UNIVERSITE PAUL VERLAINE DE METZ
UFR Mathématiques, Informatique, Mécanique
Ecole Doctorale : IAEM Lorraine
Département : Mathématiques
Spécialité : Mathématiques

Thèse

présentée pour l'obtention du titre de

Docteur en Sciences de l'Université Paul Verlaine, Metz

par **Béatrice CHEVAILLIER**

Analyse de données d'IRM fonctionnelle rénale par quantification vectorielle

Soutenue le 9 mars 2010

Membres du jury

Rapporteurs :	Françoise Peyrin	Directrice de Recherche INSERM, CREATIS
	Frédéric Richard	Maître de conférences, Université Paris Descartes
Examinateurs :	Fatiha Alabau-Boussouira	Professeur, UPVM (co-directrice de thèse)
	Frédéric Bois	Professeur, UTC (directeur de thèse)
	Agnès Desolneux	Chargée de recherche CNRS, MAP5
	Jean-François Lerallut	Professeur, UTC
	Olivier Pietquin	Enseignant-Chercheur, Supélec

Table des matières

Liste des tableaux

Table des figures

Résumé

L'évaluation de la fonction rénale est primordiale pour détecter l'insuffisance rénale. Pour calculer certains paramètres fonctionnels qui la caractérisent, il existe différentes méthodes largement reconnues et utilisées en pratique. Les plus courantes sont l'estimation de la clairance (élimination) de la créatinine naturellement présente dans le sang, ou de celle d'un radiotraceur injecté au patient. Cette estimation peut se faire, dans les deux cas, par dosage de prélèvements sanguins ou urinaires, ou bien, dans le second cas uniquement, par exploitation quantitative d'images scintigraphiques. Cependant, ces méthodes souffrent de certaines faiblesses : selon les cas, elles peuvent manquer de précision, être mal adaptées à certains types de population, s'avérer contraignantes ou nécessiter des précautions particulières liées à l'utilisation de produits radioactifs. L'utilisation de l'Imagerie par Résonance Magnétique (IRM) à rehaussement de contraste avec injection de chélates de gadolinium présente des avantages potentiels par rapport aux méthodes existantes, mais n'est pas suffisamment mature pour être utilisée en pratique clinique car certains problèmes demeurent non résolus. Une évaluation de cette méthode à grande échelle, sur une population variée de patients avec des reins sains ou pathologiques, par comparaison avec les techniques existantes, est nécessaire. L'exploitation des séquences acquises demande cependant un travail long et fastidieux aux radiologues pour effectuer le recalage de la série d'images et la segmentation des structures internes du rein, du moins si ces tâches sont exécutées « manuellement ». Notre objectif est de fournir un outil fiable et simple à utiliser qui permette d'automatiser en partie ces deux opérations, afin de rendre le processus de comparaison plus rapide.

Pour qu'une analyse correcte de la séquence soit possible, chaque voxel devrait correspondre, tout au long de l'acquisition, à une même zone anatomique. Or, les mouvements du patient, en particulier le mouvement respiratoire, qui a principalement lieu dans le plan de coupe, entraînent d'importants décalages entre les différentes images de la séquence. Un recalage est donc nécessaire, mais l'opération est particulièrement délicate à cause du rapport signal à bruit peu favorable, et surtout en raison de la cinétique de l'agent de contraste, qui induit d'importantes modifications d'intensité et qui fait apparaître des tissus différents selon l'instant d'acquisition. Des méthodes statistiques faisant appel à l'information mutuelle sont testées sur des données réelles.

Les principaux paramètres permettant classiquement d'évaluer la fonction rénale peuvent être alors estimés de manière non invasive à partir des courbes de perfusion de différentes régions d'intérêt. La segmentation de structures anatomiques rénales internes comme le cortex, la médullaire et les cavités est cruciale pour établir ces courbes. Or, l'évolution temporelle du contraste étant différente, pour des régions physiologiques, dans chacun de ces compartiments anatomiques, les voxels peuvent être classés selon leurs

courbes temps-intensité. Une stratégie de classification en deux grandes étapes est proposée : nous construisons d'abord des classificateurs de manière semi-automatique grâce à la coupe principale, c'est-à-dire celle contenant la plus grande surface rénale. Les voxels des autres coupes sont alors triés en utilisant le classificateur optimum sur la coupe principale. La première étape met en œuvre des méthodes d'apprentissage non supervisé par quantification vectorielle comme les K-moyennes ou le *Growing Neural Gas*. Cette technique est validée non seulement sur des données simulées mais aussi sur des données réelles, issues d'acquisitions sur patients. Les segmentations obtenues sont alors comparées à des segmentations manuelles faites par des radiologues : les mesures de similarité ont des valeurs comparables pour les différentes méthodes. La seconde étape s'inscrit dans le cadre de la théorie de la généralisation, qui permet de borner l'erreur de classification faite lors de son extension aux coupes voisines.

La méthode proposée procure les avantages suivants par rapport à une segmentation manuelle : un gain de temps important, une intervention de l'opérateur limitée et aisée, une bonne robustesse due à l'utilisation de l'intégralité de la séquence, au lieu de deux images seulement, et une bonne reproductibilité.

Chapitre 1

Introduction

L'évaluation de la fonction rénale est primordiale pour détecter l'insuffisance rénale, qui est un facteur de mortalité et d'augmentation du risque de pathologie cardiovasculaire [1]. L'approche la plus courante pour effectuer cette évaluation consiste à estimer le débit de filtration glomérulaire (*Glomerular filtration rate* ou GFR), qui est le volume de liquide filtré par unité de temps au niveau des glomérules (cf. paragraphe 2.1), ou bien la fonction rénale différentielle, c'est-à-dire la contribution de chacun des reins à la filtration globale. Pour estimer l'une ou l'autre de ces grandeurs, il existe différentes méthodes largement reconnues et utilisées en pratique. Cependant, elles souffrent de certaines faiblesses : selon les cas, elles peuvent manquer de précision, être mal adaptées à certains types de population, être contraignantes ou nécessiter des précautions particulières liées à l'utilisation de produits radioactifs.

L'imagerie par résonance magnétique (IRM) présente des avantages potentiels par rapport aux méthodes précédentes, mais n'est pas suffisamment mature pour être utilisée en pratique clinique car certains problèmes demeurent non résolus et demandent des études supplémentaires. Le CHU de Nancy, avec lequel nous collaborons[1], est le promoteur d'un projet national multicentrique[2] intitulé « Place de l'uro-IRM dans l'évaluation des conséquences fonctionnelles de l'obstruction urinaire de l'enfant et de l'adulte ». Ce projet vise à étudier la possibilité de remplacer les examens de médecine nucléaire par des IRM. Dans ce cadre, une évaluation de cette méthode à grande échelle, sur une population variée de patients avec des reins sains ou pathologiques, par comparaison avec les techniques existantes, est nécessaire. Afin de mieux comprendre les enjeux, commençons par dresser un bilan détaillé des techniques actuellement utilisées.

La méthode la plus couramment utilisée pour estimer le GFR est l'évaluation de la *clairance*[3] d'une substance produite à un taux pratiquement constant par le corps : la

[1]Nous remercions en particulier le Pr Jacques Felblinger, qui dirige le laboratoire IADI (Imagerie Adaptative, Diagnostique et Interventionnelle) et qui a rendu cette collaboration possible.

[2]Ce projet a fait l'objet d'un financement dans le cadre de la campagne 2005 du STIC (soutien aux techniques innovantes et coûteuses) national, et son investigateur principal est le Pr Michel Claudon.

[3]La clairance d'une substance est le volume apparent de plasma complètement épuré de cette substance par unité de temps.

créatinine. Cette dernière est entièrement évacuée par le rein dans les urines[4]. La clairance est calculée à partir des résultats du dosage de la créatinine dans le plasma et les urines. Ce procédé nécessite un recueil de la totalité des urines pendant 24 heures et une prise de sang, ce qui est assez lourd. Il est également possible de l'estimer à partir d'un seul prélèvement sanguin, mais de manière moins précise, en utilisant des formules comme celle de Cokcroft et Gault [2] ou la MDRD (*Modification of Diet in Renal Disease*) [3], adaptées à certains types de population seulement. Cependant, le débit de filtration glomérulaire ne peut pas être évalué de manière précise par cette méthode, en particulier en cas d'insuffisance rénale sévère[5] : un patient peut en effet perdre un pourcentage élevé de sa fonction rénale avant que sa concentration sanguine de créatinine ne présente des valeurs anormales [4].

Les progrès en médecine nucléaire ont permis l'utilisation de traceurs radioactifs, encore appelés radiotraceurs ou radiopharmaceutiques, qui sont des substances marquées par un élément radioactif qui perd son activité en quelques heures. Pour les radiotraceurs rénaux, l'isotope radioactif est le plus souvent du technetium ^{99m}Tc ou de l'iode ^{131}I. Le traceur, le plus souvent injecté au patient par intraveineuse, est filtré par le rein et passe dans les urines. On peut alors évaluer la clairance de ce traceur, soit en mesurant le taux de radioactivité dans des prélèvements sanguins [5, 6], soit en réalisant un examen scintigraphique, soit en couplant les deux [7]. La première méthode peut permettre d'obtenir une bonne précision dans les mesures, mais il est nécessaire de porter une grande attention aux détails de la procédure : préparation et dilution des solutions, mesure des temps d'injection et de prélèvement, précautions pour éviter la contamination des échantillons... Elle est souvent considérée comme la méthode de référence, cependant elle est longue à mettre en œuvre et elle ne donne accès qu'à la fonction rénale globale, et non à la fonction rénale différentielle.

L'utilisation de scintigraphies est devenue courante pour évaluer la fonction rénale. Cette modalité permet d'obtenir, de manière non invasive, une image de la distribution du traceur à l'intérieur du corps en enregistrant les rayonnements gamma émis par ce traceur avec une caméra extérieure, sensible à ce type de radiations et appelée gamma caméra [8]. Les rayons gamma ont une énergie suffisante pour traverser les tissus et s'échapper du corps pour atteindre les détecteurs, même s'ils sont émis par des organes situés en profondeur. Par ailleurs, les doses de traceur utilisées sont faibles et l'irradiation est beaucoup moins forte que lors d'un examen radiologique avec rayons X. Comme seuls certains rayons gamma sont sélectionnés par le collimateur de la caméra, l'image obtenue est en réalité une projection bidimensionnelle de la distribution du radiotraceur sur la surface du capteur[6] (cf. figure 1.1[7]).

[4]La majeure partie de la créatinine est filtrée dans les glomérules, l'autre est sécrétée au niveau des tubules ; elle ne subit en revanche pas de réabsorption (cf. paragraphe 2.1).

[5]La sécrétion tubulaire entraîne une surestimation de la clairance glomérulaire, négligeable tant que la fonction rénale est bonne, mais qui rend la méthode inadaptée en cas d'insuffisance rénale sévère.

[6]Il est possible de réaliser des acquisitions tridimensionnelles (tomographie), mais ce type d'examen est beaucoup moins répandu, et n'entre pas dans le cadre de notre projet.

[7]Téléchargée sur le site http ://www.givision.ch/fr/mednuc/imagesnuc/renale/renale.jpg

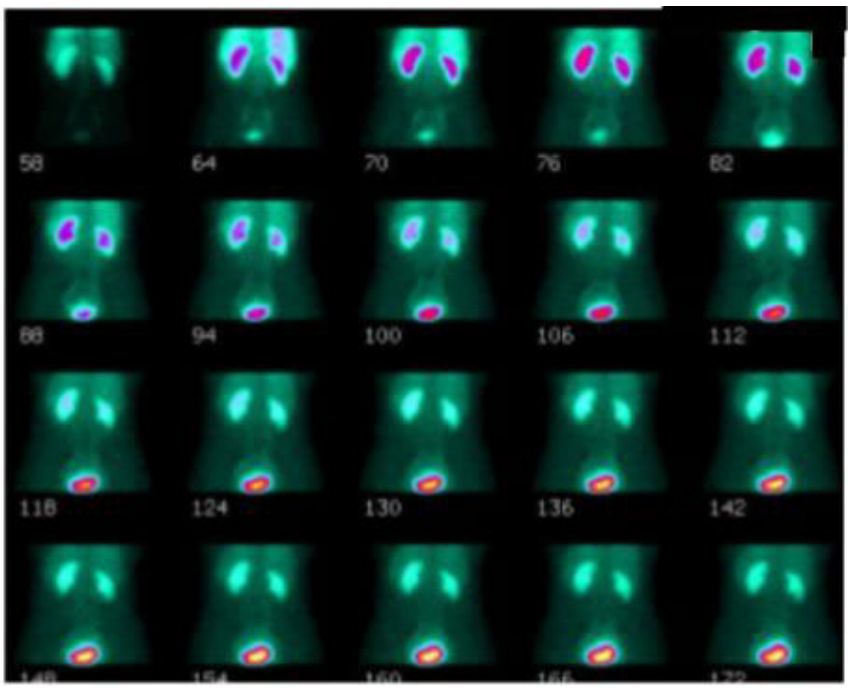


FIG. 1.1 – Exemple de scintigraphie rénale[5].

