

où $\nabla_{\theta}\mathcal{P}(\theta^i)$ est le gradient de la loi log-a posteriori au point θ^i .

Ainsi, étant donné que le gradient privilégie un déplacement de la chaîne selon la pente la plus raide, la chaîne convergera plus rapidement vers la distribution ergodique et moins d'étapes de burn-in seront nécessaires. Cependant, il est clair que la diffusion isotropique sera inefficace pour des variables θ fortement corrélées, avec des variances très différentes, car la taille du pas ϵ sera toujours adapté à la variable aléatoire avec la plus petite variance [49].

Matrice de covariance

Ce problème peut être contourné en utilisant une matrice de préconditionnement \mathbf{M} telle que [118] :

$$\theta^* = \theta^i + \frac{\epsilon^2}{2}\mathbf{M}\nabla_{\theta}\mathcal{P}(\theta^i) + \epsilon\sqrt{\mathbf{M}}\mathbf{z}^i \quad (2.17)$$

où $\sqrt{\mathbf{M}}$ peut être obtenue par diagonalisation de \mathbf{M} ou par décomposition de Cholesky de manière à ce que $\mathbf{M} = \mathbf{U}\mathbf{U}^T$ et $\sqrt{\mathbf{M}} = \mathbf{U}$.

Mais encore une fois, un niveau global de préconditionnement peut être inapproprié lorsque les régimes transitoires et stationnaires du processus de Markov sont différents [25], ou lorsque la direction des déplacements à favoriser n'est pas la même selon les différentes régions de l'espace des paramètres, comme c'est le cas ici. En effet, pour des faibles valeurs de μ , la matrice de préconditionnement doit être choisie de manière à favoriser les déplacements selon k , alors que pour des faibles valeurs de k , la matrice de préconditionnement doit permettre de favoriser les déplacements selon μ (voir fig. 2.2).

MALA sur variété Riemannienne

Nous proposons de résoudre ce problème en utilisant une méthode MCMC avancée qui utilise une matrice de préconditionnement propre à chaque position de l'espace et tenant compte des liens entre les paramètres. MALA-sur-variété ou *Manifold MALA* (mMALA) est une généralisation de MALA qui prend en compte la géométrie naturelle de la densité cible, en se basant sur la définition d'une diffusion de Langevin sur une variété Riemannienne. En d'autres termes, au lieu d'utiliser le gradient Euclidien comme dans le MALA, l'idée est de calculer le gradient dans une métrique plus appropriée en exploitant la géométrie Riemannienne de l'espace des paramètres. Lors de la simulation, la méthodologie proposée s'adaptera donc automatiquement à la structure locale de la variété Riemannienne, et offrira une convergence plus rapide et une exploration de la densité cible plus efficace.

À cet égard, Rao [114] a montré que l'espace des fonctions paramétrées de densité de probabilité est doté d'une géométrie Riemannienne naturelle. Plus précisément, la définition formelle de la distance entre deux fonctions de densité paramétrées $p(\mathbf{x}|\boldsymbol{\theta})$ et $p(\mathbf{x}|\boldsymbol{\theta} + \delta\boldsymbol{\theta})$ est définie par la forme quadratique $\delta\boldsymbol{\theta}^T \mathbf{F}(\boldsymbol{\theta}) \delta\boldsymbol{\theta}$ où $\mathbf{F}(\boldsymbol{\theta})$ est la matrice d'information de Fisher :

$$\mathbf{F}(\boldsymbol{\theta}) = -E_x \left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log\{p(\mathbf{x}|\boldsymbol{\theta})\} \right] = \begin{pmatrix} -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} \right] & -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] \\ -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] & -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} \right] \end{pmatrix} \quad (2.18)$$

avec $\boldsymbol{\theta} = \begin{pmatrix} \mu \\ k \end{pmatrix}$ et où $\mathcal{L}(\boldsymbol{\theta}^i) \triangleq \log\{p(\mathbf{x}|\boldsymbol{\theta}^i)\}$ est la log-vraisemblance.

Rao a noté que comme la matrice $\mathbf{F}(\boldsymbol{\theta})$ est par définition définie positive, elle définit une métrique locale d'une variété de Riemann [114]. Dans le cas de la fonction binomiale négative, nous avons :

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} \right] = N \left[\frac{1}{\mu} - \frac{1}{\mu + k} \right] \quad (2.19)$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} \right] = N \left[\Psi^2(k) - \frac{\mu}{k(\mu + k)} - E_x [\Psi^2(x + k)] \right] \quad (2.20)$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] = 0 \quad (2.21)$$

où Ψ^2 est la fonction trigamma $\Psi^2(x) \triangleq \frac{\partial^2}{\partial x^2} \log[\Gamma(x)]$. Plus de détails sur le calcul de $E_x [\Psi^2(x + k)]$ sont donnés dans l'annexe A.3. On obtient :

$$\mathbf{F}(\boldsymbol{\theta}) = N \begin{pmatrix} \frac{1}{\mu} - \frac{1}{\mu+k} & 0 \\ 0 & \Psi^2(k) - \frac{\mu}{k(\mu+k)} - E_x [\Psi^2(x + k)] \end{pmatrix}. \quad (2.22)$$

Comme une perspective Bayésienne est adoptée dans ce document, la probabilité jointe des données et des paramètres est utilisée pour la définition du tenseur métrique :

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta}) - \mathbf{H}(\boldsymbol{\theta}) \quad (2.23)$$

qui est l'espérance de la matrice d'information de Fisher à laquelle on rajoute l'opposé de la Hessienne propre aux lois log-a priori.

Connaissant la structure géométrique de l'espace des paramètres du modèle statistique, un schéma MCMC correct peut produire des transitions plus efficaces qui

respectent et exploitent la géométrie de la variété en utilisant un nouveau mécanisme de proposition telle que :

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^i + \frac{\epsilon^2}{2} \tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) + \epsilon \tilde{\mathbf{z}}^i. \quad (2.24)$$

où $\tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i)$ est le gradient géodésique de la log-posterior au point $\boldsymbol{\theta}^i$ [1] :

$$\tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) = \mathbf{G}^{-1}\{\boldsymbol{\theta}^i\} \nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) \quad (2.25)$$

et où le mouvement Brownien classique \mathbf{z}^i est remplacé par le mouvement Brownien sur variété de Riemann $\tilde{\mathbf{z}}^i$, pouvant être approximé par [26] :

$$\tilde{\mathbf{z}}^i \simeq \sqrt{\mathbf{G}^{-1}\{\boldsymbol{\theta}^i\}} \mathbf{z}^i \quad (2.26)$$

. Notons que la forme exacte du mouvement brownien sur une variété de Riemann contient des éléments additionnels qui sont particulièrement pertinents lorsque la courbure de la densité cible dépend de la position, mais sont négligeables dans notre cas, et le mécanisme de proposition simplifié ci-dessus, utilisé en conjonction avec la probabilité d'acceptation, définira toujours une méthode MCMC correcte qui converge vers la densité cible [49]. La fonction de proposition peut être écrite :

$$q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^i) = \mathcal{N}(\boldsymbol{\theta}^* | \tilde{\mathbf{m}}(\boldsymbol{\theta}^i, \epsilon), \epsilon^2 \mathbf{G}^{-1}\{\boldsymbol{\theta}^i\}) \quad (2.27)$$

où

$$\tilde{\mathbf{m}}(\boldsymbol{\theta}^i, \epsilon) = \boldsymbol{\theta}^i + \frac{\epsilon^2}{2} \mathbf{G}^{-1}\{\boldsymbol{\theta}^i\} \nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i). \quad (2.28)$$

Il est alors clair que le mécanisme de proposition se déplace dans \mathbb{R}^2 selon la métrique de Riemann plutôt que selon la distance Euclidienne classique. Le paramètre $\boldsymbol{\theta}^*$ sera alors accepté avec la probabilité

$$a(\boldsymbol{\theta}^* | \boldsymbol{\theta}^i) = \min(1, r(\boldsymbol{\theta}^* | \boldsymbol{\theta}^i)) \quad (2.29)$$

où

$$r(\boldsymbol{\theta}^* | \boldsymbol{\theta}^i) = \frac{P[\mathbf{x} | \boldsymbol{\theta}^*] \pi(\boldsymbol{\theta}^*) q(\boldsymbol{\theta}^i | \boldsymbol{\theta}^*)}{P[\mathbf{x} | \boldsymbol{\theta}^i] \pi(\boldsymbol{\theta}^i) q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^i)} \quad (2.30)$$

et où $P[\mathbf{x} | \cdot]$, $\pi(\cdot)$ et $q(\cdot | \cdot)$ ont été définis en (2.9), (2.14) et (2.27) respectivement.

La méthode d'échantillonnage proposée a plusieurs avantages par rapport à la méthode RWMH classique. Tout d'abord, la chaîne converge plus rapidement vers la distribution ergodique qu'avec une méthode RWMH donc moins d'itérations de

chauffage sont nécessaires. D'autre part, la taille du pas ϵ est compensée par le terme de déviation qui prend en compte la forme de la distribution cible. En d'autres termes, l'algorithme est plus souple par rapport aux différentes tailles de pas requises pour les phases transitoires et stationnaires pour le comportement différent dans chacune des directions de l'espace des paramètres. Troisièmement, la distribution de la proposition $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)$ est plus proche de la distribution a posteriori, ce qui améliore les propriétés de mélange de l'échantillonneur. Ces résultats seront illustrés par des simulations dans la partie 2.5.1.

Notons que les méthodes qui exploitent les gradients de la fonction cible, comme MALA ou MALA-sur-variété exigent une différentiabilité en chaque point de l'espace des paramètres et ne sont pas directement adaptés à des modèles non-différentiables. Malheureusement, les paramètres de la DBN sont limités à des valeurs positives et la densité cible est donc seulement différentiable à l'intérieur de l'ensemble convexe $(\mathbb{R}^+ \times \mathbb{R}^+)$. La section suivante présente un moyen efficace de remédier à ce problème.

2.4.2 MALA Proximal sur variété

Comme dit précédemment, les algorithmes MALA et MALA-sur-variété exigent une différentiabilité en chaque point de l'espace des paramètres. Toutefois, la DBN a deux paramètres restreints à des valeurs positives, $\mu \in \mathbb{R}^+$ et $k \in \mathbb{R}^+$. En pratique, simuler avec précision près de la frontière mènera souvent à des candidats qui se trouvent en dehors des limites de $\mathbb{R}^+ \times \mathbb{R}^+$, menant l'algorithme à l'échec.

Une solution simple serait de rejeter toutes ces valeurs proposées, mais dans ce cas la taille du pas peut devenir extrêmement faible, et pourrait considérablement ralentir l'algorithme.

Une meilleure solution serait de considérer un schéma d'échantillonnage à réflexion. Dans notre cas, cette technique consiste à faire *rebondir* les mouvements contre les axes ($\mu = 0$) et ($k = 0$) pour conserver les échantillons candidats successifs dans les limites (voir fig. 2.3).

Quand un nouveau $\boldsymbol{\theta}^*$ est tiré, sa valeur absolue $|\boldsymbol{\theta}^*| = (|\mu^*|, |k^*|)$ est considérée comme valeur candidate pour $\boldsymbol{\theta}$.

Le principal défaut de cette méthode est que le candidat peut ne pas se trouver dans la direction du gradient, surtout si la taille du pas est trop grande.

Dans cette partie, nous abordons ce problème en recourant à des opérateurs

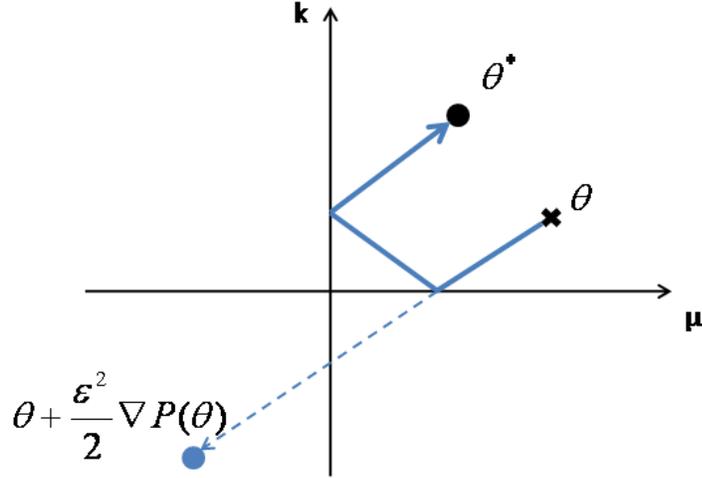


FIGURE 2.3 – Echantillonnage à réflexion

proximaux. Ceux-ci ont été largement utilisés en régularisation variationnelle [28], et ont été récemment introduit dans les méthodes Bayésiennes [106, 20].

Commençons par considérer la fonction indicatrice h :

$$h : \boldsymbol{\theta} \mapsto \begin{cases} 0, & \text{si } \boldsymbol{\theta} \in (\mathbb{R}^+)^2 \\ -\infty & \text{si } \boldsymbol{\theta} \notin (\mathbb{R}^+)^2 \end{cases} \quad (2.31)$$

et posons $f \triangleq \mathcal{P} + h$. Nous avons l'égalité $f \equiv \mathcal{P}$ sur $(\mathbb{R}^+)^2$ et contrairement à la log-posterior \mathcal{P} , f est définie sur \mathbb{R}^2 entier. Notons tout de même que f n'y est pas différentiable.

Dans [106], il est proposé une généralisation du MALA à un large éventail de distributions non-différentiables en exploitant les applications proximales d'une manière similaire à l'utilisation du gradient par le MALA classique. Ainsi, étant donné $\boldsymbol{\theta}^i$, un candidat $\boldsymbol{\theta}^*$ est proposé selon la transition :

$$\boldsymbol{\theta}^* | \boldsymbol{\theta}^i \sim \mathcal{N}(\text{prox}_f^{\delta/2}(\boldsymbol{\theta}^i), \delta \mathbb{I}_2). \quad (2.32)$$

Dans notre cas, f est convexe et son application proximale peut être efficacement approximée en utilisant la méthode de division proximale *progressive-rétrogressive* (*Forward-backward*) [103]. Ainsi l'expression $f \triangleq \mathcal{P} + h$ où \mathcal{P} est une fonction

différentiable sur $(\mathbb{R}^+)^2$ et h est une fonction possédant une application de proximité explicite, suggère l'approximation :

$$\text{prox}_f^{\delta/2}(\boldsymbol{\theta}) \approx \text{prox}_h^{\delta/2} \left(\boldsymbol{\theta} + \frac{\delta}{2} \nabla \mathcal{P}(\boldsymbol{\theta}) \right). \quad (2.33)$$

Ceci entraîne un nouveau mécanisme de proposition pour MALA :

$$\boldsymbol{\theta}^* = \text{prox}_h^{\delta/2} \left(\boldsymbol{\theta}^i + \frac{\delta}{2} \nabla \mathcal{P}(\boldsymbol{\theta}^i) \right) + \sqrt{\delta} w, w \sim N(0, \mathbb{I}_2). \quad (2.34)$$

Il est important de noter que comme h est une fonction indicatrice, son opérateur de proximité $\text{prox}_h^{\delta/2}$ peut être considéré comme un opérateur de projection sur le convexe $(\mathbb{R}^+)^2$. Ce mécanisme de proposition permet de ne proposer que des échantillons positifs.

Les résultats précédents peuvent être étendus facilement à la considération du gradient géodésique $\tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i)$.

$$\boldsymbol{\theta}^* = \text{prox}_h^{\delta/2} \left(\boldsymbol{\theta}^i + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) \right) + \sqrt{\delta} w, w \sim N(0, \mathbf{G}^{-1}\{\boldsymbol{\theta}^i\}). \quad (2.35)$$

Le paramètre $\boldsymbol{\theta}^*$ sera alors accepté avec la probabilité

$$a(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) = \min(1, r(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)) \quad (2.36)$$

où

$$r(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) = \frac{P[\mathbf{x}|\boldsymbol{\theta}^*] \pi(\boldsymbol{\theta}^*) q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*)}{P[\mathbf{x}|\boldsymbol{\theta}^i] \pi(\boldsymbol{\theta}^i) q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i)} \quad (2.37)$$

avec $P[\mathbf{x}|\cdot]$ et $\pi(\cdot)$ définis en (2.9) et (2.14) respectivement et où

$$\begin{cases} q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^i) = p_{\mathcal{N}} \left(\boldsymbol{\theta}^* | \text{prox}_h^{\delta/2} \left(\boldsymbol{\theta}^i + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^i) \right), \delta \mathbf{G}^{-1}\{\boldsymbol{\theta}^i\} \right) \\ q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^*) = p_{\mathcal{N}} \left(\boldsymbol{\theta}^i | \text{prox}_h^{\delta/2} \left(\boldsymbol{\theta}^* + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}^*) \right), \delta \mathbf{G}^{-1}\{\boldsymbol{\theta}^*\} \right) \end{cases}. \quad (2.38)$$

L'algorithme *progressif-rétrogressif* peut alors être considéré comme une étape progressive, dans la direction donnée par le gradient de f , suivie d'une étape rétrogressive où une projection est faite selon h , sur le domaine positif (voir fig. 2.4).

La généralisation du gradient par un sous-gradient permet alors de restreindre les échantillons proposés dans l'espace différentiable et donc d'explorer efficacement l'espace des paramètres.

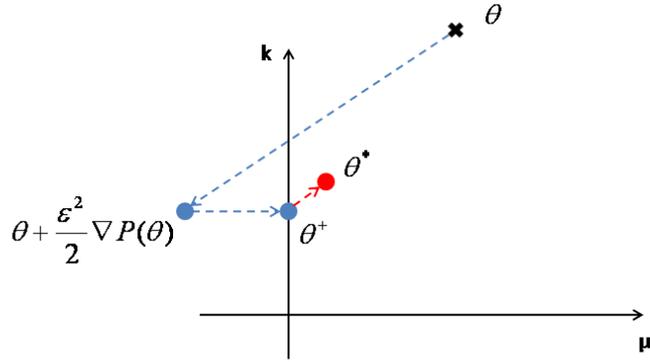


FIGURE 2.4 – Schéma progressif-rétrogressif

2.4.3 Algorithme d'échantillonnage

L'algorithme final est détaillé dans le listing 1.

Algorithm 1 Proximal MALA-sur-variété

1. Initialiser $\boldsymbol{\theta}^{(1)}$ aléatoirement
 - for** $i = 2, \dots$ **to** I **do**
 - *Mise à jour de $\boldsymbol{\theta}$* —
 - 2. Simuler $\boldsymbol{\theta}^* \sim \mathcal{N}\left(\text{prox}_h^{\delta/2}\left(\boldsymbol{\theta}^{(i-1)} + \frac{\delta}{2}\tilde{\nabla}_{\boldsymbol{\theta}}\mathcal{P}(\boldsymbol{\theta}^{(i-1)})\right), \delta\mathbf{G}^{-1}\{\boldsymbol{\theta}^{(i-1)}\}\right)$.
 - 3. Évaluer $r(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$ selon 2.37
 - 4. Tirer $u \sim \mathcal{U}(0, 1)$.
 - if** $u < r(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$ **then**
 - *accepter* —
 - 5. $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^*$
 - else**
 - *rejeter* —
 - 6. $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^{(i-1)}$
 - end if**
 - end for**
 7. Supprimer les premiers échantillons (chauffe)
-

2.5 Résultats et discussions

Des expériences ont été menées dans le cas de l'échantillonnage d'une DBN à une seule composante. La méthode proposée a été validée sur des données synthétiques et les performances de l'algorithme ont été évaluées et comparées à d'autres méthodes. Puis les expériences ont été réalisées sur une image TEP réelle pré-segmentée afin de valider le modèle proposé. Cette section présente les résultats de ces simulations et des expérimentations.

2.5.1 Estimation des paramètres sur données synthétiques

Les résultats de différentes méthodes ont été comparés sur des données synthétiques. L'influence des paramètres suivants ont été analysés :

- N (Taille des observations) $\in \{50, 100, 500, 1000\}$.
- μ théorique (Moyenne des observations) $\in \{1.0, 5.0, 10.0, 20.0\}$.
- k théorique (Paramètres de dispersion inverse) $\in \{0.5, 1.0, 2.0, 3.0\}$.

Six algorithmes d'échantillonnage ont été comparés :

GIBBS correspondant à l'estimateur Metropolis-dans-Gibbs de l'état de l'art proposé par [104] et [69].

RWMH Algorithme classique de Metropolis-Hastings de type marche aléatoire (*Random Walk Metropolis-Hastings*).

MALA Algorithm de Metropolis ajusté Langevin (*Metropolis Adjusted Langevin Algorithm*).

prox-MALA MALA proximal proposé par [106].

mMALA MALA-sur-variété (*Manifold MALA*) [49].

prox-mMALA MALA-sur-variété proximal proposé dans cette thèse.

Le nombre total d'itérations par chaîne a été fixé à 101000 pour chaque méthode, comprenant à chaque fois 1000 itérations de chauffage.

Un facteur d'échelle ϵ adapté est estimé au cours de cette période de chauffe, où les admissions et les rejets des derniers candidats sont enregistrés, et ϵ est adapté de telle sorte que le taux d'acceptation tende vers 0.5. L'initialisation des chaînes a été faite aléatoirement. A l'exception des méthodes prox-MALA et prox-mMALA, la positivité des paramètres a été assurée par un échantillonnage par réflexion comme indiqué dans 2.4.2.

Enfin, des lois a priori non informatives ont été considérées dans cette étude. Les hyper-paramètres ont été fixés à $a_\mu = 0$, $b_\mu = -0.001$, $a_k = 0$ et $b_k = -0.001$.

Convergence vers la distribution ergodique

La figure 2.5 montre les 1000 premiers pas des trois différents types de marche aléatoire, partant du même point de départ : RWMH (bleu), MALA (vert) et mMALA (rouge). On peut voir que la chaîne associée à la méthode mMALA (chemin rouge)

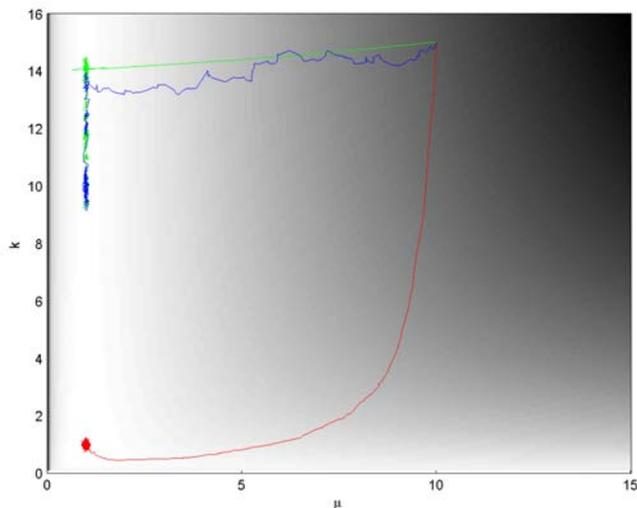


FIGURE 2.5 – Trois types de marche aléatoire initialisés au même point ($\mu_0 = 10$; $k_0 = 15$) : RWMH (bleu), MALA (vert) et mMALA (rouge). (Valeur théorique : $\mu = 1$; $k = 1$; 1000 itérations)

atteint le haut de la distribution ergodique plus rapidement que les autres chaînes. On remarque aussi que, grâce à l'influence de la pente, la chaîne MALA atteint rapidement la queue de la distribution a posteriori, mais comme la taille du pas est adaptée à la composante avec la plus petite variance, la chaîne a du mal à explorer la densité entière.

Pour une comparaison plus quantitative de la convergence des différents algorithmes, les simulations ont été répétées 20 fois pour chaque méthode, et le facteur de réduction d'échelle potentiel multivarié (*Multivariate Potential Scale Reduction Factor* MPSRF) a été mesuré sur les 20 expériences durant les premières 11000 itérations. Les figures 2.6, 2.7, 2.8 et 2.9 rapportent l'évolution du MPSRF au long des itérations pour les six méthodes, pour $N = 50$, $N = 100$, $N = 500$ et $N = 1000$ observations.

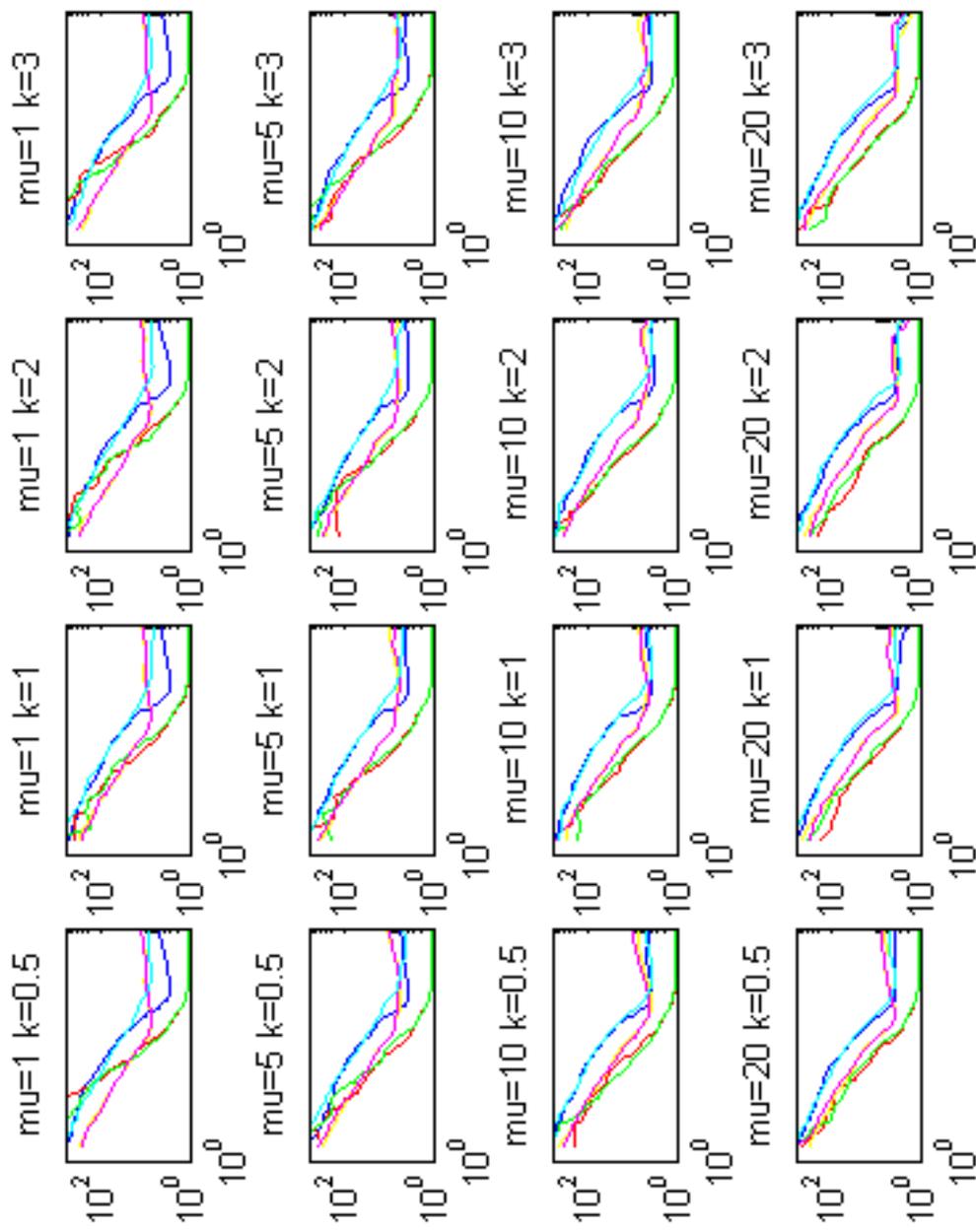


FIGURE 2.6 – Evolution du MPSRF le long des premières 11000 iterations pour $N = 50$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).

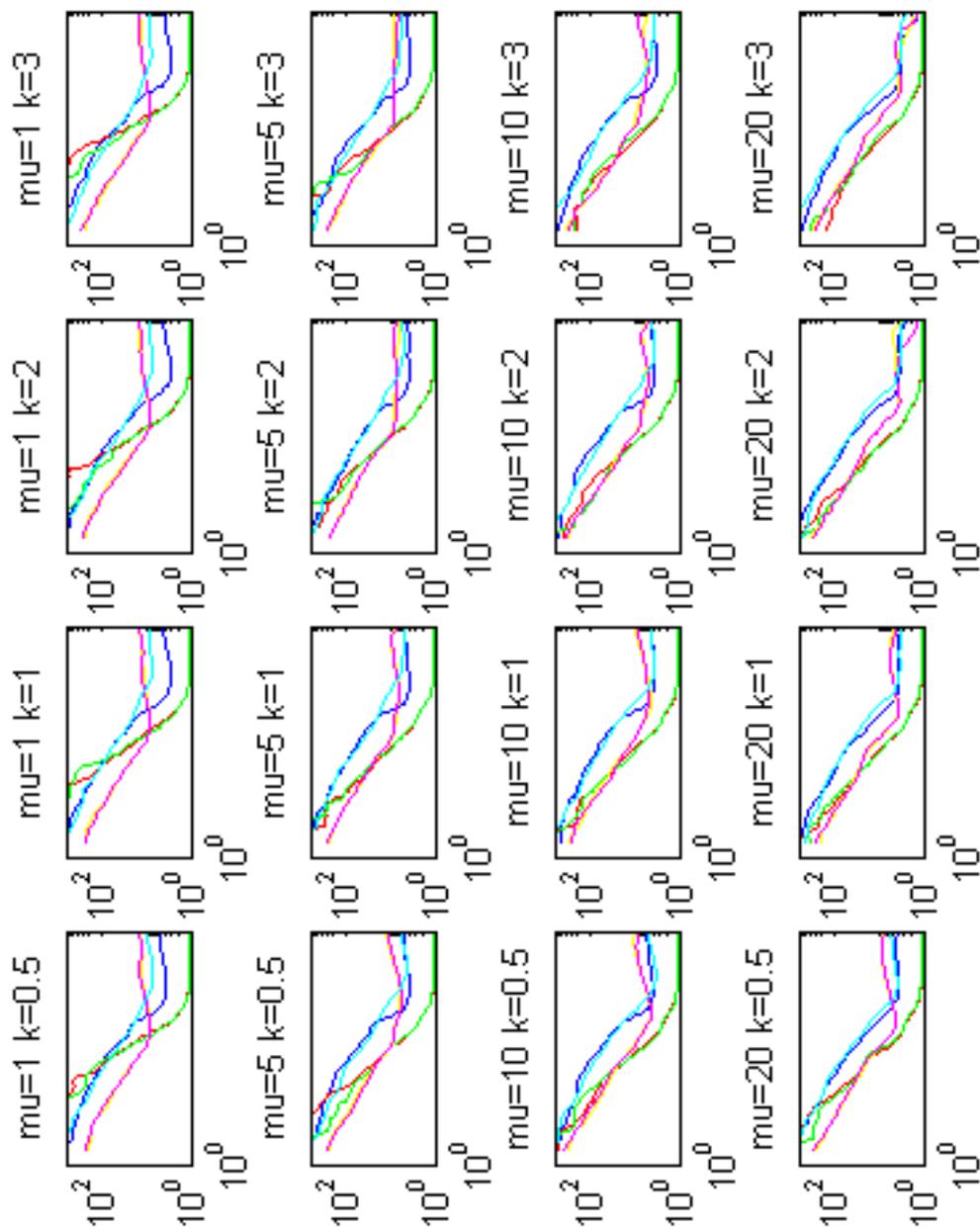


FIGURE 2.7 – Evolution du MPSRF le long des premières 11000 iterations pour $N = 100$ (échelle logarithmique). Bleu : GIBBS; cyan : RWMH; magenta : MALA; jaune : prox-MALA; vert : mMALA; rouge : prox-mMALA (échelles logarithmiques).

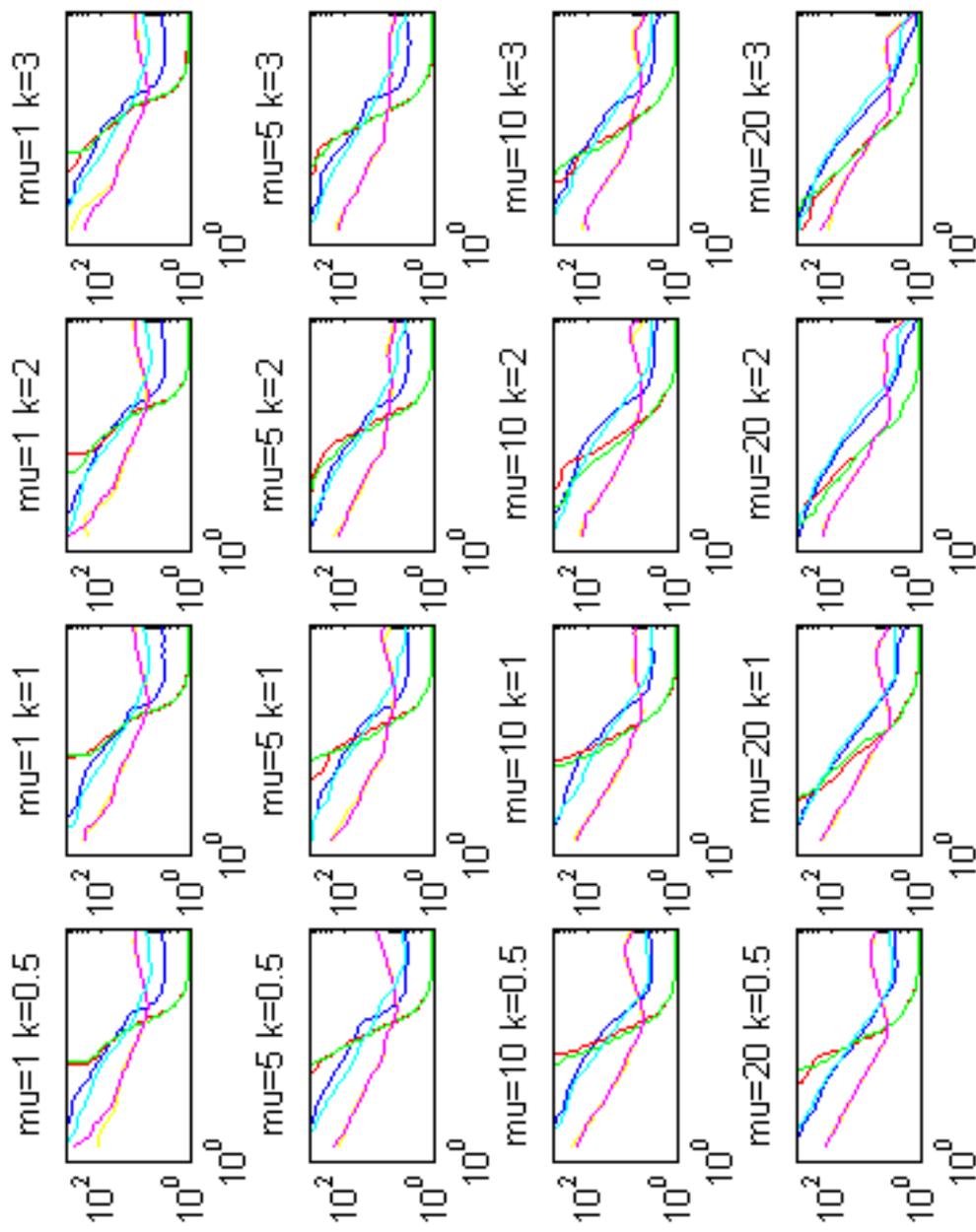


FIGURE 2.8 – Evolution du MPSRF le long des premières 11000 iterations pour $N = 500$ (échelle logarithmique). Bleu : GIBBS ; cyan : RWMH ; magenta : prox-MALA ; vert : mMALA ; rouge : prox-mMALA (échelles logarithmiques).

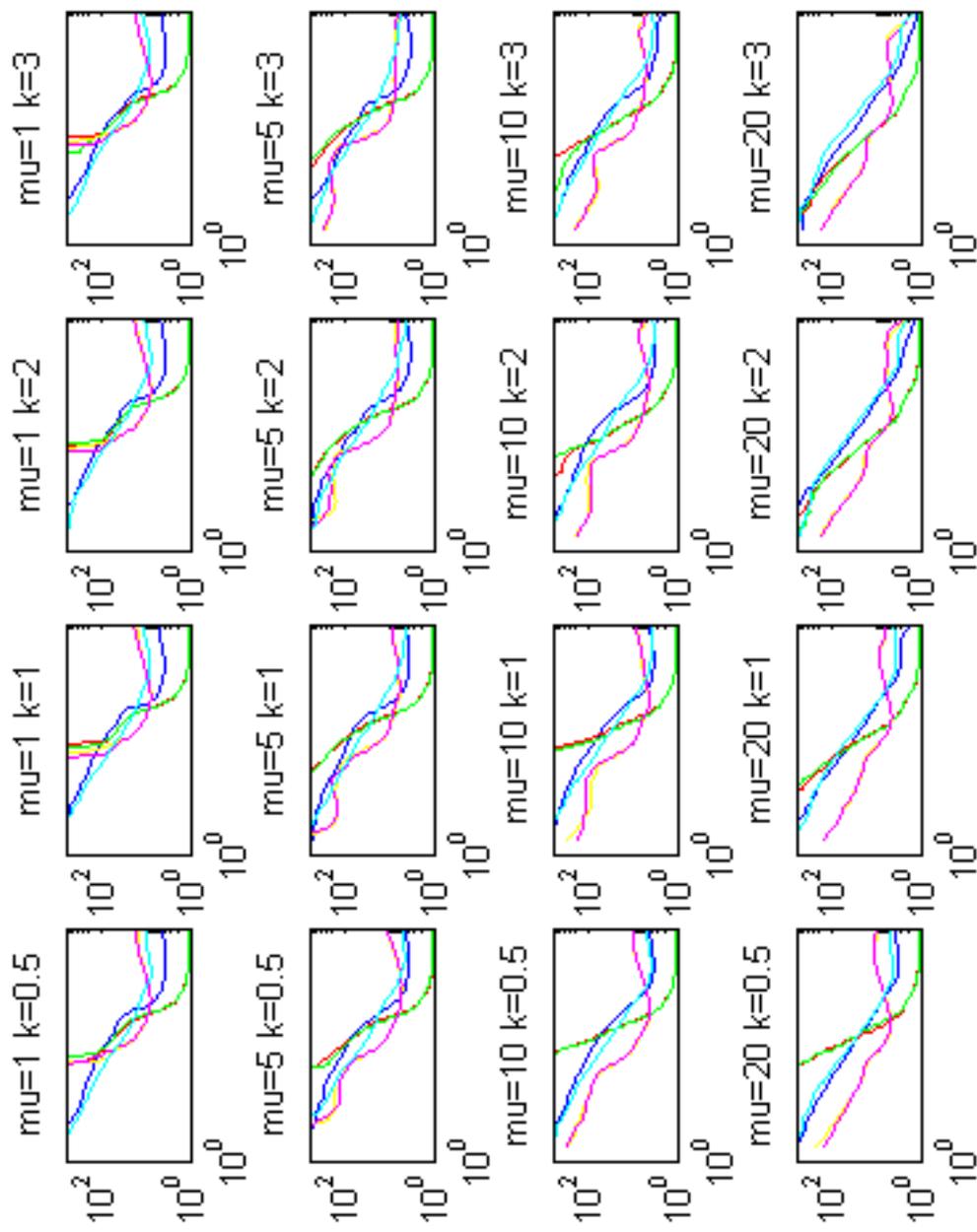


FIGURE 2.9 – Evolution du MPSRF le long des premières 11000 iterations pour $N = 1000$ (échelle logarithmique). Bleu : GIBBS; cyan : RWMH; magenta : MALA; jaune : prox-MALA; vert : mMALA; rouge : prox-mMALA (échelles logarithmiques).

Contrairement aux autres méthodes, les MPSRF associés aux méthodes mMALA et prox-mMALA deviennent faibles dès les premières itérations.

Pour plus de précision, les tableaux 2.1, 2.2, 2.3 et 2.4 indiquent le nombre d'itérations nécessaire pour atteindre un MPSRF en dessous de 1.2.

Paramètres		$N = 50$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	-	-	-	-	865	827	
	1	-	-	-	-	927	846	
	2	-	-	-	-	850	828	
	3	-	-	-	-	884	856	
5	0.5	-	-	-	-	819	848	
	1	-	-	-	-	824	833	
	2	-	-	-	-	786	790	
	3	-	-	-	-	806	816	
10	0.5	-	-	-	-	826	764	
	1	-	-	-	-	833	760	
	2	59774	-	-	-	777	740	
	3	-	96523	-	-	753	796	
20	0.5	-	-	-	-	866	856	
	1	37116	88951	97645	77349	763	768	
	2	29598	35090	26149	35690	731	760	
	3	20146	39968	32483	21088	689	722	

TABLE 2.1 – Nombre d’itérations nécessaire pour que MPSRF < 1.2 ($N = 50$). Le caractère “-” signifie que le nombre nécessaire d’itérations dépasse 101000.

Paramètres		$N = 100$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	-	-	-	-	897	942	
	1	-	-	-	-	937	943	
	2	-	-	-	-	916	950	
	3	-	-	-	-	894	930	
5	0.5	-	-	-	-	896	891	
	1	-	-	-	-	963	958	
	2	-	-	-	-	879	918	
	3	-	-	-	-	922	913	
10	0.5	-	-	-	-	890	941	
	1	71183	-	-	-	910	872	
	2	57575	-	-	-	899	870	
	3	57661	-	-	72627	908	894	
20	0.5	-	-	-	-	1082	1042	
	1	48566	80808	59490	-	979	922	
	2	31103	68547	15389	36630	864	878	
	3	16253	30144	16962	19514	900	843	

TABLE 2.2 – Nombre d'itérations nécessaire pour que $\text{MPSRF} < 1.2$ ($N = 100$). Le caractère "-" signifie que le nombre nécessaire d'itérations dépasse 101000.

Paramètres		$N = 500$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	-	-	-	-	1343	1324	
	1	-	-	-	-	1310	1363	
	2	-	-	-	-	1309	1320	
	3	-	-	-	-	1287	1330	
5	0.5	-	-	-	-	1265	1286	
	1	-	-	-	-	1300	1264	
	2	-	-	-	-	1262	1271	
	3	-	-	-	-	1235	1230	
10	0.5	-	-	-	-	1238	1250	
	1	63897	-	-	-	1232	1167	
	2	55436	97375	-	-	1217	1195	
	3	32197	90273	90185	93339	1167	1188	
20	0.5	70685	-	-	-	1478	1481	
	1	32598	82663	90236	77189	1299	1440	
	2	14533	26563	30021	36299	1208	1214	
	3	9972	22978	20186	21643	1217	1225	

TABLE 2.3 – Nombre d’itérations nécessaire pour que MPSRF < 1.2 ($N = 500$). Le caractère “-” signifie que le nombre nécessaire d’itérations dépasse 101000.

Paramètres		$N = 1000$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	-	-	-	-	1713	1758	
	1	-	-	-	-	1664	1677	
	2	-	-	-	-	1620	1630	
	3	-	-	-	-	1755	1693	
5	0.5	-	-	-	-	1525	1490	
	1	-	-	-	-	1478	1499	
	2	-	-	-	-	1520	1460	
	3	-	-	-	-	1508	1465	
10	0.5	-	-	-	-	1501	1501	
	1	87001	-	-	-	1443	1456	
	2	48070	-	-	-	1421	1416	
	3	31585	88826	95968	90122	1356	1397	
20	0.5	84641	-	-	-	1773	1915	
	1	28731	75344	80907	83390	1638	1593	
	2	17205	29557	34118	33262	1423	1484	
	3	11433	23531	20010	22112	1406	1409	

TABLE 2.4 – Nombre d’itérations nécessaire pour que $\text{MPSRF} < 1.2$ ($N = 1000$). Le caractère ”-” signifie que le nombre nécessaire d’itérations dépasse 101000.

On voit que seules quelques itérations sont nécessaires pour que les algorithmes mMALA et prox-mMALA convergent, alors que la convergence est beaucoup plus lente pour les autres algorithmes. Les méthodes GIBBS, RWMH, MALA et prox-MALA semblent moins efficaces pour des petites valeurs de μ ou de k . En particulier pour $\mu = 1$ et $\mu = 5$, aucune de ces méthodes n'a convergé vers la distribution ergodique malgré les grandes tailles de la chaîne MCMC. Les méthodes mMALA et prox-mMALA ne semblent pourtant pas suivre cette tendance et convergent efficacement quels que soient les paramètres μ et k .

Correlation entre les échantillons

Le facteur d'autocorrelation a été calculé pour chacun des algorithmes sur les 100000 itérations après le burn-in. Les coefficients ont été rapportés pour des décalages compris entre 0 et 10000 itérations.

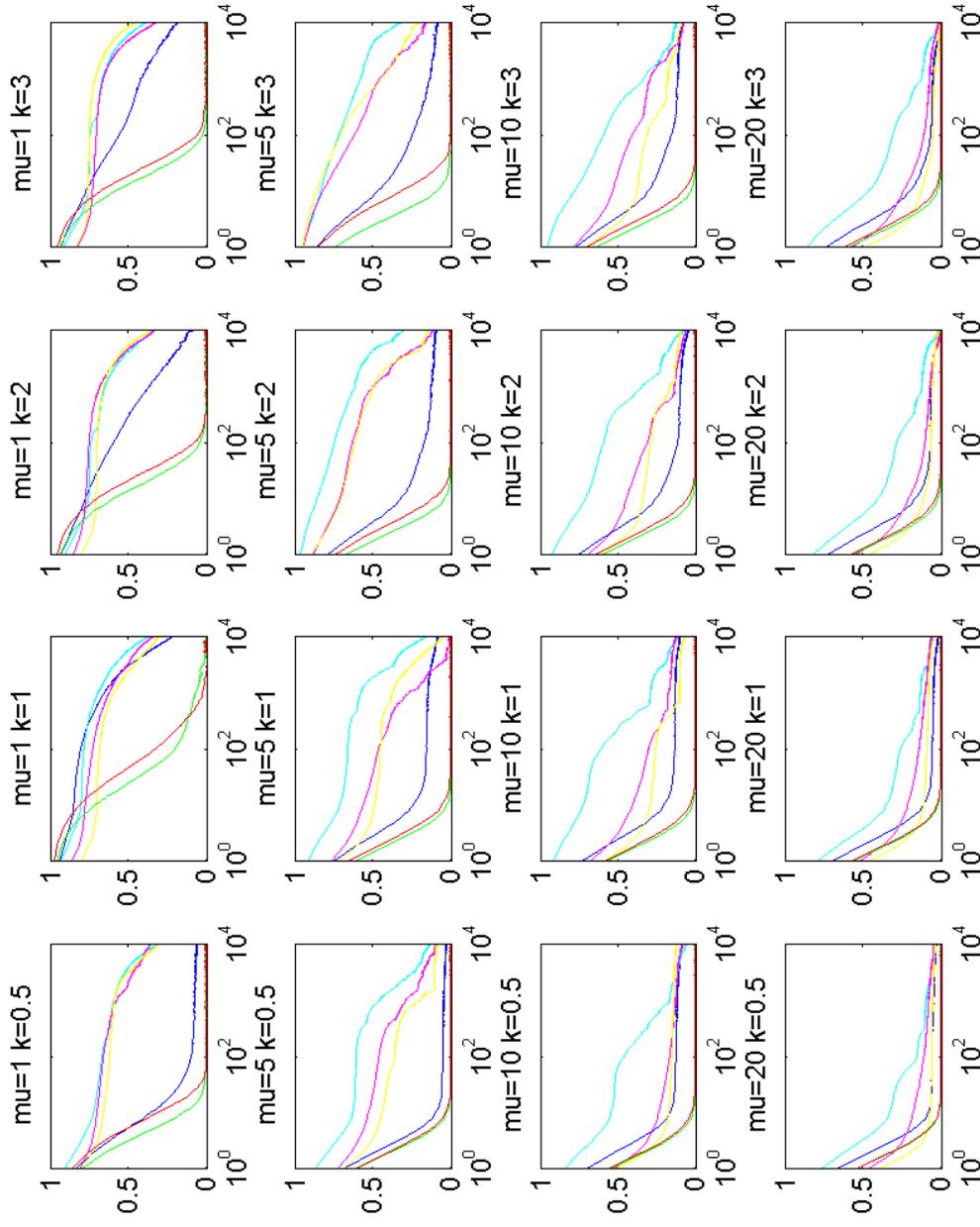


FIGURE 2.10 – Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 50$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : mMALA ; vert : prox-MALA ; rouge : prox-mMALA.

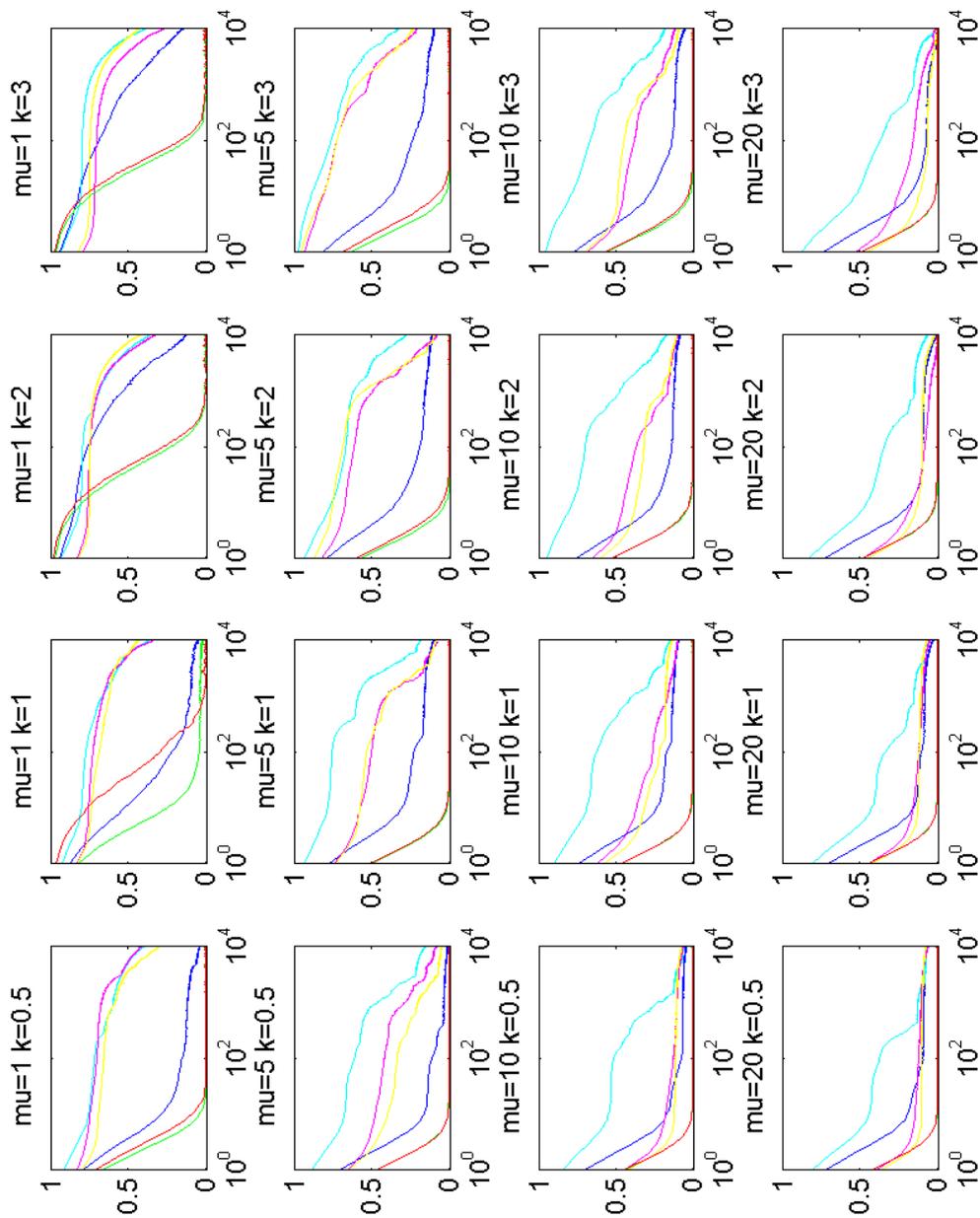


FIGURE 2.11 – Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 100$. Bleu : GIBBS ; cyan : RWMH ; magenta : prox-MALA ; jaune : mMALA ; vert : mMALA ; rouge : prox-mMALA.

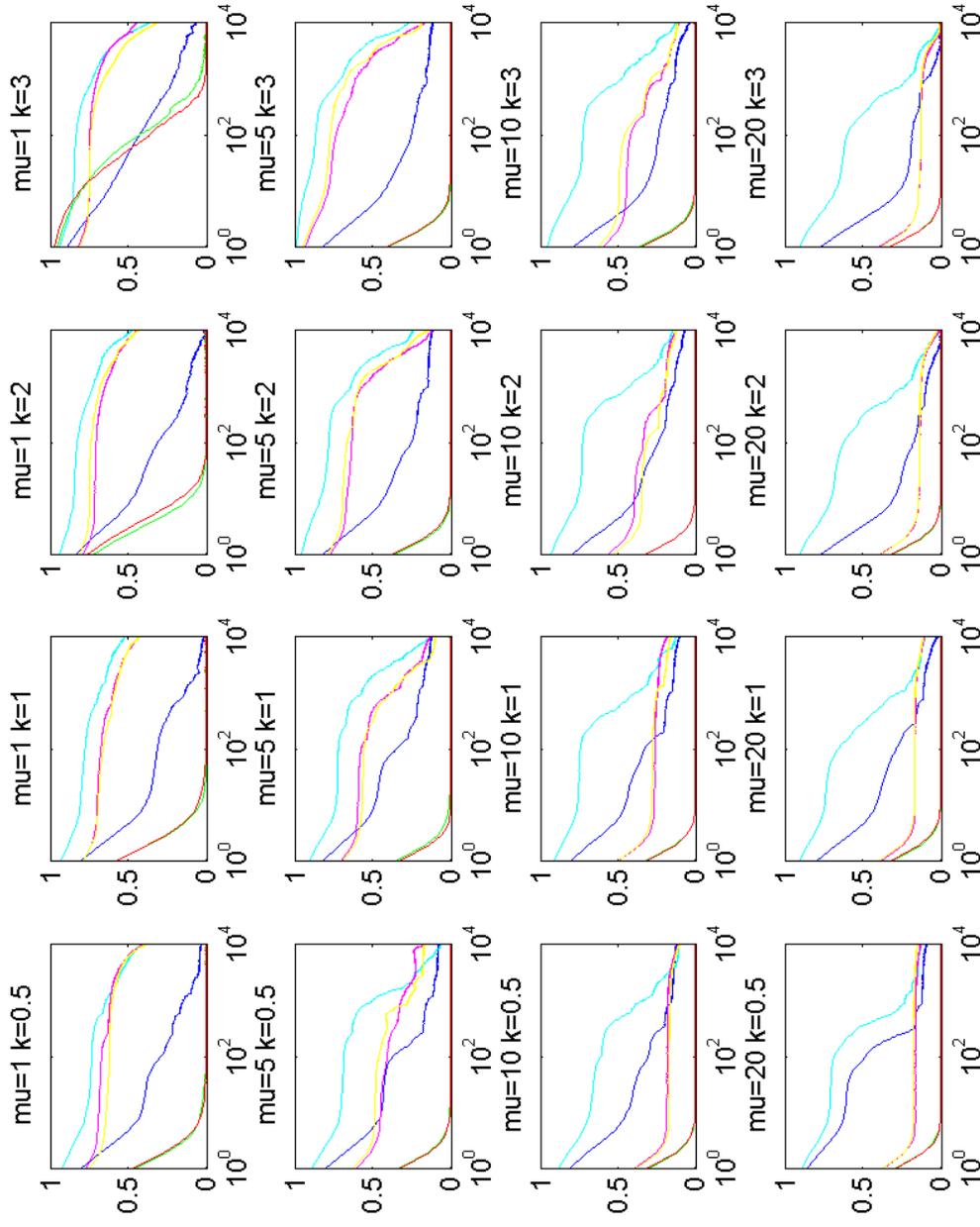


FIGURE 2.12 – Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 500$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.

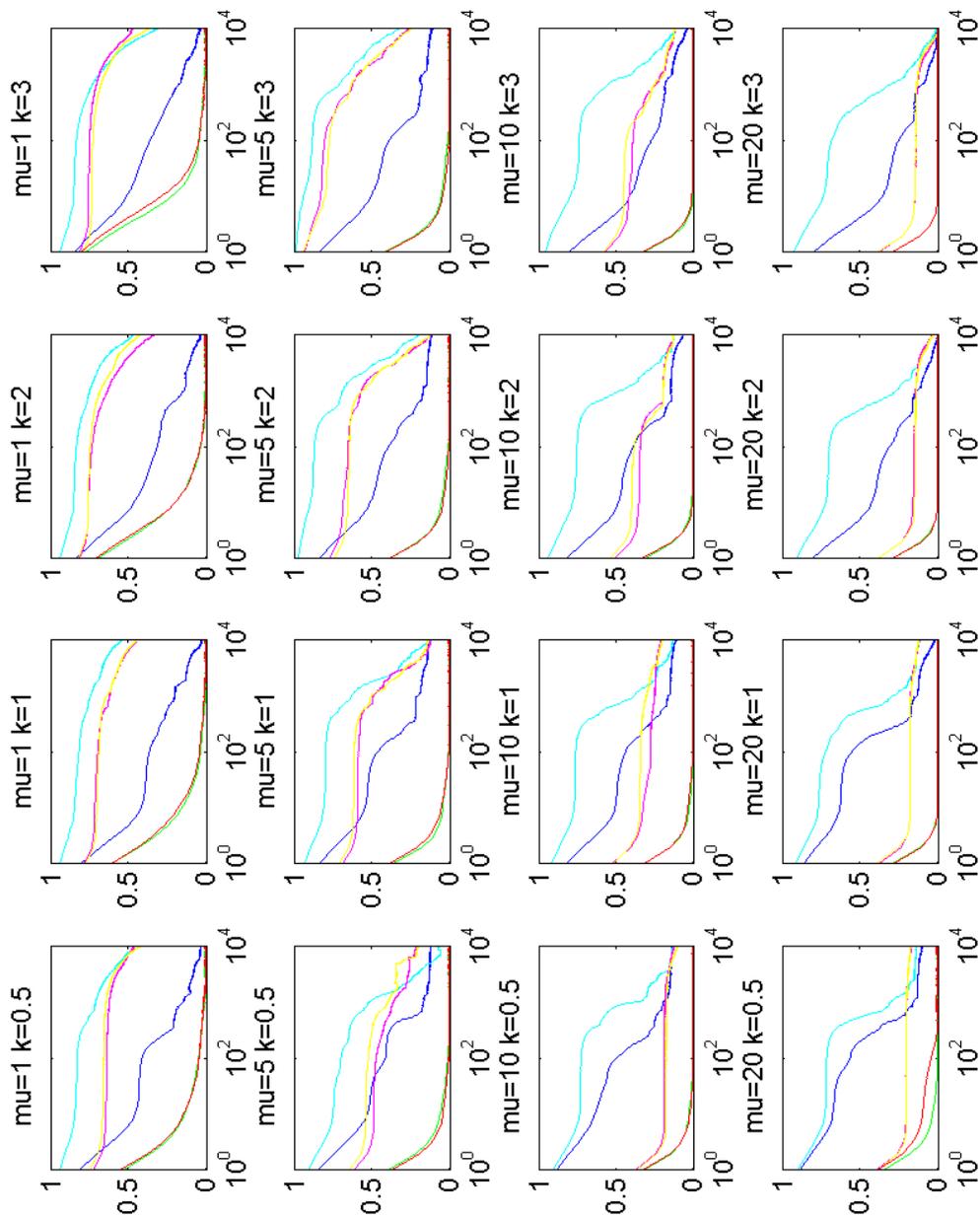


FIGURE 2.13 – Evolution de l'autocorrélation pour des décalages compris entre 0 et 10000 itérations pour $N = 1000$. Bleu : GIBBS ; cyan : RWMH ; magenta : MALA ; jaune : prox-MALA ; vert : mMALA ; rouge : prox-mMALA.

Les méthodes mMALA et prox-mMALA proposent des échantillons décorrélés après moins d'itérations que nécessaire pour les autres méthodes.

Taille effective de l'échantillon

La taille effective de l'échantillon (*effective sample size*, ESS) a été déduite des mesures d'auto-corrélation ci-dessus. Les tables 2.5, 2.6, 2.7 et 2.8 montrent le nombre d'échantillons effectifs résultant des 100000 itérations après la période de chauffe.

Paramètres		$N = 50$							
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA		
1	0.5	76	11	11	11	8848	5510		
	1	12	9	11	12	214	414		
	2	24	10	10	10	2696	2017		
	3	17	9	10	8	2489	1666		
5	0.5	147	21	30	42	16936	19030		
	1	46	14	61	26	17268	17023		
	2	47	11	19	19	16935	10020		
	3	47	11	19	16	11629	6071		
10	0.5	51	43	48	39	18440	16625		
	1	45	25	34	53	22221	18835		
	2	71	26	50	42	18249	15391		
	3	50	26	35	40	15548	15417		
20	0.5	134	95	88	121	20914	20147		
	1	164	57	65	82	18537	18093		
	2	152	64	155	149	19882	15340		
	3	244	68	119	316	16394	17381		

TABLE 2.5 – Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 50$)

Paramètres		$N = 100$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	67	10	10	11	17510	13737	
	1	61	10	10	9	170	409	
	2	17	10	10	9	776	672	
	3	17	9	11	9	1544	911	
5	0.5	216	20	32	54	23577	23538	
	1	38	15	28	28	27002	17482	
	2	39	12	21	21	17979	16584	
	3	41	10	14	14	16326	11613	
10	0.5	94	44	61	57	22298	21011	
	1	46	22	40	32	19974	24921	
	2	49	17	37	35	18057	15779	
	3	59	17	30	34	17430	19637	
20	0.5	65	56	60	59	24661	24543	
	1	85	47	68	56	30121	22826	
	2	105	42	323	88	22139	18471	
	3	251	50	99	256	16965	20891	

TABLE 2.6 – Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 100$)

Paramètres		$N = 500$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	77	10	10	10	20702	17854	
	1	85	8	10	10	20225	10823	
	2	81	8	9	9	12314	8753	
	3	36	9	9	10	225	445	
5	0.5	51	24	21	25	27962	25875	
	1	33	15	23	27	26106	30601	
	2	36	12	17	15	24889	24569	
	3	37	10	13	13	25772	23491	
10	0.5	36	28	35	38	26889	25505	
	1	38	22	23	27	26636	29453	
	2	49	20	30	32	28181	28084	
	3	54	19	32	31	26736	25481	
20	0.5	43	29	35	32	30859	27042	
	1	67	31	38	37	25346	27834	
	2	156	43	67	62	25211	26447	
	3	212	49	110	95	27753	27800	

TABLE 2.7 – Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 500$)

Paramètres		$N = 1000$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	64	9	9	9	1897	1966	
	1	48	8	9	9	3980	2326	
	2	53	8	10	9	2455	3575	
	3	53	10	9	9	1355	2429	
5	0.5	35	25	19	17	11241	21392	
	1	32	16	21	20	14589	15097	
	2	35	12	17	16	13238	11478	
	3	36	10	12	12	12386	16955	
10	0.5	33	23	34	37	25722	30258	
	1	36	23	22	21	25178	19698	
	2	43	22	30	31	24573	25068	
	3	52	19	31	30	26456	25937	
20	0.5	37	27	28	28	18717	4880	
	1	58	28	36	36	22070	27490	
	2	105	38	54	59	25797	24907	
	3	162	42	95	86	27676	25821	

TABLE 2.8 – Nombre d'échantillons effectifs parmi les 100000 itérations après la chauffe ($N = 1000$)

On constate que les algorithmes mMALA et prox-mMALA offrent la possibilité de fournir des échantillons nettement plus décorrésés que les autres algorithmes, avec un faible avantage pour les valeurs de μ grandes. Il est également intéressant de noter que l'algorithme GIBBS offre de meilleures performances que les algorithmes RWMH, MALA et prox-MALA. Ce dernier résultat peut être expliqué par le fait que dans l'algorithme de GIBBS, les propositions de chaque paramètre peuvent être acceptées ou rejetées indépendamment l'une de l'autre, permettant ainsi à la marche aléatoire de se déplacer à chaque itération dans une seule des deux directions de l'espace des paramètres. En revanche, les algorithmes RWMH, MALA et prox-MALA proposent des paires d'échantillons qui auront plus de risque d'être rejetées, étant donné l'anisotropie de la densité cible. Afin de respecter un ratio d'acceptation adéquat, la taille des pas est alors réduite, comme expliqué précédemment dans la figure 2.5, et l'algorithme produit des échantillons plus corrélés. Ce défaut n'est pas retrouvé dans les algorithmes utilisant le gradient géodésique (mMALA et prox-mMALA). On remarque également qu'étant donné une technique d'échantillonnage (MALA et prox-MALA d'une part, et mMALA et prox-mMALA d'autre part), les méthodes de réflexion et les méthodes proximales donnent ici des résultats comparables.

Nombre moyen d'échantillons indépendants tirés par seconde

A partir des résultats précédents de l'ESS et du temps de calcul nécessaire pour effectuer les simulations (sur un processeur Intel Core i7 4 coeurs, 3.8 GHz, 8Go de RAM), le nombre moyen d'échantillons indépendants tirés par seconde a été calculé pour chaque méthode. Ceux-ci sont présentés dans les tables 2.9, 2.10, 2.11 et 2.12.

Paramètres		$N = 50$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	1.62	0.12	0.11	0.22	66.77	76.36	
	1	0.25	0.13	0.12	0.23	1.62	5.74	
	2	0.51	0.11	0.10	0.19	20.34	27.96	
	3	0.37	0.09	0.09	0.17	18.78	23.09	
5	0.5	3.16	0.45	0.42	0.82	127.79	263.73	
	1	0.98	0.28	0.26	0.50	130.30	235.92	
	2	1.01	0.20	0.19	0.36	127.78	138.87	
	3	1.01	0.17	0.16	0.31	87.74	84.14	
10	0.5	1.10	0.42	0.40	0.76	139.14	230.39	
	1	0.96	0.57	0.53	1.03	167.67	261.02	
	2	1.51	0.45	0.42	0.81	137.70	213.29	
	3	1.08	0.43	0.40	0.78	117.32	213.66	
20	0.5	2.87	1.32	1.23	2.37	157.81	279.20	
	1	3.51	0.89	0.83	1.61	139.87	250.75	
	2	3.25	1.61	1.50	2.91	150.02	212.59	
	3	5.22	3.42	3.19	6.18	123.70	240.88	

TABLE 2.9 – Nombre moyen d'échantillons indépendants tirés par seconde ($N = 50$)

Paramètres		$N = 100$							
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA		
1	0.5	1.27	0.12	0.11	0.19	133.40		173.87	
	1	1.15	0.10	0.09	0.16	1.29		5.18	
	2	0.31	0.09	0.08	0.15	5.91		8.50	
	3	0.32	0.10	0.09	0.15	11.77		11.53	
5	0.5	4.09	0.58	0.51	0.91	179.62		297.92	
	1	0.72	0.30	0.27	0.48	205.72		221.27	
	2	0.74	0.23	0.20	0.35	136.97		209.91	
	3	0.78	0.15	0.13	0.23	124.38		146.99	
10	0.5	1.78	0.61	0.54	0.96	169.88		265.94	
	1	0.87	0.35	0.31	0.54	152.17		315.43	
	2	0.92	0.38	0.33	0.59	137.57		199.72	
	3	1.11	0.37	0.33	0.58	132.8		248.55	
20	0.5	1.23	0.64	0.56	1.00	187.88		310.64	
	1	1.60	0.60	0.53	0.94	229.48		288.91	
	2	1.98	0.95	0.84	1.48	168.67		233.79	
	3	4.74	2.78	2.46	4.34	129.25		264.41	

TABLE 2.10 – Nombre moyen d'échantillons indépendants tirés par seconde ($N = 100$)

Paramètres		$N = 500$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	0.70	0.08	0.07	0.10	117.62	147.13	
	1	0.78	0.08	0.07	0.10	114.91	89.19	
	2	0.74	0.07	0.06	0.09	69.96	72.13	
	3	0.33	0.08	0.07	0.10	1.28	3.67	
5	0.5	0.47	0.20	0.17	0.25	158.87	213.23	
	1	0.30	0.22	0.18	0.27	148.32	252.18	
	2	0.33	0.13	0.10	0.15	141.41	202.48	
	3	0.33	0.10	0.09	0.13	146.43	193.59	
10	0.5	0.33	0.31	0.26	0.37	152.77	210.19	
	1	0.35	0.22	0.18	0.27	151.34	242.72	
	2	0.45	0.26	0.22	0.32	160.12	231.44	
	3	0.49	0.25	0.21	0.31	151.90	209.99	
20	0.5	0.39	0.26	0.21	0.31	175.33	222.85	
	1	0.61	0.31	0.25	0.37	144.00	229.38	
	2	1.42	0.51	0.42	0.62	143.24	217.95	
	3	1.93	0.78	0.65	0.95	157.68	229.10	

TABLE 2.11 – Nombre moyen d'échantillons indépendants tirés par seconde ($N = 500$)

Paramètres		$N = 1000$						
μ	k	GIBBS	RWMH	MALA	prox-MALA	mMALA	prox-mMALA	
1	0.5	0.37	0.06	0.05	0.06	8.58	11.73	
	1	0.28	0.06	0.05	0.06	18.00	13.88	
	2	0.31	0.06	0.05	0.06	11.10	21.33	
	3	0.31	0.06	0.05	0.06	6.13	14.49	
5	0.5	0.21	0.11	0.09	0.11	50.84	127.61	
	1	0.19	0.13	0.10	0.14	65.98	90.06	
	2	0.20	0.10	0.08	0.11	59.87	68.47	
	3	0.21	0.08	0.06	0.08	56.01	101.14	
10	0.5	0.19	0.23	0.19	0.25	116.33	180.50	
	1	0.21	0.13	0.11	0.14	113.87	117.51	
	2	0.25	0.20	0.16	0.21	111.13	149.54	
	3	0.30	0.19	0.16	0.21	119.65	154.73	
20	0.5	0.22	0.18	0.14	0.19	84.65	29.11	
	1	0.33	0.23	0.18	0.24	99.81	163.99	
	2	0.61	0.37	0.30	0.40	116.66	148.58	
	3	0.94	0.55	0.44	0.58	125.16	154.03	

TABLE 2.12 – Nombre moyen d'échantillons indépendants tirés par seconde ($N = 1000$)

On constate principalement que les méthodes mMALA et prox-mMALA sont en mesure de fournir plus d'échantillons effectifs durant une période de temps donnée, même si le calcul de la matrice de Fisher ralentit l'exécution des algorithmes. Les quatre autres algorithmes donnent des résultats comparables, même si un léger avantage de l'algorithme de GIBBS est constatable.

Enfin, bien que les résultats de l'ESS montrent que l'utilisation de la méthode proximale ne donnent pas de résultats significativement meilleurs par rapport à l'échantillonnage par réflexion, il est important de remarquer que la méthode recourant aux opérateurs proximaux est plus rapide que celle basée sur un schéma réfléchitif, et fournit par conséquent des échantillons décorrés plus nombreux par intervalle de temps donné.

2.5.2 Modélisation des données TEP

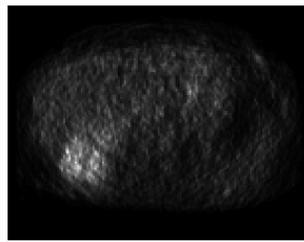
Afin de confirmer le fait que la DBN modélise correctement les données TEP, la méthode proposée a été appliquée à des images médicales réelles. La figure 2.14(a-c) montre trois coupes successives d'une région d'intérêt ($116 \times 151 \times 3$) centrée sur l'abdomen, présentant deux types de lésions couramment rencontrés en clinique : une petite tumeur et une plus large. Ces images ont été acquises par un appareil GE-Discovery ST et reconstruites par un algorithme de type OSEM. L'image a été pré-partitionnée en 6 classes selon la figure 2.14(d-f). La même figure présente également l'index associé à chacune des classes 2.14(g-l). Par ordre décroissant d'activité, les différentes classes peuvent vraisemblablement être interprétées comme suit :

- La classe 1 représente les deux zones tumorales.
- La classe 2 correspond grossièrement au foie et aux organes environnants.
- La classe 3 correspond aux tissus abdominaux de moindre activité.
- La classe 4 représente la peau et la chair.
- La classe 5 peut être interprétée comme une couche de voxels marqués par l'effet de volume partiel.
- Enfin, la classe 6 correspond au fond, où il ne devrait y avoir théoriquement aucune activité.

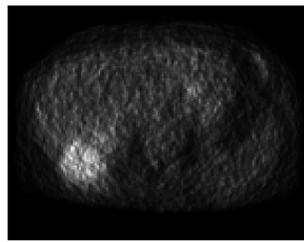
Des tests d'adéquations aux données ont été réalisés pour le modèle proposé et comparés à d'autres modèles statistiques utilisés dans la littérature. Plus précisément, la distribution de l'activité de chaque tissu a été modélisée par une distribution normale, par une distribution de Poisson et par la distribution binomiale négative proposée.

Estimation des paramètres

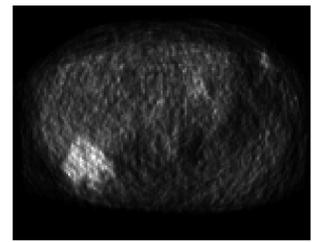
Les paramètres associés à chacune des distributions ont été estimés pour chacun des tissus. Une technique de Metropolis-Hastings par marche aléatoire (RWMH) a été utilisée pour la distribution normale et la distribution de Poisson. Pour les trois modèles, la longueur des chaînes de Monte Carlo a été fixée à 11000 itérations, dont 1000 itérations de chauffe. La valeur finale des paramètres a été choisie comme la valeur moyenne des échantillons tirés après la période de chauffage. La même expérience a été conduite 50 fois de manière à évaluer leur reproductibilité. Les estimations MMSE des paramètres de chaque modèle sont présentées dans le tableau 2.13 pour chacune des 6 classes.