La scintigraphie dite dynamique permet d'estimer le GFR et la fonction rénale différentielle ou absolue, la première étant exprimée en pourcentage de la seconde. L'examen consiste en l'acquisition d'une série d'images à différents instants après l'injection du traceur[8]. Typiquement, on enregistre une image toutes les 1 à 2 secondes pendant la minute suivant l'injection, puis toutes les 10 à 20 secondes pendant une trentaine de minutes.

Afin d'étudier l'évolution de l'activité du traceur dans le rein, on positionne généralement une région d'intérêt (*region of interest* ou ROI) englobant l'ensemble du rein[9]. On trace la courbe d'activité globale en fonction du temps, ou *rénogramme* (cf. figure 1.2)[10].

Pour obtenir une estimation fiable des paramètres représentatifs de la fonction rénale, il est nécessaire d'apporter certaines corrections au rénogramme. En effet, il convient d'éliminer le bruit dû à l'activité du traceur situé hors du rein mais qui se superpose à celle du traceur qui s'y trouve effectivement. Par ailleurs, il faut également tenir compte de la profondeur de l'organe, dont dépend la proportion de rayonnement atteignant effectivement le capteur par rapport à celui qui en est issu ; ceci est surtout vrai pour l'évaluation de la fonction absolue lorsque l'examen scintigraphique n'est pas couplé à la détermination d'une clairance rénale à partir de prélèvements urinaires ou sanguins. L'estimation du coefficient d'atténuation s'appuie sur des images anatomiques réalisées par ailleurs ou (non exclusif) des modèles mathématiques [9].

[8]Généralement du diéthylène triamine penta acide (DTPA), du mercaptoacétyltriglycine marqué au ^{99m}Tc (MAG3) ou du ^{131}I- ou ^{123}I-orthoiodohippurate (OIH), dont les mécanismes d'élimination sont différents [4]

[9]Notons qu'une analyse par zones (parenchyme et cavités) mettant mieux en évidence le transit intrarénal est possible mais nécessite la détermination de leurs contours [5]. Elle est cependant asssez difficile à mettre en œuvre et semble rarement utilisée.

[10]Téléchargée sur le site http ://www.nuclearonline.org/Newsletter/Renogram2.jpg

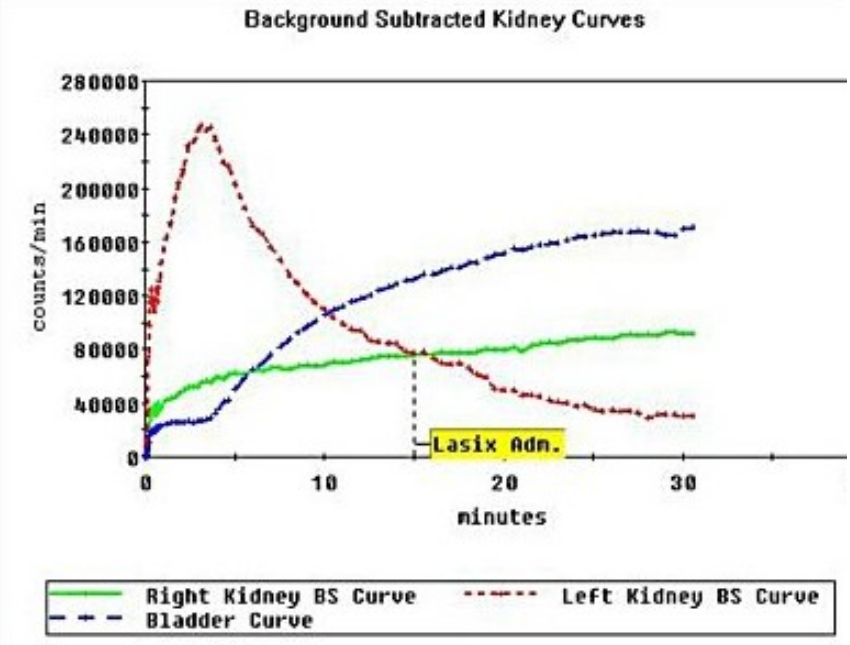


FIG. 1.2 – Exemple de rénogramme[8].

Des paramètres comme l'accumulation relative du radiopharmaceutique, le temps au pic d'activité, rapport d'activité à $t = 20$ min après l'injection et au temps du pic [4, 5] peuvent alors être estimés.

Cependant, ces méthodes présentent plusieurs inconvénients : elles nécessitent des précautions particulières liées à la radioactivité du traceur ; en outre, le risque d'allergies à l'iode est assez élevé. Certaines nécessitent des prélèvements urinaires et sanguins assez contraignants. Par ailleurs, il s'agit d'une technique d'imagerie essentiellement fonctionnelle : on n'obtient donc pas d'image anatomique, même si on peut considérer qu'on a accès à une certaine information sur le sujet. La projection est bidimensionnelle et le rein est assez mal délimité par rapport aux organes voisins. Ses structures internes étant peu visibles, une analyse fine par compartiment est assez difficile à obtenir.

Il serait intéressant de disposer d'une méthode permettant d'évaluer quantitativement la fonction rénale, qui soit plus facile à mettre en œuvre, qui donnerait des résultats précis et reproductibles, qui ne nécessiterait pas de prélèvement urinaire ou sanguin, tout en apportant des informations morphologiques sur le rein [10]. Une technique plus récente que la scintigraphie, et qui est susceptible de remplir ces conditions, a fait son apparition : l'IRM dynamique de perfusion.

L'IRM de perfusion est utilisée pour d'autres organes que le rein, en particulier le cœur et le cerveau. Sur une série d'images, on arrive à visualiser certaines anomalies fonctionnelles : pour différencier tissus perfusés et non perfusés, soit on réalise un marquage des noyaux d'hydrogène par des impulsions radiofréquences, soit on utilise un produit de contraste (ce qui est majoritairement le cas en IRM rénale). Il est ainsi possible de repérer des zones nécrosées, des tumeurs, certains œdèmes, des inflammations,... Un autre objectif, plus ambitieux, est d'évaluer, toujours à partir d'une série d'images, des volumes

sanguins et des paramètres temporels (temps jusqu'au pic de contraste ou autres), afin d'en déduire le débit sanguin qui irrigue l'organe exploré [11, 12].

En ce qui concerne le rein, l'IRM de perfusion n'est cependant pas utilisée couramment en pratique clinique car les mesures de paramètres fonctionnels comme le GFR ne sont pas suffisamment corrélées à celles obtenues par des méthodes de référence [10]. Plusieurs facteurs peuvent expliquer ceci. Alors qu'en scintigraphie, le signal mesuré est approximativement proportionnel à la concentration du radiopharmaceutique, le lien entre la concentration de l'agent de contraste et le signal mesuré en IRM est beaucoup plus complexe (cf. paragraphe 2.2.2). Par ailleurs, l'exploitation des résultats en IRM nécessite de délimiter de manière précise le contour extérieur du rein ; or, son contraste est voisin de celui des organes limitrophes pendant une bonne partie de la perfusion, ainsi ce contour n'est pas très bien défini et l'extraction n'est donc pas aisée. En revanche, en scintigraphie, on se contente souvent de placer une région d'intérêt (ROI) entourant largement le rein, ce qui est plus facile et plus rapide. En outre, l'examen d'IRM dure plusieurs minutes ; une acquisition synchronisée avec la respiration n'est pas envisageable car la résolution temporelle serait trop faible par rapport à la vitesse des phénomènes d'intérêt. Le mouvement respiratoire qui la perturbe doit donc être corrigé par un post-traitement de recalage particulièrement délicat en IRM en raison des changements de contraste pendant la perfusion (cf. paragraphe 2.4). Le mode d'exploitation de la scintigraphie rend sans doute les résultats moins dépendants des mouvements respiratoires, en raison de la taille de la ROI utilisée. Certains modèles mathématiques utilisés pour évaluer des paramètres caractérisant la fonction rénale nécessitent la connaissance de la fonction d'entrée artérielle (*arterial input function* ou AIF), c'est-à-dire la concentration du radiotraceur ou de l'agent de contraste dans l'aorte ; alors que, dans le cas d'un examen par imagerie nucléaire, elle peut être évaluée par dosage d'une série de prélèvements sanguins suffisamment rapprochés, elle doit être estimée directement sur les images pour l'IRM, ce qui entraîne des problèmes de reproductibilité essentiellement fonctions de la taille de la ROI utilisée pour la déterminer [13]. Enfin, il existe une grande variété de protocoles différents, que ce soit au niveau des paramètres et séquences d'IRM utilisées, de la dose d'agent de contraste injectée, du post-traitement (recalage et segmentation) et des modèles utilisés, ce qui rend les comparaisons difficiles.

Il n'est donc pas encore sûr que la fonction rénale puisse être quantifiée de manière précise à partir de ce type d'images [10]. Une étape de validation importante est nécessaire avant que l'on puisse savoir si l'IRM de perfusion pourra remplacer la scintigraphie pour l'évaluation de la fonction rénale [14], alors que celle-ci est actuellement considérée comme la méthode de référence.

Le projet national intitulé « Place de l'uro-IRM dans l'évaluation des conséquences fonctionnelles de l'obstruction urinaire de l'enfant et de l'adulte », mené par le CHU de Nancy en partenariat avec 17 autres centres, vise à étudier la possibilité de remplacer en pratique clinique les scintigraphies par des examens d'IRM dynamique de perfusion avec injection d'un agent de contraste. Il s'agit non seulement de comparer les valeurs de paramètres caractérisant la fonction rénale obtenus par IRM à des résultats issus de scintigraphies, afin de valider la méthode [15, 16, 17], mais aussi de voir si les avantages potentiels de l'IRM peuvent effectivement être exploités. En particulier, une étude diffé-

renciée par compartiment serait rendue possible parce qu'en plus des informations sur la cinétique du traceur, on peut obtenir des informations morphologiques meilleures qu'avec la scintigraphie [18], à condition qu'on arrive facilement à segmenter les structures rénales internes (même si l'analyse par zones peut éventuellement être pratiquée en scintigraphie, à notre connaissance, on se limite au parenchyme et aux cavités, alors que nous espérons pouvoir extraire un compartiment supplémentaire, à savoir la médullaire). Des tests doivent être effectués sur une large base de données, afin de faire une évaluation statistique sur une population suffisamment nombreuse et représentative (adultes ou enfants, reins sains ou pathologiques...).

Mener ce projet à bien demande un travail long et fastidieux aux radiologues pour effectuer le recalage de la série d'images et la segmentation des structures internes du rein, du moins si ces tâches sont exécutées « manuellement » ; une fois ces étapes réalisées, l'estimation des paramètres caractéristiques de la fonction rénale peut se faire de manière quasiment automatique. Cependant, il s'avère que les segmentations obtenues peuvent varier de manière assez importante d'un radiologue à l'autre. En effet, on combine les difficultés liées à la complexité de la forme des compartiments à segmenter, comme en IRM de perfusion cérébrale, et les brusques changements de contraste et les mouvements rencontrés par exemple en IRM de perfusion myocardique.