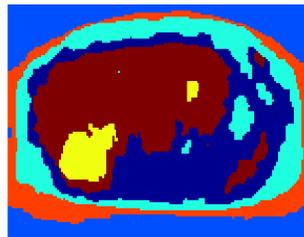
(a) Originale (coupe 1)



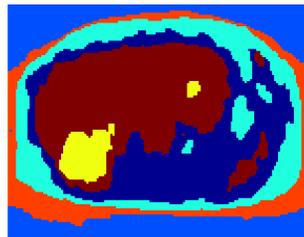
(b) Originale (coupe 2)



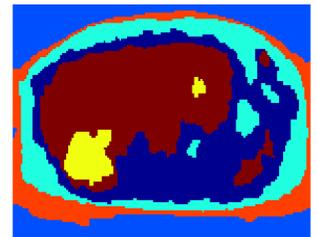
(c) Originale (coupe 3)



(d) Partition z (slice 1)



(e) Partition z (slice 2)



(f) Partition z (slice 3)



(g) Classe 1



(h) Classe 2



(i) Classe 3



(j) Classe 4



(k) Classe 5



(l) Classe 6

FIGURE 2.14 – Image originale (a-c), partition (d-f) et indice des différentes classes (g-l).

Classe	\mathcal{N}		σ	\mathcal{P}		\mathcal{BN}				
	μ			λ	μ	k				
1	14461.1	± 5.53	4752.4	± 2.53	14461.4	± 0.0640	14451.0	± 1.442	9.71	± 0.0045
2	6209.9	± 0.68	1912.2	± 0.38	6209.9	± 0.0162	6209.7	± 0.228	10.95	± 0.0021
3	3173.6	± 0.42	1101.5	± 0.22	3173.5	± 0.0112	3173.5	± 0.131	8.74	± 0.0016
4	1459.3	± 0.25	572.9	± 0.15	1459.3	± 0.0074	1459.3	± 0.080	6.57	± 0.0013
5	253.4	± 0.08	156.8	± 0.04	253.4	± 0.0049	253.4	± 0.026	2.95	± 0.0008
6	41.0	± 0.01	29.3	± 0.01	40.9	± 0.0015	40.9	± 0.004	1.93	± 0.0004

TABLE 2.13 – Paramètres estimés pour chacun des modèles de distribution : Normale (\mathcal{N}), Poisson (\mathcal{P}) et binomiale négative (\mathcal{BN})

Tout d'abord, il est intéressant de noter que conformément aux attentes, le paramètre μ de la loi normale, le paramètre λ de la loi Poisson et le paramètre μ de la loi binomiale négative sont très proches. Comme précédemment prouvé sur des données synthétiques, l'algorithme d'estimation des paramètres présenté dans ce chapitre offre une grande reproductibilité. La reproductibilité des algorithmes utilisés pour estimer les paramètres des autres modèles est également assurée, compte tenu des faibles écarts des estimations d'une expérience à l'autre. De plus, les valeurs du PSRF obtenues sur les 50 chaînes pour chaque paramètre de chaque modèle à l'issue des 11000 itérations montrent que toutes les chaînes ont bien convergé (voir tableau 2.14). Ces deux résultats montrent que les paramètres utilisés dans la suite de l'étude ont été estimés par des algorithmes précis et robustes.

Classe	\mathcal{N}		\mathcal{P}	\mathcal{BN}	
	μ	σ	λ	μ	k
1	1.0012	1.0005	1.0002	1.0000	1.0000
2	1.0008	1.0005	1.0002	1.0001	1.0001
3	1.0009	1.0004	1.0002	1.0000	1.0001
4	1.0008	1.0006	1.0001	1.0000	1.0000
5	1.0007	1.0003	1.0002	1.0000	1.0001
6	1.0008	1.0004	1.0003	1.0000	1.0001

TABLE 2.14 – PSRF

Adéquation aux données

Les paramètres présentés ont été utilisés pour modéliser les données par les différentes distributions. La figure 2.15 montre l'histogramme des données et les courbes associées aux différents modèles.

La figure 2.15 montre que la densité de probabilité estimée en utilisant le modèle de distribution binomiale négative proposé épouse correctement l'histogramme des données. Ce résultat est particulièrement marqué pour les classes 1 à 4, c'est à dire pour la tumeur et les différents tissus. L'adéquation aux données dans le cas de la classe représentant l'effet de volume partiel (5) et la classe correspondant au fond (6) est moins significative. Le modèle de distribution normale offre des résultats sensiblement moins bons alors que le modèle Poissonien ne correspond visiblement pas aux données. Ces résultats d'adéquation aux données sont repris quantitativement dans les tableaux suivants. Le tableau 2.15 rapporte les résultats des tests de Kolmogorov-

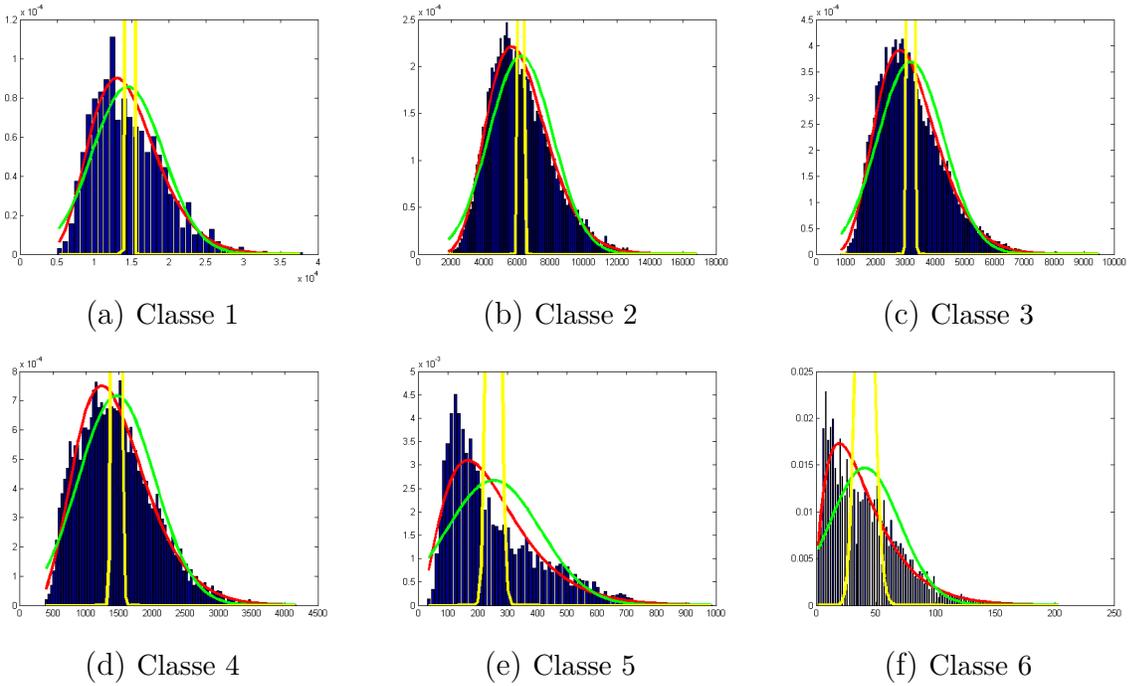


FIGURE 2.15 – Adéquation aux données pour la loi normale (en vert), la loi Poisson (en jaune) et la loi binomiale négative (en rouge)

Smirnov (KS), de Cramer Von Mises (CvM) et du χ^2 pour chacun des modèles et pour chacune des 6 classes.

Classe	Kolmogorov-Smirnov (KS)		Cramer von Mises (CvM) ($\cdot 10^{-3}$)		χ^2					
	\mathcal{N}	\mathcal{BN}	\mathcal{N}	\mathcal{P}	\mathcal{N}	\mathcal{P}				
1	0.0865	0.5481	0.0408	0.0408	35.4	1537.8	5.3	0.823	Inf	0.050
2	0.0618	0.5023	0.0201	0.0201	62.2	3182.4	4.7	1.238	Inf	0.016
3	0.0700	0.5034	0.0190	0.0190	74.3	2925.8	5.3	3.789	Inf	0.020
4	0.0640	0.4786	0.0214	0.0214	74.1	2944.0	3.7	0.174	Inf	0.022
5	0.1794	0.5303	0.0705	0.0705	364.5	2647.8	54.1	0.436	Inf	0.106
6	0.1554	0.4052	0.0510	0.0510	374.4	2227.8	22.4	2.419	Inf	0.066

TABLE 2.15 – Test de Kolmogorov-Smirnov (KS), de Cramer von Mises (CvM) et du χ^2

Le tableau suivant 2.16 présente également les critères d'information d'Akaike (AIC) et Bayésien (BIC) pour chacun des modèles et pour chacune des 6 classes.

Classe	AIC ($\cdot 10^3$)			BIC ($\cdot 10^3$)		
	\mathcal{N}	\mathcal{P}	\mathcal{BN}	\mathcal{N}	\mathcal{P}	\mathcal{BN}
1	35.01	Inf	34.81	35.03	Inf	34.82
2	240.56	Inf	239.25	240.58	Inf	239.26
3	204.60	Inf	203.06	204.62	Inf	203.08
4	139.40	Inf	138.39	139.42	Inf	138.41
5	73.92	547.04	71.89	73.93	547.05	71.90
6	101.18	268.38	97.62	101.19	268.38	97.63

TABLE 2.16 – Critère d'Information d'Akaike (AIC) et Critère d'Information Bayésien(BIC)

Les différents tests confirment les résultats visualisés sur la figure 2.15. Le modèle de distribution binomiale négative surpasse les deux autres modèles quelque soit le critère utilisé.

Discussions

Les résultats des tests de Kolmogorov-Smirnov, de Cramer von Mises et du χ^2 montrent que l'adéquation aux données est globalement meilleure pour les tissus présentant une forte activité, donc spécialement la tumeur (classe 1), quelque soit le modèle. Les classes 5 et 6 représentent des régions dues à l'effet de volume partiel et au fond respectivement, et ne sont bien modélisées par aucune des trois distributions.

Bien que, en théorie, la modélisation des données par une distribution Poisson est plus légitime que la modélisation par des lois Gaussiennes, le modèle Poissonien ne concorde pas avec les données dans cet exemple. Cela peut venir du fait que la méthode de reconstruction utilisée ne conserve pas la statistique Poissonnienne des données (zéros artificiels, lissages, mises à l'échelle, ajustement sur le contraste). Il est alors intéressant de constater que la distribution binomiale négative est robuste face à ces effets indésirables apparaissant lors de la reconstruction.

En conclusion, nous observons que le modèle statistique proposé modélise correctement l'activité des deux tumeurs, mais également celle des autres tissus biologiques (foie, peau, ...) et surpasse les modèles Gaussien et Poissonnien en terme d'adéquation aux données.

Bayesian segmentation of tumors in PET with a negative binomial mixture model

Introduction

Existing segmentation methods used in clinics are mainly based on thresholding techniques. A number of research works have reported their limitations. In addition, the heterogeneity of the activity in tumoral tissues is now an established fact. Unfortunately, existing methods do not consider this phenomenon. Segmenting tissues that exhibit large intra-tissue variability of activity requires appropriate data models. Here we develop a segmentation method that takes into account this phenomenon.

PET data are considered as a mixture of negative binomial distributions. The parameters of the mixture and the segmentation are carried out jointly in Bayesian framework.

Problem statement

PET data are here considered as a stationary process based on the negative binomial distribution whose parameters are piecewise constant. Thus, the density of the entire image is represented by a mixture of negative binomial distributions (see (3.2)).

This chapter presents a method for joint segmentation and parameter estimation.

Bayesian model

The problem of segmentation and estimation of the model parameters is formulated as the search for the maximum a posteriori (see (3.3)). A hierarchical Bayesian model is established.

Likelihood

We consider that given a single tissue, the activity of one of its voxels follows a negative binomial distribution (see (3.4)).

The likelihood of the entire image can be expressed as a finite mixture of negative binomial distributions (see (3.5)).

Priors on the parameters

Gamma distributions priors were chosen for both mean μ_z and inverse dispersion parameters k_z (see (3.6)).

Assuming that the parameters are independent, the joint prior distribution of the vector $\boldsymbol{\theta}$ is expressed in (3.8).

Prior on the labels

We introduce a Potts MRF as prior to consider the spatial coherence of the mixture components.

The probability that a voxel n is associated with the label z_n knowing the label of the other voxels of the image can be expressed by (3.10).

A 3-D neighborhood structure of the first order is considered (see figure 3.2).

This results in a prior for \mathbf{z} as expressed in (3.11).

To summarize the relationships between the various parameters and hyper-parameters, figure 3.3 presents the proposed hierarchical Bayesian model as a directed acyclic graph.

Posterior distribution $(\boldsymbol{\theta}, \mathbf{z})$

Considering that the unknown parameter vector $\boldsymbol{\theta}$ and labels \mathbf{z} are independent, and using Bayes theorem, the posterior distribution of the parameter vector $(\mathbf{z}, \boldsymbol{\theta})$ can be expressed as (3.12).

Hybrid Gibbs sampler

This section presents the method to sample the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$. As the distribution is sampled in a high-dimensional space, a Gibbs sampler is used. The samples are drawn according to the conditional distributions of the posterior distribution (3.12).

Conditional distribution $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$

The conditional distribution $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ can be developed as in (3.13).

The parameter vectors $\boldsymbol{\theta}$ are updated according to the proximal Manifold MALA (prox-mMALA) detailed in chapter 2.

Because the proposal distributions (3.14) are not symmetric, the acceptance ratio a is expressed as in (3.15).

Conditional distribution $P[\mathbf{z}|\boldsymbol{\theta}, \mathbf{x}]$

The sampling of the whole vector \mathbf{z} may itself be split into several Gibbs steps component by component.

The conditional distribution of the discrete label z_n of each voxel can then be fully characterized by (3.18).

As (3.18) defines a Markov Random Field, the samples are drawn according to this conditional distribution by drawing integers in the finite set $\{1, \dots, Z\}$ with probabilities defined in (3.18).

Summarized algorithm

The algorithm summarizing the overall sampling method is presented in the pseudo-code 2.

Experimental results

The proposed method was validated on synthetic data to evaluate the performance of the algorithm to find the model parameters while correctly estimating the labels of each voxel. Then, tests were performed on real PET images to test the proposed negative binomial mixture model and evaluate the ability of the algorithm to discriminate the tissues.

Synthetic data

A 3-D image composed of 3 spatially coherent classes was generated. Voxels of each class were assigned a value drawn accordingly to a negative binomial distribution.

Segmentation results

Figure 3.4 shows the theoretical labels, the generated data and the segmentation results of the algorithm for different values of β .

Table 3.1 presents the confusion matrices. Regardless of the value of β , the rate of misclassified voxels does not exceed 1%.

Parameter estimation

Figure 3.5 shows the histograms of the generated samples for each parameter.

We note that the samples have been generated around the correct values and the average of these samples can be used as maximum a posteriori.

The same experiment was conducted 50 times in order to measure the accuracy and the reproducibility of the algorithm.

Accuracy

The MMSE estimates of each parameter and the corresponding standard deviations are shown in table 3.2. The algorithm provides high accuracy in estimating the parameters.

Reproducibility

To measure the robustness of the algorithm, the Potential Scale Reduction Factor (PSRF) was measured. Figure 3.6 shows the evolution of the PSRF during the burn-in for each parameter. After only 1000 sampling iterations, a PSRF below 1.001 was obtained for each parameter of each class, showing the robustness of the algorithm to jointly estimate the labels and the model parameters.

Real PET data

This section has two objectives : i) show that the negative binomial mixture model fits the PET images, and ii) evaluate the performance of the algorithm to jointly classify the tissues and estimate the model parameters. The method was applied to abdominal real PET images. Figure 3.8 shows three successive slices of the test image.

Segmentation results

Figure 3.8 shows the results of the segmentation on the data volume.

We observe that the algorithm not only delineated the two tumors, but also affected appropriate labels to other biological tissues (liver, skin, flesh, etc.) whose biological characteristics are distinct.

Goodness of fit

The activity distribution of the image was modeled by the proposed negative binomial mixture model as well as by the other mixture models, Gaussian and Poisson. The associated segmentation results are shown in figure 3.9.

The goodness of fit results are presented in tables 3.3 and 3.4.

For graphical results, figure 3.10 represents the histogram of the data and the curves associated with the tested models.

We note that the proposed negative binomial mixture fits the data better than other mixture models.

The same experiment was conducted 50 times in order to measure the accuracy and the reproducibility of the algorithm.

Precision

The MMSE estimates of each parameter and the corresponding standard deviations are shown table 3.5. We notice that the standard deviations are small relatively to the average value. The algorithm provides high accuracy in estimating the parameters.

Reproducibility

To measure the robustness of the algorithm, the Potential Scale Reduction Factor (PSRF) value at the end of the burn-in is reported in table 3.6. A PSRF below 1.002 was obtained for each parameter, demonstrating a good repeatability.

Influence of the number of classes Z

Figure 3.11 shows the segmentation results for varying numbers of classes Z . Table 3.7 presents the associated goodness of fit results.

It is important to note that the tumor segmentation becomes stable for Z over 6. This result supports the robustness of the model.

Influence of the granularity β

Figure 3.12 shows the segmentation results for varying granularity coefficient β . Table 3.8 presents the associated goodness of fit results.

We see that even without spatial a priori ($\beta = 0$), most of the voxels are correctly labeled. This supports that the proposed model is particularly suitable.

The spatial a priori not only provides better visual outcome, but helps the algorithm to correctly fit the model to the data.

Chapitre 3

Segmentation Bayésienne de tumeurs en TEP par mélange de lois binomiales négatives.

3.1 Introduction

La segmentation précise des images TEP est un préalable à la décision médicale en oncologie. En effet, la précision du diagnostic repose sur la qualité des mesures effectuées sur les tumeurs (e.g. SUV, volume, etc.). Cependant, la segmentation des tumeurs est un problème difficile. Les méthodes classiques ne sont pas adaptées car la tumeur n'est pas un objet géométrique ayant des frontières bien nettes.

En routine clinique, on utilise souvent une méthode semi-manuelle. Ces méthodes dépendent directement des valeurs des pixels qui sont très peu significatives individuellement. De plus, l'effet de la variabilité inter-opérateurs humains donne des résultats très peu reproductibles.

En imagerie TEP, la segmentation est rendue plus difficile par le faible rapport signal sur bruit et le manque de contraste dû à l'absence d'information anatomique. L'hétérogénéité de l'activité dans les tumeurs est un autre facteur qui complexifie la segmentation.

Plusieurs méthodes de seuillage fixe ou adaptatif ont été largement étudiées dans la littérature [37]. Mais ces approches de seuillage ont souvent besoin d'une sélection manuelle initiale et nécessitent souvent une calibration sur des acquisitions fantômes. Ces techniques sont donc dépendantes du système, et l'étalonnage sur des fantômes n'est pas réaliste car l'activité de volumes du fantôme ne reflète pas la non-uniformité des tumeurs réelles. En outre, Hatt et al. [56] déclarent que l'hétérogénéité d'absorp-

tion du traceur a un impact sur les méthodes à seuil. Ils concluent que le seuillage ne doit pas être utilisé pour la délimitation de grosses tumeurs, car ces méthodes ont tendance à sous-estimer largement l'étendue spatiale de la tumeur à cause de l'hétérogénéité. Werner-Wasik et al. [141] ont comparé la segmentation par seuillage et par gradient et concluent à la faiblesse du seuillage et à la nécessité d'adapter une méthode plus précise.

Récemment, d'autres algorithmes de segmentation plus avancées ont été proposées dans la littérature tels que le clustering [74] et les contours actifs [73]. Parmi ces algorithmes avancés, des travaux récents ont adressé le problème de la segmentation TEP dans un cadre Bayésien. Par exemple, [23] a proposé une méthode de segmentation basée sur un champ de Markov. Des modèles Gauss-Markov-Potts ont également été étudiés dans [92, 91] et un champ de Potts a aussi été utilisé par [133] où un algorithme résolvant les problèmes de reconstruction et de segmentation est proposé en se basant sur un modèle Poissonien. Le lecteur intéressé pourra trouver plusieurs états de l'art sur la segmentation d'images TEP dans la littérature [136, 146, 56].

Les approches de segmentation ne tiennent habituellement pas compte de la complexité de structure des tissus due à leur hétérogénéité. Or, il a été constaté de fortes variations de l'activité dans les tumeurs [62, 61]. La tumeur est donc largement hétérogène en terme d'activité ce qui produit une grande variabilité de l'intensité dans les régions correspondantes dans les images TEP.

Ainsi, lorsqu'une tumeur présente de fortes variations d'absorption du traceur, un algorithme de segmentation simple (de type seuillage) détectera une région limitée à une petite zone où l'activité est la plus forte au lieu de la tumeur entière. De ce fait, il paraît important de baser la segmentation sur les statistiques des tissus plutôt que les valeurs de l'activité. La modélisation des données et la segmentation doivent donc être réalisées conjointement.

Récemment, des techniques de segmentation ont été étudiées pour tenir compte de l'hétérogénéité dans les tissus. Notons qu'habituellement, l'hétérogénéité est considérée comme l'imbrication de quelques zones (trois en général), chacune correspondant à un stade de développement de la tumeur. Parmi ces travaux, une approche floue appelée FLAB [58, 60, 57, 59] a été proposée permettant d'utiliser un maximum de trois sous-classes pour distinguer les zones d'hétérogénéité. Une autre segmentation robuste à l'hétérogénéité a été proposée dans [9] en modifiant l'algorithme Fuzzy C-Means par l'intégration des fonctionnalités supplémentaires adaptées à la nature de l'imagerie TEP et l'utilisation de la transformée en ondelettes *à trous* pour permettre l'évaluation de l'absorption de lésions hétérogènes. Enfin, une méthode pour corriger l'effet de volume partiel dû à l'hétérogénéité tumorale a également été proposée par

le moyen d'un algorithme *espérance-maximisation* dans [7].

Dans cette thèse, nous considérons l'hétérogénéité au sens large. Un tissu hétérogène en terme d'activité est considéré comme l'entrelacement de réseaux cellulaires différents. Vu la résolution faible de l'imagerie TEP, il est évidemment impossible d'identifier et discriminer ces réseaux. Nous considérons donc que l'hétérogénéité s'exprime comme une variabilité spatiale de l'activité dans la tumeur.

A la base de cette hypothèse, nous développons une méthode de segmentation prenant en considération la variation de l'activité. Cette méthode discrimine les tissus présentant des statistiques différentes. Le processus de segmentation repose sur le modèle de données présenté dans le chapitre précédent. Les données d'une image TEP représentant divers tissus sont ainsi décrites par un *mélange* de distributions binomiales négatives. Ce modèle a l'avantage de respecter essentiellement les variations de l'activité à l'intérieur des tissus. La dépendance des paramètres du modèle et de la segmentation est formulée comme un modèle Bayésien hiérarchique. Dans ce modèle, nous introduisons un champ de Markov Potts comme a priori pour prendre en considération la cohérence spatiale des composants du mélange. Le problème de segmentation et d'estimation des paramètres du modèle est formulée comme la recherche du maximum a posteriori d'une distribution a posteriori. Ce problème d'optimisation est résolu par un algorithme Monte-Carlo par chaîne de Markov (MCMC).

Ce chapitre est organisé comme suit. Le problème de la segmentation et d'estimation des paramètres conjointement est introduit dans la section 3.2 ainsi que le modèle de donnée. La section 3.3 présente le modèle Bayésien hiérarchique et présente en particulier le champ de Markov-Potts utilisé comme a priori spatial. La méthode de résolution est présentée dans la section 3.4, où un échantillonneur de type Metropolis-dans-Gibbs est détaillé. Enfin, la section 3.5 rapporte les résultats effectués sur des données synthétiques d'une part, afin de caractériser l'efficacité de l'algorithme, ainsi que sur des données réelles d'autre part, afin de vérifier les performances de l'estimation du modèle et de la segmentation.

3.2 Position du problème

Soit $\mathbf{x} \in \mathbb{N}^N$ une image TEP 3-D (N étant le nombre total de voxels). Les $x_n \in \mathbb{N}$ correspondent aux observations, soit la quantité de radioactivité reçue par le $n^{\text{ème}}$ voxel de \mathbf{x} .

Ce chapitre considère le cas où l'image \mathbf{x} est partitionnée en Z tissus biologiques distincts $\{\mathcal{Z}_1, \dots, \mathcal{Z}_Z\}$, chacun ayant sa propre activité biologique caractéristique et par conséquent sa propre distribution statistique.

Il a été montré dans le chapitre précédent que la distribution binomiale négative pouvait modéliser correctement l'activité d'un tissu individuel sur une image TEP. Il est alors légitime de considérer l'image \mathbf{x} comme un processus stationnaire fondé sur la loi négative binomiale dont les paramètres sont constants par morceaux :

$$\forall x_n \in \mathcal{Z}_z, x_n \sim \mathcal{BN}(\mu_z, k_z) \quad (3.1)$$

où μ_z et k_z sont les paramètres associés au tissu \mathcal{Z}_z .

Il est alors possible de représenter la distribution de l'image entière par un mélange de lois binomiales négatives.

$$x_n \propto \sum_{z=1}^Z \omega_z \mathcal{BN}(\mu_z, k_z) \quad (3.2)$$

où ω_z représente la proportion de voxels associés à la classe \mathcal{Z}_z avec $\sum_{z=1}^Z \omega_z = 1$.

Le nombre de classes Z est supposé connu dans cette étude. Cette hypothèse pourrait être relâchée en utilisant une stratégie de recherche de dimension, comme un algorithme MCMC à saut réversible [53]. Le nombre de classes peut également être déterminé par des cliniciens après inspection visuelle.

Nous introduisons donc explicitement un vecteur caché d'étiquettes $\mathbf{z} = \{z_1, \dots, z_N\}$ associé au vecteur d'observation $\mathbf{x} = \{x_1, \dots, x_N\}$ tel que $z_n = z$ si $x_n \in \mathcal{Z}_z$. Ces étiquettes serviront à discriminer les différents tissus.

Il est ensuite possible de poser le problème de segmentation et d'estimation jointe des paramètres de la manière suivante :

$$(\hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}) = \underset{(\boldsymbol{\theta}, \mathbf{z})}{\operatorname{argmax}} p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x}) \quad (3.3)$$

où $\boldsymbol{\theta}$ est un vecteur de paramètres inconnu associé à un modèle paramétrique de formation d'image $p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z})$.

Ce chapitre résout le problème de segmentation en se basant sur les statistiques des tissus, dans le sens où deux voxels seront associés à une même classe s'ils ont des statistiques similaires. Le résultat de cette segmentation fournit également les paramètres du mélange de lois binomiales négatives utilisé pour modéliser l'image.

Ces paramètres pourraient servir à la caractérisation des tissus, notamment à l'hétérogénéité des tumeurs.

3.3 Modèle Bayésien

Nous posons le problème de la segmentation dans un cadre Bayésien. La résolution de (3.3) est considérée comme une évaluation du *maximum a posteriori*.

En se basant sur le théorème de Bayes, l'estimation de la densité a posteriori des paramètres inconnus $\boldsymbol{\theta}$ et \mathbf{z} nécessite alors l'expression de la vraisemblance $p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})$ et l'introduction de la densité a priori $p(\mathbf{z}, \boldsymbol{\theta})$ pour les étiquettes et les paramètres du modèle. Ils sont définis dans les sections suivantes.

3.3.1 Vraisemblance

En raison de la physique nucléaire intrinsèque au processus de formation de l'image, et afin de modéliser correctement l'hétérogénéité intra-tissulaire, nous considérons que l'activité d'une voxel x_n appartenant à la classe \mathcal{Z}_z suit une distribution binomiale négative de paramètre $\boldsymbol{\theta}_z = (\mu_z, k_z)^T$:

$$\begin{aligned} P[x_n|\boldsymbol{\theta}_z, z_n = z] &= \mathcal{BN}(x_n|\mu_z, k_z) \\ &= \binom{x_n + k_z - 1}{x_n} \left(\frac{\mu_z}{\mu_z + k_z} \right)^{x_n} \left(\frac{k_z}{\mu_z + k_z} \right)^{k_z} \end{aligned} \quad (3.4)$$

Ce choix a été justifié dans le chapitre précédent (Chap 2).

Puisqu'à chaque classe correspond une distribution binomiale négative, il s'ensuit facilement que la vraisemblance de l'image entière peut être exprimée comme un mélange fini de distributions binomiales négatives.

En posant $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_Z^T)^T$, le vecteur de paramètres inconnus du problème et en supposant que les observations sont indépendantes entre elles, la vraisemblance s'exprime comme suit :

$$P[\mathbf{x}|\boldsymbol{\theta}, \mathbf{z}] = \prod_{z=1}^Z \prod_{x_n \in \mathcal{Z}_z} P[x_n|\boldsymbol{\theta}_z, z_n = z] \quad (3.5)$$

où $P[x_n|\boldsymbol{\theta}_z, z_n = z]$ est défini dans (3.4),

3.3.2 A priori sur les paramètres

Des distributions gamma ont été choisies comme distributions a priori à la fois pour les paramètres de moyenne μ_z que pour les paramètres de dispersion inverse k_z :

$$\begin{aligned} \mu_z &\sim \Gamma \left(1 + a_\mu, \frac{-1}{b_\mu} \right), \forall z = 1, \dots, Z \\ k_z &\sim \Gamma \left(1 + a_k, \frac{-1}{b_k} \right), \forall z = 1, \dots, Z \end{aligned} \quad (3.6)$$

où a_μ et b_μ d'une part, et a_k et b_k d'autre part sont les hyperparamètres associés respectivement aux variables μ_z et k_z . Ces hyperparamètres sont considérés communs pour toutes les classes. Les lois log-a priori s'écriront donc :

$$\begin{aligned}\log [\pi(\mu_z)] &= a_\mu \log(\mu_z) + b_\mu \mu_z + c, \forall z = 1, \dots, Z \\ \log [\pi(k_z)] &= a_k \log(k_z) + b_k k_z + c, \forall z = 1, \dots, Z\end{aligned}\tag{3.7}$$

où c est une constante.

En supposant que les paramètres sont indépendants, la distribution a priori jointe du vecteur $\boldsymbol{\theta}$ est :

$$\begin{aligned}\pi(\boldsymbol{\theta}) &= \prod_{z=1}^Z \pi(\boldsymbol{\theta}_z) \\ &= \prod_{z=1}^Z \pi(\mu_z) \pi(k_z)\end{aligned}\tag{3.8}$$

où le logarithme de $\pi(\mu_z)$ et de $\pi(k_z)$ ont été définis dans (3.7). Voir chap 2.3.1 pour plus de détails.

3.3.3 A priori sur les étiquettes

Dans les images où le rapport signal sur bruit est faible, les valeurs d'activité observées peuvent être insuffisantes pour effectuer une segmentation précise. L'approche Bayésienne nous permet d'incorporer au modèle statistique des informations a priori sur la répartition spatiale des différentes classes. Il est en effet normal de considérer que la classe d'un voxel est corrélée à la classe de ses pixels voisins. L'a priori $\pi[\mathbf{z}]$ sur le vecteur d'étiquettes devrait alors favoriser la cohérence spatiale inhérente aux tissus biologiques.

Depuis les travaux fondateurs de [46], les champs de Markov (MRF) sont devenus un choix courant de lois a priori pour exprimer la corrélation spatiale dans les images. Les champs de Markov considère que la distribution d'un voxel conditionnellement aux autres voxels de l'image peut être approximée par la distribution d'un voxel conditionnellement à ses voisins :

$$P[z_n | \mathbf{z}_{-n}] = P[z_n | \mathbf{z}_{\mathcal{V}(n)}]\tag{3.9}$$

où $\mathbf{z}_{-n} = (z_1, \dots, z_{n-1}, z_{n+1}, \dots, z_N)$ et $\mathcal{V}(\cdot)$ désigne une structure de voisinage. Il existe plusieurs structures de voisinage qui ont été utilisées dans la littérature. Dans le cas bi-dimensionnel, le voisinage le plus fréquemment utilisé est défini par les quatre

(premier ordre) ou huit (second ordre) voxels les plus proches (voir figure 3.1). De même, dans le cas tridimensionnel, les voisinages les plus couramment utilisés sont définis par les six (premier ordre) ou quatorze (second ordre) voxels les plus proches (voir figure 3.2).

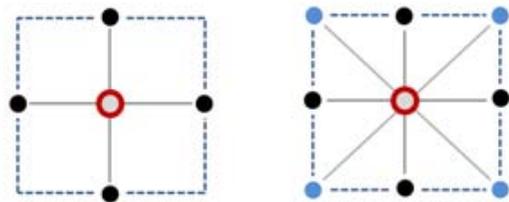


FIGURE 3.1 – Structure de voisinage 2-D du premier ordre (à gauche) et du second ordre (à droite)

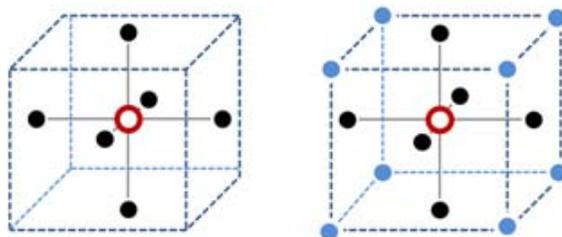


FIGURE 3.2 – Structure de voisinage 3-D du premier ordre (à gauche) et du second ordre (à droite)

Dans cette étude, nous considérons un champ de Potts 3-D [145] comme distribution a priori pour \mathbf{z} . La probabilité a priori qu'un voxel n soit associé au label z_n connaissant le label des autres voxels de l'image peut être formulé par :

$$\begin{aligned}
 P[z_n | \mathbf{z}_{-n}] &= P[z_n | \mathbf{z}_{\mathcal{V}(n)}] \\
 &\triangleq \frac{1}{C(\beta)} \exp \left[\sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'}) \right]
 \end{aligned}
 \tag{3.10}$$

où β est le coefficient de granularité, $C(\beta)$ est une constante de normalisation ou fonction de partition, $\delta(\cdot)$ est la fonction de Kronecker et $\mathcal{V}(\cdot)$ désigne une structure de voisinage 3-D du premier ordre.

Ce qui donne comme a priori sur \mathbf{z} :

$$\begin{aligned}\pi[\mathbf{z}] &= \prod_{z=1}^Z P[z_n | \mathbf{z}_{-n}] \\ &= \frac{1}{C(\beta)} \exp \left[\sum_{n=1}^N \sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'}) \right]\end{aligned}\tag{3.11}$$

Le paramètre de granularité β joue un rôle important car il représente le degré d'homogénéité de la segmentation finale. En effet, en choisissant un β petit, la partition obtenue sera de fréquence relativement élevée, avec des régions moins connexes, alors que le choix d'un β large débouchera sur un résultat présentant des régions plus connexes.

Pour résumer les relations entre les différents paramètres et hyper-paramètres, la figure 3.3 représente le modèle Bayésien hiérarchique proposé sous la forme d'un graphe orienté acyclique.

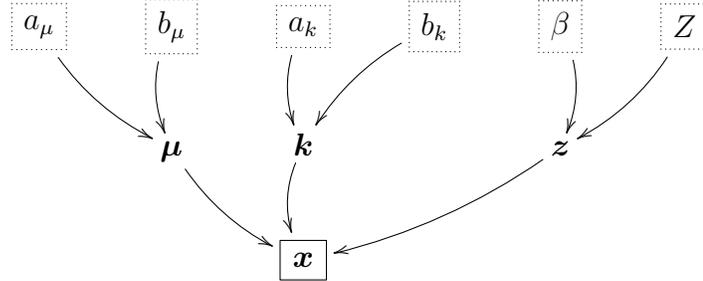


FIGURE 3.3 – Graphe orienté acyclique pour le modèle Bayésien proposé. Les hyperparamètres fixes sont représentés dans des cases en pointillé.

3.3.4 Distribution a posteriori de $(\boldsymbol{\theta}, \mathbf{z})$

En considérant que le vecteur de paramètres inconnus $\boldsymbol{\theta}$ et les étiquettes \mathbf{z} sont indépendants et en utilisant le théorème de Bayes, la distribution a posteriori du vecteur de paramètre $(\mathbf{z}, \boldsymbol{\theta})$ peut être exprimée comme :

$$p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x}) \propto P[\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}] \pi(\boldsymbol{\theta}) \pi[\mathbf{z}]\tag{3.12}$$

où la vraisemblance $P[\mathbf{x}|\boldsymbol{\theta}, \mathbf{z}]$ et les distributions a priori $\pi(\boldsymbol{\theta})$ et $\pi[\mathbf{z}]$ ont été définies en (3.5), (3.8) et (3.10), respectivement.

Déduire une solution explicite au problème de la segmentation (3.3) n'est pas possible en raison de la complexité de la distribution a posteriori (3.12). Pour cela, nous avons recours à un algorithme MCMC pour générer des échantillons qui sont asymptotiquement distribués selon la loi a posteriori (3.12) (voir [116] pour plus de détails). Ensuite, ces échantillons peuvent être utilisés pour évaluer le maximum a posteriori (MAP).

3.4 Echantillonneur de Gibbs hybride

Cette section présente un moyen d'échantillonner la loi a posteriori jointe $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$. Pour échantillonner une distribution dans des espaces de grandes dimensions, on a souvent recours à un échantillonnage de type Gibbs. Cet échantillonneur consiste à considérer la distribution jointe comme composée d'un couple de variables aléatoires individuelles et de choisir un nouvel échantillon pour chaque sous-espace alternativement, plutôt que de choisir un échantillon dans l'espace entier. La génération de ces échantillons est obtenue par tirages successifs selon les distributions conditionnelles de la distribution a posteriori (3.12). Dans ce chapitre, on se propose de générer alternativement $\boldsymbol{\theta}$ selon $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ et \mathbf{z} selon $P[\mathbf{z}|\boldsymbol{\theta}, \mathbf{x}]$. Il est important de préciser à ce stade que la faible efficacité de l'algorithme de Gibbs rapporté dans le chapitre précédent était obtenue lorsque les différentes composantes de $\boldsymbol{\theta}$ lui-même étaient échantillonnées alternativement. Ici, $\boldsymbol{\theta}$ n'est pas échantillonné par un schéma de Gibbs mais par une marche aléatoire de type Metropolis-Hastings (MH) selon la distribution conditionnelle $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$. L'algorithme global résulte donc en un échantillonneur de Gibbs hybride [116, p.317].

Les distributions conditionnelles ainsi que la méthode pour les échantillonner sont présentés dans les prochaines sections.

3.4.1 Distribution conditionnelle de $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$

La distribution conditionnelle $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ peut être développée ainsi :

$$p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x}) \propto P[\mathbf{x}|\boldsymbol{\theta}, \mathbf{z}]\pi(\boldsymbol{\theta}) \quad (3.13)$$

où $P[\mathbf{x}|\boldsymbol{\theta}, \mathbf{z}]$ et $\pi(\boldsymbol{\theta})$ ont été définies dans (3.5) et (3.8) respectivement.

Cependant, l'échantillonnage selon $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ est difficile. Pour pallier ce problème, un algorithme de type Metropolis-Hastings (MH) est utilisé, donnant ainsi un échantillonneur

de Gibbs hybride. Plus précisément, $\boldsymbol{\theta}$ est mis à jour selon la version proximale de la marche aléatoire MALA-sur-variété (prox-mMALA) détaillée dans le chapitre précédent (chap. 2). La loi de proposition est donc formulée :

$$\begin{aligned}\boldsymbol{\theta}_z^* &\sim q(\boldsymbol{\theta}_z^* | \boldsymbol{\theta}_z^{(i-1)}) \\ &\triangleq \mathcal{N}(\text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^{(i-1)} + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}_z} \mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})), \sqrt{\delta} \mathbf{G}^{-1} \{ \boldsymbol{\theta}_z^{(i-1)} \})\end{aligned}\quad (3.14)$$

où $\boldsymbol{\theta}_z^{(i-1)}$ est la valeur actuelle de la chaîne, $\tilde{\nabla}_{\boldsymbol{\theta}_z} \mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})$ est le gradient géodésique de la densité cible au point $\boldsymbol{\theta}_z^{(i-1)}$ et δ est estimé durant la période de chauffe de telle sorte que le taux d'acceptation tende vers $\frac{1}{2}$, comme recommandé dans [117].

Comme les lois de proposition ne sont pas symétriques, le rapport d'acceptation a prend alors la forme :

$$a = \min \left\{ 1, \prod_{n|z_n=z} \frac{P[x_n | \boldsymbol{\theta}_z^*, z_n = z]}{P[x_n | \boldsymbol{\theta}_z^{(i-1)}, z_n = z]} \frac{\pi(\boldsymbol{\theta}_z^*)}{\pi(\boldsymbol{\theta}_z^{(i-1)})} \frac{q(\boldsymbol{\theta}_z^{(i-1)} | \boldsymbol{\theta}_z^*)}{q(\boldsymbol{\theta}_z^* | \boldsymbol{\theta}_z^{(i-1)})} \right\} \quad (3.15)$$

où les vraisemblances $P[x_n | \boldsymbol{\theta}_z^{(\cdot)}, \mathbf{z}]$ et les lois a priori $\pi(\boldsymbol{\theta}_z^{(\cdot)})$ ont été définies dans (3.4) et (3.8) respectivement et où les lois de propositions sont :

$$\begin{cases} q(\boldsymbol{\theta}_z^* | \boldsymbol{\theta}_z^{(i-1)}) = p_{\mathcal{N}}(\boldsymbol{\theta}_z^* | \text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^{(i-1)} + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}_z} \mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})), \delta \mathbf{G}^{-1} \{ \boldsymbol{\theta}_z^{(i-1)} \}) \\ q(\boldsymbol{\theta}_z^{(i-1)} | \boldsymbol{\theta}_z^*) = p_{\mathcal{N}}(\boldsymbol{\theta}_z^{(i-1)} | \text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^* + \frac{\delta}{2} \tilde{\nabla}_{\boldsymbol{\theta}_z} \mathcal{P}(\boldsymbol{\theta}_z^*)), \delta \mathbf{G}^{-1} \{ \boldsymbol{\theta}_z^* \}) \end{cases} \quad (3.16)$$

3.4.2 Distribution conditionnelle de $P[\mathbf{z} | \boldsymbol{\theta}, \mathbf{x}]$

L'échantillonnage de \mathbf{z} entier peut lui-même être décomposé coordonnée par coordonnée en plusieurs étapes de Gibbs. La distribution conditionnelle associée à l'étiquette d'un voxel z_n peut alors être entièrement caractérisée par les probabilités :

$$P[z_n = z | \mathbf{z}_{-n}, \boldsymbol{\theta}_z, x_n] \propto P[x_n | \boldsymbol{\theta}_z, z_n = z] P[z_n | \mathbf{z}_{-n}] \quad (3.17)$$

où \mathbf{z}_{-n} représente le vecteur \mathbf{z} auquel le $n^{\text{ème}}$ élément a été enlevé. D'après les équations (3.5) et (3.10), il est alors possible d'exprimer ces probabilités comme

suit :

$$\begin{aligned}
P[z_n = z | \mathbf{z}_{-n}, \mu_z, k_z, x_n] \\
&\propto \binom{x_n + k_z - 1}{x_n} \left(\frac{\mu_z}{\mu_z + k_z}\right)^{x_n} \left(\frac{k_z}{\mu_z + k_z}\right)^{k_z} \\
&\times \exp \left[\sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'}) \right]
\end{aligned} \tag{3.18}$$

Il est intéressant de noter que la densité a posteriori du vecteur d'étiquettes \mathbf{z} dans (3.18) définit un champ aléatoire de Markov (MRF). Par conséquent, échantillonner selon cette distribution conditionnelle peut être réalisée en tirant des variables aléatoires dans l'ensemble fini $\{1, \dots, Z\}$ avec les probabilités (3.18).

3.4.3 Algorithme d'échantillonnage

L'algorithme résumant la méthode d'échantillonnage globale est présentée dans le pseudo-code 2.

Algorithm 2 Echantillonneur de Gibbs hybride proposé

Initialisation :

- Echantillonner $\boldsymbol{\theta}_z^0$ ($z = \{1, \dots, Z\}$) selon (3.8).
- Générer $z_1^0, z_2^0, \dots, z_N^0$ tels que $P[z_n^0 = z] = \frac{1}{Z}$.

for $i = 1, 2, \dots, I$ **do**

— *Mise à jour de $\boldsymbol{\theta}$* —

for $z = 1, 2, \dots, Z$ **do**

1. Proposer $\boldsymbol{\theta}_z^*$ selon (3.14).
2. Calculer le rapport d'acceptation a selon l'expression (3.15).
3. Tirer $u \sim \mathcal{U}(0, 1)$.

if ($u < a$) **then**

4. Fixer $\boldsymbol{\theta}_z^{(i)} = \boldsymbol{\theta}_z^*$.

else

5. Fixer $\boldsymbol{\theta}_z^{(i)} = \boldsymbol{\theta}_z^{(i-1)}$.

end if

end for

— *Mise à jour de \mathbf{z}* —

for $n = 1, 2, \dots, N$ **do**

6. Tirer z_n dans $\{1, \dots, Z\}$ selon (3.18).

end for

end for

3.5 Résultats et discussions

La méthode proposée a été validée sur des images synthétiques afin d'évaluer les performances de l'algorithme à retrouver les paramètres du modèle associés aux observations générées, tout en estimant correctement les labels de chaque voxel. Des tests ont ensuite été réalisés sur des images TEP réelles pour tester le modèle de mélange de loi binomiales négatives proposé et pour évaluer la capacité de l'algorithme à discriminer les tissus et en particulier les tumeurs. Cette section présente les résultats de ces simulations et expérimentations.

3.5.1 Données synthétiques

Afin de mesurer la faculté de l'algorithme à retrouver correctement les valeurs des paramètres et à segmenter les différentes classes, l'algorithme a été appliqué à des données synthétiques.

Une image 3-D ($100 \times 100 \times 3$) composée de 3 classes spatialement cohérentes a été générée. Les voxels de chacune de ces classes ont été affectés d'une valeur tirée selon une distribution binomiale négative propre à la classe avec $\boldsymbol{\mu} = [1500, 3500, 7000]$ et $\boldsymbol{k} = [7, 11, 6]$. Ces valeurs correspondent à l'ordre de grandeur des valeurs d'activité mesurées couramment sur des images réelles.

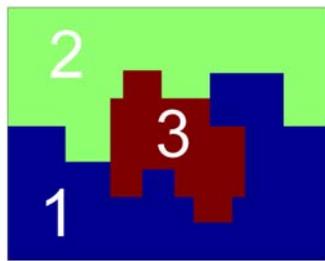
Les labels ont été initialisés de manière à ce qu'il y ait la même quantité de voxels dans chacune des classes et triés en trois groupes de valeurs. Une initialisation aléatoire des labels donne également de bons résultats en pratique mais nécessite une période de burn-in plus longue.

Les chaînes MCMC comprennent 60000 itérations (dont 10000 itérations de chauffe). La valeur finale des paramètres a été choisie comme étant la valeur moyenne des échantillons tirés après la période de chauffe.

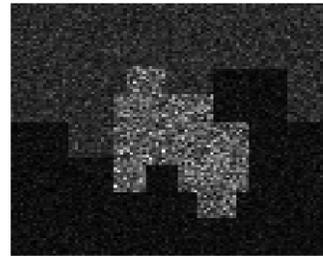
La section suivante présente les résultats concernant l'aptitude de l'algorithme à affecter les bons labels aux voxels ainsi que sa précision et sa reproductibilité quant à l'estimation des paramètres du modèle.

Résultats de la segmentation

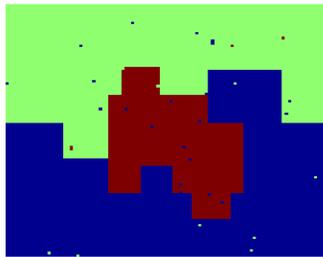
La figure 3.4 montre les étiquettes théoriques \mathbf{z} , les données générées ainsi que les résultats de segmentation de l'algorithme pour différentes valeurs de β .



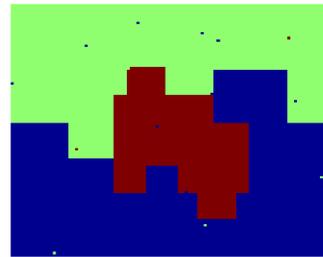
(a) Vrais labels



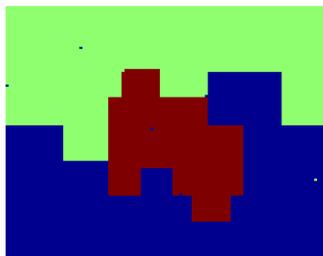
(b) Valeurs générées



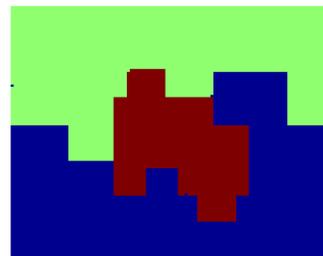
(c) $\beta = 0.8$



(d) $\beta = 0.9$



(e) $\beta = 1.0$



(f) $\beta = 1.1$

FIGURE 3.4 – Etiquettes théoriques (a), Valeurs générées (b) et estimations a posteriori des labels pour différentes valeurs de granularité (c-f) pour un mélange de 3 classes.

On constate que quelque soit la valeur de β , les classes ont été globalement correctement identifiées. En effet, très peu de voxels ont été étiquetés faussement, et ces voxels sont distribués de manière éparses. Le paramètre de granularité β joue ici un rôle important. Nous remarquons qu'un β grand aura tendance à lisser les

étiquettes.

Pour des résultats plus quantitatifs, le tableau suivant 3.1 présente les matrices de confusion pour les différentes valeurs de β testées.

		Classe estimée					
		$\beta = 0.8$			$\beta = 0.9$		
		1	2	3	1	2	3
Classe réelle	1	13 158	45	0	13 177	26	0
	2	53	11 736	31	35	11 772	13
	3	42	19	4916	24	15	4 938
Mal classifiés		0.63%			0.38%		

		$\beta = 1.0$			$\beta = 1.1$		
		1	2	3	1	2	3
		Classe réelle	1	13 192	11	0	13 197
2	20		11 793	7	15	11 801	4
3	15		13	4 949	8	12	4 957
Mal classifiés		0.22%			0.15%		

TABLE 3.1 – Matrice de confusion pour différentes valeurs de β

De manière générale, le taux de voxels mal classifiés ne dépasse pas les 1%. Conformément à la constatation visuelle (fig. 3.4), le nombre de voxels mal classifiés décroît lorsque β augmente. L'a priori sur les labels doit cependant être utilisé finement sur des images réelles dont la fréquence est plus élevée.

Estimation des paramètres

La figure 3.5 montre les histogrammes des échantillons générés par les chaînes MCMC pour chaque paramètre pour $\beta = 0.9$.

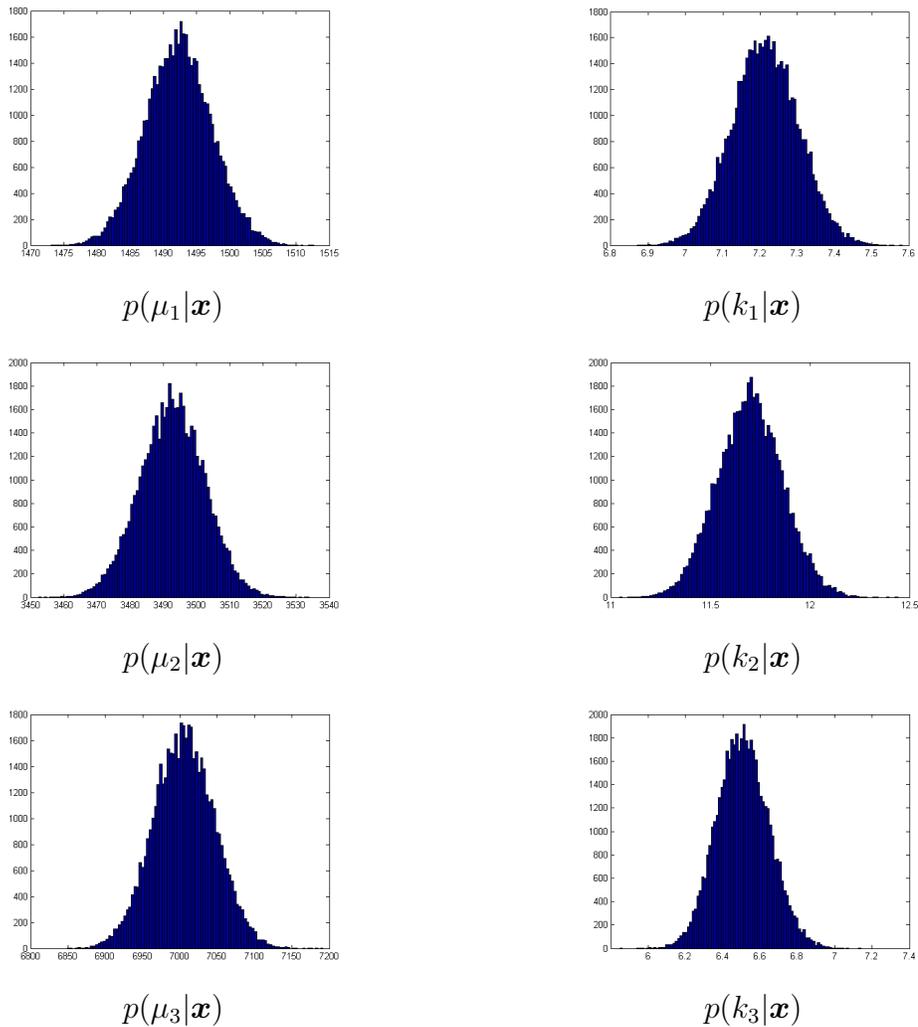


FIGURE 3.5 – Histogramme des échantillons générés par les chaînes MCMC après la période de chauffe. Valeurs théoriques $\boldsymbol{\mu} = [1500, 3500, 7000]$ et $\mathbf{k} = [7, 11, 6]$.

Nous constatons que les échantillons ont bien été générés autour des bonnes valeurs et que la moyenne des ces échantillons pour chaque paramètre peut bien être utilisée comme maximum a posteriori.

La même expérience a été conduite 50 fois de manière à mesurer la précision ainsi que la reproductibilité de l'algorithme sur des chaînes plus courtes de 6000 itérations (dont 1000 itérations de chauffe).

Précision

Les estimations MMSE de chaque paramètre ainsi que les écarts-types correspondants sont présentés table 3.2.

	valeur théorique	moyenne MMSE	écart-type
μ_1	1500	1492.3	0.0931
k_1	7	7.2138	0.0018
μ_2	3500	3492.6	0.2158
k_2	11	11.6983	0.0032
μ_3	7000	7005.9	1.0016
k_3	6	6.5084	0.0031

TABLE 3.2 – Estimation des paramètres calculée sur 50 chaînes indépendantes.

On constate que les estimations de moyenne concordent globalement avec les valeurs théoriques. Les résultats plus pertinents sont les écarts-types mesurés sur les 50 chaînes qui sont très faibles. L’algorithme offre ainsi une grande précision quant à l’estimation des paramètres. Il est important de préciser que la différence entre la valeur théorique et les valeurs estimées peuvent également être dues aux erreurs de génération des données. Les valeurs estimées représentent alors plutôt les valeurs empiriques des observations.

Reproductibilité

Pour mesurer la robustesse de l’algorithme, le facteur de réduction d’échelle potentiel (PSRF - *Potential Scale Reduction Factor*) a été mesuré sur les 50 expériences. La figure 3.6 montre l’évolution du PSRF du burn-in pour chaque paramètre.

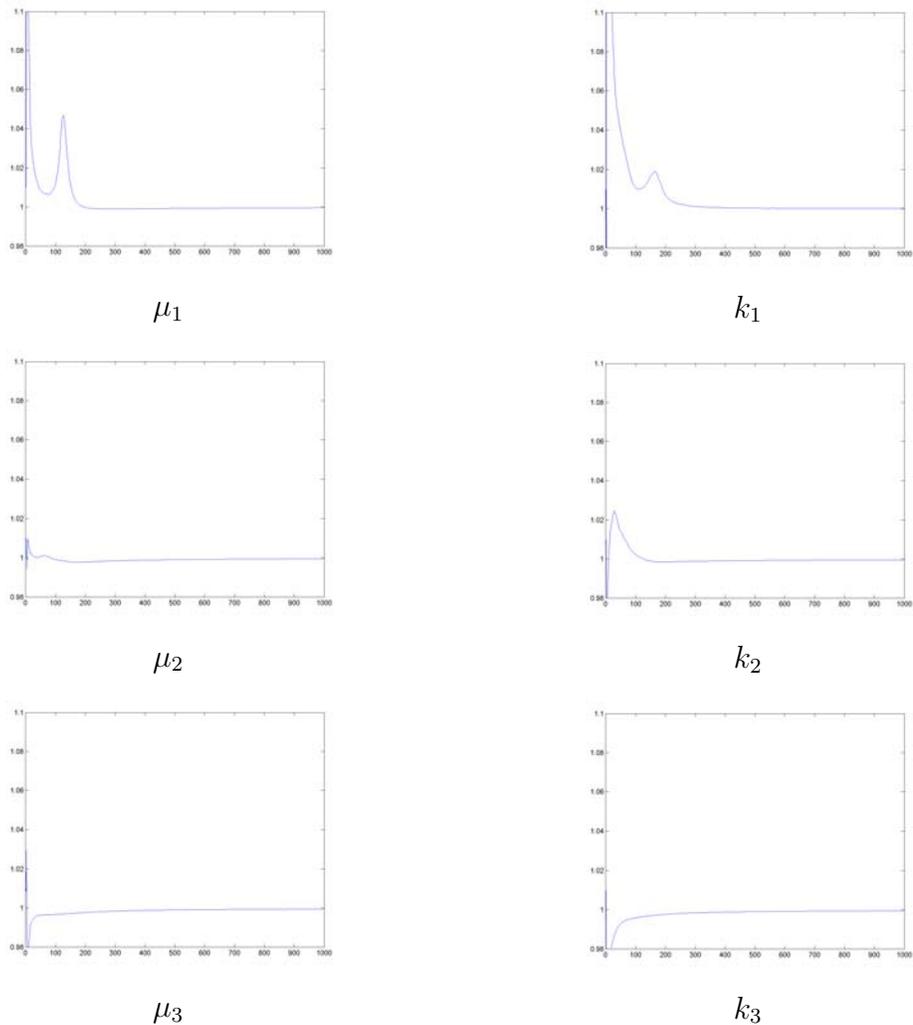


FIGURE 3.6 – Evolution du PSRF lors du burn-in pour chaque paramètre.

Le PSRF paraît globalement instable durant les 200 premières itérations. Cela peut-être dû à un problème d’identifiabilité. En effet, le fait d’échantillonner selon la loi jointe $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$ donne naissance à des modes locaux dans la densité cible. Plus concrètement la chaîne MCMC peut prendre différentes directions depuis son départ et peut nécessiter un saut important correspondant à un inversement des labels. Il faut donc compter quelques centaines d’itérations pour que l’algorithme converge grossièrement vers les bonnes valeurs. Après ces quelques itérations, le PSRF chute rapidement à 1. Après seulement 1000 itérations d’échantillonnage, un PSRF au

dessous de 1.001 a été obtenu pour chaque paramètre de chaque classe. Cela justifie la robustesse de l’algorithme à estimer les labels et les paramètres du modèle de manière jointe.

3.5.2 Données TEP réelles

Cette section a pour but de confirmer que le modèle de mélange de distributions binomiales négatives modélise bien les images TEP et d’évaluer les performances de l’algorithme à effectuer conjointement la classification de tissus et l’estimation des paramètres.

Pour cela, la méthode a été appliquée à une image TEP abdominale réelle. Ces données ont été acquises par un scanner GE Discovery ST et l’image a été reconstruite par un algorithme de type OSEM. L’image a été rognée sur une région d’intérêt ($116 \times 151 \times 3$). La figure 3.7 montre trois coupes successives de l’image de test. On peut y voir deux tumeurs, une grosse en bas à gauche, et une plus petite, en haut à droite. Dans le reste de cette section, des chaînes MCMC comprenant 6000 itérations (dont 1000 itérations de chauffe) ont été utilisées.

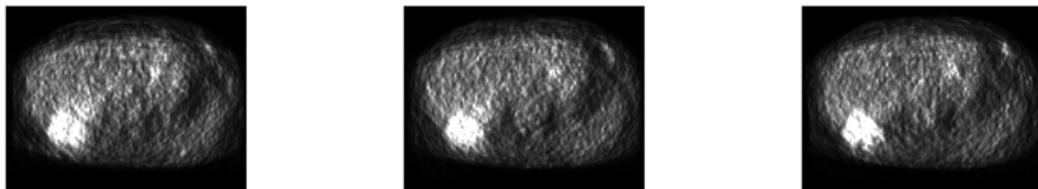


FIGURE 3.7 – 3 coupes consécutives de l’image TEP réelle.

Résultats de la segmentation

La figure 3.8 suivante montre les résultats de la segmentation sur le volume de données. Une granularité de $\beta = 0.9$ a été utilisée et $Z = 6$ classes ont été recherchées.

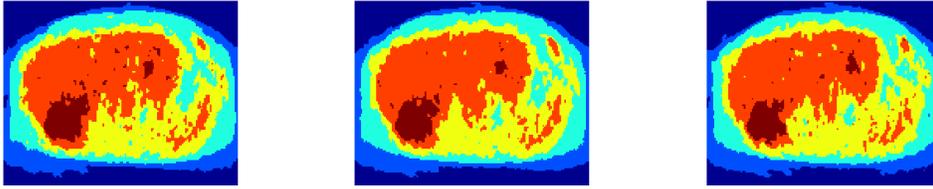
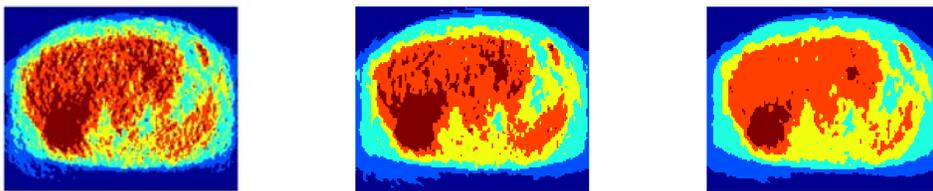


FIGURE 3.8 – Labels estimés sur les trois coupes ($Z = 6, \beta = 0.9$)

Nous remarquons que les deux tumeurs ont été segmentées, en accord avec l'évaluation d'un médecin. Ce résultat illustre la robustesse de la méthode proposée quand à la segmentation de lésions dans lesquelles l'activité est hétérogène malgré le contraste faible propre aux images TEP. De plus, nous observons que non seulement l'algorithme a segmenté les deux tumeurs, mais il a également affecté des labels appropriés aux autres tissus biologiques (foie, peau, chair, etc.) dont les caractéristiques biologiques sont distinctes.

Adéquation aux données

Pour mesurer à quel point le modèle de mélange de lois binomiales négatives modélise les données réelles, des tests d'adéquation aux données (*goodness of fit*) tels que celui de Kolmogorov-Smirnov, de Cramer von Mises et du χ^2 ont été effectués. La distribution des activités de l'image réelle a été modélisée par le modèle de mélange de distributions binomiales négatives proposé ainsi que par d'autres modèles de mélange ayant été proposés dans la littérature, à savoir le modèle de mélange Gaussien et le modèle de mélange de distributions Poisson. Les résultats de segmentation associés sont représentés figure 3.9.



Mélange Gaussien

Mélange Poisson

Mélange BN

FIGURE 3.9 – Segmentation obtenue en modélisant les données par différents mélanges ($Z = 6, \beta = 0.9$) pour la coupe 2.

Les résultats des tests d'adéquation aux données sont présentés dans le tableau 3.3.

Mélange de distributions	Kolmogorov-Smirnov	Cramer von Mises	χ^2
Gaussiennes	0.12422	0.15337	381.9175
Poisson	0.24246	0.59478	Inf
Binomiales négatives	0.088113	0.082125	0.078718

TABLE 3.3 – Tests d'adéquation aux données, obtenus pour différents modèles de mélange

Les valeurs des critères d'information d'Akaike (AIC) et Bayésien (BIC) pour chacun des modèles est également présenté sur la table 3.4.

Mélange de distributions	AIC	BIC
Gaussiennes	5108.4939	5214.9277
Poisson	43623.0957	43676.3126
Binomiales négatives	4021.8675	4128.3013

TABLE 3.4 – Critère d'Information d'Akaike (AIC) et Critère d'Information Bayésien(BIC)

On remarque que le modèle de mélange proposé de distributions binomiales négatives modélise mieux les données que les autres modèles de mélange. Les résultats du χ^2 sont particulièrement pertinents. Les tests d'adéquation aux données ne sont pas si mauvais pour le mélange de Gaussiennes, car les lois Gaussiennes sont relativement souples et peuvent facilement s'adapter aux données. Par contre, les résultats de segmentation montrent que le partitionnement obtenu n'est pas bon. En effet, les classes sont très dispersées alors que le même coefficient de granularité a été utilisé. La segmentation donnée par le modèle de mélange de distributions Poisson est plus intéressant que celui des lois Gaussiennes, mais les résultats des tests quantitatifs sont très mauvais. Le mélange de lois binomiales négatives offre à la fois une bonne segmentation, et un modèle fin pour les données.

Pour des résultats graphiques, la figure 3.10 représente l'histogramme des données ainsi que les courbes associées aux différents modèles testés, sur une échelle linéaire et logarithmique.

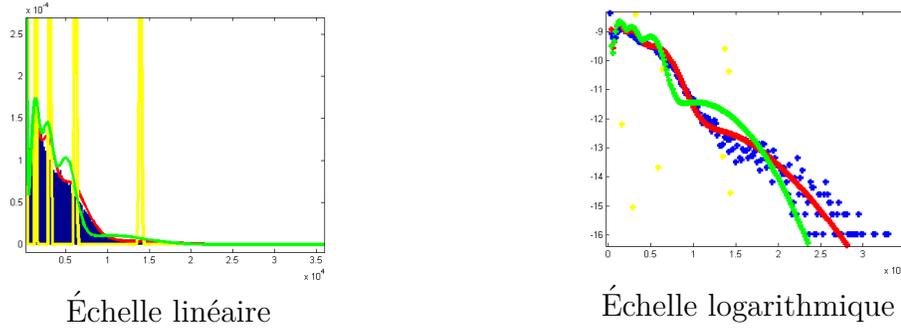


FIGURE 3.10 – Ajustement aux données des modèles de mélange testés sur une échelle linéaire. En bleu : histogramme des données ; en rouge : mélange de DBN ; en vert : mélange de distributions Gaussiennes ; en jaune : mélange de distributions Poisson.

On constate que la densité estimée en utilisant le modèle de mélange proposé (en rouge) concorde de manière fine avec les données empiriques. Cette concordance est particulièrement représentative pour les fortes valeurs d'activité. Le modèle proposé sera donc d'autant plus efficace pour modéliser les classes tumorales (les tumeurs présentant les plus fortes activités). Nous voyons ici clairement que le modèle de mélange Poisson n'est pas adapté. Chacune des composantes du mélange est tenue à avoir la moyenne égale à la variance. Une simple renormalisation des données par l'algorithme de reconstruction peut en être la cause.

Précision

Cette même expérience, pour $Z = 6$ et $\beta = 0.9$ a été reproduite 50 fois. Les estimations MMSE de chaque paramètre ainsi que les écarts-types correspondants sont rapportés dans le tableau 3.5

	MMSE	écart-type
μ_1	38	0.0329
k_1	2.0598	0.0014
μ_2	239	0.4864
k_2	2.8102	0.0047
μ_3	1444	1.4661
k_3	7.3396	0.0122
μ_4	3132	2.2365
k_4	12.6628	0.0139
μ_5	6181	1.7271
k_5	14.0360	0.0105
μ_6	13905	5.9256
k_6	9.9645	0.0128

TABLE 3.5 – Estimations MMSE pour chaque paramètre, calculés sur 50 chaînes indépendantes de 5000 itérations.

On remarque que les écarts types sont faibles par rapport à la valeur de la moyenne, aussi bien pour le paramètre de moyenne μ que pour le paramètre de dispersion inverse k . Ceci atteste d'une bonne précision de l'algorithme.

Reproductibilité

Pour juger de la reproductibilité de l'algorithme, la valeur du PSRF à la fin du burn-in est présentée dans la table 3.6.

	PSRF
μ_1	1.0002
k_1	0.9997
μ_2	1.0016
k_2	0.9998
μ_3	1.0018
k_3	0.9996
μ_4	1.0006
k_4	0.9995
μ_5	0.9996
k_5	0.9995
μ_6	0.9996
k_6	0.9998

TABLE 3.6 – PSRF atteint à la fin du burn-in, pour chaque paramètre, calculés sur 50 chaînes indépendantes.

Un PSRF autour de 1 et inférieur à 1.002 a été trouvé pour chaque paramètre, attestant d'une bonne répétabilité.

Influence du nombre de classes Z

La figure 3.11 montre les résultats de la segmentation selon le nombre de classes (Z) recherchées. La granularité β est fixée à 0.9.

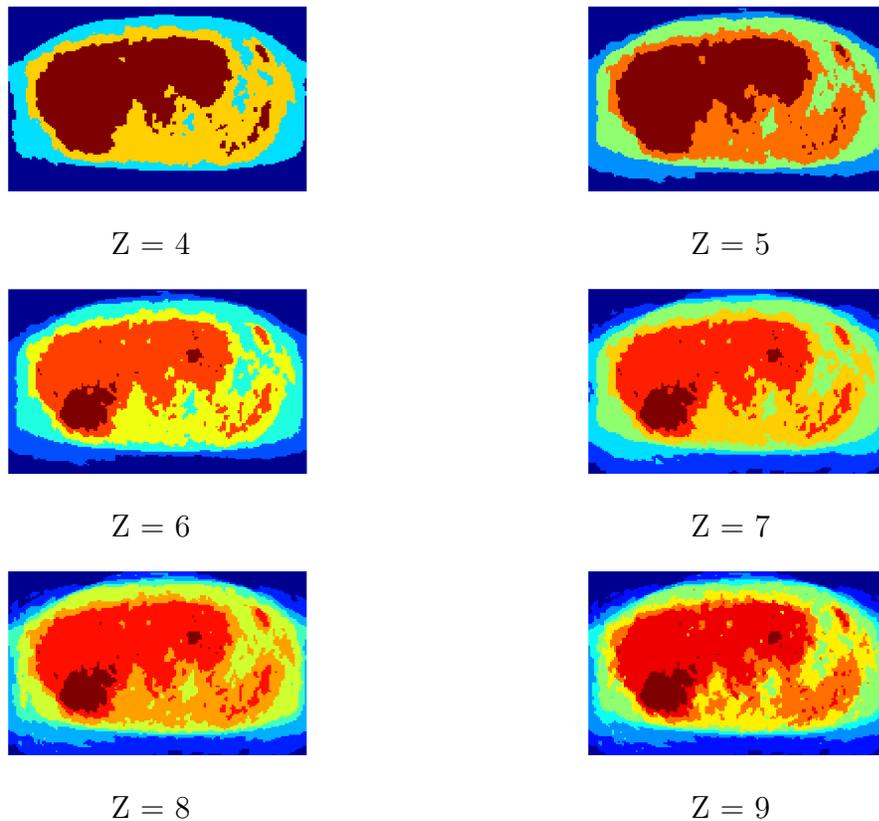


FIGURE 3.11 – Résultats de la segmentation selon le nombre de classes recherchées.

Le nombre de classes Z n'est pas estimé par l'algorithme. Il pourra cependant être choisi visuellement par les cliniciens. Une perspective à cette étude serait d'estimer conjointement le nombre de classes en l'intégrant dans le modèle Bayésien. Une piste serait d'utiliser un algorithme MCMC à saut réversible [53].

Il est cependant important de remarquer que le choix du nombre de classes à rechercher n'est pas si significatif car la segmentation des tumeurs devient stable. En effet, on remarque sur les images que la tumeur est segmentée grossièrement de la même manière pour Z supérieur à 6. La tumeur est bien représentée par une distribution binomiale négative unique, donc une classe sera facilement représentative. Mais ce n'est pas autant le cas pour le fond, comme les voxels du fond sont moins bien représentés par une distribution binomiale négative unique, l'algorithme préférera modéliser les voxels appartenant au fond par plusieurs classes.

Il est également intéressant de constater que l'algorithme affecte une classe propre

aux voxels situés sur la périphérie des tissus, vraisemblablement parce qu'ils sont victimes de l'effet de volume partiel. On peut donc en déduire qu'il est important de corriger l'effet de volume partiel si l'on souhaite segmenter avec plus de précision les différents tissus [64, 123, 129].

Le tableau 3.7 présente les résultats d'attache aux données pour les différentes valeurs de Z .

Nombre de classes	Kolmogorov-Smirnov	Cramer von Mises	χ^2
Z=4	0.030821	0.025961	1.2367
Z=5	0.086707	0.092967	0.9274
Z=6	0.088113	0.082125	0.0787
Z=7	0.032851	0.010095	0.0245
Z=8	0.073447	0.077858	0.0494
Z=9	0.078677	0.086413	0.0505

TABLE 3.7 – Résultats des tests de Kolmogorov-Smirnov et de Cramer von Mises obtenues pour différents nombre de classes Z

Les différents tests semblent donner une meilleure modélisation pour $Z = 7$. Choisir un nombre de classes trop bas peut être problématique car des tissus différents devront être modélisés par une même classe, et les paramètres du modèle de cette classe devra donc s'adapter à ces différents tissus. Il est intéressant de remarquer également que choisir un nombre de classes plus élevé a plutôt tendance à enfreindre la modélisation. Cependant, on note que l'augmentation du nombre de classes n'altère pas la cohérence spatiale des régions segmentées et notamment des tumeurs. Ce résultat soutient la robustesse du modèle.