Notre objectif est de fournir un outil fiable et simple à utiliser qui permette d'automatiser en totalité le recalage et en partie l'étape de segmentation, afin de rendre le processus de comparaison plus rapide et moins dépendant de l'opérateur, tout en laissant à ce dernier la possibilité d'intervenir. Nous proposons en particulier de réaliser une segmentation dite fonctionnelle en classant les voxels [11] rénaux en fonction de leur évolution temporelle. Ce procédé se distingue des méthodes plus classiques où la segmentation est réalisée sur une ou deux images de la série, alors que nous exploitons cette dernière en totalité. Nous espérons que les résultats ainsi obtenus soient plus reproductibles et moins sensibles au bruit et aux erreurs résiduelles de recalage. Le fait de ne pas disposer de vérité-terrain pour valider les résultats est l'une des difficultés majeures que nous avons rencontrées. Cela a d'abord orienté notre choix pour la méthode de segmentation, en nous faisant éliminer les méthodes d'apprentissage supervisé. Nous avons par ailleurs été amenés à construire un modèle de rein le plus réaliste possible afin de tester nos méthodes par des simulations, ce qui, à notre connaissance, n'a jamais été fait auparavant. Enfin, nous avons dû proposer une manière de valider nos résultats sur des données réelles.

Terminons cette introduction par une brève description des différents chapitres de cette thèse, l'ensemble des étapes du traitement étant résumé sur le schéma-bloc de la figure 1.3. Les chapitres 2 et 3 sont consacrés à l'état de l'art respectivement sur l'application et sur le cadre mathématique, en particulier sur les modèles statistiques utilisés. Les chapitres 4 et 5 concernent notre contribution.

Dans le chapitre 2, nous présentons différentes notions sur la physiologie rénale, les mécanismes de filtration, et expliquons comment l'IRM de perfusion avec injection de

[11] Un voxel (contraction de « *volumetric pixel* ») est l'équivalent d'un pixel en dimension 3. Les IRM dont nous disposons sont des acquisitions 3D, donc nous parlerons plutôt de voxels, sauf quand nous ne considérerons qu'une seule coupe : nous emploierons alors l'un ou l'autre terme.

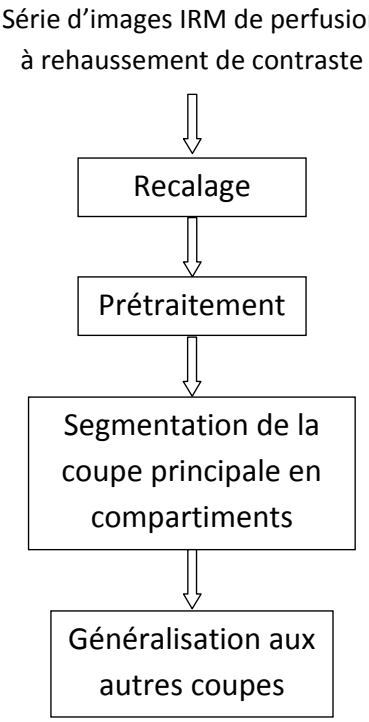


FIG. 1.3 – Traitements à effectuer sur les données.

produit de contraste permet de les étudier. Nous ne décrivons pas en détails les principes de l'imagerie IRM, mais donnons des éléments pour comprendre à quoi correspondent les paramètres de réglage des séquences utilisées, le rôle de l'agent de contraste ainsi que pour interpréter ce que l'on observe sur les images en lien avec la filtration rénale. Nous continuons par un bilan des modèles existants utiles à l'évaluation de la fonction rénale, qui permettent d'exploiter les courbes temps-intensité tirées des séries d'images.

L'exploitation de la plupart de ces modèles nécessite l'extraction de différents compartiments rénaux. Nous répertorions les méthodes de segmentation utilisées en IRM de perfusion rénale, qui doivent tenir compte des particularités de cette modalité : en particulier, les images sont floues, très bruitées et soumises à de très forts changements de contraste durant la perfusion. De plus, un pré-traitement de recalage est nécessaire pour éliminer le mouvement respiratoire : la fin de ce chapitre est dédiée à l'état de l'art des techniques spécifiquement testées dans ce domaine.

Le chapitre 3 est consacré au cadre mathématique dans lequel nous nous plaçons pour effectuer la segmentation fonctionnelle du rein sur toutes les coupes acquises. Dans l'introduction de ses deux parties, nous expliquons brièvement l'intérêt des notions introduites pour notre application. La première partie de ce chapitre concerne la quantification vectorielle par les algorithmes des K-moyennes ou du *Growing Neural Gas* (GNG) : pour regrouper les voxels selon leur évolution temporelle, nous procédons d'abord à une

quantification vectorielle des courbes temps-intensité de la coupe rénale principale. Nous pouvons non seulement en déduire une partition, ou *clustering*, des voxels rénaux de cette coupe mais également classer ceux de toutes les autres coupes. Nous aboutissons donc à une segmentation tridimensionnelle du rein en compartiments fonctionnels. La validité de cette extension est étudiée dans la seconde partie du chapitre.

Les chapitres 4 et 5 reprennent en détails l'intégralité de la méthode, en faisant le lien avec le chapitre 3.

Le chapitre 4 est plus spécifiquement consacré aux données simulées ou réelles, dont la nature influence tant le choix d'un pré-traitement de recalage et de la stratégie de segmentation que la validation des résultats. Notre objectif principal est de réaliser des tests sur des données réelles, mais nous ne disposons pas de vérité-terrain associée. Nous proposons donc un modèle rénal destiné à tester nos méthodes de segmentation. Nous espérons être en mesure de mettre en évidence et d'identifier l'origine de certains phénomènes fréquemment rencontrés lors des segmentations sur les données réelles. Ces dernières doivent subir un pré-traitement de recalage destiné à éliminer au mieux le mouvement respiratoire qui perturberait leur exploitation. Nous dédions une partie de ce chapitre au choix d'une méthode pour effectuer cette opération, rendue délicate pour les mêmes raisons que la segmentation (citées ci-dessus). Il ne s'agit pas de la partie la plus importante de notre travail, d'autres partenaires du projet étant davantage impliqués dans ce domaine. Cependant, pour des questions de temps, nous avons été amenés à développer notre propre méthode, puisque nous en avions besoin pour effectuer les segmentations. Nous terminons ce chapitre par une réflexion sur la validation des résultats sur les données réelles, par comparaison aux « pseudo vérités-terrain » que constituent des segmentations réalisées manuellement par des experts.

Le chapitre 5 est consacré aux méthodes de segmentation fonctionnelles testées et aux résultats obtenus, tant sur des données simulées que sur des données réelles. Nous proposons différentes variantes, dont nous étudions les avantages et les défauts.

Nous sommes conscients du fait qu'il existe certaines redites, en particulier entre le chapitre 3 d'une part, et les chapitres 4 et 5 d'autre part. Nous avons ainsi souhaité nous adapter à la sensibilité des différents lecteurs : dans le premier chapitre, la présentation est plutôt orientée vers les mathématiques, alors qu'elle l'est davantage vers l'application dans les deux suivants.

Le chapitre 6 tient lieu de conclusion. Dans l'annexe A se trouvent un ensemble de figures représentant les résultats de l'étape de recalage, tant sur des données simulées que des données réelles. Dans l'annexe B, nous reprenons quelques résultats qui permettent de comparer les méthodes d'optimisation de Newton ou de descente de gradient ordinaire ou stochastique. En effet, l'algorithme des K-moyennes peut être considéré comme une optimisation de Newton pour la fonction de coût, alors que le *Neural Gas*, voisin du *Growing Neural Gas*, consiste en une descente de gradient stochastique (cf. annexe C).

Chapitre 2

L'IRM de perfusion rénale

L'imagerie fonctionnelle rénale par IRM dynamique à rehaussement de contraste est actuellement considérée comme une possible alternative aux examens scintigraphiques pour étudier les pathologies rénales, en particulier pour évaluer le débit de filtration glomérulaire (*Glomerular filtration rate* ou GFR) et la fonction rénale différentielle. Elle a donné lieu à de nombreuses publications ces dix dernières années. Par rapport aux méthodes scintigraphiques, l'absence de risque d'exposition à un rayonnement radioactif, tant pour le patient que pour le personnel soignant, est un avantage certain. De plus, l'agent de contraste utilisé en IRM est moins allergène que les radiotraceurs. En outre, l'IRM pourrait présenter des possibilités supplémentaires qui sont encore en cours d'étude. Elle apporte en particulier des informations à la fois morphologiques et fonctionnelles ; il est certainement possible de faire une étude plus fine par compartiment, en raison d'une meilleure résolution spatiale [18], d'où l'idée d'étudier la possibilité de remplacer, à terme, les examens scintigraphiques par des IRM [14, 10].

Dans ce chapitre, nous présenterons les différentes notions qui permettent d'analyser les séquences d'images obtenues, en particulier les courbes temps-intensité qui peuvent en être déduites pour évaluer la fonction rénale. Nous commencerons par des éléments d'anatomie et d'histologie rénale, puis expliquerons brièvement le mécanisme de filtration des urines. Afin de montrer comment ce mécanisme peut être étudié par le biais de l'IRM, nous aborderons ensuite le principe de formation des images pour cette modalité d'imagerie, en insistant sur les éléments caractéristiques des séquences dynamiques à rehaussement de contraste. Le rôle particulier du produit de contraste, dont la séquence permet de suivre le devenir, sera mis en évidence.

2.1 Eléments d'anatomie et d'histologie du rein

Les reins sont les organes qui aident à maintenir une composition fixe du sang en en éliminant par filtration et par excrétions active les solutés qui y sont en sur-concentration. Le sous-produit de cette excrétion est l'urine. Ils sont en forme de haricot, et leur dimension approximative chez l'homme est d'environ 11 à 12 cm de long, 5 à 7 cm de large, et 2 à 3 cm d'épaisseur [19]. Leur structure interne est très complexe (cf. figure 2.1[1]). On peut

[1] http ://www.bartleby.com/107/illus1127.html

cependant distinguer trois parties principales, ou compartiments : le cortex, la médullaire et les cavités, représentés sur la figure 2.2[2]. Les cavités sont formées du pelvis et des calices rénaux, et constituent la partie supérieure des voies urinaires (dans la suite de notre étude, nous ne considérerons pas la totalité du pelvis mais nous nous limiterons à la partie incluse dans le sinus rénal).

Le cortex est situé en périphérie du rein, alors que la médullaire en est une partie interne, formée des pyramides dites de Malpighi dont les bases sont tournées vers la partie corticale [20]. Pour comprendre le fonctionnement du rein, il est nécessaire de décrire plus finement chacune de ces parties.

Le cortex et la médullaire contiennent des néphrons (de l'ordre d'un million), dans lesquels le sang est filtré, et du tissu interstitiel. Un néphron est représenté sur la figure 2.3 avec une partie du système vasculaire qui irrigue le rein : il comprend un glomérule et un tubule rénal, et se trouve à cheval sur les compartiments cortical et médullaire. Chaque glomérule contient un réseau de petites artères issues des artères rénales.

Filtration du sang et évacuation des urines

Le mécanisme de filtration du sang et d'évacuation des urines est décrit dans [21]. Le sang arrive par l'artère rénale jusqu'au réseau d'artérioles des glomérules. A ce niveau, de l'eau, les composants sanguins de petite taille contenus dans le plasma peuvent traverser une membrane, la capsule de Bowman, qui agit comme un filtre. L'urine primitive ainsi formée de l'autre côté est ensuite évacuée par les tubules rénaux, puis par les tubes collecteurs, où une partie de l'eau est réabsorbée et une régulation fine de la concentration sanguine de certains ions essentiels s'opère. L'urine provenant des pyramides de Malpighi est ensuite récupérée dans les calices puis évacuée vers l'uretère et la vessie.

Ce mécanisme peut être étudié et la fonction rénale évaluée à partir de modalités d'imagerie dites fonctionnelles, dont fait partie l'IRM dynamique à rehaussement de contraste.

2.2 L'IRM dynamique à rehaussement de contraste

L'imagerie fonctionnelle fournit des informations complémentaires à celles obtenues par l'imagerie morphologique classique : son utilisation est surtout très développée en imagerie cérébrale, et dans une moindre mesure, en imagerie cardiaque, pulmonaire, hépatique et rénale.

Il existe dans ce cadre plusieurs techniques visant à faciliter l'étude de la fonction rénale, comme l'angiographie par résonance magnétique (MRA), l'IRM de diffusion (*diffusion-weighted* MRI)... [22] L'IRM dynamique par rehaussement de contraste avec injection de chélates de gadolinium est l'une des plus couramment utilisées [23]. Nous commencerons par rappeler brièvement quelques principes de base de l'IRM avant de souligner les particularités de l'IRM dynamique par rehaussement de contraste. Nous commenterons alors les courbes temps-intensité typiques qui peuvent en être déduites.