Influence du paramètre de granularité β

La figure 3.12 affiche les résultats du processus de segmentation pour différentes valeurs de granularité β .

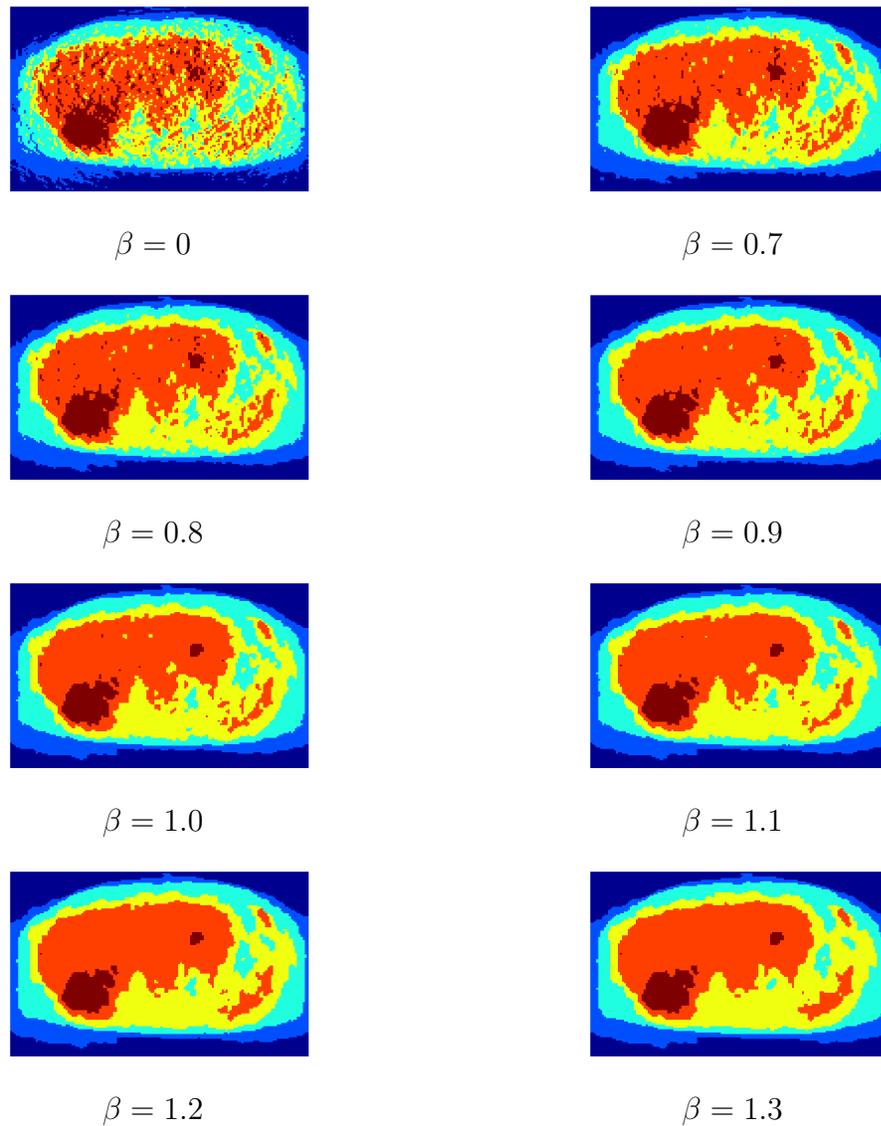


FIGURE 3.12 – Influence de la granularité

Nous constatons que même sans a priori spatial ($\beta = 0$), la majeure partie des voxels a été correctement étiquetée. Cela revendique le fait que le modèle proposé est particulièrement bien adapté car il est responsable de la majeure partie du résultat. Un β de forte valeur aura tendance à lisser l'image, et à retourner un partitionnement où les différentes classes sont plus connexes.

Le tableau 3.8 présente les résultats d'attache aux données pour les différentes valeurs de β .

Granularité	Kolmogorov-Smirnov	Cramer von Mises	χ^2
$\beta = 0$	0.055126	0.05148	0.06403
$\beta = 0.7$	0.089383	0.08331	0.08763
$\beta = 0.8$	0.088886	0.08294	0.08316
$\beta = 0.9$	0.088113	0.08213	0.07872
$\beta = 1.0$	0.087075	0.08094	0.07462
$\beta = 1.1$	0.086269	0.07934	0.07151
$\beta = 1.2$	0.085492	0.07986	0.06903
$\beta = 1.3$	0.085295	0.07791	0.06822

TABLE 3.8 – Résultats des tests de Kolmogorov-Smirnov et de Cramer von Mises obtenues pour différentes granularité β

Les meilleurs résultats sont obtenus lorsqu'aucun a priori spatial n'a été utilisé. Ce résultat est conforme à nos attentes car le fait de ne pas utiliser d'a priori spatial élimine une contrainte, et chaque voxel peut être rattaché à la classe qui le modélise le mieux.

On remarque cependant qu'hormis ce cas, les données sont mieux modélisées pour des fortes valeurs de granularité β , même si la considération de la cohérence spatiale peut contraindre la qualité des tests d'adéquation aux données. Ainsi l'a priori spatial ne permet pas seulement d'obtenir un meilleur résultat visuel, en dépit d'une bonne modélisation, mais aide visiblement l'algorithme à modéliser correctement les données. En d'autre termes, conformément à la problématique fixée, l'algorithme ne fournit pas la meilleure solution visuelle de segmentation, ni le meilleur modèle pour les données, mais bien le meilleur compromis entre ces deux aspects.

3.5.3 Comparaison avec les méthodes de seuillage

La figure 3.13 représente 3 coupes successives, centrées sur la tumeur, de différents jeux de données réelles, reflétant les différents types de tumeurs rencontrées en clinique. Chaque ligne correspond à un patient différent, à l'exception des deux premières lignes, représentant deux tumeurs d'un même patient. Trois délimitations sont également représentées :

en bleu, une délimitation manuelle établie par un expert.

en vert, la délimitation obtenue par un seuillage adaptatif, tel qu'utilisé couramment en clinique.

en rouge, la délimitation estimée par l'algorithme proposé.

La valeur du seuil utilisée dans la technique de seuillage adaptatif a été fixée en utilisant la méthode de Daisne [33, 71], selon la courbe de calibration (3.19) ci-dessous :

$$s_{\text{adapt}} = 95.6 * \exp(-SB/2.4) + 32.1 \quad (3.19)$$

où SB est le rapport du signal maximum dans la tumeur sur le signal moyen dans l'environnement et où les paramètres mis en jeu ont été estimés à partir d'expériences sur fantômes, en utilisant la méthode itérative de Jentzen [72].

On remarque que toutes les tumeurs ont bien été délimitées. On constate que la délimitation obtenue en utilisant la méthode de seuillage adaptatif ne délimite que la partie la plus active des tumeurs. Cela est problématique lorsque la tumeur est hétérogène, comme c'est le cas pour le patient A-T1. Les méthodes ne considérant pas l'hétérogénéité des tumeurs auront tendance à sous-estimer leur volume.

Bien que la segmentation manuelle ne soit pas une référence absolue, le tableau 3.9 présente les coefficients DICE obtenus en comparant la segmentation manuelle à la segmentation estimée par notre méthode, ainsi qu'aux résultats de segmentation obtenus par différents seuillages. Les valeurs des seuils adaptatifs sont présentés dans le tableau 3.10.

Jeu	Seuil 30%	Seuil 35%	Seuil 40%	Seuil adaptatif	Méthode proposée
Patient A-T1	0.8260	0.7170	0.6151	0.7895	0.8652
Patient A-T2	0.3893	0.5185	0.6065	0.5581	0.7055
Patient B	0.7987	0.8209	0.7712	0.5842	0.8636
Patient C	0.4836	0.4321	0.3344	0.4699	0.8871
Patient D	0.7247	0.6191	0.5180	0.6607	0.8800

TABLE 3.9 – Coefficients DICE obtenus en comparant la segmentation manuelle à la segmentation estimée par notre méthode, ainsi qu'à la segmentation obtenue par différents seuillages

On note en premier lieu que la segmentation obtenue en utilisant la méthode proposée offre des résultats plus proche d'une segmentation manuelle qu'une segmentation résultant d'une méthode de seuillage. Il est également intéressant de noter que le seuil à utiliser pour obtenir une segmentation plus proche de la segmentation

manuelle varie selon l'image utilisée. En effet, les délimitations obtenues sur les patients A-T1, C et D sont plus fidèles à la délimitation manuelle en utilisant un seuil de 30% alors qu'un seuil de 35% est plus représentatif sur le patient B et 40% pour le patient A-T2. Les résultats obtenus par la méthode proposée concorde avec les résultats obtenus manuellement quelque soit le type de tumeur.

Le tableau 3.10 présente le seuil nécessaire pour une technique de seuillage afin que tous les pixels obtenus par la méthode proposée soient considérés. A titre de comparaison, les seuils utilisés en clinique selon une technique de seuillage adaptatif sont également présentés.

Jeu	Seuil nécessaire	Seuil adaptatif estimé
Patient A, T1	16.8605%	32.6685%
Patient A, T2	26.6173%	35.5100%
Patient B	20.3397%	49.6897%
Patient C	10.6146%	32.1152%
Patient D	17.4423%	33.3576%

TABLE 3.10 – Seuil nécessaire pour considérer tous les voxels segmentés par la méthode proposée et seuil adaptatif estimé

Les seuils obtenus sont très faibles. Ces résultats montrent que la délimitation habituelle utilisant une méthode de seuillage a tendance à sous-estimer le volume de la tumeur par rapport à la méthode proposée. On peut déduire de ces résultats que la méthode proposée est moins sensible au bruit que les méthodes se basant directement sur l'activité mesurée dans chaque pixel.

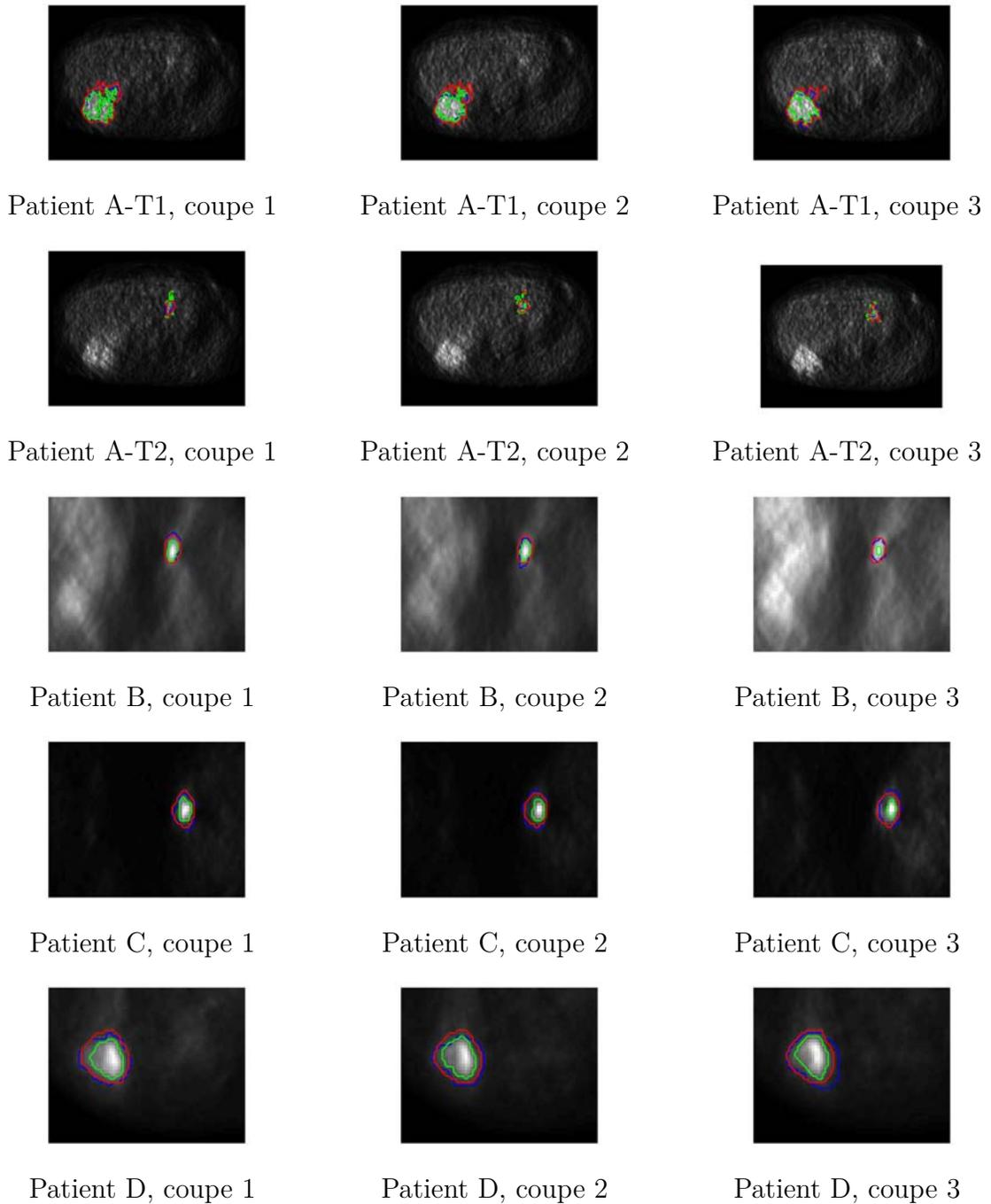


FIGURE 3.13 – 3 coupes TEP successives de différents patients et délimitations obtenues par la méthode proposée (en rouge), par une technique de seuillage adaptatif (en vert) et manuellement (en bleu).

Spatiotemporal fusion of PET-CT data for tumor tracking

Introduction

The anatomical and functional information are complementary and are therefore potentially interesting to combine.

In this chapter, an automatic spatio-temporal segmentation technique for 4-D bimodal PET-CT images is presented. The temporal dimension corresponds here to the different stages of the respiratory cycle.

The proposed delimitation process allows the possibility to locate and track the tumors.

Problem Statement

To reduce the blurring effects in PET images due to respiratory movements, different pairs of PET and CT images have been reconstructed at various stages of the respiratory cycle.

We propose to model the bimodal PET-CT data of a unique class by a bivariate negative binomial-normal (BN-N) distribution. The data of an image with various tissues are then described by a mixture of BN-N distributions (see (4.5)).

The problem of joint segmentation and parameter estimation is set as a *maximum a posteriori* (MAP) estimation (see (4.6)).

Hierarchical Bayesian model

Likelihood (data model)

By expressing the likelihood of a voxel from a specific class as a product (4.7) and by assuming that all observations are independent in each channel, the likelihood of the whole image can be defined by (4.8).

Prior on the model parameters θ

Gamma distributions are chosen as prior for each parameter (see (4.9)).

Assuming that the mixture parameters are independent, the joint prior distribution for θ is given by (4.10).

Prior on the labels \mathbf{z}

In this study, a 4-D Potts-Markov random field is proposed as a prior density for \mathbf{z} (see (4.12)). Each voxel is connected to its 6 spatial neighbors, and its 2 temporal neighbors (see figure 4.1).

Posterior distribution $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$

Considering that the parameter vector $\boldsymbol{\theta}$ and the label vector \mathbf{z} are independent, using Bayes theorem, the posterior distribution can be expressed as in (4.13).

Hybrid Gibbs sampler

A Metropolis-within-Gibbs sampler is proposed to solve the problem of segmentation and parameter estimation. For this, we alternatively propose $\boldsymbol{\theta}$ and \mathbf{z} according to the associated conditional distributions.

Conditional distribution $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$

The conditional distribution $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ can be developed as in (4.14).

The parameter vectors $\boldsymbol{\theta}$ are generated according to the proximal Manifold MALA (prox-mMALA) detailed in chapter 2.

Because the proposal distributions (4.17) are not symmetric, the acceptance ratio a is expressed as in (4.16).

Conditional distribution $p(\mathbf{z}|\boldsymbol{\theta}, \mathbf{x})$

Sampling \mathbf{z} is done voxel by voxel. The conditional distribution of the discrete label z_n of each voxel can then be fully characterized by (4.19).

As (4.19) defines a Markov random field, the samples are drawn according to this conditional distribution by drawing integers in the set $\{1, \dots, Z\}$ with probabilities defined in (4.19).

Summarized algorithm

The pseudo-code (3) summarizes the proposed proximal Manifold MALA-within-Gibbs algorithm.

Experimental results

Experiments were performed on real 4-D PET-CT images of the thorax.

Figure 4.2 shows the original PET and CT images, as well as the segmentation results of each modality separately using a univariate mixture model.

The proposed algorithm was applied to the data using different mixture models :

- $\mathcal{N} \times \mathcal{N}$: Mixture model of normal distributions, assuming the PET and CT channels independent.
- $\text{Biv}\mathcal{N}$: Mixture model of bivariate normal distributions, without independence assumption.
- $\mathcal{P} \times \mathcal{P}$: Mixture model of Poisson distributions, assuming the PET and CT channels independent.
- $\text{Biv}\mathcal{P}$: Mixture model of bivariate Poisson distributions, without independence assumption.
- $\mathcal{BN} \times \mathcal{N}$: Mixture model of negative binomial - normal distributions proposed in this chapter.

Segmentation results

Figures (4.3.a) to (4.3.e) present the segmentation results obtained by the five mixture models.

The visual segmentation results with distributions $(\mathcal{BN} \times \mathcal{N})$, $(\mathcal{P} \times \mathcal{P})$ and $(\text{Biv}\mathcal{P})$ are better than normal distributions. With the model $(\mathcal{BN} \times \mathcal{N})$, anatomical structures are better respected than with the other models. In addition, a single class is assigned to the tumor, except for some voxels with very high biological activity.

Influence of the number of classes

Figures (4.4.a) to (4.4.e) present the segmentation results obtained with different number of classes Z . It is important to note that beyond $Z = 5$, the delineation of the tumor remains constant, which indicates the robustness of the method.

Influence of the granularity

Figures (4.5.a) to (4.5.d) report the segmentation results obtained with different coefficients of granularity β .

We note that in all cases (except the case $\beta = 0$), the tumor was discriminated and its volume remains the same, indicating that a poor choice of β has no significant impact on the delineation of the tumor.

Tumor tracking

Figure 4.6 presents the segmentation results of the 6 temporal bins, corresponding to the different phases of a respiratory motion.

The labeling process gives good regularity between the temporal slices.

Chapitre 4

Fusion spatio-temporelle de données TEP-TDM pour le suivi de tumeur

4.1 Introduction

L'imagerie fonctionnelle permet parfois de détecter des changements pathologiques à des stades où aucun changement n'est encore visible en imagerie anatomique. Les structures des organes sont toutefois difficilement localisables voire absentes. La modalité anatomique offre quant à elle une image de meilleure résolution permettant d'obtenir des informations précises sur les caractéristiques géométriques des tumeurs telles que leur localisation spatiale, leur taille, leur volume ou leur forme. Les informations anatomiques et fonctionnelles sont complémentaires et donc potentiellement intéressantes à combiner. Leur fusion permet d'améliorer le diagnostic clinique dans de nombreux examens à visée oncologique [2, 127]. De nos jours, les systèmes TEP sont souvent couplés à des tomodesitomètres (TDM) permettant ainsi d'obtenir des informations anatomiques et fonctionnelles lors du même examen.

La superposition d'images anatomique et fonctionnelle implique un recalage précis entre les deux modalités. Les acquisitions TEP et TDM étant souvent séquentielles, la qualité du recalage est toutefois perturbée par les mouvements physiologiques du patient comme la respiration, les battements cardiaques, etc. En particulier, la lenteur de l'acquisition TEP résulte en une image représentant généralement une moyenne du signal reçu sur plusieurs cycles respiratoires. Pour réduire ces artefacts dus aux mouvements de respiration, des travaux récents ont proposé une technique de reconstruction 4D PET-TDM permettant de partitionner le cycle respiratoire

en bins temporels et de reconstruire des images TEP et TDM pour chacun de ces bins [102, 35]. Ce type de reconstruction fournit des images 4-D (une série temporelle d'images 3-D) bimodales recalées spatio-temporellement. La segmentation spatio-temporelle de telles images 4D PET-TDM cohérentes offre l'avantage de pouvoir localiser et de suivre le déplacement des tumeurs, ce qui est particulièrement intéressant en radiothérapie, où une segmentation précise est nécessaire à chaque instant afin d'épargner les tissus sains.

En pratique clinique, les tumeurs sont délimitées manuellement ou semi-automatiquement en TEP et en TDM séparément. De nombreuses méthodes de segmentation ont été proposées dans chacune des deux modalités. Cependant, les efforts de recherche se sont concentrés davantage sur les techniques de segmentation propre à une seule des deux modalités. Seules quelques études récentes ont été consacrées à la segmentation automatique des deux modalités. Wojak et al.[144] ont proposé une approche variationnelle en deux phases où les données TEP sont utilisées pour guider la segmentation de l'images TDM. Ballangan et al.[6] ont proposé une technique où la région tumorale obtenue à partir de l'image TEP est utilisée comme masque pour effectuer une segmentation dans l'image TDM associée, en résolvant un modèle de mélange Gaussien à deux composantes . Ces méthodes utilisent l'image TEP pour informer la segmentation de l'image TDM mais ne tirent pas entièrement profit de la bimodalité.

Han et al. [55] ont formulé le problème de segmentation comme la minimisation de l'énergie d'un champ de Markov (MRF) binaire en pénalisant la différence de segmentation entre les images TEP et TDM. Ce procédé nécessite cependant une interaction de l'utilisateur. Plus récemment, Bagci et al. [4] ont présenté une méthode automatisée basée sur les graphes pour segmenter simultanément les structures fonctionnelles et anatomiques. Gribben et al.[54] ont proposé une technique de segmentation multimodale dans un cadre Bayésien. Une représentation vectorielle est utilisée pour les données TEP et TDM et les intensités de voxels sont modélisées par une distribution Gaussienne bivariée. Un modèle de Potts 3-D est également proposé comme a priori spatial. La loi a posteriori est ensuite maximisée par un algorithme de type Espérance-Maximisation (EM). La mise en oeuvre suppose toutefois que les tumeurs soient isolées à l'intérieur d'une région d'intérêt et la dimension temporelle n'est pas prise en compte. Récemment, Bai et al.[5] ont proposé une segmentation 4-D d'images TEP-TDM en utilisant un champ aléatoire de Markov (MRF) et une régularisation entre la TEP et la TDM. Leur méthode consiste à segmenter l'image TDM à l'aide d'un algorithme d'optimisation du flot maximum en utilisant les données TEP comme information a priori.

Dans ce chapitre, une technique automatique de segmentation d’images bimodales est présentée. Pour éviter les problèmes de recalage dus aux mouvements respiratoires, le processus de délimitation proposé permet de réaliser une segmentation spatio-temporelle sur des images 4-D. Cette segmentation vise à discriminer les tissus présentant des statistiques différentes et repose donc sur un modèle statistique des données. Pour cela, un nouveau modèle de représentation des images TEP-TDM est proposé. Les données bimodales sont représentées comme un mélange de distributions bi-dimensionnelles Négative Binomiale-Normale (BN-N). Ce modèle a l’avantage de pouvoir représenter l’hétérogénéité des tumeurs sur les images TEP, comme justifié au chapitre 2. Un modèle Bayésien hiérarchique est proposé dans lequel l’estimation des paramètres du modèle et le processus de segmentation sont réalisés conjointement. Un champ aléatoire de Markov (MRF) 4-D est incorporé au modèle pour modéliser la corrélation spatio-temporelle entre les composantes du mélange. Ce problème d’optimisation est résolu par un algorithme original de type Monte-Carlo par chaîne de Markov (MCMC).

Le chapitre est organisé comme suit. Le modèle des données et le problème de segmentation spatio-temporelle est énoncé section 4.2. La section 4.3 présente le modèle Bayésien hiérarchique et en particulier le champs de Markov-Potts 4D utilisé comme a priori spatio-temporel. L’algorithme MCMC utilisé pour estimer conjointement les paramètres du modèle et les étiquettes des voxels est présenté dans la section 4.4. Enfin, la section 4.5 rapporte les résultats obtenus sur des données bimodales 4-D réelles.

4.2 Position du problème

Soit \mathbf{x} une image 4-D TEP-TDM (une série temporelle d’images 3-D). La valeur du n ième voxel $\mathbf{x}_n \in \mathbb{N}^2$ est un vecteur à deux composantes correspondant aux données TEP et TDM : $\mathbf{x}_n = (x_n^{\text{TEP}}, x_n^{\text{TDM}})^T$. On admettra que les canaux TEP et TDM ont été correctement recalés spatialement. De plus, pour réduire les effets de flous dus aux mouvements respiratoires dans l’image TEP, différents couples d’images (TEP et TDM) ont été reconstruits à différentes étapes du cycle respiratoire selon la méthode présentée dans [102, 35]. La dimension temporelle correspond donc ici aux différents stades du cycle respiratoire. On supposera que l’image \mathbf{x} est composée de Z tissus biologiques distincts $\{\mathcal{Z}_1, \dots, \mathcal{Z}_Z\}$ selon lesquels les voxels suivent une distribution statistique caractéristique (Z étant supposé connu dans cette étude).

Ce chapitre résout le problème de segmentation en se basant sur les distributions statistiques des tissus, dans le sens où deux voxels seront associés à une même classe s’ils partagent des statistiques similaires. Pour cela, un nouveau modèle de données

est proposé.

Dans [54], un mélange de lois Gaussiennes bivariées a été utilisé pour modéliser les données TEP-TDM bimodales. Cependant, comme les lois marginales des lois Gaussiennes multivariées sont elles-mêmes Gaussiennes, cela implique que l'activité TEP sera modélisée par une loi Gaussienne. Or le chapitre 2 montre la meilleure efficacité des lois binomiales négatives pour modéliser les données TEP, notamment afin de prendre en compte l'hétérogénéité des tumeurs. Étant donné une classe \mathcal{Z}_z , nous souhaitons donc que la variable aléatoire associées aux données TEP X_z^{TEP} suive une loi binomiale négative :

$$X_z^{\text{TEP}} \sim \mathcal{BN}(\mu_z, k_z) \quad (4.1)$$

Concernant la modélisation de coefficients d'atténuation en TDM, la représentation classique par loi normale [122, 6, 54] est admise dans ce chapitre :

$$X_z^{\text{TDM}} \sim \mathcal{N}(m_z, \sigma_z) \quad (4.2)$$

Étant donné les deux lois marginales (4.1) et (4.2), et en supposant l'indépendance entre les deux modalités, nous proposons de modéliser les données bimodales TEP-TDM d'une même classe \mathcal{Z}_z par une distribution bivariée *binomiale négative-normale* (BN-N) :

$$(X_z^{\text{TEP}}, X_z^{\text{TDM}}) \sim \mathcal{BN}(\mu_z, k_z) \times \mathcal{N}(m_z, \sigma_z) \triangleq \mathcal{BNN}(\theta_z) \quad (4.3)$$

où $\theta_z \triangleq (\mu_z, k_z, m_z, \sigma_z)$ sont les paramètres associés au tissu \mathcal{Z}_z .

On a donc :

$$\forall x_n \in \mathcal{Z}_z, x_n \sim \mathcal{BNN}(\theta_z) \quad (4.4)$$

Les données d'une image \mathbf{x} présentant divers tissus seront ainsi décrites par *mélange* de distributions \mathcal{BNN} .

$$x_n \propto \sum_{z=1}^Z \omega_z \mathcal{BNN}(\theta_z) \quad (4.5)$$

où ω_z représente la proportion de voxels associés à la classe \mathcal{Z}_z avec $\sum_{z=1}^Z \omega_z = 1$.

Ce chapitre résout le problème de la segmentation d'images bimodales TEP-TDM à 4 dimensions en se basant sur les statistiques des tissus. Le modèle de données proposé permet de prendre en compte les variations d'activité à l'intérieur d'un même tissu en TEP. Le résultat de cette segmentation fournit également les paramètres du mélange de lois utilisés pour modéliser l'image.

Pour cela, le problème de segmentation et d'estimation conjointe des paramètres est posé comme une recherche de *maximum a posteriori* (MAP) :

$$(\hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}) = \underset{p(\boldsymbol{\theta}, \mathbf{z})}{\operatorname{argmax}} p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x}) \quad (4.6)$$

où $\boldsymbol{\theta}$ est le vecteur de paramètres inconnu associé à un modèle paramétrique de formation d'image. Ce problème est résolu dans un modèle Bayésien hiérarchique présenté dans la section suivante 4.3.

4.3 Modèle Bayésien hiérarchique

Le théorème de Bayes nous permet d'exprimer la distribution a posteriori (4.6) à partir de la vraisemblance $p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x})$ et de la loi a priori $\pi(\boldsymbol{\theta}, \mathbf{z})$. Leur expression est définie dans les sections suivantes.

4.3.1 Vraisemblance (modèle des données)

Comme justifié dans la section 4.2, nous proposons d'exprimer la vraisemblance d'un voxel x_n d'une classe \mathcal{Z}_z comme le produit :

$$\begin{aligned} p(x_n | \boldsymbol{\theta}_z, z_n = z) &= p_{\mathcal{BN}}(x_n^{\text{TEP}} | \mu_z, k_z) \cdot p_{\mathcal{N}}(x_n^{\text{TDM}} | m_z, \sigma_z) \\ &= \binom{x_n^{\text{TEP}} + k_z - 1}{x_n^{\text{TEP}}} \left(\frac{\mu_z}{\mu_z + k_z} \right)^{x_n^{\text{TEP}}} \left(\frac{k_z}{\mu_z + k_z} \right)^{k_z} \\ &\quad \cdot \frac{1}{\sigma_z \sqrt{2\pi}} e^{-\frac{(x_n^{\text{TDM}} - m_z)^2}{\sigma_z^2}} \end{aligned} \quad (4.7)$$

avec $\boldsymbol{\theta}_z = (\mu_z, k_z, m_z, \sigma_z)$

En supposant que toutes les observations sont indépendantes dans chacun des canaux, la vraisemblance peut être alors définie par :

$$p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}) = \prod_{z=1}^Z \prod_{n \in \mathcal{I}_z} p(x_n | \boldsymbol{\theta}_z, z_n = z) \quad (4.8)$$

où \mathcal{I}_z est l'ensemble des voxels appartenant à la classe \mathcal{Z}_z et $\boldsymbol{\theta} = \{\boldsymbol{\theta}_z\}_{z=1..Z}$.

4.3.2 A priori sur les paramètres du modèle θ

La matrice θ est l'ensemble des paramètres inconnus pour le modèle proposé. Des distributions gamma sont choisies comme distribution a priori pour chaque paramètre $\theta_{z,i}$

$$\theta_{z,i} \sim \Gamma(1 + a, \frac{-1}{b}), \quad \forall z = 1 \dots Z, \forall i = 1 \dots 4, \quad (4.9)$$

où les hyperparamètres a et b sont ici communs pour toutes les classes et définissent un a priori non-informatif.

En supposant que les paramètres du mélange sont indépendants, la distribution a priori jointe pour θ est écrite :

$$\pi(\theta) = \prod_{z=1}^Z \prod_{i=1}^4 \pi(\theta_{z,i}) \quad (4.10)$$

4.3.3 A priori sur les étiquettes z

Il est naturel de considérer que l'étiquette d'un voxel est corrélée avec celles de ses voisins. Dans cette étude, un champ aléatoire de Markov 4-D de type Potts est proposé comme densité a priori pour z .

Chaque voxel est connecté à ses 6 voisins spatiaux, et à ses 2 voisins temporels (voir la figure 4.1). La densité a priori des étiquettes correspond donc à la distribution

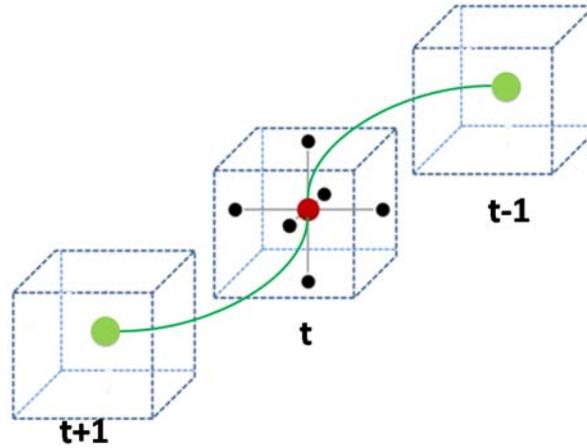


FIGURE 4.1 – 4-D neighborhood

de Potts :

$$\pi(\mathbf{z}) = \frac{1}{C(\beta)} \exp \left[\sum_{n=1}^N \left(\beta_s \sum_{n' \in \mathcal{V}_s(n)} \delta(z_n - z_{n'}) + \beta_t \sum_{n' \in \mathcal{V}_t(n)} \delta(z_n - z_{n'}) \right) \right] \quad (4.11)$$

où $C(\beta)$ est la fonction de partition, $\delta(\cdot)$ est la fonction de Kronecker et où $\mathcal{V}_s(n)$ et $\mathcal{V}_t(n)$ désignent respectivement les voisinages spatiaux et temporels avec les coefficients de granularité correspondants β_s et β_t .

Nous considérons ici que la même valeur de granularité β peut être utilisée dans les dimensions spatiales et dans la dimension temporelle :

$$\pi(\mathbf{z}) = \frac{1}{C(\beta)} \exp \left[\sum_{n=1}^N \sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'}) \right] \quad (4.12)$$

où $\mathcal{V}(n) = \mathcal{V}_s(n) \cup \mathcal{V}_t(n)$

4.3.4 Distribution a posteriori $p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x})$

En considérant le vecteur de paramètres $\boldsymbol{\theta}$ et le vecteur d'étiquettes \mathbf{z} indépendants, en utilisant le théorème de Bayes, la distribution a posteriori du vecteur $(\mathbf{z}, \boldsymbol{\theta})$ peut être exprimée comme :

$$p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x}) \propto p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}) \pi(\boldsymbol{\theta}) \pi(\mathbf{z}) \quad (4.13)$$

où la vraisemblance $p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z})$ et les distributions a priori $\pi(\boldsymbol{\theta})$ et $\pi(\mathbf{z})$ ont été respectivement définies dans (4.8), (4.10) et (4.12).

4.4 Échantillonneur de Gibbs Hybride

Un échantillonneur Metropolis-dans-Gibbs est proposé pour résoudre le problème de segmentation et d'estimation des paramètres. Pour cela, on propose alternativement $\boldsymbol{\theta}$ et \mathbf{z} . Les échantillons sont tirés itérativement selon les densités conditionnelles $p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x})$ et $p(\mathbf{z} | \boldsymbol{\theta}, \mathbf{x})$ de la loi a posteriori jointe $p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{x})$. Les distributions conditionnelles sont présentées dans les prochaines sections.

4.4.1 Distribution conditionnelle $p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x})$

La distribution conditionnelle $p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x})$ peut être développée ainsi :

$$p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x}) \propto p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}) \pi(\boldsymbol{\theta}) \quad (4.14)$$

où $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{z})$ et $\pi(\boldsymbol{\theta})$ ont été définies dans (4.8) et (4.10) respectivement.

Pour générer des échantillons asymptotiquement distribués selon $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$, une marche aléatoire de type RWMH ne suffit pas à cause de l'anisotropie de l'espace des paramètres. En effet, les résultats du chapitre 2 montrent que le sous-espace relatif aux paramètres μ et k est fortement anisotrope, et l'ajout des dimensions relatives aux paramètres m et σ ne fait qu'accentuer le problème. Pour résoudre ce problème, on se propose de générer des vecteurs de paramètre $\boldsymbol{\theta}$ selon la version proximale de la marche aléatoire MALA-sur-variété (prox-mMALA) détaillée dans le chapitre (chap. 2). La loi de proposition est donc formulée :

$$\begin{aligned} \boldsymbol{\theta}_z^* &\sim q(\boldsymbol{\theta}_z^*|\boldsymbol{\theta}_z^{(i-1)}) \\ &\triangleq \mathcal{N}(\text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^{(i-1)} + \frac{\delta}{2}\tilde{\nabla}_{\boldsymbol{\theta}_z}\mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})), \sqrt{\delta}\mathbf{G}^{-1}\{\boldsymbol{\theta}_z^{(i-1)}\}) \end{aligned} \quad (4.15)$$

où $\boldsymbol{\theta}_z^{(i-1)}$ est la valeur actuelle de la chaîne, $\tilde{\nabla}_{\boldsymbol{\theta}_z}\mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})$ est le gradient géodésique de la densité cible au point $\boldsymbol{\theta}_z^{(i-1)}$ et δ est estimé durant la période de chauffe de telle sorte que le taux d'acceptation tende vers $\frac{1}{2}$, comme recommandé dans [117].

Comme les lois de proposition ne sont pas symétriques, le rapport d'acceptation a prend alors la forme :

$$a = \min \left\{ 1, \prod_{n|z_n=z} \frac{p(x_n|\boldsymbol{\theta}_z^*, z_n=z)}{p(x_n|\boldsymbol{\theta}_z^{(i-1)}, z_n=z)} \frac{\pi(\boldsymbol{\theta}_z^*)}{\pi(\boldsymbol{\theta}_z^{(i-1)})} \frac{q(\boldsymbol{\theta}_z^{(i-1)}|\boldsymbol{\theta}_z^*)}{q(\boldsymbol{\theta}_z^*|\boldsymbol{\theta}_z^{(i-1)})} \right\} \quad (4.16)$$

où les vraisemblances $p(x_n|\boldsymbol{\theta}_z^{(\cdot)}, \mathbf{z})$ et les lois a priori $\pi(\boldsymbol{\theta}_z^{(\cdot)})$ ont été définies dans (4.7) et (4.10) respectivement et où les lois de propositions sont définies par :

$$\begin{cases} q(\boldsymbol{\theta}_z^*|\boldsymbol{\theta}_z^{(i-1)}) = p_{\mathcal{N}}(\boldsymbol{\theta}_z^*|\text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^{(i-1)} + \frac{\delta}{2}\tilde{\nabla}_{\boldsymbol{\theta}_z}\mathcal{P}(\boldsymbol{\theta}_z^{(i-1)})), \delta\mathbf{G}^{-1}\{\boldsymbol{\theta}_z^{(i-1)}\}) \\ q(\boldsymbol{\theta}_z^{(i-1)}|\boldsymbol{\theta}_z^*) = p_{\mathcal{N}}(\boldsymbol{\theta}_z^{(i-1)}|\text{prox}_h^{\delta/2}(\boldsymbol{\theta}_z^* + \frac{\delta}{2}\tilde{\nabla}_{\boldsymbol{\theta}_z}\mathcal{P}(\boldsymbol{\theta}_z^*)), \delta\mathbf{G}^{-1}\{\boldsymbol{\theta}_z^*\}) \end{cases} \quad (4.17)$$

4.4.2 Distribution conditionnelle $p(\mathbf{z}|\boldsymbol{\theta}, \mathbf{x})$

L'échantillonnage de \mathbf{z} se fait voxel par voxel. La distribution conditionnelle du label discret z_n de chaque voxel peut être alors entièrement caractérisée par :

$$p(z_n = z|x_n, \boldsymbol{\theta}_z, \mathbf{z}_{-n}) \propto p(x_n|\boldsymbol{\theta}_z, z_n = z)p(z_n|\mathbf{z}_{-n}) \quad (4.18)$$

où \mathbf{z}_{-n} représente le vecteur \mathbf{z} dont le $n^{\text{ième}}$ élément a été supprimé.

En reprenant les équations (4.8) et (4.12), on obtient :

$$\begin{aligned}
p(z_n = z | x_n, \theta_z, \mathbf{z}_{-n}) &\propto \binom{x_n^{\text{TEP}} + k_z - 1}{x_n^{\text{TEP}}} \left(\frac{\mu_z}{\mu_z + k_z} \right)^{x_n^{\text{TEP}}} \left(\frac{k_z}{\mu_z + k_z} \right)^{k_z} \\
&\times \frac{1}{\sigma_z \sqrt{2\pi}} e^{-\frac{(x_n^{\text{TDM}} - m)^2}{\sigma_z^2}} \\
&\times \exp \left[\sum_{n' \in \mathcal{V}(n)} \beta \delta(z_n - z_{n'}) \right]
\end{aligned} \tag{4.19}$$

Comme (4.19) définit un champ aléatoire de Markov, les échantillons sont générés selon cette distribution conditionnelle en tirant aléatoirement des entiers dans l'ensemble $\{1, \dots, Z\}$ avec les probabilités définies dans (4.19).

4.4.3 Résumé de l'algorithme

Le pseudo-code (3) résume l'algorithme MALA sur variété proximal - dans - Gibbs proposé.

Algorithm 3 Échantillonneur MALA sur variété proximal - dans - Gibbs proposé

Initialisation :

- Générer $\boldsymbol{\theta}^{(0)}$ selon (4.10).
- Tirer $z_1^{(0)}, z_2^{(0)}, \dots, z_N^{(0)}$ aléatoirement.

for $t = 1, 2, \dots, T$ **do**

– *Mise à jour de $\boldsymbol{\theta}$* –

for $z = 1, 2, \dots, Z$ **do**

1. Proposer $\boldsymbol{\theta}_z^*$ selon (4.15).
2. Calculer le ratio d'acceptation a (voir (4.16)).
3. Tirer $u \sim \mathcal{U}(0, 1)$.
- if** ($u < a$) **then**
 4. Fixer $\boldsymbol{\theta}_z^{(t)} = \boldsymbol{\theta}_z^*$.
- else**
 5. Fixer $\boldsymbol{\theta}_z^{(t)} = \boldsymbol{\theta}_z^{(t-1)}$.

end if

end for

– *Mise à jour de \mathbf{z}* –

for $n = 1, 2, \dots, N$ **do**

7. Tirer z_n dans $\{1, \dots, Z\}$ selon (4.19).

end for

end for

4.5 Résultats expérimentaux

Des expériences ont été réalisées sur des images TEP-TDM 4-D réelles du thorax, reconstruites selon la méthode décrite dans [102, 35]. Les images originales TEP et TDM ont été acquises par un scanner Discovery ST de General Electric, sur un patient souffrant d'un cancer du poumon. Six images ont été reconstruites dans chaque modalité, correspondant aux différents bins temporels du cycle respiratoire. Les images TDM ont été redimensionnées par interpolation linéaire afin de correspondre aux dimensions des images TEP. Enfin, les images ont été rognées sur des région d'intérêt de taille $(56 \times 91 \times 3)$.

La figure 4.2 suivante montre la coupe centrée sur la tumeur des images originales TEP (Fig.4.2.a) et TDM (Fig.4.2.b) ainsi que les résultats de segmentation sur chacun des canaux TEP (Fig.4.2.c) et TDM (Fig.4.2.d) séparément en utilisant un modèle de mélange univarié.

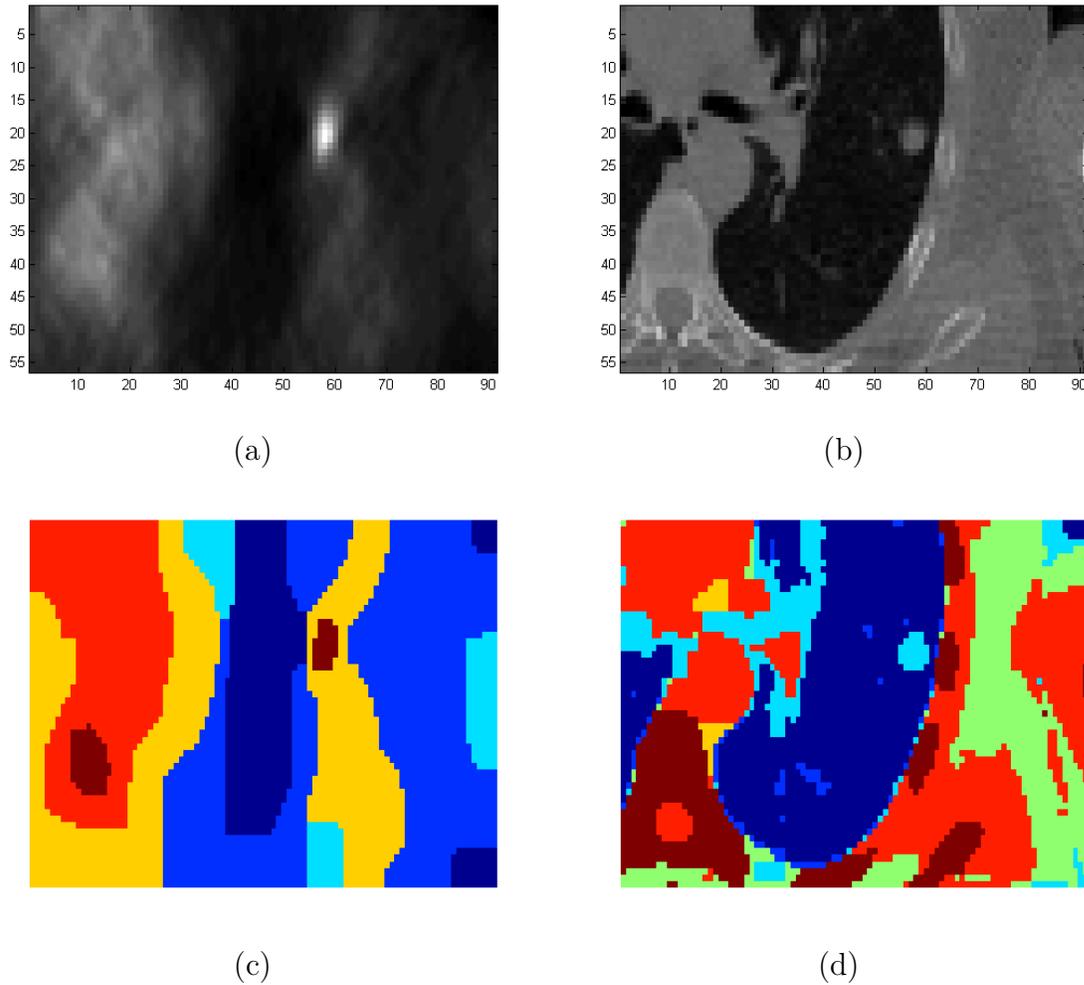


FIGURE 4.2 – (a) Image TEP originale; (b) Image TDM originale; (c) Segmentation sur le canal TEP seul, basée sur un modèle de mélange de distributions binomiales négatives; (d) Segmentation sur le canal TDM seul, basée sur un modèle de mélange de loi normales.

Les résultats TEP montrent des classes présentant des incohérences anatomiques à cause de la faible résolution et du manque d'information anatomique. Dans l'image TDM, des mauvaises classifications sont visibles (des tissus différents associés à la même classe). Par exemple, des voxels du fond et des voxels appartenant à la tumeur sont injustement associés à la même classe.

L'algorithme proposé a été appliqué sur les données en considérant des modèles

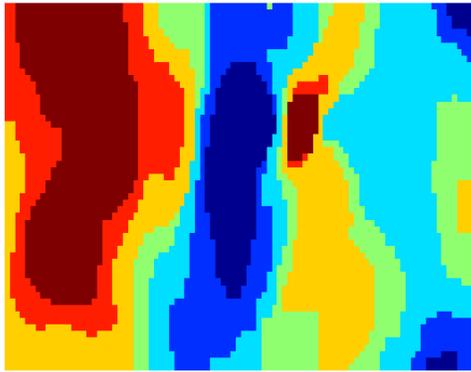
de mélange de différentes distributions statistiques :

- $\mathcal{N} \times \mathcal{N}$: Mélange de distributions normales en supposant les canaux TEP et TDM indépendants.
- $\text{Biv}\mathcal{N}$: Mélange de distributions normales bivariées sans hypothèse d'indépendance.
- $\mathcal{P} \times \mathcal{P}$: Mélange de distributions Poisson en supposant les canaux TEP et TDM indépendants.
- $\text{Biv}\mathcal{P}$: Mélange de distributions Poisson bivariées sans hypothèse d'indépendance [68] (voir annexe C).
- $\mathcal{BN} \times \mathcal{N}$: Mélange de distributions binomiale négative - normale proposé dans ce chapitre.

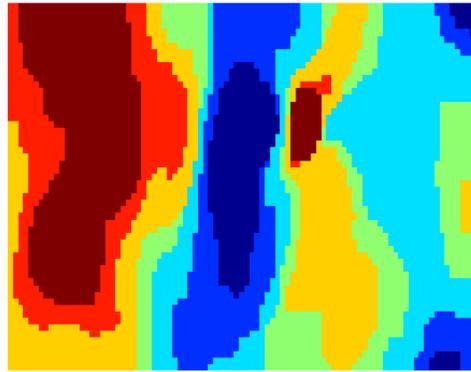
La valeur du MAP pour chaque paramètre a été fixée comme la moyenne sur les 2000 échantillons (après 5000 itérations de chauffe) générés par la chaîne de Monte-Carlo, pour $Z = 7$ classes. Afin de limiter le risque d'enfermer les chaînes dans un mode local, la valeur de granularité β du champ de Potts a été initialisée à 0 au début de l'échantillonnage et augmente progressivement jusqu'à atteindre sa valeur fixe finale à la fin de l'étape de chauffe. La granularité β a été fixée à 1.3 dans les expériences présentées dans ce chapitre.

4.5.1 Résultats de la la segmentation

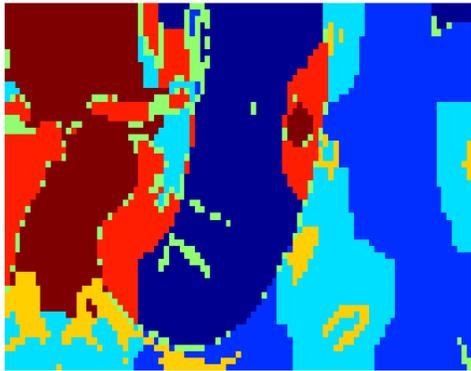
Les figures (4.3.a) à (4.3.e) présentent les résultats de segmentation obtenus par les cinq modèles de mélange.



(a) $\mathcal{N} \times \mathcal{N}$



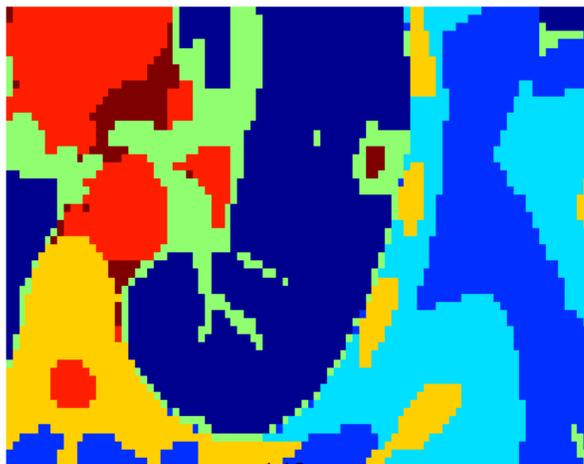
(b) Biv \mathcal{N}



(c) $\mathcal{P} \times \mathcal{P}$



(d) Biv \mathcal{P}



(e) $\mathcal{BN} \times \mathcal{N}$

FIGURE 4.3 – Résultats de la segmentation pour les différents modèles de mélange testés.

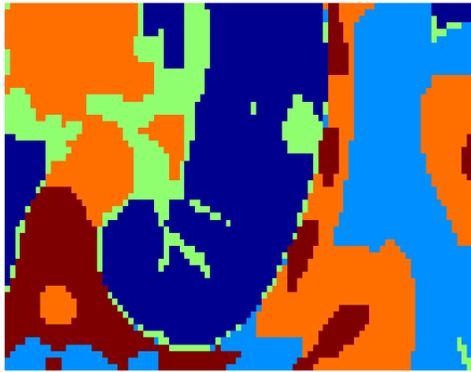
Il est visuellement évident que les résultats de segmentation des distributions $(\mathcal{BN} \times \mathcal{N})$, $(\mathcal{P} \times \mathcal{P})$ et $(\text{Biv}\mathcal{P})$ sont meilleurs que ceux des distributions normales. Les formes des classes sont en meilleure conformité avec les structures anatomiques. On peut remarquer que dans ces trois cas, les voxels correspondant aux os (voir les zones blanches sur les images TDM originales) sont associés à la même classe dans le résultat de la segmentation (Fig.4.3. c-e). Il est aussi intéressant de constater que les régions montrant des activités fonctionnelles différentes (voir l'image originale PET) ont des étiquettes distinctes. Ceci indique l'avantage de la fusion bimodale TEP-TDM lors de la segmentation.

De plus, les modèles bivarié et univarié pour chaque canal donnent des résultats très similaires (Fig.4.3. a-b, Fig.4.3. c-d). Cela indique que le niveau de dépendance entre les données TEP et TDM est négligeable.

Dans le modèle $(\mathcal{BN} \times \mathcal{N})$, les structures anatomiques sont mieux respectées que pour les autres modèles. De plus, une classe propre est affectée à la tumeur à l'exception de quelques voxels dont l'activité biologique est également très élevée.

4.5.2 Influence du nombre de classes

Les figures (4.4.a) à (4.4.e) présentent les résultats de segmentation obtenus en fonction du nombre de classes Z recherchées.



(a) $Z = 5$



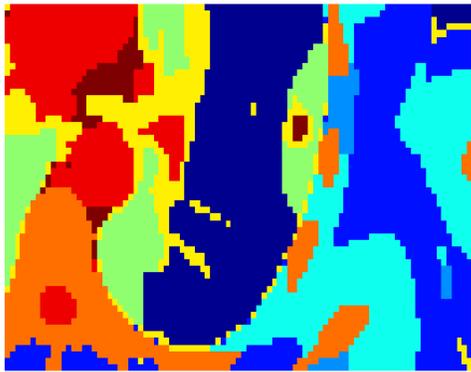
(b) $Z = 6$



(c) $Z = 7$



(d) $Z = 8$



(e) $Z = 9$

FIGURE 4.4 – Influence du nombre de classes Z recherchées

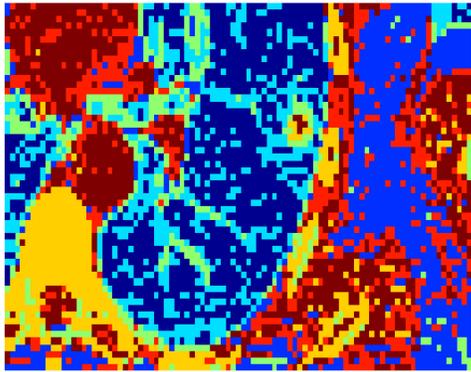
On remarque que les structures sont globalement les mêmes. En effet, l'augmentation du nombre de classes recherchées semble scinder les classes trouvées pour des Z plus petits, plutôt que de ré-organiser les structures. En outre, on remarque que les résultats pour $Z = 8$ et $Z = 9$ sont très similaires, ce qui montre qu'à partir de $Z > 8$, le nombre de classes recherchées a peu d'influence sur le résultat de la segmentation.

La nouvelle classe créée entre $Z = 7$ et $Z = 8$ à l'intérieur du poumon semble provenir de l'important effet de volume partiel de l'image TEP (voir fig.4.2.a).

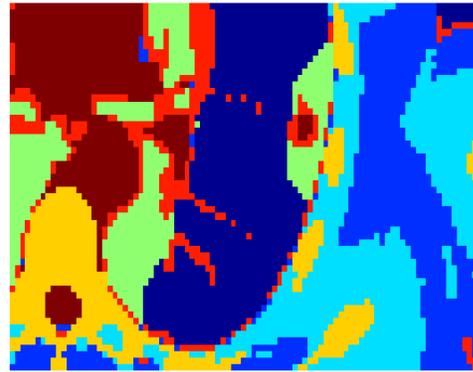
Enfin, il est important de noter qu'à partir de $Z = 5$, la délimitation de la tumeur reste constante, ce qui témoigne de la robustesse de la méthode.

4.5.3 Influence de la granularité

Les figures (4.5.a) à (4.5.d) rapportent les résultats de segmentation obtenus en fonction du coefficient de granularité β .



(a) $\beta = 0$



(b) $\beta = 0.8$



(c) $\beta = 1.0$



(d) $\beta = 1.3$

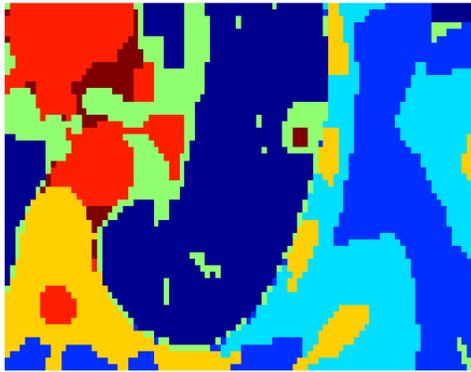
FIGURE 4.5 – Influence de la granularité β .

On constate que la granularité β joue un rôle majeur dans le processus de segmentation. Comme anticipé, une valeur de β plus élevée donne des régions plus connexes. Lorsque $\beta = 0$, c'est à dire lorsqu'aucun a priori spatio-temporel n'est utilisé, la pixelisation de l'image rend impraticable toute exploitation.

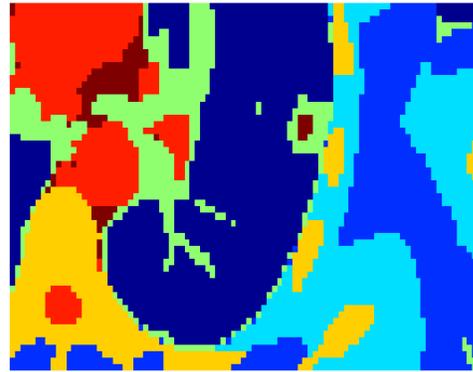
On remarque enfin que dans tous les cas (hormis le cas $\beta = 0$), la tumeur a été isolée et son volume reste identique, ce qui montre qu'un mauvais choix de β n'a pas d'incidence importante sur la délimitation de la tumeur.

4.5.4 Suivi

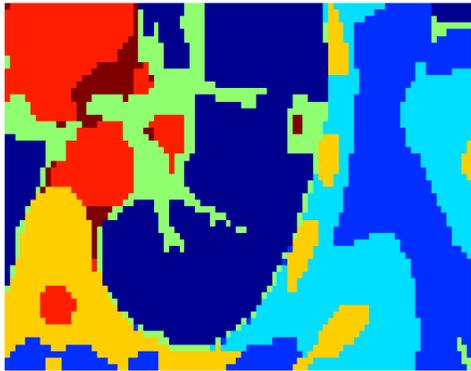
Finalement, la figure 4.6 présente les résultats de la segmentation des 6 bins temporels, correspondant aux différents stades du mouvement respiratoire.



(a) Bin 1



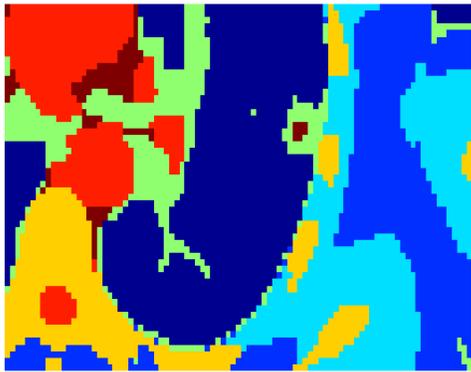
(b) Bin 2



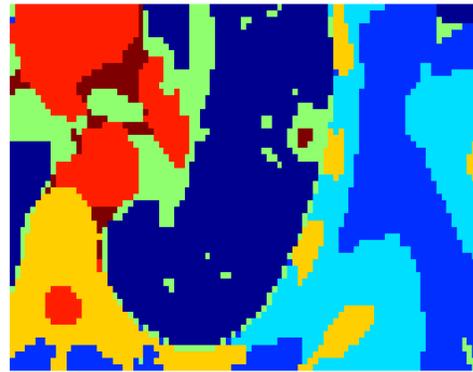
(c) Bin 3



(d) Bin 4



(e) Bin 5



(f) Bin 6

FIGURE 4.6 – Segmentation des 6 bins temporels, correspondant aux différents stades du mouvement respiratoire.

Le processus d'étiquetage offre une bonne régularité entre les coupes temporelles.

Les figures 4.7, 4.8 et 4.9 présentent, sur le plan coronal, des jeux d'images associés à trois patients différents. Le patient A correspond au jeu de données présenté dans les résultats précédents. Les deux premières lignes correspondent aux images TEP, et les deux dernières lignes aux images TDM correspondantes. Chacune des images présente les 6 bins temporels de l'image originale et les résultats de segmentation estimés par la méthode proposée (en rouge), par une technique de seuillage adaptatif tel qu'utilisée habituellement en clinique (en vert) ainsi que par une délimitation manuelle établie par un expert (en bleu).

La valeur du seuil utilisée dans la technique de seuillage adaptatif a été fixée en utilisant la méthode de Daisne [33, 71], selon la courbe de calibration (4.20) ci-dessous :

$$s_{\text{adapt}} = 95.6 * \exp(-SB/2.4) + 32.1 \quad (4.20)$$

où SB est le rapport du signal maximum dans la tumeur sur le signal moyen dans l'environnement et où les paramètres mis en jeu ont été estimés à partir d'expériences sur fantômes, en utilisant la méthode itérative de Jentzen [72].

On remarque dans un premier temps que toutes les tumeurs ont été détectées, qu'elles soient de petite taille (patient A), de taille moyenne (patient B) ou de grande taille (patient C). Il est particulièrement intéressant de remarquer que le processus de segmentation ne privilégie pas une modalité par rapport à l'autre. En effet, on constate que le résultat de segmentation obtenu sur le patient B est plus proche de l'image TDM que de l'image TEP originale (voir également fig.4.11). En revanche, la tumeur du patient C n'est pas visible en TDM, et la délimitation s'inspire plus des données TEP. Cela montre l'intérêt de traiter les deux canaux de l'image bi-modale de façon symétrique, plutôt que d'effectuer la segmentation dans une modalité, quitte à utiliser l'autre modalité comme a priori.

Le tableau 4.1 présente les coefficients DICE obtenus en comparant à la délimitation manuelle la segmentation proposée ainsi que la segmentation obtenue par différents seuillages. Les valeurs des seuils adaptatifs sont présentés dans le tableau 4.2. Les résultats sont calculés pour chaque bin temporel, ainsi que sur l'image 4-D entière.

Jeu de données		Seuil 30%	Seuil 35%	Seuil 40%	Seuil adaptatif	Méthode proposée
Patient A	bin 1	0.6479	0.7188	0.8000	0.7692	0.7241
	bin 2	0.7619	0.8421	0.9275	0.9063	0.7532
	bin 3	0.6667	0.7385	0.7586	0.8400	0.7419
	bin 4	0.7436	0.8116	0.8308	0.8333	0.8788
	bin 5	0.7105	0.7941	0.8710	0.8667	0.7595
	bin 6	0.7407	0.8000	0.8406	0.8571	0.7941
	Total	0.7140	0.7866	0.8413	0.8481	0.7756
Patient B	bin 1	0.7629	0.7609	0.7381	0.7789	0.6452
	bin 2	0.7395	0.6903	0.6481	0.7069	0.7154
	bin 3	0.7451	0.7200	0.6737	0.7327	0.7547
	bin 4	0.7460	0.6833	0.6496	0.7154	0.7143
	bin 5	0.7885	0.7800	0.7391	0.7647	0.6602
	bin 6	0.8381	0.8200	0.7473	0.8462	0.6337
	Total	0.7688	0.7392	0.6951	0.7551	0.6902
Patient C	bin 1	0.8524	0.7180	0.6305	0.7600	0.8579
	bin 2	0.8547	0.7770	0.6489	0.8014	0.8448
	bin 3	0.8101	0.7648	0.6937	0.7914	0.8348
	bin 4	0.8596	0.7680	0.6581	0.7977	0.8629
	bin 5	0.8309	0.7302	0.6009	0.7638	0.8522
	bin 6	0.8145	0.7164	0.6438	0.7541	0.8378
	Total	0.8378	0.7462	0.6458	0.7785	0.8486

TABLE 4.1 – Coefficients DICE obtenus en comparant la segmentation manuelle à la segmentation estimée par notre méthode, ainsi qu’à la segmentation obtenue par différents seuillages

On note que les résultats de segmentation obtenus sur des grosses tumeurs (patient C) sont plus proches d’une segmentation manuelle qu’en utilisant une méthode de seuillage. Lorsque le volume des tumeurs est plus petit (patients A et B), la délimitation manuelle est très grossière et les résultats sont moins significatifs. On remarque cependant que pour certaines tumeurs, une segmentation plus proche d’une segmentation manuelle est obtenue avec un seuil à 40% (patient A), alors que d’autres sont obtenues avec un seuil à 30% (Patients B et C). La méthode de segmentation proposée permet d’obtenir des résultats convenables quelque soit le type de tumeur.

Le tableau 4.2 présente pour chaque jeu de données le seuil nécessaire afin que

tous les pixels obtenus par la méthode proposée soient considérés si une technique de seuillage est utilisée. A titre de comparaison, les seuils utilisés en clinique selon une technique de seuillage adaptatif sont également présentés.

Jeu	Seuil nécessaire	Seuil adaptatif estimé
Patient A	18.9097%	43.2581%
Patient B	11.7672%	32.1253%
Patient C	12.2277%	33.4136%

TABLE 4.2 – Seuil nécessaire pour considérer tous les voxels délimités par la méthode proposée et seuil adaptatif mesuré, et seuil adaptatif estimé.

Nous constatons que les seuils mesurés sont bien inférieurs aux seuils habituellement utilisés en clinique. L'utilisation d'une technique de seuillage dont la valeur du seuil est aussi basse n'est pas envisageable car beaucoup de faux-positifs seraient considérés et le volume des tumeurs serait lourdement sur-estimé. Cela témoigne de la souplesse de l'algorithme proposé par rapport aux méthodes de seuillage.

Les figures 4.10, 4.11 et 4.12 suivantes présentent les résultats de segmentation précédents centrés sur la tumeur. La première ligne correspond aux données TEP et la seconde ligne aux données TDM. Les 6 colonnes correspondent aux 6 bins temporels.

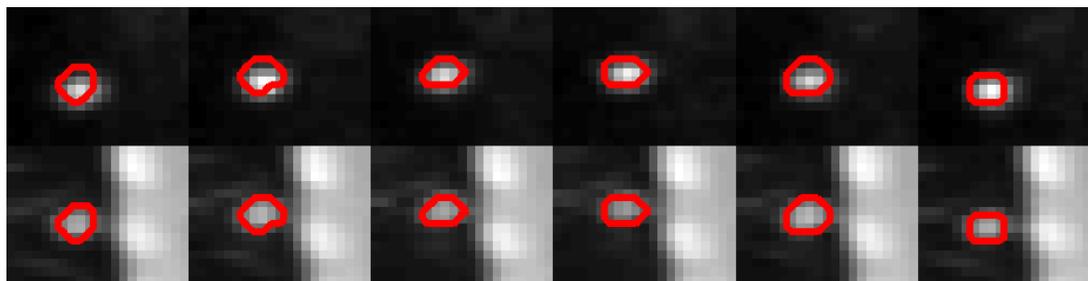


FIGURE 4.10 – Suivi temporel : Patient A

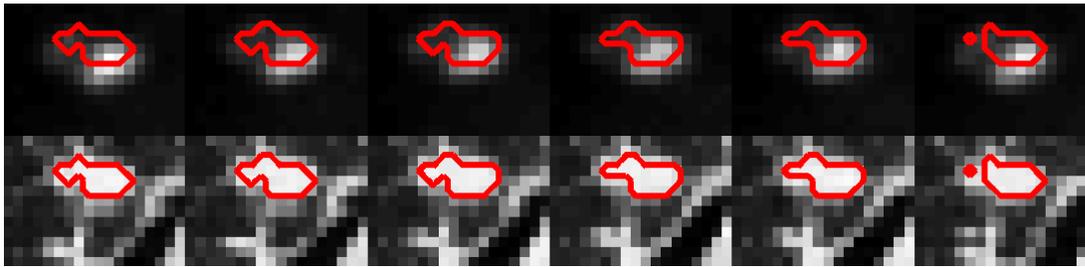


FIGURE 4.11 – Suivi temporel : Patient B

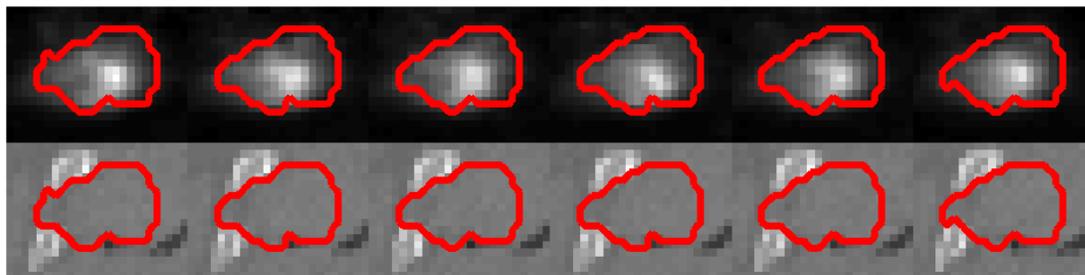


FIGURE 4.12 – Suivi temporel : Patient C

Nous remarquons que la régularité temporelle permet bien de suivre la tumeur d'un bin à l'autre. Le déplacement axial dû aux mouvement respiratoires est particulièrement visible sur des tumeurs de petite taille (patient A : la tumeur monte puis redescend de quelques pixels). Le déplacement de la tumeur du patient B n'est pas significatif mais la tumeur semble déformée par les mouvement respiratoires. La tumeur du patient C semble ne pas être affectée par le mouvement respiratoire. La possibilité d'obtenir une délimitation spatio-temporelle de la tumeur permet de mieux prendre en compte son déplacement et ses déformations au cours des différentes phases respiratoires, et peut être très intéressante en radiothérapie pour irradier à chaque instant le volume cible tout en épargnant les tissus sains voisins.

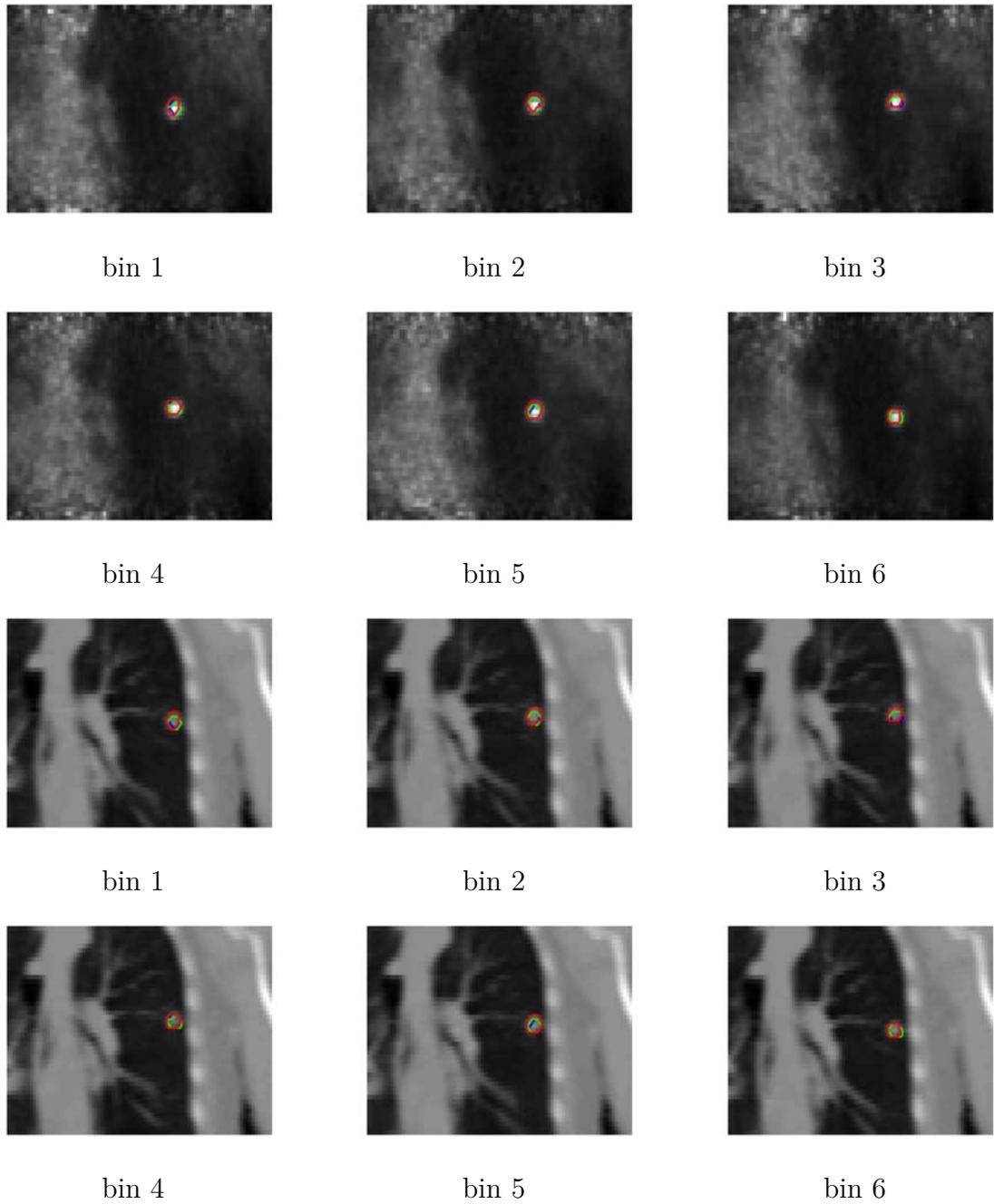


FIGURE 4.7 – Patient A : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée; vert : segmentation par seuillage adaptatif; bleu : segmentation manuelle.