[2] http ://fr.wikipedia.org/wiki/Fichier :Kidney_nephron.png

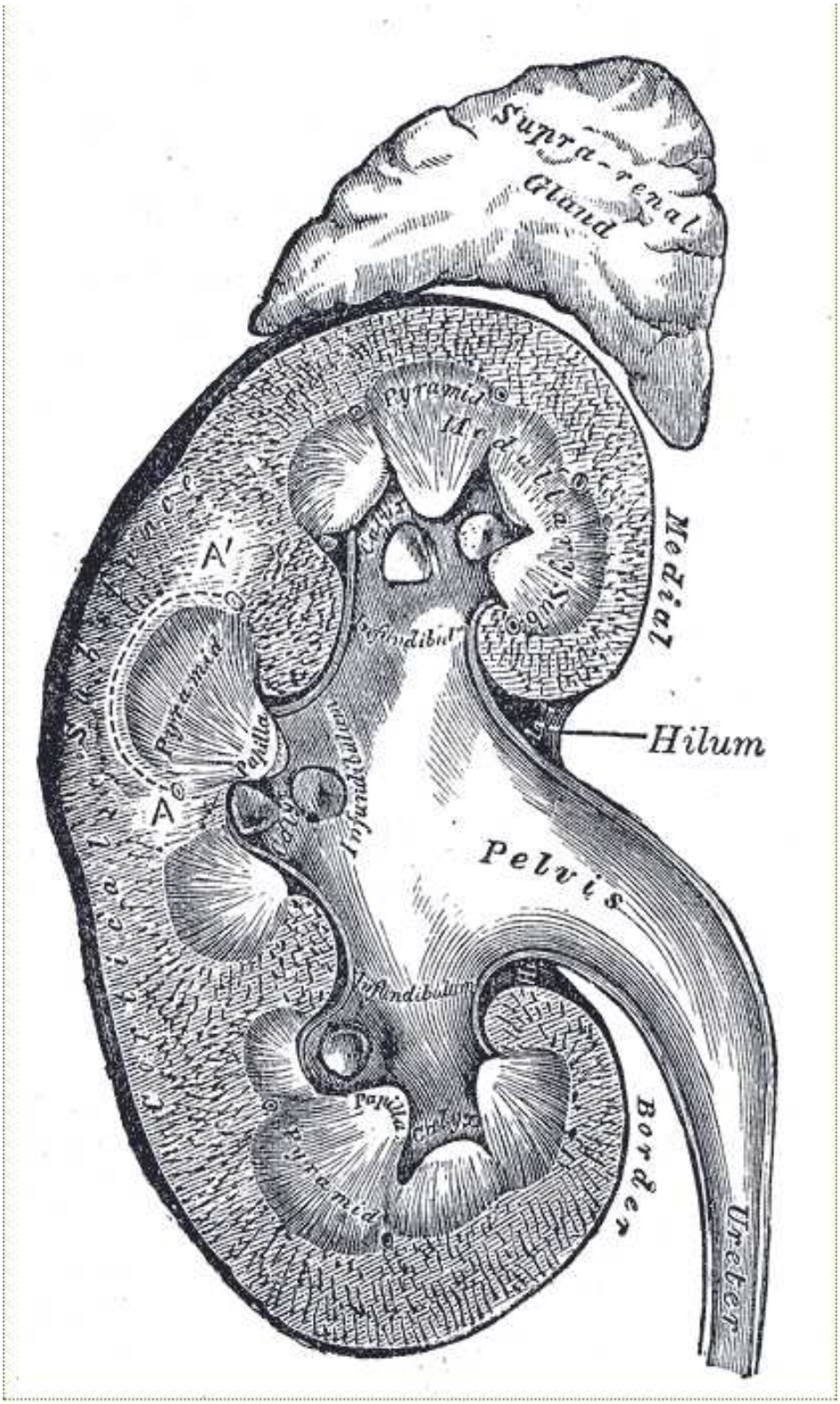
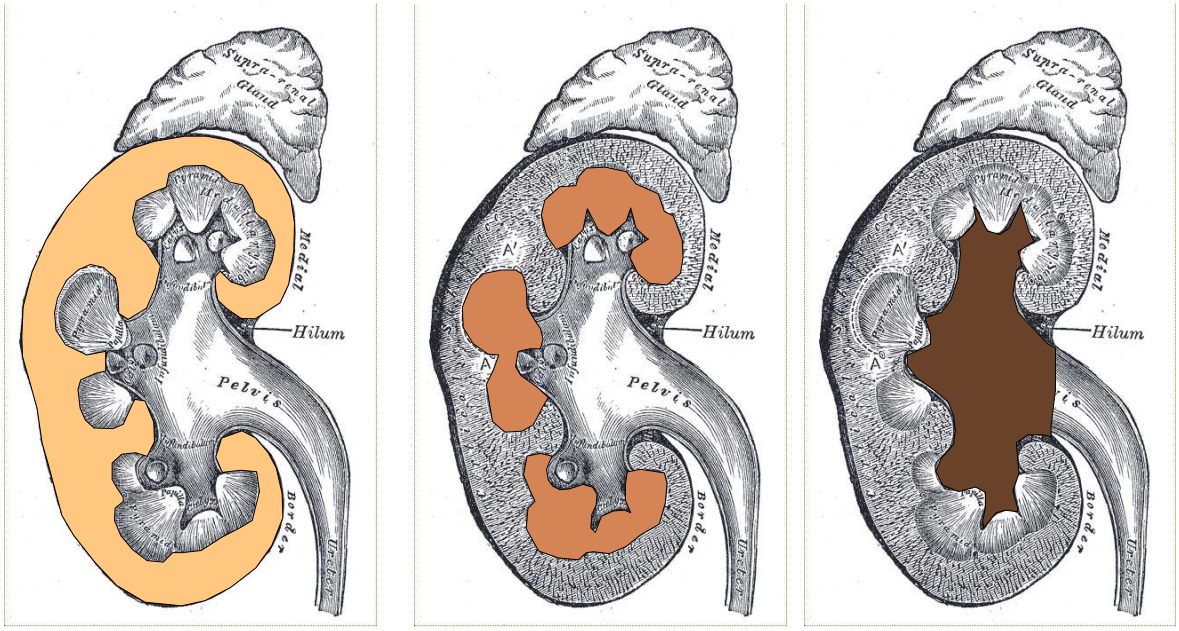


FIG. 2.1 – Anatomie du rein (Reproduction d'une lithographie extraite de Gray's Anatomy of the Human Body).



(a) (b) (c)

FIG. 2.2 – Anatomie du rein : cortex (a), médullaire (b) et cavités (c).

2.2.1 Formation des images IRM

Il est naturellement hors de propos de présenter l'intégralité des mécanismes complexes de formation d'une image en IRM, en particulier l'étude statistique des propriétés d'une population de spins ou le codage spatial (le lecteur se reportera par exemple à [11] pour davantage de détails). Nous exposerons uniquement quelques notions qui nous semblent utiles pour comprendre la particularité de l'IRM dynamique à rehaussement de contraste et les quelques termes couramment employés pour caractériser les séquences correspondantes.

Le signal acquis en IRM est essentiellement émis par les protons des molécules d'eau placées dans un champ magnétique adéquat. Dans le cas où ces protons sont soumis à un champ magnétique $\mathbf{B_0}$ constant (classiquement de $1,5$ T en IRM), leurs spins sont à l'origine d'une aimantation tissulaire macroscopique correspondant à un état d'équilibre statistique. Leur excitation par une onde radiofréquence à la fréquence dite de Larmor, qui dépend du champ $\mathbf{B_0}$, provoque un basculement de cette aimantation, d'un angle θ qui dépend de l'intensité et de la durée de l'onde radiofréquence (voir figure 2.4). Quand l'émission radiofréquence est interrompue, débute une phase de relaxation ramenant progressivement l'ensemble des spins à l'état d'équilibre macroscopique initial.

L'aimantation macroscopique peut s'écrire comme la somme d'une aimantation longitudinale parallèle au champ $\mathbf{B_0}$ et d'une composante transversale qui lui est perpendiculaire.

$$\mathbf{M} = \mathbf{M_r} + \mathbf{M_z}. \tag{2.1}$$

Son évolution peut être ainsi également décomposée en deux phénomènes distincts, correspondant à des mécanismes microscopiques différents. La phase de basculement se traduit

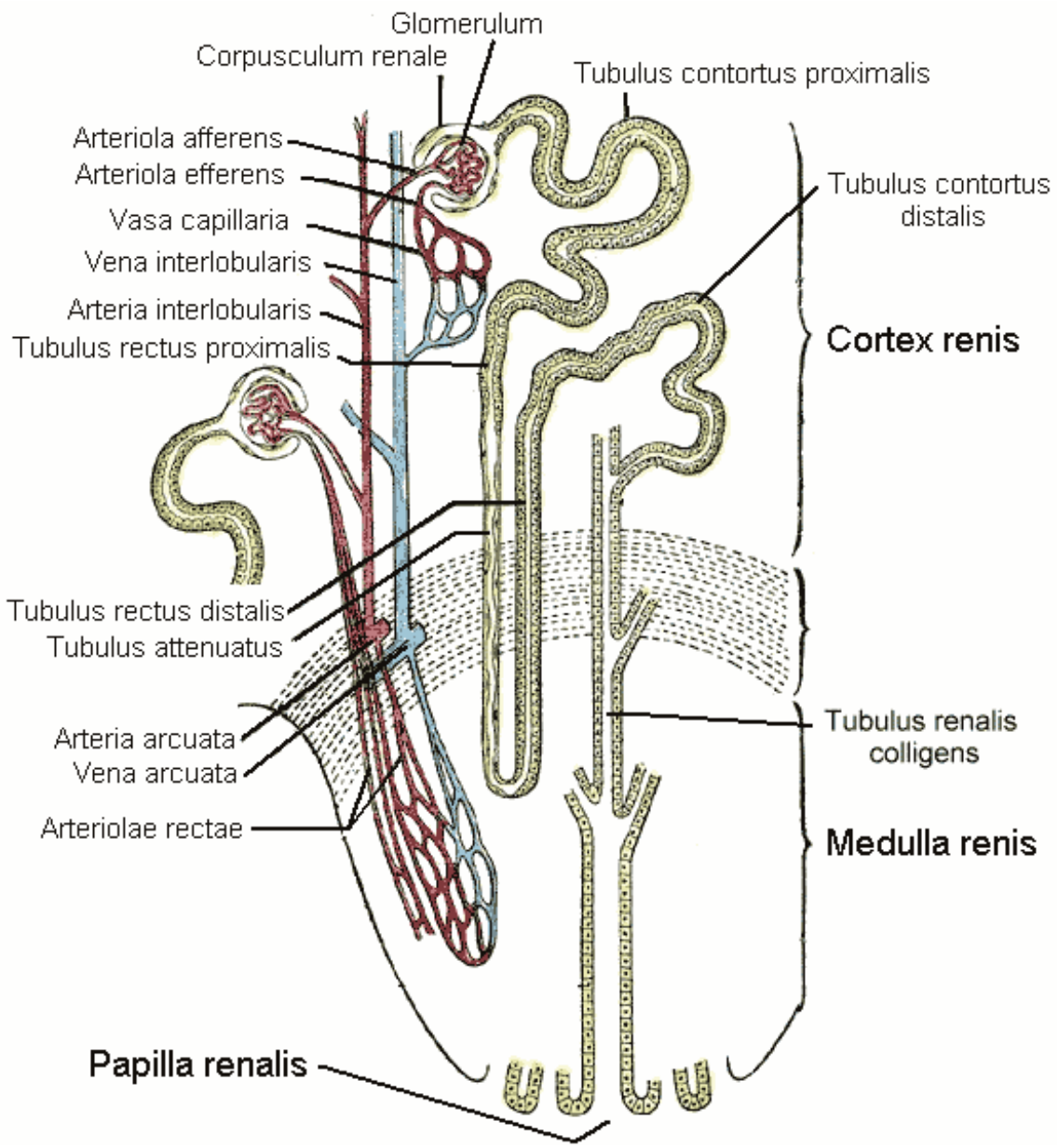


FIG. 2.3 – Structure d'un néphron (Reproduction d'une lithographie extraite de Gray's Anatomy of the Human Body).

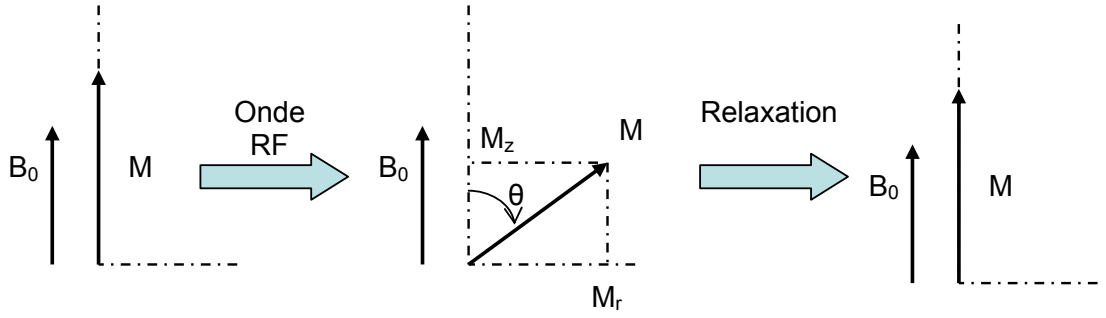


FIG. 2.4 – Evolution de l'aimantation macroscopique d'un ensemble de spins placés dans un champ magnétique et soumis à une onde radiofréquence à la fréquence de Larmor.

par une augmentation de la norme de $\mathbf{M_r}$, et une diminution de celle de $\mathbf{M_z}$ (sauf pour un angle de 180°). Lors de la phase de retour à l'équilibre, la repousse de $\mathbf{M_z}$ et la décroissance de $\mathbf{M_r}$ se font selon des évolutions exponentielles à des vitesses différentes, qui sont représentées respectivement sur les figures 2.5 et 2.6, pour un angle de bascule de 90°. Les ordres de grandeur des temps en abscisses sont tout à fait typiques. Le temps T_1 est par définition le temps nécessaire pour que l'aimantation longitudinale reprenne 63 % de sa valeur finale. Le temps T_2 est le temps nécessaire pour que l'aimantation transversale revienne à 37 % de sa valeur initiale. Ces constantes de temps sont caractéristiques d'un tissu. T_2 est toujours inférieur à T_1, qui varie lui-même entre 200 ms et 3 s.

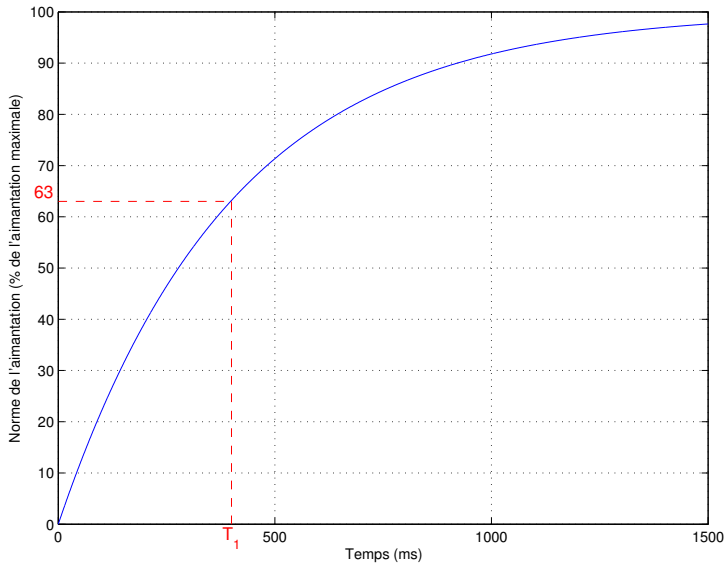


FIG. 2.5 – Relaxation de la composante longitudinale M_z de l'aimantation macroscopique pour un angle de bascule de 90° ; définition de la constante T_1.

Le signal est recueilli pendant la phase de relaxation. Le mode d'acquisition en IRM est cependant très particulier dans la mesure où c'est la transformée de Fourier de l'image finale qui est acquise (données dites brutes), et non directement l'image elle-même, qui

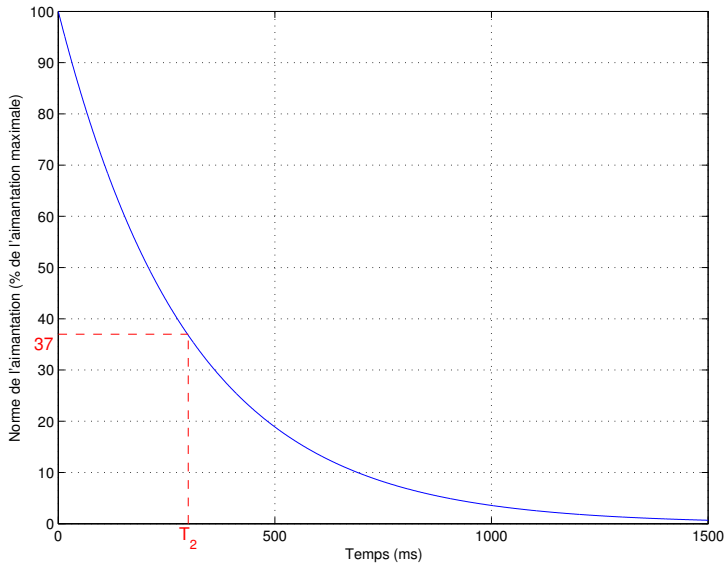


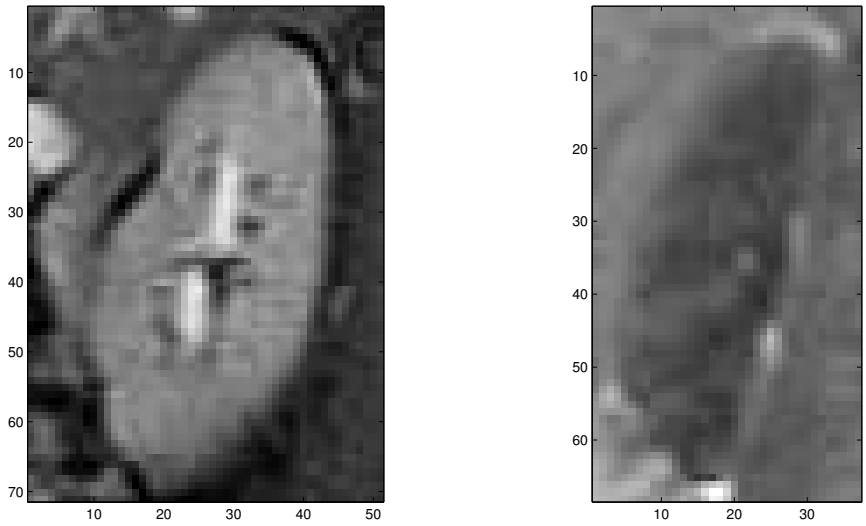
FIG. 2.6 – Relaxation de la composante transversale M_r de l'aimantation macroscopique pour un angle de bascule de 90° ; définition de la constante T_2.

est reconstruite par la suite grâce à une transformée de Fourier inverse. L'image finale est en général le carré du module de cette transformée inverse. Les données brutes sont acquises dans un espace appelé couramment l'espace K, qui est donc un plan de Fourier. Le mode de remplissage de l'espace K détermine en particulier la résolution spatiale de l'image finale.

Pour acquérir les données permettant de reconstruire une image complète, il est nécessaire d'envoyer plusieurs impulsions successives, correspondant à un angle de bascule de l'aimantation donné, et permettant chacune de remplir une partie de l'espace K. Ces impulsions sont séparées par un intervalle de temps appelé temps de répétition T_R. Le temps d'écho T_E est le temps séparant une impulsion de l'acquisition partielle correspondante. Ces temps dépendent des réglages effectués par l'opérateur, et en particulier du type de séquence utilisé. En ajustant convenablement l'angle de bascule, les rapports entre ces temps T_E, T_R, il est possible d'obtenir, pour des tissus différents, des signaux d'amplitudes différentes qui dépendent de ces temps. Il existe par exemple les images dites pondérées T_1 ou pondérées T_2. Certains tissus ne peuvent être différenciés qu'avec certains types de pondération. Sur la figure 2.7(a), différentes parties du rein apparaissent, alors qu'on ne peut les distinguer sur la figure 2.7(b).

Séquences utilisées

Les séquences utilisées dans le cadre de notre étude sont de type écho de gradient ultrarapides, pondérées T_1. L'angle de bascule de l'aimantation est faible (de l'ordre de 10°), ce qui permet d'aboutir à un retour à l'équilibre plus rapide et à des temps de répétition et d'écho courts, donc à des résolutions temporelles assez élevées : le temps d'acquisition pour une coupe est de l'ordre de la seconde. En revanche, la résolution spatiale est assez limitée.



(a) (b)

FIG. 2.7 – Images IRM de reins en séquence FIESTA et LAVA (GE).

2.2.2 Cas particulier de l'IRM dynamique à rehaussement de contraste : rôle du produit de contraste

L'IRM dynamique à rehaussement de contraste permet de suivre le devenir d'un produit de contraste injecté dans l'organisme. Le produit n'est pas lui-même visible directement, mais modifie les temps de relaxation des protons des molécules d'eau voisines, donc le contraste des tissus avoisinants. Il existe deux grands types d'agents de contraste [24] :

1. les chélates de gadolinium, paramagnétiques, qui raccourcissent le temps T_1 (aux faibles concentrations utilisées en pratique clinique, l'effet T_2 est généralement considéré comme négligeable devant l'effet T_1 [17], donc qui augmentent le signal en pondération T_1 [25]
2. les produits à base d'oxyde de fer super-paramagnétiques (SPIO et USPIO), qui diminuent le T_2, donc affaiblissent le signal en pondération T_2.

Les séries d'images que nous utiliserons pour nos tests seront réalisées après injection de chélates de gadolinium. Les acquisitions sont répétées sur plusieurs minutes pour obtenir une séquence complète. Trois images issues d'une même séquence sont présentées figure 2.8, mettant en évidence la modification d'intensité due au produit de contraste.

Lien entre la concentration d'agent de contraste et le signal IRM Pour quantifier les paramètres fonctionnels comme le GFR, il est nécessaire d'estimer les modifications temporelles de la concentration d'agent de contraste à partir des images de la séquence IRM. Si la relation entre la concentration [Gd] de chélate de gadolinium et l'intensité sur l'image n'est pas linéaire, celle entre la concentration et le taux de relaxation en T_1, qui est l'inverse du temps de relaxation T_1, l'est [23] :

$$R_1 = R_1^0 + r[\mathrm{Gd}], \tag{2.2}$$

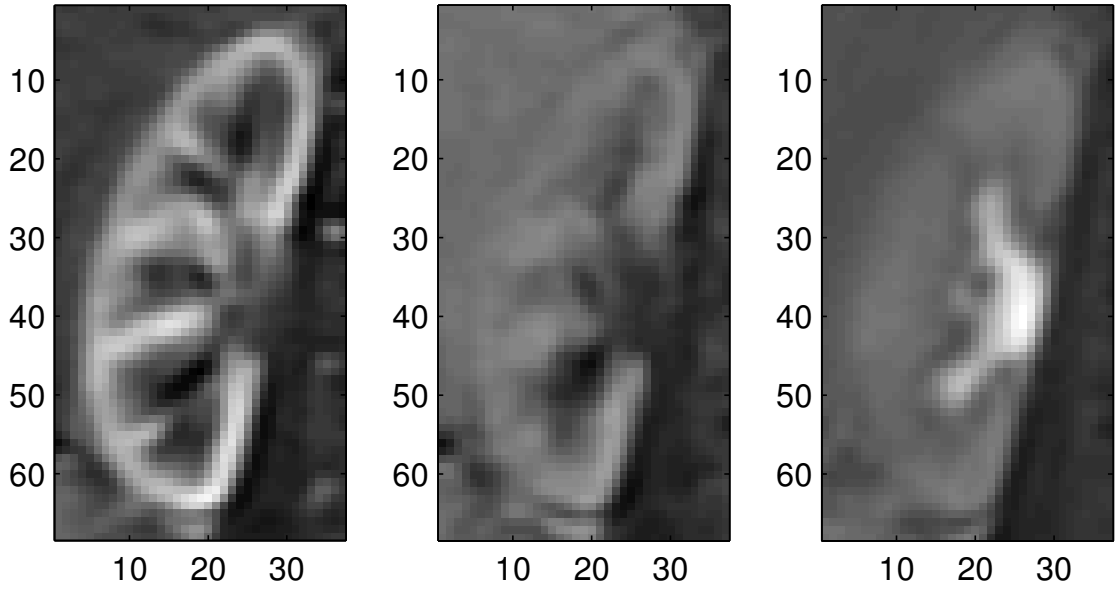


FIG. 2.8 – Exemples d'images extraites d'une séquence d'IRM dynamique au pic artériel (à gauche), pendant la filtration (au milieu) et la phase tardive (à droite).

en notant R_1^0 le taux de relaxation T_1 dans le tissu en l'absence d'agent de contraste, et r la relaxivité de cet agent. La relaxivité représente le changement du taux de relaxation T_1 pour une concentration unitaire et caractérise donc la capacité de l'agent à modifier le signal émis par les spins qui l'entourent. Elle dépend du champ magnétique externe et de la température. La conversion entre intensité du signal et concentration [Gd] est donc possible. On peut

- soit acquérir des images de tubes contenant des solutions de chélates de gadolinium de différentes concentrations connues avec la même séquence que celle utilisée pour le patient, puis tracer une courbe de calibration. Ceci suppose cependant que la relaxivité de l'agent dans les solutions et *in vivo* soient identiques.
- soit utiliser directement la relation entre l'intensité sur l'image et R_1, qui dépend de la séquence utilisée ; la conversion peut s'avérer complexe.