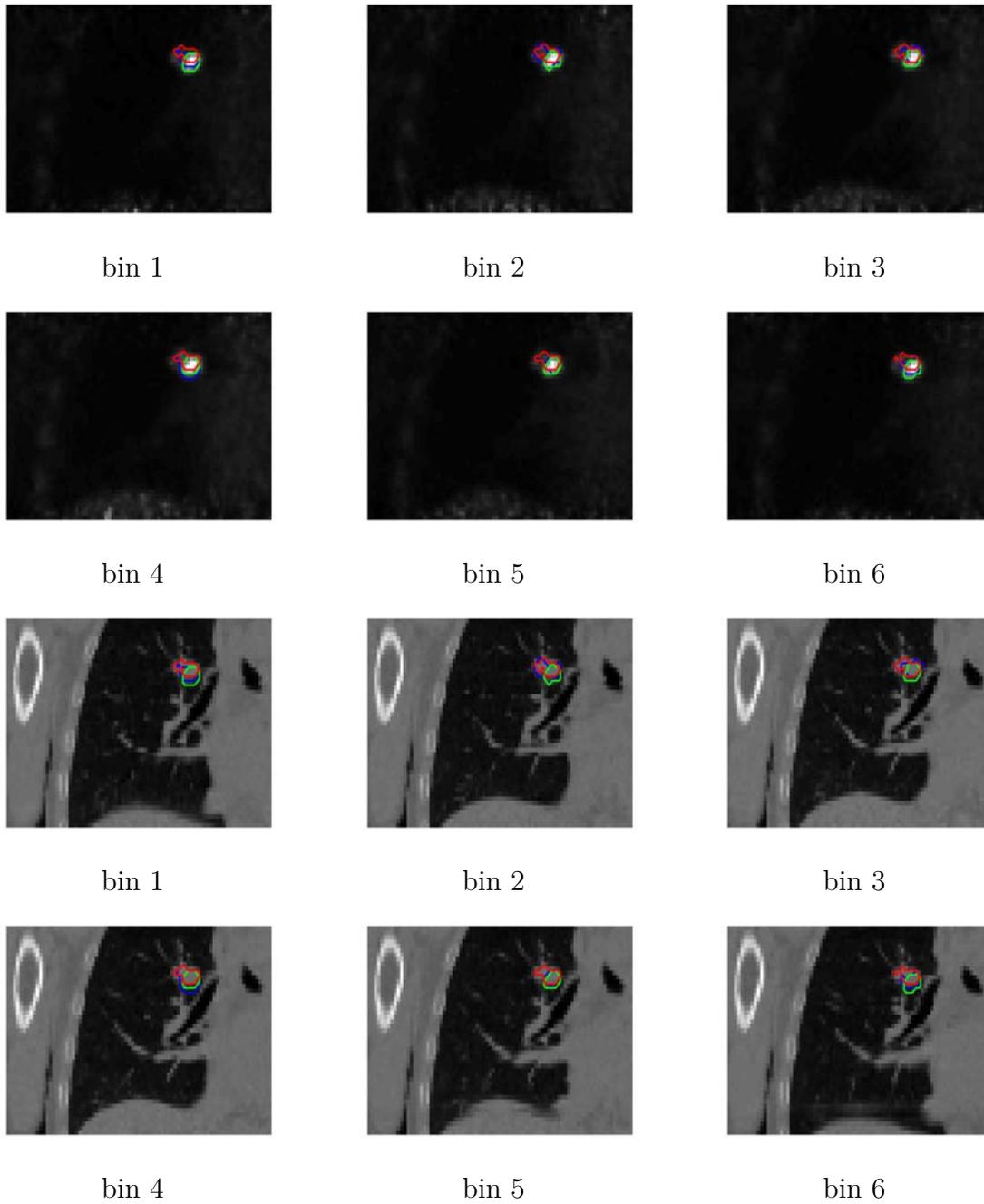


FIGURE 4.8 – Patient B : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée ; vert : segmentation par seuillage adaptatif ; bleu : segmentation manuelle.

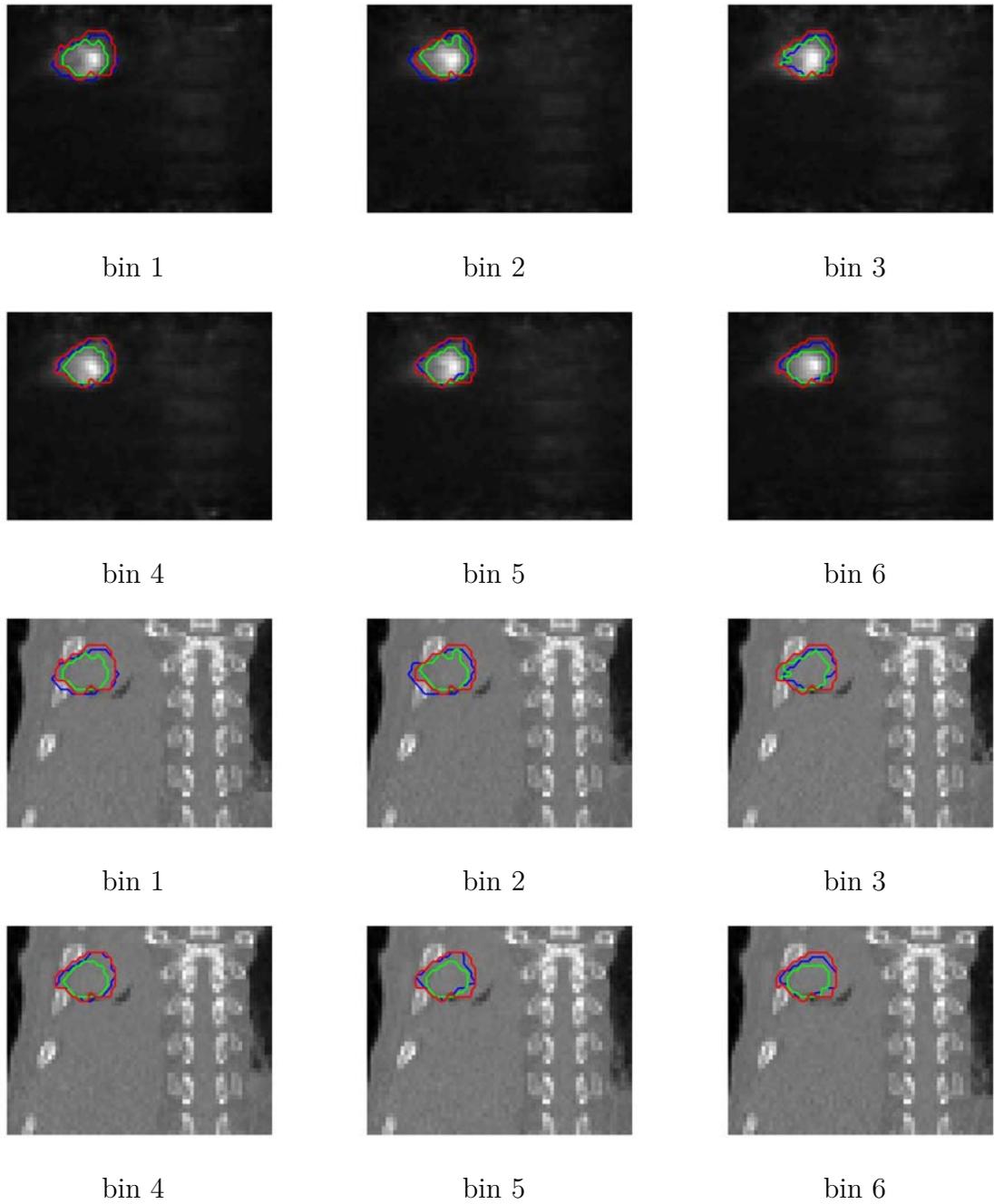


FIGURE 4.9 – Patient C : Résultats de segmentation sur le plan coronal en TEP et en TDM pour chaque bin temporel. Rouge : méthode proposée; vert : segmentation par seuillage adaptatif; bleu : segmentation manuelle.

Conclusion

This thesis presented methods of image processing developed to provide an automatic segmentation method to clinicians to better delineate and assess tumors in PET images. Specifically, Bayesian methods and MCMC algorithms were used to segment the tissues. Three main methodological contributions have been proposed in Chapters 2, 3 and 4.

Chapter 2 studied the statistical distribution of a single tissue in PET images. In the accordance with the image formation model, a negative binomial distribution was proposed to consider the heterogeneity of the activity within a single tissue. A robust Bayesian method based on an original sampling technique is presented to estimate the parameters of this distribution. First, a manifold-MALA system has been proposed to remedy the problems of anisotropy, allowing the sampler to draw less correlated samples. Moreover, the use of proximity operators offers the possibility of addressing the positivity constraint effectively. The estimation algorithm has been validated on synthetic data and provides accurate and robust results when the conventional method of Metropolis-within-Gibbs fails. Experimental results of goodness of fit on real data also confirm the choice of the proposed model.

Chapter 3 addressed the problem of multi-tissue 3D segmentation in PET. The segmentation performed by this algorithm discriminates tissues with different statistics. To take into account the heterogeneity of the tissues, a finite mixture of spatially coherent negative binomial distributions was proposed to model the voxels' activity. To cope with the strong correlation between the process of segmentation and the model parameters estimation, this chapter has presented a way to solve both tasks jointly. A hierarchical Bayesian model was used in which the spatial correlation is modeled by a Potts-Markov random field. A Markov Chain Monte Carlo (MCMC) mechanism was presented to jointly estimate the parameters of the mixture and the voxels' label. The effectiveness of the algorithm has been illustrated on simulated data and on PET images obtained under real clinical conditions. This illustrates the flexibility of the proposed algorithm when it comes to segment tissues whose biological activity is heterogeneous. The proposed technique correctly segmented lesions,

as well as other organs, and outperforms segmentation results obtained using other mixture models.

This thesis also proposed to integrate anatomical information in the segmentation process to obtain a more precise delimitation. Images of computed tomography (CT) that are acquired during the same examination are considered. Chapter 4 proposes a mixture model of bivariate negative binomial - normal distributions to represent the fused PET and CT images. In addition to the proposed model, this chapter presents an unsupervised technique for spatiotemporal segmentation of dynamical bimodal images. A hierarchical Bayesian model was developed including a four dimensional Potts-Markov field to deal with the spatial and temporal coherence of a time series of PET-CT images. An MCMC algorithm based on a Metropolis-within-Gibbs sampler has been developed to jointly estimate the parameters of the Bayesian model while segmenting the fused image. The method was tested on real 4-D PET-CT images from a patient having a lung tumor. The obtained segmentation results are visually consistent with the anatomical structures and enable the tracking and delineation of the tumor.

The work presented in this thesis leads to several perspectives.

First, the segmentation process described in Chapters 3 and 4 depend on the hyper-parameter β , which controls the amount of spatial correlation introduced by the Potts model. This parameter plays an important role and was chosen by cross validation. The parameter β of the Potts field could be integrated into the Bayesian scheme to be estimated jointly with the unknown parameters of the model, as proposed in [105].

PET imaging suffers from an important partial volume effect that significantly disrupts the segmentation process. It is necessary to correct this artifact before further processing of the image. However, because of the complex geometry of the acquisition system, the point spread function (PSF) varies spatially and are only known in some points, which makes the traditional deconvolution algorithms unusable. We proposed in [67] a method for estimating a variable PSF at any location in space by an original interpolation technique. This interpolation is formulated as a problem of optimal transport and is based on the geometric variability of the PSF and not on its photometric variability. The estimation results of the PSF at any point by the proposed method are promising and suggest the possibility of developing a complete partial volume correction technique in PET and other imaging modality.

In chapter 4, we assumed that the dependence between PET and CT channels is negligible, while some research works take it into account [54]. Statistical modeling of bimodal PET-CT images by distributions taking into account the correlation

should be investigated in greater detail. To consider the heterogeneity of the biological activity of the tissues and to remain faithful to the modeling of PET images by a negative binomial distribution, the bivariate negative binomial distribution model [86, 131] can be investigated.

On a medical level, segmentation methods proposed in Chapters 3 and 4 have not been tested extensively. Future work should measure their performance in a clinical context by applying them to a large number of images and by comparing the resulting segmentations with annotations by several experts.

The parameters μ and k of the negative binomial distribution provide a physical interpretation to characterize different tissues. Indeed, μ is the mean activity, and $\alpha \triangleq 1/k$ can be interpreted as the distance of the data distribution to the Poisson distribution and reflects the over-dispersion of the data. In addition to performing a tissue segmentation, the proposed algorithms estimate parameters that can be used for tissue characterization. Many recent studies focus particularly on indicators of the heterogeneity of the activity of biological tissues [14, 134, 15, 24]. The parameter $\alpha \triangleq 1/k$ has the advantage of being based on the data statistics, and not directly on the measured activity. It could be interesting to study this indicator in clinical setting.

Finally, it should be noted that the techniques presented in this thesis can be generalized to other modalities, by incorporating the appropriate statistical models.

Conclusion

Cette thèse a présenté des méthodes de traitement d'image développées pour proposer aux cliniciens une méthode automatique de segmentation d'image TEP afin de mieux délimiter les lésions et de mieux les quantifier. Plus précisément, des méthodes Bayésiennes et des algorithmes MCMC ont été utilisés pour segmenter les tissus. Trois contributions méthodologiques principales ont été proposées dans les chapitres 2, 3 et 4 .

Le chapitre 2 a étudié la distribution statistique d'une région mono-tissulaires dans les images TEP. À partir du modèle de formation de l'image, un modèle de distribution binomiale négative a été proposé, permettant de considérer l'hétérogénéité dans l'activité d'un même tissu. Une méthode Bayésienne robuste reposant sur une technique originale d'échantillonnage est présentée afin d'estimer les paramètres de cette loi. D'une part, un système MALA-sur-variété a été proposé pour atténuer les problèmes d'anisotropie, permettant à l'échantillonneur de tirer des échantillons moins corrélés. D'autre part, l'utilisation du cadre proximal offre la possibilité de traiter le problème de la contrainte de positivité d'une manière efficace. L'algorithme d'estimation a été validé sur des données synthétiques et offre des résultats précis et robustes lorsque la méthode classique de Metropolis-dans-Gibbs échoue. Des résultats expérimentaux des tests d'ajustement effectués sur des données réelles confirment également le choix du modèle proposé.

Le chapitre 3 a abordé le problème de la segmentation multi-tissus d'images TEP en 3D. La segmentation réalisée par cet algorithme discrimine les tissus présentant des statistiques différentes. Pour prendre en considération l'hétérogénéité des tissus, un mélange fini et spatialement cohérent de distributions binomiales négatives a été proposé pour modéliser l'activité des voxels. Pour faire face à la forte corrélation existant entre le processus de segmentation et l'estimation des paramètres du modèle, ce chapitre a présenté un moyen de résoudre ces deux tâches de manière conjointe. Un modèle Bayésien a été mis en oeuvre dans lequel la corrélation spatiale est modélisée par un champ aléatoire de Potts-Markov. Un mécanisme de Monte-Carlo par Chaîne de Markov (MCMC) a été présenté pour estimer conjointement les paramètres du

mélange et les labels des voxels. L'efficacité de l'algorithme a été illustrée sur des données simulées et sur des images TEP réalisées dans des conditions cliniques réelles. Cela illustre la flexibilité de l'algorithme proposé quand il s'agit de segmenter des tissus dont l'activité biologique est hétérogène. La technique proposée segmente correctement les lésions, voire les autres organes, et surpasse les résultats de segmentation obtenus en utilisant d'autres modèles de mélange.

Cette thèse a également proposé d'intégrer des informations anatomiques dans le processus de segmentation afin d'obtenir une délimitation plus précise. On considère pour cela les images de tomodensitométrie (TDM ou CT) acquises lors du même examen. Le chapitre 4 propose un modèle de mélange de distributions bivariées binomiale négative - normale pour représenter les images TEP et TDM fusionnées. En plus du modèle proposé, ce chapitre présente une technique non-supervisée pour la segmentation spatio-temporelle d'images bimodales dynamiques. Un modèle Bayésien hiérarchique a été élaboré comprenant un champ de Potts-Markov à quatre dimensions pour respecter la cohérence spatiale et temporelle d'une série temporelle d'images PET-TDM. Un algorithme MCMC reposant sur un échantillonneur Metropolis-dans-Gibbs a été développé pour estimer conjointement les paramètres du modèle Bayésien tout en segmentant l'image fusionnée. La méthode a été appliquée à des images en quatre dimensions réelles TEP-TDM d'un patient ayant une tumeur au poumon. Les résultats de segmentation obtenus sont visuellement en concordance avec les structures anatomiques et permettent la délimitation et le tracking de la tumeur.

Le travail présenté dans cette thèse ouvre plusieurs perspectives.

En premier lieu, le processus de segmentation décrit dans les chapitres 3 et 4 dépendent de l'hyper-paramètre β , qui contrôle la quantité de corrélation spatiale introduite par le modèle de Potts. Ce paramètre joue un rôle important et a été choisi dans ces travaux de manière heuristique par validation croisée. Le paramètre β de champ de Potts pourrait être intégrée au schéma Bayésien afin de l'estimer conjointement avec les paramètres inconnus du modèle, comme proposé dans [105].

L'imagerie TEP est victime d'un effet de volume partiel important qui perturbe significativement le processus de segmentation. Il est nécessaire d'effectuer une correction de cet artefact avant tout autre traitement de l'image. Cependant, à cause de la géométrie complexe du système d'acquisition, la fonction d'étalement du point (PSF) varie spatialement et n'est connue qu'en certains points, ce qui rend les algorithmes classiques de déconvolution inutilisables. Nous avons proposé dans [67] une méthode d'estimation d'une PSF variable en tout point de l'espace selon une technique d'interpolation originale. Cette interpolation est formulée comme un problème

de transport optimal et repose sur la variabilité géométrique de la PSF et non sur sa variabilité photométrique. Les résultats d'estimation de la PSF en tout point par la méthode proposée semblent prometteurs et suggèrent la possibilité de développer une technique de correction de volume partiel complète en TEP et pour d'autres modalités d'imagerie.

Dans le chapitre 4, nous avons supposé négligeable la dépendance entre les canaux TEP et TDM, alors que certains travaux la prennent en compte [54]. La modélisation statistique des images bimodales TEP-TDM par des distributions statistiques prenant en considération cette corrélation devrait être étudiée plus amplement. Pour considérer l'hétérogénéité de l'activité biologique des tissus et pour rester fidèle à la modélisation des images TEP par une distribution binomiale négative, le modèle de distribution binomial négative bivarié [86, 131] peut s'avérer intéressant.

Sur un plan plus médical, les méthodes de segmentation proposées dans les chapitres 3 et 4 n'ont pas été testées extensivement. Des travaux futurs devraient mesurer leurs performances dans un contexte clinique en les appliquant à un grand nombre d'images et en comparant les segmentations résultantes avec les annotations de plusieurs experts.

Les paramètres μ et k de la distribution binomiale négative offrent une interprétation physique pour caractériser les différents tissus. En effet, μ correspond à l'activité moyenne, et $\alpha \triangleq 1/k$ peut être interprétée comme la distance de la distribution des données à la distribution de Poisson et témoigne de la sur-dispersion des données. En plus d'effectuer une segmentation des tissus, les algorithmes présentés dans cette thèse estiment des paramètres pouvant être proposés aux cliniciens pour caractériser certains cancers. De nombreux travaux récents s'intéressent en particulier à un indicateur de l'hétérogénéité de l'activité biologique d'un tissu [14, 134, 15, 24]. Le paramètre $\alpha \triangleq 1/k$ a l'avantage de reposer sur les statistiques des données, et non sur l'activité directement mesurée. Il semble intéressant d'étudier cet indicateur en clinique.

Enfin, il est à noter que les techniques présentées dans cette thèse peuvent être généralisées à d'autres modalités en y intégrant les modèles statistiques adaptés.

Annexe A

Détails des calculs du MALA sur variété pour la loi binomiale négative

La probabilité qu'une observation générée par une loi binomiale négative de paramètre $\boldsymbol{\theta} = (\mu, k)$ obtienne la valeur x_i est définie par la fonction de densité de probabilité :

$$P[X = x_i | \mu, k] = \binom{x_i + k - 1}{x_i} \left(\frac{\mu}{\mu + k}\right)^{x_i} \left(\frac{k}{\mu + k}\right)^k \quad (\text{A.1})$$

En supposant que les observations sont mutuellement indépendantes, la vraisemblance $p(\mathbf{x} | \boldsymbol{\theta})$ peut donc être exprimée comme le produit des probabilités relatives à chacune des observations :

$$\begin{aligned} P[\mathbf{x} | \mu, k] &= \prod_{i=1}^N P[x_i | \mu, k] \\ &= \prod_{i=1}^N \binom{x_i + k - 1}{x_i} \left(\frac{\mu}{\mu + k}\right)^{x_i} \left(\frac{k}{\mu + k}\right)^k \\ &= \prod_{i=1}^N \left[\frac{\Gamma(x_i + k)}{\Gamma(x_i + 1)\Gamma(k)} \mu^{x_i} k^k (\mu + k)^{-(k+x_i)} \right] \end{aligned} \quad (\text{A.2})$$

où Γ est la fonction gamma.

La log-vraisemblance $\mathcal{L}(\boldsymbol{\theta})$ prend alors la forme :

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &\triangleq \log\{P[\mathbf{x}|\boldsymbol{\theta}]\} \\ &= \sum_{i=1}^N [\log\{\Gamma(x_i + k)\} - \log\{\Gamma(x_i + 1)\} - \log\{\Gamma(k)\} \\ &\quad + x_i \log(\mu) + k \log(k) \\ &\quad - (k + x_i) \log(\mu + k)]\end{aligned}\tag{A.3}$$

A.1 MALA

MALA nécessite l'expression du gradient de la fonction log-vraisemblance. Nous calculons pour cela les dérivées partielles :

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{i=1}^N \left[\frac{x_i}{\mu} - \frac{x_i + k}{\mu + k} \right]\tag{A.4}$$

$$\frac{\partial \mathcal{L}}{\partial k} = \sum_{i=1}^N \left[\Psi(x_i + k) - \Psi(k) + 1 + \log(k) - \log(\mu + k) - \frac{x_i + k}{\mu + k} \right]\tag{A.5}$$

où Ψ est la fonction psi, ou digamma ($\Psi(x) \triangleq \frac{\partial}{\partial x} \log(\Gamma(x))$).

Le gradient de la log-vraisemblance $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ peut donc s'exprimer :

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mu} \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial k} \end{bmatrix} = \begin{bmatrix} \sum_i \left[\frac{x_i}{\mu} - \frac{x_i + k}{\mu + k} \right] \\ \sum_i \left[\Psi(x_i + k) - \Psi(k) + 1 - \frac{x_i + k}{\mu + k} - \log\left(1 + \frac{\mu}{k}\right) \right] \end{bmatrix}\tag{A.6}$$

Comme une démarche Bayésienne est utilisée, il est également nécessaire de considérer les informations a priori. D'après la loi de Bayes, la loi log a posteriori $\mathcal{P}(\boldsymbol{\theta}) \triangleq \log\{p(\boldsymbol{\theta}|\mathbf{x})\}$ est définie par :

$$\mathcal{P}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \log\{\pi(\boldsymbol{\theta})\}\tag{A.7}$$

et par linéarité du gradient nous obtenons :

$$\nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) + \begin{bmatrix} \frac{\partial \pi(\boldsymbol{\theta})}{\partial \mu} \\ \frac{\partial \pi(\boldsymbol{\theta})}{\partial k} \end{bmatrix}\tag{A.8}$$

$$= \begin{bmatrix} \sum_i \left[\frac{x_i}{\mu} - \frac{x_i + k}{\mu + k} \right] + \frac{a_\mu}{\mu} + b_\mu \\ \sum_i \left[\Psi(x_i + k) - \Psi(k) + 1 - \frac{x_i + k}{\mu + k} - \log\left(1 + \frac{\mu}{k}\right) + \frac{a_k}{k} + b_k \right] \end{bmatrix}\tag{A.9}$$

A.2 MALA sur variété

Nous considérons maintenant les dérivées partielles secondes de la log-vraisemblance nécessaires au calcul de la matrice de Fisher :

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} = \sum_i \left[-\frac{x_i}{\mu^2} + \frac{x_i + k}{(\mu + k)^2} \right] \quad (\text{A.10})$$

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} = \sum_i \left[\Psi^2(x_i + k) - \Psi^2(k) + \frac{x_i - \mu}{(\mu + k)^2} + \frac{\mu}{k(\mu + k)} \right] \quad (\text{A.11})$$

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} = \sum_i \left[\frac{x_i - \mu}{(\mu + k)^2} \right] \quad (\text{A.12})$$

où Ψ^2 est la fonction trigamma $\Psi^2(x) \triangleq \frac{\partial^2}{\partial x^2} \log(\Gamma(x))$.

Les valeurs de la matrice de Fisher sont alors les opposées des espérances de ces dérivées partielles secondes :

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} \right] = N \left[\frac{1}{\mu} - \frac{1}{\mu + k} \right] \quad (\text{A.13})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} \right] = N \left[\Psi^2(k) - \frac{\mu}{k(\mu + k)} - E_x [\Psi^2(x + k)] \right] \quad (\text{A.14})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] = 0 \quad (\text{A.15})$$

Plus de détails sur le calcul de $E_x [\Psi^2(x + k)]$ seront donnés dans l'annexe A.3.

Nous obtenons donc la matrice de Fisher suivante :

$$\mathbf{F}(\boldsymbol{\theta}) = -E_x \left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log\{P[\mathbf{x}|\boldsymbol{\theta}]\} \right] \quad (\text{A.16})$$

$$= \begin{pmatrix} -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} \right] & -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] \\ -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] & -E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} \right] \end{pmatrix} \quad (\text{A.17})$$

$$= N \begin{pmatrix} \frac{1}{\mu} - \frac{1}{\mu+k} & 0 \\ 0 & \Psi^2(k) - \frac{\mu}{k(\mu+k)} - E_x [\Psi^2(x + k)] \end{pmatrix} \quad (\text{A.18})$$

On considère alors la matrice G , matrice de Fisher relative à la densité log a

posteriori, en ajoutant à F l'opposé de la matrice Hessienne des densités log a priori :

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta}) \quad (\text{A.19})$$

$$= \mathbf{F}(\boldsymbol{\theta}) - \begin{pmatrix} \frac{a_\mu}{\mu^2} & 0 \\ 0 & \frac{a_k}{k^2} \end{pmatrix} \quad (\text{A.20})$$

$$= N \begin{pmatrix} \frac{1}{\mu} - \frac{1}{\mu+k} - \frac{a_\mu}{\mu^2} & 0 \\ 0 & \Psi^2(k) - \frac{\mu}{k(\mu+k)} - E_x[\Psi^2(x+k)] - \frac{a_k}{k^2} \end{pmatrix} \quad (\text{A.21})$$

Le gradient géodésique de la densité a posteriori au point $\boldsymbol{\theta}$ est alors donné par l'expression :

$$\tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}) = \mathbf{G}^{-1}\{\boldsymbol{\theta}\} \nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}) \quad (\text{A.22})$$

A.3 Estimation de $E_x[\Psi^2(x+k)]$

A notre connaissance, il n'existe pas de moyen simple de calculer analytiquement le terme $E_x[\Psi^2(x+k)]$ nécessaire dans la matrice de Fisher. Pour résoudre ce problème en un temps relativement court, une table de consultation est a été créée. La valeur de $E_x[\Psi^2(x+k)]$ a été calculée sur une grille de paramètres (k, μ) , en moyennant les valeurs d'une population de 10000 observations générées. Les valeurs de $E_x[\Psi^2(x+k)]$ pour les paramètres ne se situant pas sur les nœuds de la grille sont alors estimées par interpolation linéaire.

Annexe B

Détails des calculs du MALA sur variété pour la loi bivariée binomiale négative - normale

La probabilité qu'une observation suivant une loi bivariée binomiale négative - normale de paramètre $\boldsymbol{\theta} = (\mu, k, m, \sigma)$ obtienne la valeur x_i est définie par la fonction de densité de probabilité :

$$\begin{aligned} p(x_n|\mu, k, m, \sigma) &= p_{\mathcal{BN}}(x_n^{\text{TEP}}|\mu, k) \cdot p_{\mathcal{N}}(x_n^{\text{TDM}}|m, \sigma) \\ &= \binom{x_n^{\text{TEP}} + k - 1}{x_n^{\text{TEP}}} \left(\frac{\mu}{\mu + k}\right)^{x_n^{\text{TEP}}} \left(\frac{k}{\mu + k}\right)^k \\ &\quad \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n^{\text{TDM}} - m)^2}{\sigma^2}} \end{aligned}$$

En supposant que les observations sont mutuellement indépendantes, la vraisemblance $p(\mathbf{x}|\boldsymbol{\theta})$ peut donc être exprimée comme le produit des probabilités relatives à chacune des observations :

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= \prod_{i=1}^N p(x_i|\mu, k, m, \sigma) \\ &= \prod_{i=1}^N \left[\frac{\Gamma(x_i^{\text{TEP}} + k)}{\Gamma(x_i^{\text{TEP}} + 1)\Gamma(k)} \mu^{x_i^{\text{TEP}}} k^k (\mu + k)^{-(k+x_i^{\text{TEP}})} \right. \\ &\quad \left. \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i^{\text{TDM}} - m)^2}{\sigma^2}} \right] \end{aligned} \tag{B.1}$$

où Γ est la fonction gamma.

La log-vraisemblance $\mathcal{L}(\boldsymbol{\theta})$ prend alors la forme :

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &\triangleq \log\{P[\mathbf{x}|\boldsymbol{\theta}]\} \\ &= \sum_{i=1}^N [\log\{\Gamma(x_i + k)\} - \log\{\Gamma(x_i + 1)\} - \log\{\Gamma(k)\} \\ &\quad + x_i \log(\mu) + k \log(k) - (k + x_i) \log(\mu + k) \\ &\quad - \frac{1}{2} \log(\sigma) - \frac{1}{2} \log(2\pi) - \frac{(x_i - m)^2}{2\sigma^2}]\end{aligned}$$

B.1 MALA

MALA nécessite l'expression du gradient de la fonction log-vraisemblance. Nous calculons pour cela les dérivées partielles :

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{i=1}^N \left[\frac{x_i}{\mu} - \frac{x_i + k}{\mu + k} \right] \quad (\text{B.2})$$

$$\frac{\partial \mathcal{L}}{\partial k} = \sum_{i=1}^N \left[\Psi(x_i + k) - \Psi(k) + 1 + \log(k) - \log(\mu + k) - \frac{x_i + k}{\mu + k} \right] \quad (\text{B.3})$$

$$\frac{\partial \mathcal{L}}{\partial m} = \sum_{i=1}^N \left[\frac{x_i - m}{\sigma^2} \right] \quad (\text{B.4})$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \sum_{i=1}^N \left[-\frac{1}{\sigma} + \frac{(x_i - m)^2}{\sigma^3} \right] \quad (\text{B.5})$$

où Ψ est la fonction psi, ou digamma ($\Psi(x) \triangleq \frac{\partial}{\partial x} \log(\Gamma(x))$).

Le gradient de la log-vraisemblance $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ peut donc s'exprimer :

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mu} \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial k} \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial m} \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \sum_i \left[\frac{x_i}{\mu} - \frac{x_i + k}{\mu + k} \right] \\ \sum_i \left[\Psi(x_i + k) - \Psi(k) + 1 - \frac{x_i + k}{\mu + k} - \log\left(1 + \frac{\mu}{k}\right) \right] \\ \sum_i \left[\frac{x_i - m}{\sigma^2} \right] \\ \sum_i \left[-\frac{1}{\sigma} + \frac{(x_i - m)^2}{\sigma^3} \right] \end{bmatrix} \quad (\text{B.6})$$

Comme une démarche Bayésienne est utilisée, il est également nécessaire de considérer les informations a priori. D'après la loi de Bayes, la loi log a posteriori

$\mathcal{P}(\boldsymbol{\theta}) \triangleq \log\{p(\boldsymbol{\theta}|\mathbf{x})\}$ est définie par :

$$\mathcal{P}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \log\{\pi(\boldsymbol{\theta})\} \quad (\text{B.7})$$

et par linéarité du gradient nous obtenons :

$$\begin{aligned} \nabla_{\boldsymbol{\theta}}\mathcal{P}(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}) + \begin{bmatrix} \frac{\partial\pi(\boldsymbol{\theta})}{\partial\mu} \\ \frac{\partial\pi(\boldsymbol{\theta})}{\partial k} \\ \frac{\partial\pi(\boldsymbol{\theta})}{\partial m} \\ \frac{\partial\pi(\boldsymbol{\theta})}{\partial\sigma} \end{bmatrix} \\ &= \begin{bmatrix} \sum_i \left[\frac{x_i}{\mu} - \frac{x_i+k}{\mu+k} \right] + \frac{a_\mu}{\mu} + b_\mu \\ \sum_i \left[\Psi(x_i+k) - \Psi(k) + 1 - \frac{x_i+k}{\mu+k} - \log\left(1 + \frac{\mu}{k}\right) + \frac{a_k}{k} + b_k \right] \\ \sum_i \left[\frac{x_i-m}{\sigma^2} + \frac{a_m}{k} + b_m \right] \\ \sum_i \left[-\frac{1}{\sigma} + \frac{(x_i-m)^2}{\sigma^3} + \frac{a_\sigma}{k} + b_\sigma \right] \end{bmatrix} \end{aligned} \quad (\text{B.8})$$

B.2 MALA sur variété

Nous considérons maintenant les dérivées partielles secondes de la log-vraisemblance nécessaires au calcul de la matrice de Fisher :

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\mu^2} = \sum_i \left[-\frac{x_i}{\mu^2} + \frac{x_i+k}{(\mu+k)^2} \right] \quad (\text{B.10})$$

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial k^2} = \sum_i \left[\Psi^2(x_i+k) - \Psi^2(k) + \frac{x_i-\mu}{(\mu+k)^2} + \frac{\mu}{k(\mu+k)} \right] \quad (\text{B.11})$$

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial m^2} = \sum_i \left[-\frac{1}{\sigma^2} \right] \quad (\text{B.12})$$

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\sigma^2} = \sum_i \left[\frac{1}{\sigma^2} - \frac{3(x_i-m)^2}{\sigma^4} \right] \quad (\text{B.13})$$

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\mu\partial k} = \sum_i \left[\frac{x_i-\mu}{(\mu+k)^2} \right] \quad (\text{B.14})$$

$$\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial m\partial\sigma} = \sum_i \left[-2\frac{x_i-m}{\sigma^3} \right] \quad (\text{B.15})$$

où Ψ^2 est la fonction trigamma $\Psi^2(x) \triangleq \frac{\partial^2}{\partial x^2} \log(\Gamma(x))$.

Les valeurs de la matrice de Fisher sont alors les opposées des espérances de ces dérivées partielles secondes :

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu^2} \right] = N \left[\frac{1}{\mu} - \frac{1}{\mu + k} \right] \quad (\text{B.16})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial k^2} \right] = N \left[\Psi^2(k) - \frac{\mu}{k(\mu + k)} - E_x [\Psi^2(x + k)] \right] \quad (\text{B.17})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial m^2} \right] = N \left[\frac{1}{\sigma^2} \right] \quad (\text{B.18})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma^2} \right] = N \left[\frac{2}{\sigma^2} \right] \quad (\text{B.19})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial k} \right] = 0 \quad (\text{B.20})$$

$$-E_x \left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial m \partial \sigma} \right] = 0 \quad (\text{B.21})$$

Plus de détails sur le calcul de $E_x [\Psi^2(x + k)]$ sont donnés dans l'annexe A.3.

Nous obtenons donc la matrice de Fisher suivante :

$$\mathbf{F}(\boldsymbol{\theta}) = -E_x \left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log\{P[\mathbf{x}|\boldsymbol{\theta}]\} \right] \quad (\text{B.22})$$

$$= N \begin{pmatrix} \frac{1}{\mu} - \frac{1}{\mu+k} & 0 & 0 & 0 \\ 0 & \Psi^2(k) - \frac{\mu}{k(\mu+k)} - E_x [\Psi^2(x + k)] & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma} & 0 \\ 0 & 0 & 0 & \frac{2}{\sigma} \end{pmatrix} \quad (\text{B.23})$$

On considère alors la matrice G , matrice de Fisher relative à la densité log a

posteriori, en ajoutant à F l'opposé de la matrice Hessienne des densités log a priori :

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta}) \quad (\text{B.24})$$

$$= \mathbf{F}(\boldsymbol{\theta}) - \begin{pmatrix} \frac{a_\mu}{\mu^2} & 0 & 0 & 0 \\ 0 & \frac{a_k}{k^2} & 0 & 0 \\ 0 & 0 & \frac{a_m}{m^2} & 0 \\ 0 & 0 & 0 & \frac{a_\sigma}{\sigma^2} \end{pmatrix} \quad (\text{B.25})$$

$$= N \begin{pmatrix} \frac{1}{\mu} - \frac{1}{\mu+k} - \frac{a_\mu}{\mu^2} & 0 & 0 & 0 \\ 0 & \Psi^2(k) - \frac{\mu}{k(\mu+k)} - E_x[\Psi^2(x+k)] - \frac{a_k}{k^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma} - \frac{a_m}{m^2} & 0 \\ 0 & 0 & 0 & \frac{2}{\sigma} - \frac{a_\sigma}{\sigma^2} \end{pmatrix} \quad (\text{B.26})$$

Le gradient géodésique de la densité a posteriori au point $\boldsymbol{\theta}$ est alors donné par l'expression :

$$\tilde{\nabla}_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}) = \mathbf{G}^{-1}\{\boldsymbol{\theta}\} \nabla_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}) \quad (\text{B.27})$$

Annexe C

Modèle de mélange de distributions Poisson bivariées

C.1 Modèle des données / Vraisemblance

Le modèle de mélange de lois Poisson a été considéré comme plus approprié pour modéliser les projections TEP ainsi que les images TEP reconstruites [113, 109]. Les projections acquises en tomodensitométrie peuvent également être considérées comme suivant un mélange de distributions Poisson [42]. Nous ferons ici l'hypothèse que les images TEP et CT suivent toutes les deux des mélanges de distributions de Poisson.