Propriétés du chélate de gadolinium Le chélate de gadolinium est filtré au niveau du glomérule sans sécrétion tubulaire ni réabsorption (cf. paragraphe 2.1 page 29) et peut donc être considéré comme un traceur glomérulaire. Son trajet explique l'allure des courbes temps-intensité typiques par compartiment décrites au paragraphe 2.2.3 page 37, qui peuvent être utilisées pour déterminer le débit de filtration glomérulaire.

2.2.3 Courbes temps-intensité typiques pour le rein sain

Dans le paragraphe 2.1 page 29, nous avons défini une structure anatomique comprenant trois parties, ou compartiments : le cortex, la médullaire et les cavités. Une courbe temps-intensité moyenne pour chaque partie peut être déduite d'une séquence : cette courbe traduit l'évolution de la concentration du produit de contraste dans chacun de

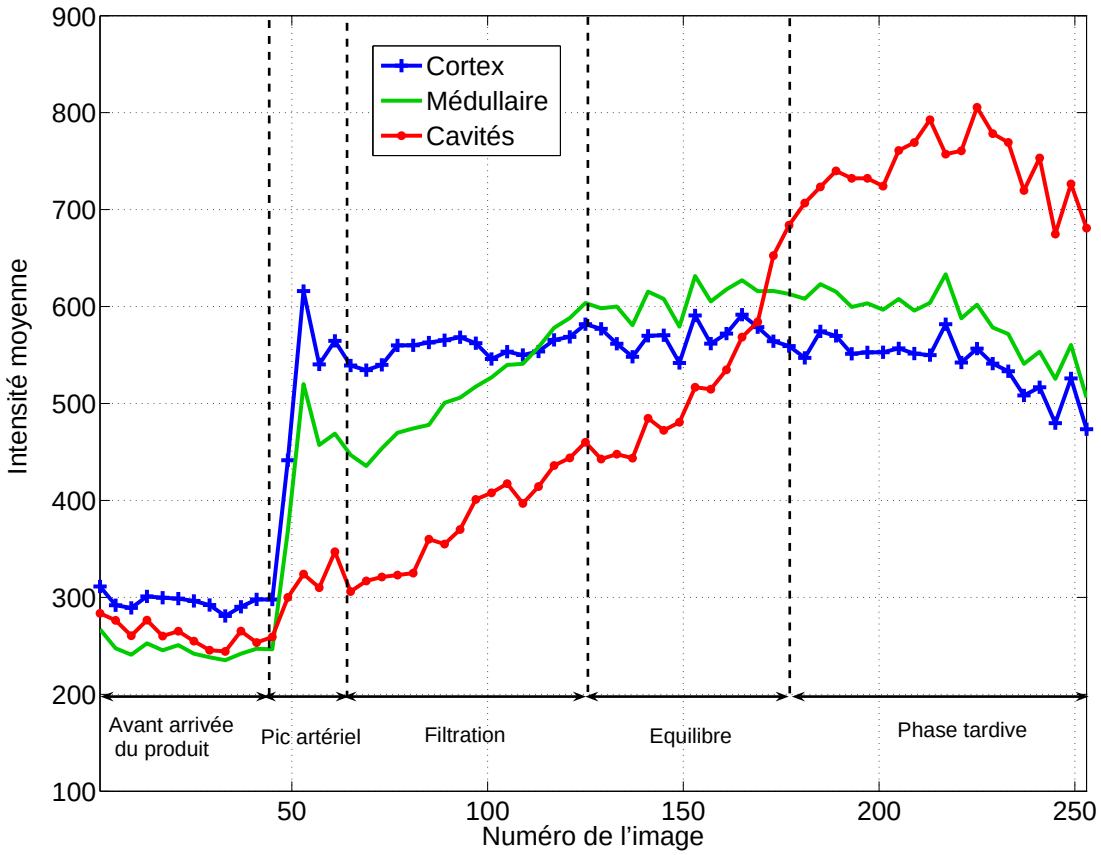


FIG. 2.9 – Exemple de courbes temps-intensité typiques pour le cortex, la médullaire et les cavités.

ces trois compartiments. Cinq phases successives, reliées au mécanisme de filtration et d'évacuation des urines décrit ci-dessus, peuvent être observées sur la figure 2.9[3] :

1. Baseline (environ 60 s) : l'agent de contraste n'a pas encore atteint la zone rénale ; le cortex possède un niveau de signal plus élevé car il contient une quantité d'eau plus importante ; la médullaire et les cavités ont des signaux assez comparables de sorte qu'on ne peut les distinguer durant cette phase.
2. Pic artériel (environ 15 s) : ce moment correspond à l'arrivée du produit de contraste par les vaisseaux sanguins. Le cortex présente théoriquement le pic artériel de plus grande amplitude, mais en raison de la vascularisation importante de l'ensemble du rein, des pics apparaissent également dans les autres compartiments.
3. Filtration (environ 100 s) :
   – dans le cortex, le signal reste à un niveau élevé ; une légère montée peut éventuellement être observée.
   – le signal de la médullaire augmente également avec un léger retard par rapport au cortex, la pente de la courbe en étant plus raide,
   – le signal des cavités augmente un peu.
4. Équilibre (environ 150 s) : les signaux du cortex et de la médullaire deviennent difficiles à distinguer, les cavités sont encore peu rehaussées.
5. Phase tardive (environ 5 min) : l'urine est évacuée ; alors que les signaux du cortex et de la médullaire restent semblables et diminuent légèrement, les cavités se remplissent et l'intensité du signal correspondant devient très élevée.

2.3 Modèles pour l'évaluation quantitative de la fonction rénale

Il existe dans la littérature plusieurs types de modèles utiles à l'évaluation de la fonction rénale. Leur exploitation nécessite souvent l'utilisation de courbes de perfusion [26]. Ils sont souvent inspirés de ceux utilisés pour l'analyse des scintigraphies. Voir aussi [25].

2.3.1 Représentations basées sur l'évolution du contraste

Des paramètres comme l'intensité maximale, l'intervalle de temps entre l'injection du produit de contraste et le pic artériel, la pente et l'aire sous la courbe permettent d'évaluer la fonction rénale de manière semi-quantitative (cf. par exemple [27, 28]). Ils sont essentiellement utilisés pour l'évaluation de la fonction rénale différentielle [29] pour les patients souffrant d'une pathologie unilatérale, ce qui entre dans le cadre du projet du CHU.

[3]Remarquons que l'échelle des temps est graduée en numéro d'images, et correspond donc à un temps réel non linéaire, les acquisitions dans la deuxième moitié étant plus espacées que dans la première

2.3.2 Représentations externes

Le rein est considéré comme un système avec une ou plusieurs entrées (dose de produit de contraste injecté, fonction d'entrée artérielle...) et des sorties (quantité d'agent de contraste éliminé, courbes de perfusion...). Plusieurs méthodes permettent d'évaluer les taux de perfusion et de filtration glomérulaire avec ce modèle. Il est également possible d'estimer la réponse impulsionnelle du système, voire d'identifier les paramètres de ce modèle, mais il est difficile de choisir *a priori* une structure de modèle adaptée.

2.3.3 Représentations internes

Dans ce type de représentations, le rein est considéré comme un ensemble de compartiments. Malgré l'arrangement complexe des néphrons et des vaisseaux dans le rein, on ne considère généralement qu'un nombre assez limité de compartiments (cf. [30]). Par exemple, pour un modèle à deux compartiments, le rein est supposé contenir un compartiment vasculaire et un compartiment tubulaire, et le traceur pénètre par l'artère rénale et est évacué par filtration glomérulaire. Ces modèles simples sont basés sur une cinétique du premier ordre. Les flux sont unidirectionnels, et les taux qui décrivent le transport du produit de contraste sont considérés comme constants. Ces modèles ne sont ainsi valables que pendant une certaine phase de la perfusion dont les limites ne sont pas toujours aisées à déterminer. Le modèle de Patlak-Rutland [31] par exemple, qui est l'un des plus couramment utilisés, entre dans cette catégorie : notons respectivement $K(t)$ et $b(t)$ la quantité de gadolinium dans le rein et la concentration de gadolinium dans l'aorte à l'instant t. Supposons que

- la quantité de gadolinium dans l'espace vasculaire est proportionnelle à $b(t)$,
- la quantité de gadolinium filtrée dans le néphron à l'intégrale de la courbe de concentration de gadolinium $b(t)$ dans l'aorte,

alors on montre que l'égalité suivante, dite équation du tracé de Patlak-Rutland, est vérifiée :

$$\frac{K(t)}{b(t)} = \alpha + \mathrm{GFR} \times \frac{\int_0^t b(x)\mathrm{d}x}{b(t)}. \tag{2.3}$$

Le GFR est donc la pente de la droite obtenue en traçant $\frac{K(t)}{b(t)}$ en fonction de $\frac{\int_0^t b(x)\mathrm{d}x}{b(t)}$, et peut être évalué par une méthode des moindres carrés par exemple. D'autres modèles sont décrits dans [32].

Les modèles comprenant un nombre élevé de compartiments sont plus rarement utilisés dans ce cadre [17]. Notons que l'utilisation de nombre de ces modèles nécessite la connaissance de la fonction d'entrée artérielle (*arterial input fonction* ou AIF), ce qui est vrai aussi pour les modèles utilisés en scintigraphie.

2.3.4 Conclusion

Les modèles décrits ci-dessus peuvent selon les cas être appliqués au rein entier, ou seulement au parenchyme, soit dans son intégralité, soit en distinguant cortex et médullaire, ou bien être adaptés seulement au cortex [25]. Ils doivent encore être validés sur des populations suffisamment nombreuses. Leur exploitation à partir des séries d'images

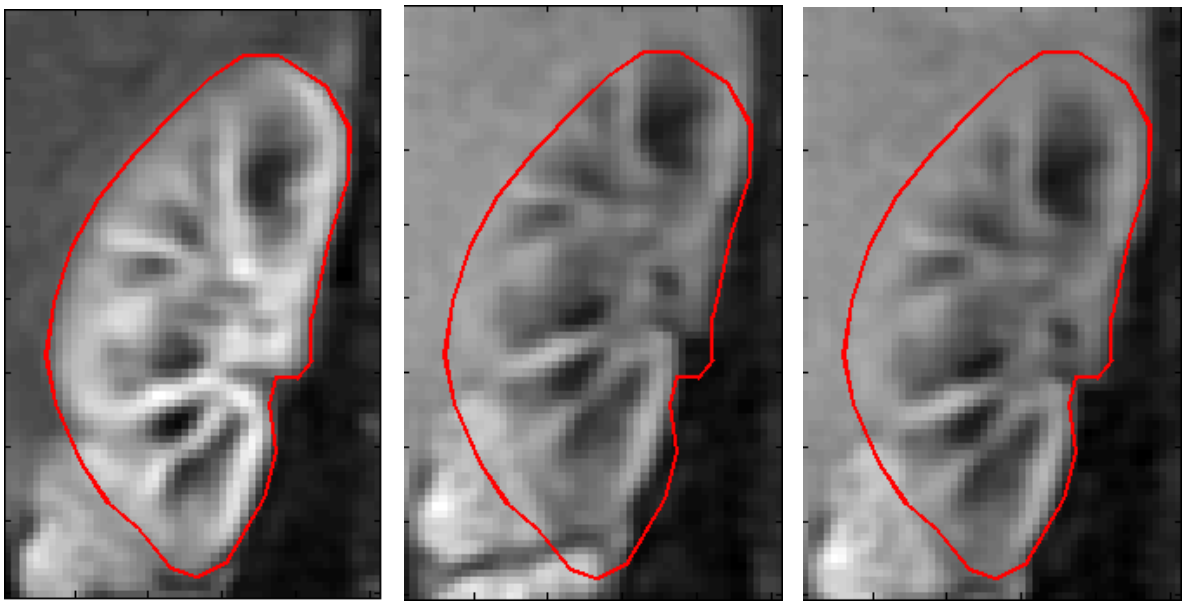
d'IRM de perfusion acquises nécessite en particulier un recalage des images pour éliminer le mouvement respiratoire et une segmentation des structures rénales internes (cortex, médullaire et cavités). Ce travail est long et fastidieux s'il est effectué « manuellement » et rend difficile une étude sur une population nombreuse. C'est pourquoi nous nous proposons d'automatiser en partie ces étapes. Nous commençons par dresser un état de l'art sur le recalage et sur les méthodes de segmentation dans les paragraphes 2.4 et 2.5.

2.4 Pré-traitement : recalage de séquences d'IRM de perfusion rénale

Après avoir expliqué pourquoi un recalage s'avère nécessaire sur les séries d'acquisitions exploitées, nous ferons un point sur les méthodes existantes et choisirons celles qui nous paraissent les mieux adaptées à notre problème, selon les contraintes spécifiques à nos images. Nous n'avons pas développé de nouvelle méthode. Des tests sur données réelles permettront de valider notre choix et d'avoir une idée des erreurs qui peuvent subsister.

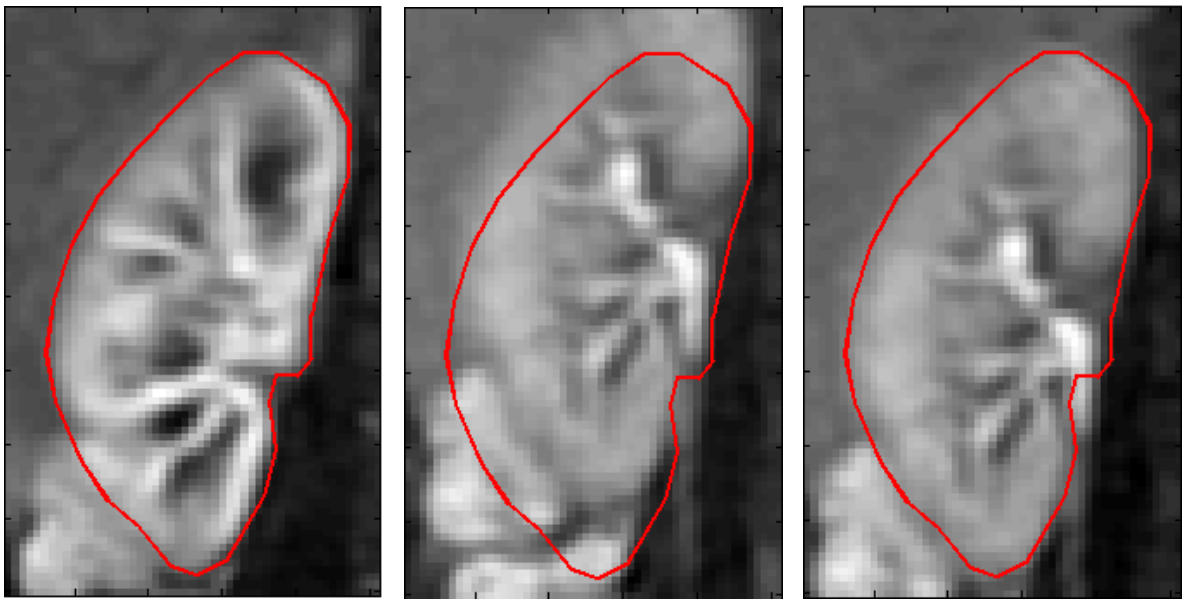
2.4.1 Nécessité d'un recalage

L'étude de l'évolution temporelle par voxel ou par compartiment nécessite, pour qu'elle soit cohérente, que chaque voxel corresponde tout au long de l'acquisition à la même zone anatomique. Or, les mouvements du patient conduisent à des situations semblables à celle présentée sur la figure 2.10(a) et (b), où la région d'intérêt entourée par le trait rouge matérialise le rein entier (délimité sur l'image (a) et reporté sur les autres).



(a) (b) (c)

FIG. 2.10 – Exemples d'images extraites d'une séquence d'IRM dynamique : image de référence (a), image suivante avant recalage (b) et après recalage (c).



(a) (b) (c)

FIG. 2.11 – Exemples d'images extraites d'une séquence d'IRM dynamique : image de référence (a), image en phase tardive avant recalage (b) et après recalage (c).

Différents types de mouvements peuvent être distingués :

- les mouvements respiratoires, qui restent à peu près dans le plan de coupe en raison de l'inclinaison adéquate choisie pour l'acquisition. Ils peuvent être corrigés dans une certaine mesure.
- les mouvements intempestifs, qui impliquent souvent un changement de plan de coupe et qui sont difficilement corrigibles, même en se servant éventuellement des autres coupes, à cause de la faible résolution dans la direction perpendiculaire aux coupes.

Notons que les mouvements sont de plus grande amplitude qu'en IRM de perfusion cérébrale. Par ailleurs, l'examen ne peut être réalisé en apnée puisqu'il dure une dizaine de minutes ; il n'est pas non plus possible de synchroniser l'acquisition avec la respiration car cela diminuerait beaucoup trop la résolution temporelle, par rapport à la rapidité des phénomènes d'intérêt.

Nous essaierons donc de corriger au mieux les mouvements respitoires, pour obtenir par exemple, à partir des images (a) et (b) de la figure 2.10 ou 2.11, l'image (c) de ces mêmes figures. Nous supposerons que les images avec mouvements intempestifs de grande ampleur ont été éliminées de la série, ce qui se fait très aisément par simple visualisation de la « vidéo » correspondante. Nous souhaitons qu'il n'y ait pas d'autre intervention de la part d'un opérateur.

Difficultés particulières Notons que, pour les séquences d'IRM de perfusion à rehaussement de contraste, les modifications locales d'intensité entre deux images peuvent avoir quatre causes principales [33] :

– un changement rapide et non uniforme du contraste à l'intérieur des différents compartiments rénaux lors du passage du bolus[4],
– le bruit d'acquisition, sachant qu'il est fréquent d'avoir un faible rapport signal à bruit avec les séquences ultra-rapides utilisées,
– un mouvement résiduel hors plan de coupe,
– un mouvement dans le plan de coupe qui subsisterait après le recalage.

Seules les deux dernières devraient donner lieu à une correction, mais la difficulté à les distinguer des deux autres expliquera certains de nos choix. Nous devrons également tenir compte du fait que l'on cherche à recaler une série entière (256 images pour nos données réelles) et pas seulement deux images. Par ailleurs se pose la question de la validation en l'absence de vérité-terrain.

2.4.2 Méthodes de recalage

Une étude sur les méthodes spécifiquement utilisées pour recalage d'images médicales est faite dans [34]. Il s'agit de trouver la transformation optimale à appliquer à l'image à recaler I_T pour qu'elle ressemble le plus possible à une image de référence I_R. Dans ce paragraphe, parallèlement à une présentation générale des méthodes de recalage, nous donnons des éléments pour choisir une méthode adéquate dans notre cas. Deux questions principales se posent :

– Quel type de transformation autoriser ?
– Comment mesurer la ressemblance entre deux images ?

Nous ferons aussi au fur et à mesure un bilan des méthodes qui ont été spécifiquement appliquées à l'IRM de perfusion rénale.

Transformations autorisées

Dans notre cas, toutes les transformations envisagées sont bidimensionnelles dans le plan de coupe. On peut les classer en trois grandes catégories :

1. les transformations globales rigides, qui sont des compositions de translations et de rotations,
2. les transformations globales non rigides, qui incluent les transformations affines, avec contraintes éventuelles comme la dilatation pure ou au contraire avec conservation des surfaces,
3. les transformations locales élastiques,
4. les transformations de type fluide pour les grandes déformations [35].

Il est aussi possible de traiter d'abord l'image par morceaux en effectuant un recalage rigide de chaque bloc (*block matching*) puis en cherchant une transformation non rigide lissée à partir de l'ensemble des transformations trouvées précédemment.

Pour l'IRM de perfusion rénale, le rein est la plupart du temps considéré comme un organe rigide subissant uniquement des translations [36, 37], combinées éventuellement avec une rotation [38, 39, 40, 33]. Dans [41], une translation verticale uniforme pour tous

[4]Un bolus est l'injection rapide d'une dose élevée d'une substance dans un vaisseau sanguin.

les pixels d'une ligne, mais variant d'une ligne à l'autre, est recherchée : si une amélioration du recalage est observée pour des IRM cardiaques, on ne trouve pas de conclusion pour l'IRM rénale. Dans [42, 43], des transformations affines sont envisagées, mais il est difficile de mesurer leur apport par rapport à des transformations rigides. Dans [44] cependant, les résultats de recalages rigides et non rigides sont comparés : la conclusion est que le recalage non rigide ne semble pas apporter d'amélioration sensible.

L'observation attentive des séquences tests dont nous disposons, qui concernent des reins sains, laisse effectivement supposer que le mouvement du rein est essentiellement une translation rigide, limitée à une quinzaine de pixels dans la direction verticale, à trois pixels dans la direction horizontale. De légères rotations peuvent être remarquées, ainsi que de faibles déformations.

En ce qui concerne la précision des transformations, les translations recherchées peuvent être au pixel près [43, 36, 37] ou subpixel [38, 40, 39, 41, 33]. Les données concernant les rotations sont plus vagues : elles sont au degré près dans [33]. Il faut cependant noter que, très souvent, seules les translations au pixel près sont validées (cf. paragraphe 2.4.4). Quant aux transformations élastiques, nous n'avons pas trouvé de renseignement sur la précision ou sur les contraintes dans la littérature.

Mesure de ressemblance

On peut distinguer les mesures de ressemblance basées sur la comparaison de points remarquables et les mesures de similarité entre pixels.

Comparaison de points de référence La méthode de recalage associée consiste à aligner ces points et à minimiser leur distance euclidienne entre leur position dans l'image à recaler et dans l'image de référence. Elle nécessite un pré-traitement pour définir ces points.

Alignement de points remarquables anatomiques Si cette méthode est largement utilisées pour d'autres organes (cerveau, poumon, cœur ou genou), de tels points sont difficiles à définir sur le rein.

Alignement de contours Dans [39, 38], une détection de contours est réalisée grâce à une transformation en ondelettes. Cependant, les organes voisins (foie, rate) ont des évolutions de contraste semblables à celle du rein, de sorte que les contours externes de ce dernier sont peu marqués (cf. figure 2.10(c)), alors que des structures internes apparaissent ou disparaissent pendant certaines phases de la perfusion (cf. figure 2.10(a) et (c)). C'est pourquoi les contours restent d'une part difficiles à extraire d'une bonne partie des images de toutes les séries dont nous disposons. D'autre part, il y a un risque non négligeable d'aligner des contours qui en réalité ne se correspondent pas, étant issus d'organes différents. Cette méthode de comparaison ne semble pas véritablement adaptée à notre problème.