Étant donné une classe \mathcal{Z}_z , soient X_z^{TEP} et X_z^{TDM} les variables aléatoires associées respectivement aux données TEP et TDM. En considérant que chaque modalité suit une distribution de Poisson,

$$\begin{aligned} X_z^{\text{TEP}} &\sim \mathcal{P}(\lambda_z^{\text{TEP}}) \\ X_z^{\text{TDM}} &\sim \mathcal{P}(\lambda_z^{\text{TDM}}) \end{aligned} \tag{C.1}$$

Si les deux canaux sont indépendants, la vraisemblance peut être exprimée comme le produit des deux distributions :

$$x_z | (\lambda_z^{\text{TEP}}, \lambda_z^{\text{TDM}}) \sim \mathcal{P}(X_z^{\text{TEP}} | \lambda_z^{\text{TEP}}) \times \mathcal{P}(X_z^{\text{TDM}} | \lambda_z^{\text{TDM}}) \tag{C.2}$$

Or certains travaux considèrent les deux modalités comme intrinsèquement corrélées [54]. De ce fait, nous proposons de modéliser les données bimodales par un mélange de distributions de Poisson bivariées, qui a l'avantage de considérer la dépendance

entre les deux canaux. On peut interpréter les données comme étant le résultats du processus suivant :

$$\begin{aligned} X_z^{\text{TEP}} &= X_z^1 + X_z^0 \\ X_z^{\text{TDM}} &= X_z^2 + X_z^0 \end{aligned} \quad (\text{C.3})$$

où les X_z^i sont des variables aléatoires suivant chacune une loi de Poisson :

$$\forall i = \{0, 1, 2\}, X_z^i \sim \mathcal{P}(\theta_z^i) \quad (\text{C.4})$$

et où X_z^0 représente la dépendance entre les intensités TEP et TDM.

Cela implique les distributions marginales de Poisson suivantes :

$$\begin{aligned} X_z^{\text{TEP}} &\sim \mathcal{P}(\theta_z^1 + \theta_z^0) \\ X_z^{\text{TDM}} &\sim \mathcal{P}(\theta_z^2 + \theta_z^0) \end{aligned} \quad (\text{C.5})$$

Par conséquent, les données bimodales TEP/TDM extraites de la classe \mathcal{Z}_z suivent une distribution de Poisson bivariée :

$$(X_z^{\text{TEP}}, X_z^{\text{TDM}}) \sim \mathcal{BP}(\theta_z^1, \theta_z^2, \theta_z^0) \quad (\text{C.6})$$

ou

$$\begin{aligned} P[x_n | \boldsymbol{\theta}_z, z_n = z] &= e^{-(\theta_z^1 + \theta_z^2 + \theta_z^0)} \frac{(\theta_z^1)^{x_n^{\text{TEP}}}}{x_n^{\text{TEP}}!} \frac{(\theta_z^2)^{x_n^{\text{TDM}}}}{x_n^{\text{TDM}}!} \\ &\cdot \sum_{j=0}^{\min(x_n^{\text{TEP}}, x_n^{\text{TDM}})} \binom{x_n^{\text{TEP}}}{j} \binom{x_n^{\text{TDM}}}{j} j! \left(\frac{\theta_z^0}{\theta_z^1 \theta_z^2} \right)^j \end{aligned} \quad (\text{C.7})$$

En supposant que toutes les observations sont indépendantes dans chacun des canaux, la vraisemblance peut être définie par :

$$P[\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}] = \prod_{z=1}^Z \prod_{n \in \mathcal{I}_z} P[x_n | \boldsymbol{\theta}_z, z_n = z] \quad (\text{C.8})$$

où \mathcal{I}_z est l'ensemble des voxels appartenant à la classe \mathcal{Z}_z et $\boldsymbol{\theta}_z = (\theta_z^1, \theta_z^2, \theta_z^0)$.

C.2 Distribution conditionnelle $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$ dans un échantillonneur de Gibbs hybride.

Chatelain2006 et al. [22] ont étudié récemment des méthodes de moments pour estimer les paramètres d'un mélange de distributions Poisson multivariées. Nous inscrivons ici l'estimation des paramètres dans un cadre Bayésien. Plus précisément, un échantillonneur Metropolis-dans-Gibbs est proposé pour résoudre le problème conjoint de segmentation et d'estimation des paramètres. Les échantillons sont tirés itérativement selon les densités conditionnelles de la densité cible. Contrairement à la distribution binomiale négative, la distribution de Poisson bivariée n'engendre pas de problème d'anisotropie ou de corrélation entre ses paramètres. Un échantillonneur basique suffit pour générer des échantillons asymptotiquement distribués selon la distribution conditionnelle $p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{x})$.

Le vecteur de paramètre, $\boldsymbol{\theta}$ est mis à jour au moyen d'un algorithme reposant sur une marche aléatoire de type Metropolis-Hastings (ou Random Walk Metropolis-Hastings RWMH) [116, p. 245] avec la loi de proposition suivante :

$$\boldsymbol{\theta}_z^* \sim \mathcal{TN}(\boldsymbol{\theta}_z^{(t-1)}, \epsilon \mathbb{I}_3). \quad (\text{C.9})$$

où \mathcal{TN} est la distribution normale tri-variée, $\boldsymbol{\theta}_z^{(t-1)}$ est la dernière valeur de la chaîne et le pas ϵ est choisi de manière à ce que le taux d'acceptation tende vers 1/2. De plus, en considérant le fait que la loi de proposition est symétrique, le ratio d'acceptation peut être exprimé comme le produit du rapport de vraisemblance et du rapport des a priori :

$$a = \min \left\{ 1, \prod_{n \in \mathcal{I}_z} \frac{P[x_n | \boldsymbol{\theta}_z^*, z_n = z]}{P[x_n | \boldsymbol{\theta}_z^{(t-1)}, z_n = z]} \frac{\pi(\boldsymbol{\theta}_z^*)}{\pi(\boldsymbol{\theta}_z^{(t-1)})} \right\}$$

où $\pi(\boldsymbol{\theta}_z)$ est la distribution a priori de $\boldsymbol{\theta}_z$.

Bibliographie

- [1] Shun'ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191. AMS Bookstore, 2000.
- [2] Gerald Antoch, Jörg Stattaus, Andre T. Nemat, Simone Marnitz, Thomas Beyer, Hilmar Kuehl, Andreas Bockisch, Jörg F. Debatin, and Lutz S. Freudenberg. Non-small cell lung cancer : Dual-modality PET/CT in preoperative staging. *Radiology*, 229(2) :526–533, 2003. PMID : 14512512.
- [3] M. Aristophanous, B.C. Penney, M.K. Martel, and C. A. Pelizzari. A gaussian mixture model for definition of lung tumor volumes in positron emission tomography. *Med. Phys.*, 34(11) :4223–4235, 2007.
- [4] Ulas Bagci, JayaramK. Udupa, Jianhua Yao, and DanielJ. Mollura. Co-segmentation of functional and anatomical images. In Nicholas Ayache, Hervé Delingette, Polina Golland, and Kensaku Mori, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 459–467. Springer Berlin Heidelberg, 2012.
- [5] Junjie Bai, Qi Song, Sudershan K. Bhatia, and Xiaodong Wu. Globally optimal lung tumor co-segmentation of 4D CT and PET images. *Proc. SPIE*, 8669 :86690W–86690W–7, 2013.
- [6] C. Ballangan, Xiuying Wang, M. Fulham, S. Eberl, and Dagan Feng. Lung tumor delineation in PET-CT images using a downhill region growing and a Gaussian mixture model. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2173–2176, Sept 2011.
- [7] David L. Barbee, Ryan T. Flynn, James E. Holden, Robert J. Nickles, and Robert Jeraj. A method for partial volume correction of PET-imaged tumor heterogeneity using expectation maximization with a spatially varying point spread function. *Physics in Medicine and Biology*, 55(1) :221–236, JAN 7 2010.
- [8] Sandip Basu, Thomas C. Kwee, Robert Gatenby, Babak Saboury, Drew A. Torigian, and Abass Alavi. Evolving role of molecular imaging with PET

- in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(6) :987–991, JUN 2011.
- [9] Saoussen Belhassen and Habib Zaidi. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med. Phys.*, 37(3) :1309–1324, 2010.
- [10] Thierry Berghmans, Michele Dusart, Marianne Paesmans, Claude Hossein-Foucher, Irene Buvat, Catherine Castaigne, Arnaud Scherpereel, Céline Masciaux, Michel Moreau, Martine Roelandts, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC) : a systematic review and meta-analysis (MA) by the european lung cancer working party for the IASLC lung cancer staging project. *Journal of Thoracic Oncology*, 3(1) :6–12, 2008.
- [11] Q. Black, I. Grills, L. Kestin, C. Wong, J. Wong, A. Martinez, and D. Yan. Defining a radiotherapy target with positron emission tomography. *Int. J. Radiat. Oncol. Biol. Phys.*, 60 :1272–1282, 2004.
- [12] C. I. Bliss and R. A. Fisher. Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2) :pp. 176–200, 1953.
- [13] Eric T. Bradlow, Bruce G. S. Hardie, and Peter S. Fader. Bayesian inference for the negative binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, 11(1) :pp. 189–201, 2002.
- [14] Frank J Brooks. Area under the cumulative SUV-volume histogram is not a viable metric of intratumoral metabolic heterogeneity : further comments. *European Journal of Nuclear Medicine and Molecular Imaging*, 40(12) :1926–1927, 2013.
- [15] Frank J Brooks. On some misconceptions about tumor heterogeneity quantification. *European Journal of Nuclear Medicine and Molecular Imaging*, 40(8) :1292–1294, 2013.
- [16] Quentin L. Burrell. Using the gamma-poisson model to predict library circulations. *Journal of the American Society for Information Science*, 41(3) :164–170, 1990.
- [17] Irène Buvat. Les limites du SUV. *Médecine Nucléaire*, 31(4) :165 – 172, 2007. 9è Conférence Internationale de l’ACOMEN L’instrumentation en Médecine Nucléaire.

- [18] N.G. Cadigan and Jared Tobin. Estimating the negative binomial dispersion parameter with highly stratified surveys. *Journal of Statistical Planning and Inference*, 140(7) :2138 – 2147, 2010.
- [19] Vincent P. Carroll, Hau L. Lee, and Ambar G. Rao. Implications of salesforce productivity heterogeneity and demotivation : A navy recruiter case study. *Management Science*, 32(11) :1371–1388, 1986.
- [20] Lotfi Chaari, Hadj Batatia, Caroline Chaux, and Jean-Yves Tournet. Sparse signal and image recovery using a proximal bayesian algorithm. *arXiv preprint arXiv :1401.3988*, 2014.
- [21] KSC Chao, WR Bosch, S Mutic, JS Lewis, F Dehdashti, MA Mintun, JF Dempsey, CA Perez, JA Purdy, and MJ Welch. A novel approach to overcome hypoxic tumor resistance : Cu-ATSM-guided intensity-modulated radiation therapy. *International Journal of Radiation Oncology*Biophysics*, 49(4) :1171–1182, MAR 15 2001.
- [22] F. Chatelain, A. Ferrari, and J.-Y. Tournet. Parameter estimation for multivariate mixed Poisson distributions. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 3, page III, may 2006.
- [23] J. L. Chen, S. R. Gunn, M. S. Nixon, and R. N. Gunn. Markov random field models for segmentation of PET images. In *Proc. Inf. Proc. Med. Im.*, volume 2082, pages 468–474, 2001.
- [24] Sugama Chicklore, Vicky Goh, Musib Siddique, Arunabha Roy, Paul K Marsden, and Gary JR Cook. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *European Journal of Nuclear Medicine and Molecular Imaging*, 40(1) :133–140, 2013.
- [25] Ole F. Christensen, Gareth O. Roberts, and Jeffrey S. Rosenthal. Scaling limits for the transient phase of local metropolis-hastings algorithms. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 67(2) :253–268, 2005.
- [26] K.L. Chung. *Lectures from Markov processes to Brownian motion*. Springer, New York, 1982.
- [27] Suzanne J. Clark and Joe N. Perry. Estimation of the negative binomial parameter κ by maximum quasi-likelihood. *Biometrics*, 45(1) :pp. 309–316, 1989.
- [28] PatrickL. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In Heinz H. Bauschke, Regina S. Burachik, Patrick L.

- Combettes, Veit Elser, D. Russell Luke, and Henry Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 185–212. Springer New York, 2011.
- [29] Arthur H Compton. A quantum theory of the scattering of X-rays by light elements. *Physical Review*, 21(5) :483, 1923.
- [30] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9) :2722–2727, 1963.
- [31] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. ii. *Journal of Applied Physics*, 35(10) :2908–2913, 1964.
- [32] Jean-François Daisne, Mérence Sibomana, Anne Bol, Guy Cosnard, Max Lonneux, and Vincent Grégoire. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients : accuracy, reproducibility and consistency. *Radiotherapy and Oncology*, 69(3) :237–245, 2003.
- [33] Jean-François Daisne, Mérence Sibomana, Anne Bol, Thomas Doumont, Max Lonneux, and Vincent Grégoire. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios : influence of reconstruction algorithms. *Radiotherapy and Oncology*, 69(3) :247 – 250, 2003.
- [34] John J. Deely and Adrian F. M. Smith. Quantitative refinements for comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 161(1) :pp. 5–12, 1998.
- [35] D. Didierlaurent, S. Ribes, H. Batatia, C. Jaudet, L.O. Dierickx, S. Zerdoud, S. Brillouet, O. Caselles, and F. Courbon. The retrospective binning method improves the consistency of phase binning in respiratory-gated PET/CT. *Physics in Medicine and Biology*, 57(23) :7829–7841, novembre 2012.
- [36] David Didierlaurent. *Mise en cohérence spatiale et temporelle de données TEP-TDM acquises en respiration libre pour l'évaluation du cancer pulmonaire*. PhD thesis, 2013. Thèse de doctorat dirigée par Courbon, Frédéric et Batatia, Hadj Radio-physique et imagerie médicale Toulouse 3 2013.
- [37] L. Drever, W. Roa, A. McEwan, and D. Robinson. Iterative threshold segmentation for PET target volume delineation. *Med. Phys.*, 34 :1253–1265, 2007.
- [38] Laura Drever, Don M Robinson, Alexander McEwan, and Wilson Roa. A local contrast based approach to threshold segmentation for PET target volume delineation. *Medical physics*, 33(6) :1583–1594, 2006.

- [39] Janet F. Eary, Finbarr O’Sullivan, Janet O’Sullivan, and Ernest U. Conrad. Spatial Heterogeneity in Sarcoma (18)F-FDG Uptake as a Predictor of Patient Outcome. *Journal of Nuclear Medicine*, 49(12) :1973–1979, DEC 2008.
- [40] Andrew Samuel Christopher Ehrenberg. *Repeat-Buying : Facts, Theory, and Applications*. Griffin New London, 1988.
- [41] I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, Richard Laforest, W. L. Thorstad, and J. O. Deasy. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognition*, 42(6) :1162–1171, JUN 2009.
- [42] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imaging*, 21(2) :89–99, Feb 2002.
- [43] J. Erdi, O. Mawlawi, S. Larson, M. Imbriaco, H. Yeung, R. Finn, and J. Humm. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer*, 80 :2505–2509, 1997.
- [44] R. A. Fisher. The negative binomial distribution. *Annals of Human Genetics*, 11(1) :182–187, 1941.
- [45] S. Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Verlag, 2006.
- [46] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6(6) :721–741, Nov 1984.
- [47] General Electric Healthcare. General electric healthcare. http://www3.gehealthcare.com/en/Products/Categories/GoldSeal_-_Refurbished_Systems/GoldSeal_PET-CT/GoldSeal_Discovery_ST.
- [48] Peter Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1) :105 – 117, 1972.
- [49] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(2) :123–214, 2011.
- [50] Globocan 2012. globocan.iarc.fr. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
- [51] Richard Gordon, Robert Bender, and Gabor T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3) :471 – 481, 1970.

- [52] Olga Y. Gorlova, Shih-Feng Weng, Yiqun Zhang, Christopher I. Amos, and Margaret R. Spitz. Aggregation of cancer among relatives of never-smoking lung cancer patients. *International Journal of Cancer*, 121(1) :111–118, 2007.
- [53] P. J. Green. Reversible jump Markov chain Monte Carlo methods computation and Bayesian model determination. *Biometrika*, 82(4) :711–732, December 1995.
- [54] H. Gribben, P. Miller, G.G. Hanna, K.J. Carson, and A.R. Hounsell. MAP-MRF segmentation of lung tumours in PET/CT images. In *Biomedical Imaging : From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, pages 290–293, June 2009.
- [55] Dongfeng Han, John Bayouth, Qi Song, Aakant Taurani, Milan Sonka, John Buatti, and Xiaodong Wu. Globally optimal tumor segmentation in PET-CT images : A graph-based co-segmentation method. In Gábor Székely and HorstK. Hahn, editors, *Information Processing in Medical Imaging*, volume 6801 of *Lecture Notes in Computer Science*, pages 245–256. Springer Berlin Heidelberg, 2011.
- [56] Mathieu Hatt, Catherine Cheze Le Rest, Nidal Albarghach, Olivier Pradier, and Dimitris Visvikis. PET functional volume delineation : a robustness and repeatability study. *European Journal of Nuclear Medicine and Molecular Imaging*, 38 :663–672, 2011. 10.1007/s00259-010-1688-6.
- [57] Mathieu Hatt, Catherine Cheze le Rest, A. Turzo, C. Roux, and Dimitris Visvikis. A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET. *Medical Imaging, IEEE Transactions on*, 28(6) :881–893, june 2009.
- [58] Mathieu Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, Catherine Cheze-Le Rest, and Dimitris Visvikis. Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET. *Phys. Med. Biol.*, 52 :3467–3491, 2007.
- [59] Mathieu Hatt, Catherine Cheze le Rest, Patrice Descourt, André Dekker, Dirk De Ruysscher, Michel Oellers, Philippe Lambin, Olivier Pradier, and Dimitris Visvikis. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *International Journal of Radiation Oncology*Biology*Physics*, 77(1) :301 – 308, 2010.
- [60] Mathieu Hatt, C. Roux, and Dimitris Visvikis. A segmentation algorithm for heterogeneous tumor automatic delineation in PET. In *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*, volume 5, pages 3939–3945, 26 2007-nov. 3 2007.

- [61] Gloria H. Heppner. Tumor heterogeneity. *Cancer Research*, 44(6) :2259–2265, 1984.
- [62] Gloria H. Heppner and Bonnie E. Miller. Tumor heterogeneity : biological implications and therapeutic consequences. *Cancer and Metastasis Reviews*, 2 :5–23, 1983. 10.1007/BF00046903.
- [63] Corneline J. Hoekstra, Otto S. Hoekstra, Sigrid G. Stroobants, Johan Vans-teenkiste, Johan Nuyts, Egbert F. Smit, Maarten Boers, Jos W.R. Twisk, and Adriaan A. Lammertsma. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. *Journal of Nuclear Medicine*, 43(10) :1304–1309, 2002.
- [64] Nikie Hoetjes, Floris van Velden, Otto Hoekstra, Corneline Hoekstra, Nanda Krak, Adriaan Lammertsma, and Ronald Boellaard. Partial volume correction strategies for quantitative FDG PET in oncology. *European Journal of Nuclear Medicine and Molecular Imaging*, 37 :1679–1687, 2010. 10.1007/s00259-010-1472-7.
- [65] G. N. Hounsfield. Computerized transverse axial scanning (tomography) : Part 1. description of system. *The British Journal of Radiology*, 46(552) :1016–1022, 1973. PMID : 4757352.
- [66] H.M. Hudson and R.S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *Medical Imaging, IEEE Transactions on*, 13(4) :601–609, Dec 1994.
- [67] Zacharie Irace and Hadj Batatia. Motion-based interpolation to estimate spatially variant PSF in positron emission tomography. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5, Sept 2013.
- [68] Zacharie Irace and Hadj Batatia. Bayesian spatiotemporal segmentation of combined PET-CT data using a bivariate Poisson mixture model. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European (in press)*, Sept 2014.
- [69] Zacharie Irace, Marcelo Pereyra, Nicolas Dobigeon, and Hadj Batatia. Bayesian segmentation of chest tumors in PET scans using a Poisson-gamma mixture model. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 809–812, June 2011.
- [70] Zacharie Irace, Marcelo Pereyra, Nicolas Dobigeon, and Hadj Batatia. Un modèle Bayésien de mélange de lois Poisson-Gamma pour segmenter des images TEP. In *Actes de la conférence RFIA 2012*, pages 978–2–9539515–2–3, Lyon, France, January 2012. Session "Posters".

- [71] Cyril Jaudet, David Didierlaurent, Julia Nalis, Lawrence O. Dierickx, Olivier Caselles, and Frederic Courbon. Quality controls and delineation protocol of PET/CT gated acquisition in function of the movement amplitude, size of spheres and signal over background ratio. *Proc. SPIE*, 7961 :796143–796143–12, 2011.
- [72] Walter Jentzen, Lutz Freudenberg, Ernst G. Eising, Melanie Heinze, Wolfgang Brandau, and Andreas Bockisch. Segmentation of PET volumes by iterative image thresholding. *Journal of Nuclear Medicine*, 48(1) :108–114, 2007.
- [73] N. Joshi and M. Brady. Non-parametric mixture model based evolution of level sets and application to medical images. *Int. J. Comp. Vis.*, 88 :52–68, 2010.
- [74] M. E. Kamasak and B. Bayraktar. Clustering dynamic PET images on the projection domain. *IEEE Trans. Nucl. Sci.*, 53(3) :496–503, 2007.
- [75] John W. Keyes. SUV : Standard uptake or silly useless value? *Journal of Nuclear Medicine*, 36(10) :1836–1839, 1995.
- [76] Elizabeth A. Kidd and Perry W. Grigsby. Intratumoral metabolic heterogeneity of cervical cancer. *Clinical Cancer Research*, 14(16) :5236–5241, AUG 15 2008.
- [77] S.A. Klugman. *Bayesian statistics in actuarial science : with emphasis on credibility*, volume 15. Springer, 1992.
- [78] Didier Lardinois, Walter Weder, Thomas F. Hany, Ehab M. Kamel, Stephan Korom, Burkhardt Seifert, Gustav K. von Schulthess, and Hans C. Steinert. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *New England Journal of Medicine*, 348(25) :2500–2507, 2003. PMID : 12815135.
- [79] Jerald F. Lawless. Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 15(3) :pp. 209–225, 1987.
- [80] Ho Yun Lee, Ji Yun Jeong, Kyung Soo Lee, Hyo Jin Kim, Joung-ho Han, Byung-Tae Kim, Jhingook Kim, Young Mog Shim, Jae-Hun Kim, and Inyoung Song. Solitary pulmonary nodular lung adenocarcinoma : Correlation of histopathologic scoring and patient survival with imaging biomarkers. *Radiology*, 264(3) :884–893, 2012. PMID : 22829686.
- [81] Quanzheng Li, E. Asma, Sangtae Ahn, and R.M. Leahy. A fast fully 4-D incremental gradient reconstruction algorithm for list mode PET data. *Medical Imaging, IEEE Transactions on*, 26(1) :58–67, jan. 2007.
- [82] A. Lindén and S. Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7) :1414–1421, 2011.

- [83] CC Ling, J Humm, S Larson, H Amols, Z Fuks, S Leibel, and JA Koutcher. Towards multidimensional radiotherapy (MD-CRT) : Biological imaging and biological conformality. *International Journal of Radiation Oncology*Biology*Physics*, 47(3) :551–560, JUN 1 2000.
- [84] James O. Lloyd-Smith. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS ONE*, 2(2) :e180, 02 2007.
- [85] Dominique Lord and Luis F. Miranda-Moreno. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes : A Bayesian perspective. *Safety Science*, 46(5) :751 – 770, 2008.
- [86] M.J. Maher. A bivariate negative binomial model to explain traffic accident migration. *Accident Analysis & Prevention*, 22(5) :487 – 498, 1990.
- [87] Nataliya V. Malyskina, Fred L. Mannering, and Andrew P. Tarko. Markov switching negative binomial models : An application to vehicle accident frequencies. *Accident Analysis & Prevention*, 41(2) :217 – 226, 2009.
- [88] Medical Illustration. [medicalillustration.com. http://mrrsinc.medicalillustration.com/generateexhibit.php?ID=38806.](http://mrrsinc.medicalillustration.com/generateexhibit.php?ID=38806)
- [89] C. Michel, M. Sibomana, A Boi, X. Bernard, M. Lonneux, M. Defrise, C. Comtat, P.E. Kinahan, and D.W. Townsend. Preserving poisson characteristics of PET data with weighted OSEM reconstruction. In *Nuclear Science Symposium, 1998. Conference Record. 1998 IEEE*, volume 2, pages 1323–1329 vol.2, 1998.
- [90] S B Mohr, C F Garland, E D Gorham, W B Grant, and F C Garland. Could ultraviolet B irradiance and vitamin D be associated with lower incidence rates of lung cancer? *Journal of Epidemiology and Community Health*, 62(1) :69–74, 2008.
- [91] D. W. G. Montgomery, A. Amira, and H. Zaidi. Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. *Med. Phys.*, 34(2) :722–736, 2007.
- [92] D.W.G. Montgomery, A. Amira, and F. Murtagh. An automated volumetric segmentation system combining multiscale and statistical reasoning. In *Proc. IEEE Int. Symp. Circ. and Syst.*, volume 4, pages 3789–3792, 2005.
- [93] Ranjini Natarajan and Charles E. McCulloch. Gibbs sampling with diffuse proper priors : A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics*, 7(3) :pp. 267–277, 1998.

- [94] Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3) :pp. 705–741, 2003.
- [95] Ursula Nestle, Stephanie Kremp, Andrea Schaefer-Schuler, Christiane Sebastian-Welsch, Dirk Hellwig, Christian Rube, and Carl-Martin Kirsch. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. *Journal of Nuclear Medicine*, 46(8) :1342–1348, 2005.
- [96] Ursula Nestle, Andrea Schaefer-Schuler, Stephanie Kremp, Andreas Groeschel, Dirk Hellwig, Christian Rube, and Carl-Martin Kirsch. Target volume definition for 18F-FDG PET-positive lymph nodes in radiotherapy of patients with non-small cell lung cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 34(4) :453–462, 2007.
- [97] T.E. Nichols, Jinyi Qi, E. Asma, and R.M. Leahy. Spatiotemporal reconstruction of list-mode PET data. *Medical Imaging, IEEE Transactions on*, 21(4) :396–404, apr. 2002.
- [98] Stanislas Nicol. *Etude et construction d’un tomographe TEP/TDM pour petits animaux, combinant modules phoswich à scintillateurs et detecteur à pixels hybrides*. PhD thesis, 2010. These de doctorat dirigee par Morel, Christian et Karkar, Sonia Instrumentation Aix Marseille 2 2010.
- [99] F O’Sullivan, S Roy, and J Eary. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics*, 4(3) :433–448, JUL 2003.
- [100] F O’Sullivan, S Roy, J O’Sullivan, C Vernon, and J Eary. Incorporation of tumor shape into an assessment of spatial heterogeneity for human sarcomas imaged with FDG-PET. *Biostatistics*, 6(2) :293–301, APR 2005.
- [101] Z. Ouksili. *Compensation du mouvement respiratoire dans les images TEP/TDM thoraciques*. PhD thesis, University of Toulouse, 2010.
- [102] Z. Ouksili and H. Batatia. 4D CT image reconstruction based on interpolated optical flow fields. In *IEEE Int. Conf. on Image Processing (ICIP 2010)*, pages 633 – 636, Hong Kong, Sep 2010.
- [103] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3) :123–231, 2013.
- [104] Byung-Jung Park and Dominique Lord. Application of finite mixture models for vehicle crash data analysis. *Accident Analysis & Prevention*, 41(4) :683 – 691, 2009.

- [105] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret. Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm. *IEEE Trans. Image Processing*, 22(6) :2385–2397, June 2013.
- [106] Marcelo Pereyra. Proximal Markov chain Monte Carlo algorithms. *arXiv preprint arXiv :1306.0187*, 2013.
- [107] Walter W. Piegorsch. Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, 46(3) :pp. 863–867, 1990.
- [108] Pneumocourlancy. pneumocourlancy.fr. http://pneumocourlancy.fr/popup/anat_poumon.html.
- [109] N. Pustelnik, C. Chaux, and J.-C. Pesquet. A wavelet-based quadratic extension method for image deconvolution in the presence of poisson noise. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 701 –704, april 2009.
- [110] Jinyi Qi and R.H. Huesman. List mode reconstruction for PET with motion compensation : a simulation study. In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pages 413–416, 2002.
- [111] Jinyi Qi and Richard M Leahy. Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine and Biology*, 51(15) :R541, 2006.
- [112] Johann Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser manningfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsische Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physikalische Klasse Radon*, 69 :262–277, 1917.
- [113] A. Rangarajan, I.T. Hsiao, and G. Gindi. A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *J. Math. Im. Vis.*, 12 :199–217, 2000.
- [114] C. Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of Cal. Math. Soc.*, 37(3) :81–91, 1945.
- [115] A.J. Reader, S. Ally, F. Bakatselos, R. Manavaki, R.J. Walledge, A.P. Jeavons, P.J. Julyan, Sha Zhao, D.L. Hastings, and J. Zweit. One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays. *Nuclear Science, IEEE Transactions on*, 49(3) :693–699, Jun 2002.
- [116] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.

- [117] 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 259–273. Chapman & Hall, London, 1996.
- [118] G. O. Roberts and O. Stramer. Langevin Diffusions and Metropolis-Hastings Algorithms. *Methodology and Computing in Applied Probability*, 4 :337–357, 2002.
- [119] Gareth O. Roberts and Jeffrey S. Rosenthal. Markov-chain Monte Carlo : Some practical implications of theoretical results. *Canadian Journal of Statistics*, 26(1) :5–20, 1998.
- [120] Anthony Joseph Rockmore and Albert Macovski. A maximum likelihood approach to emission image reconstruction from projections. *Nuclear Science, IEEE Transactions on*, 23(4) :1428–1432, 1976.
- [121] Krishna Saha and Sudhir Paul. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, 61(1) :pp. 179–185, 2005.
- [122] S. Sanjay-Gopal and T.J. Hebert. Bayesian pixel classification using spatially variant finite mixtures and the generalized em algorithm. *Image Processing, IEEE Transactions on*, 7(7) :1014–1028, Jul 1998.
- [123] P. Santago and H.D. Gage. Statistical models of partial volume effect. *Image Processing, IEEE Transactions on*, 4(11) :1531 –1540, November 1995.
- [124] V. Savani and A. A. Zhigljavsky. Efficient estimation of parameters of the negative binomial distribution. *Communications in Statistics - Theory and Methods*, 35(5) :767–783, 2006.
- [125] P. J. Schluter, J. J. Deely, and A. J. Nicholson. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(3) :pp. 293–316, 1997.
- [126] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *Medical Imaging, IEEE Transactions on*, 1(2) :113–122, Oct 1982.
- [127] Sung Shine Shim, Kyung Soo Lee, Byung-Tae Kim, Myung Jin Chung, Eun Jung Lee, Joungho Han, Joon Young Choi, O Jung Kwon, Young Mog Shim, and Seonwoo Kim. Non-small cell lung cancer : Prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology*, 236(3) :1011–1019, 2005. PMID : 16014441.
- [128] Donald L Snyder, Lewis J Thomas, and Michel M Ter-Pogossian. A mathematical model for positron-emission tomography systems having time-of-flight measurements. *Nuclear Science, IEEE Transactions on*, 28(3) :3575–3583, 1981.

- [129] Marine Soret, Stephen L. Bacharach, and Irene Buvat. Partial-volume effect in PET tumor imaging. *J Nucl Med*, 48(6) :932–945, 2007.
- [130] Kuan-Hao Su, Jay S. Chen, Jih-Shian Lee, Chi-Min Hu, Chi-Wei Chang, Yuan-Hwa Chou, Ren-Shyan Liu, and Jyh-Cheng Chen. Image segmentation and activity estimation for microPET 11C-raclopride images using an expectation-maximum algorithm with a mixture of Poisson distributions. *Computerized Medical Imaging and Graphics*, 35(5) :417 – 426, 2011.
- [131] Kocherlakota Subrahmaniam and Kathleen Subrahmaniam. On the estimation of the parameters in the bivariate negative binomial distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 131–146, 1973.
- [132] WA Tome and JF Fowler. Selective boosting of tumor subvolumes. *International Journal of Radiation Oncology*Biography*Physics*, 48(2) :593–599, SEP 1 2000.
- [133] D. Van de Sompel and M. Brady. Simultaneous reconstruction and segmentation algorithm for positron emission tomography and transmission tomography. In *Biomedical Imaging : From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 1035 –1038, may 2008.
- [134] Wouter van Elmpt, Catharina ML Zegers, Marco Das, and Dirk De Ruyscher. Imaging techniques for tumour delineation and heterogeneity quantification of lung cancer : overview of current possibilities. *Journal of thoracic disease*, 6(4) :319, 2014.
- [135] Floris H. P. van Velden, Patsuree Cheebsumon, Maqsood Yaquub, Egbert F. Smit, Otto S. Hoekstra, Adriaan A. Lammertsma, and Ronald Boellaard. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(9) :1636–1647, SEP 2011.
- [136] S. Vauclin, I. Gardin, K. Doyeux, S. Hapdey, A. Edet-Sanson, and P. Vera. Segmentation des images TEP au 18F-FDG. principe et revue de la littérature. *Médecine Nucléaire*, 34(6) :358 – 369, 2010.
- [137] Lise Vaudor. Estimation du paramètre de distribution de la distribution binomiale négative : A priori, effort d'échantillonnage, et information. In *42èmes Journées de Statistique*, Marseille, France, France, 2010.
- [138] Daphné Wallach. *Compensation du mouvement respiratoire en TEP/TDM à l'aide de la super-résolution*. These, Université de Bretagne occidentale - Brest, October 2011.

- [139] Yining Wang. Estimation problems for the two-parameter negative binomial distribution. *Statistics & Probability Letters*, 26(2) :113 – 114, 1996.
- [140] Otto Warburg et al. On the origin of cancer cells. *Science*, 123(3191) :309–314, 1956.
- [141] Maria Werner-Wasik, Arden D. Nelson, Walter Choi, Yoshio Arai, Peter F. Faulhaber, Patrick Kang, Fabio D. Almeida, Ying Xiao, Nitin Ohri, Kristin D. Brockway, Jonathan W. Piper, and Aaron S. Nelson. What is the best way to contour lung tumors on PET scans? multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *International Journal of Radiation Oncology*Biology*Physics*, 82(3) :1164 – 1171, 2012.
- [142] Catharine ML West, Terry Jones, and Pat Price. The potential of positron-emission tomography to study anticancer-drug resistance. *Nature Reviews Cancer*, 4(6) :457–469, 2004.
- [143] Linda J Willson, J Leroy Folks, and JH Young. Multistage estimation compared with fixed-sample-size estimation of the negative binomial parameter k . *Biometrics*, 40(1) :109–117, 1984.
- [144] J. Wojak, E.D. Angelini, and I. Bloch. Joint variational segmentation of CT-PET data for tumoral lesions. In *Biomedical Imaging : From Nano to Macro, 2010 IEEE International Symposium on*, pages 217–220, April 2010.
- [145] F. Y. Wu. The Potts model. *Rev. Mod. Phys.*, 54(1) :235–268, Jan 1982.
- [146] Habib Zaidi and Issam El Naqa. PET-guided delineation of radiation therapy treatment volumes : a survey of image segmentation techniques. *European Journal of Nuclear Medicine and Molecular Imaging*, 37 :2165–2187, 2010. 10.1007/s00259-010-1423-3.
- [147] Yunlong Zhang, Zhirui Ye, and Dominique Lord. Estimating dispersion parameter of negative binomial distribution for analysis of crash data : Bootstrapped maximum likelihood method. *Transportation Research Record : Journal of the Transportation Research Board*, 2019 :15–21, 2007.
- [148] Haiyuan Zhu and Hassan Lakkis. Sample size calculation for comparing two negative binomial rates. *Statistics in Medicine*, 33(3) :376–387, 2014.