Similarité des pixels La mesure de ressemblance s'appuie sur l'ensemble de la zone à recaler : il peut s'agir d'un rectangle comprenant le rein ; cependant, dans la presque

totalité des publications citées ci-dessus, à l'exception de [42], un masque global du rein est d'abord tracé manuellement, et les organes voisins ne sont alors pas pris en compte. Dans [44], des essais sont faits avec ou sans masque, les résultats étant considérablement meilleurs dans le premier cas.

Considérons des pixels censés représenter les mêmes points anatomiques sur l'image de référence X et sur l'image à aligner Y. Deux hypothèses peuvent être faites.

– On suppose que ces pixels sont de même intensité (cf. figure 2.10) : la ressemblance peut alors se mesurer avec des moindres carrés F ou par la fonction de corrélation γ. En notant x_{ij} les niveaux de gris des pixels de X et y_{ij} ceux de Y, ces deux fonctions s'écrivent respectivement

$$F(X,Y) = \sum_{i,j} (x_{ij} - y_{ij})^2 \tag{2.4}$$

et

$$\gamma(X,Y) = \sum_{i,j} x_{ij} y_{ij}. \tag{2.5}$$

En IRM rénale avec injection de produit de contraste, les changements d'intensité souvent très brusques font que cette hypothèse peut être valable à relativement court terme, mais il arrive qu'elle ne soit pas vérifiée, même pour deux images successives, dans certaines phases de perfusion. Une fonction d'énergie dont l'un des termes correspond à la somme des carrés de la différence d'intensité sur tous les pixels entre l'image de référence et l'image recalée est minimisée dans [41], les autres termes pouvant être considérés comme des termes de régularisation imposant des contraintes de lissage temporel sur l'ensemble de la série à recaler.

– On suppose qu'il existe une dépendance entre les intensités des pixels correspondants, mais on ne la spécifie pas : la similarité entre les deux images peut se mesurer par l'information mutuelle (cf. figure 2.11). Cette métrique semble mieux adaptée à notre cas que la précédente.

Information mutuelle L'information mutuelle est une notion issue de la théorie de l'information [45]. Elle permet de mesurer des relations statistiques entre deux images A et B sans spécifier la dépendance entre leurs intensités. Chaque pixel de l'image A (respectivement B) est considéré comme la réalisation d'une variable aléatoire de densité de probabilité $p(a)$ (respectivement $p(b)$). Notons $p(a,b)$ la probabilité jointe de ces deux variables aléatoire, c'est-à-dire la probabilité de l'événement « un pixel de l'image A a un niveau de gris a et le pixel correspondant sur l'image B a un niveau de gris b ». L'information mutuelle peut être définie de la manière suivante :

$$I(A,B) = \sum_a \sum_b p(a,b) \log \frac{p(a,b)}{p(a)p(b)}. \tag{2.6}$$

On peut la considérer comme une mesure de la distance entre la distribution jointe des niveaux de gris des deux images, $p(a,b)$, et la distribution jointe dans le cas où les deux images sont indépendantes (dans ce dernier cas, $p(a,b) = p(a)p(b)$). Elle est supposée

maximale quand les deux images sont convenablement recalées. L'implantation des méthodes de recalage basées sur l'information mutuelle nécessite en particulier de choisir une méthode d'interpolation pour estimer les niveaux de gris de l'image transformée, et la densité de probabilité jointe de l'image à recaler et de l'image de référence.

Une variante de l'information mutuelle est la mesure de similarité de points (*point similarity measure*) ; elle nécessite également l'évaluation de densités de probabilité. Bien que cette mesure ait plutôt été développée pour le recalage non rigide, elle peut aussi être utilisée pour les transformations rigides. Elle est décrite en détails dans [46] et est appliquée à l'IRM de perfusion rénale dans [44].

Comparaison des transformées de Fourier Dans [36], un recalage en translation au pixel près est effectué en se basant sur l'information tirée de la différence des phases des transformées de Fourier des images. Dans [38], la rotation est également évaluée, par minimisation d'une fonction d'énergie faisant intervenir la différence d'amplitude des transformées de Fourier. Dans [39, 40], une méthode semblable est appliquée après un prétraitement de détection de contours. D'après [33], ces méthodes semblent plus robustes que celles basées sur les niveaux de gris pour estimer les translations, mais la précision obtenue pour les rotations demeure insuffisante.

Mesure de similarité entre les gradients des images Dans [41, 37, 33], on utilise un critère de similarité entre les gradients de l'image à recaler et de l'image de référence, qui tient compte à la fois des contours et de l'intensité de leurs zones limitrophes. Ce critère est indépendant des changements de contraste brusques, au sens où l'orientation des contours reste à peu près la même pendant la perfusion.

2.4.3 Recalage de la série complète

En IRM de perfusion rénale, il s'agit de recaler une série complète d'images (256 dans notre cas), et non seulement deux images. On peut imaginer de recaler toutes les images de la série sur une image de référence ; c'est la technique adoptée dans [42, 33, 43, 44]. Le principal désavantage est que plus les images sont éloignées temporellement, moins elles se ressemblent (cf. figure 2.11(a) et (c)). Cependant, le fait qu'une image de la série ne soit pas convenablement recalée ne fausse pas le traitement des autres images de la série. L'image de référence semble être le plus souvent choisie proche du pic cortical [43, 44], qui offre un bon contraste.

En revanche, si on recale chaque image de la série sur celle qui la précède, il est nécessaire de faire des compositions de transformations, avec accumulation des erreurs [33] ; nous avons effectivement constaté ce phénomène lors de nos propres tests. Comme un seul recalage défectueux affecte celui de toutes les autres images qui suivent dans la série, la surveillance d'un opérateur pour détecter et corriger d'éventuelles dérives est requise dans [36], mais soulignons que le recalage est fait uniquement en translation au pixel près. La tâche de l'opérateur se révélerait beaucoup plus ardue pour des rotations, voire impossible pour des transformations élastiques. Ce choix ne semble donc pas très avantageux, même si les images à recaler se ressemblent davantage que dans le cas précédent.

Notons que l'aspect temporel est cependant pris en compte dans deux méthodes : dans [38, 37], un recalage au pixel près est suivi d'une segmentation puis d'un recalage et d'une segmentation plus fins ; dans [41], un terme de régularisation lissant temporellement les transformations est ajouté à la fonction d'énergie globale à minimiser, laquelle fait intervenir simultanément toutes les images de la série.

En revanche, on ne trouve pas d'essais de recalage sur un prototype construit à partir de la série d'images elle-même [47], peut-être en raison de la difficulté à estimer un tel modèle sur le type d'images que nous traitons.

2.4.4 Validation sur données réelles

La validation du recalage sur données réelles se heurte à la difficulté suivante : il n'existe pas de vérité terrain à laquelle on pourrait comparer les images recalées. La plupart du temps, dans la littérature, les résultats sont validés de la manière suivante :

- qualitativement, en visionnant par exemple la « vidéo » de l'acquisition [43, 44],
- quantitativement par comparaison avec un recalage manuel effectué par des experts, du moins pour les translations [37, 38], voire pour les rotations [38, 39], mais pas pour les transformations non rigides,
- dans [33], soulignant que le fait de considérer les recalages manuels comme une référence est contestable, en particulier parce qu'il dépend de l'opérateur, les auteurs proposent d'évaluer et de comparer les variances autour de la droite du modèle de Patlak (cf. 2.3) pour les données recalées ou non. Dans le même ordre d'idées, dans [42], les déviations standard des courbes temps-intensité de plusieurs régions d'intérêt délimitées manuellement sont évaluées et comparées pour deux méthodes de recalage différentes.

2.5 Méthodes de segmentation de séquences d'IRM dynamique rénale à rehaussement de contraste

Dans cette section, nous dressons un état de l'art des techniques de segmentation utilisées en IRM rénale à rehaussement de contraste. Beaucoup, dont la segmentation manuelle, n'utilisent qu'une ou deux images de la série complète et peuvent souffrir d'un recalage déficient. En revanche, celles basées sur les profils fonctionnels des voxels peuvent faire intervenir l'intégralité de la séquence et présenter une meilleure robustesse.

2.5.1 Segmentation manuelle par un radiologue

Actuellement, la segmentation anatomique est faite par des radiologues. Le procédé est long et fastidieux, surtout si le nombre de coupes à traiter est élevé [48]. L'opération est aussi délicate car les images sont floues et très bruitées. Par ailleurs, les différents compartiments ne sont pas tous visibles pendant la même phase de perfusion pour des raisons physiologiques, donc ils ne peuvent être segmentés sur une image unique extraite de la série. Les radiologues doivent examiner l'intégralité de la séquence pour choisir les deux images les mieux adaptées, l'une vers le pic cortical pour le cortex, l'autre en

phase tardive pour les cavités, la médullaire étant la partie restante. En cas de recalage déficient ou de mouvement hors plan de coupe, la segmentation peut être incohérente et l'analyse fonctionnelle du rein peut varier de manière importante. La procédure est ainsi sujette à d'importantes variations intra- et inter-observateurs. Le fait de considérer ces segmentations comme la référence pour la validation, ce qui est généralement le cas dans la littérature, peut soulever des questions, tout comme nous l'avons déjà souligné pour le recalage dans le paragraphe précédent. Nous parlerons plutôt de « pseudo vérités-terrain ». Nous reviendrons sur ce point au paragraphe 4.4.

2.5.2 Méthodes automatiques ou semi-automatiques

Etant donné la complexité des structures à segmenter et la qualité médiocre des images, les méthodes entièrement automatiques semblent difficilement utilisables. Il est également important que le processus permette de profiter de la compétence de médecins.

Méthodes de segmentation sur une image

Assez peu de méthodes classiques de segmentation en vogue dans le domaine de l'imagerie médicale ont été testées sur des acquisitions IRM de perfusion rénale à rehaussement de contraste [26]. Un aperçu de l'ensemble des méthodes citées dans un cadre non limité à la perfusion rénale peut être trouvé dans [49]. Il est parfois difficile de classer les méthodes proposées car elles comprennent souvent plusieurs étapes faisant intervenir des notions différentes. Dans l'aperçu qui suit, nous nous limiterons aux méthodes qui ont réellement été testées en IRM rénale.

Techniques de seuillage Tous les pixels dont les niveaux de gris font partie d'un intervalle donné sont supposés appartenir au même compartiment anatomique. Le périmètre du rein est d'abord délimité manuellement, puis certaines structures internes sont alors segmentées par seuillage dans différentes phases fonctionnelles. Par exemple, dans [50], le cortex de reins de porcs est extrait par un simple seuillage d'intensité pendant la phase corticale. Cependant, même si la technique se rapproche de la méthode de segmentation manuelle, la précision est limitée par les faibles rapports signal à bruit dans le parenchyme, ce qui rend difficile la séparation du cortex et de la médullaire.

Approche région : agrégation de pixels ou croissance de régions Ces méthodes exploitent l'homogénéité de certains attributs d'ensembles de pixels connexes. Dans [51], un algorithme de croissance de régions avec un seuil adaptatif produit un contour de rein tridimensionnel et une segmentation des structures rénales internes.

Approche frontière : contours actifs et courbes ou surfaces de niveau (*level-set*) Ces méthodes utilisent les variations d'intensité correspondant aux frontières de régions associées aux contours d'objets dans l'image. Dans [41], une première segmentation des frontières rénales est obtenue par une méthode *level-set*. Cependant la méthode ne convient pas pour délimiter les structures internes. Le principe reposant sur la minimisation d'une fonction de coût contenant différents termes liés à la fois à l'image et à la forme du

contour, l'optimisation peut aboutir à un minimum local. Il peut être en outre difficile de régler les poids relatifs associés à chaque terme. Vu la qualité des images et l'influence des organes périphériques, les contours sont parfois vraiment difficiles à distinguer sur certaines images.

Morphologie mathématique Dans [52], différents seuillages sont suivis par des opérations morphologiques (fermeture et érosion) visant à éliminer les structures externes restantes.

Segmentation fonctionnelle utilisant plusieurs images

Pour des raisons physiologiques, l'évolution temporelle du contraste est différente dans chaque compartiment anatomique (cf. 2.2.3). Certaines méthodes de segmentation utilisent cette propriété. Ainsi, dans [41], la segmentation des structures rénales internes est réalisée en minimisant une fonctionnelle d'énergie qui prend en compte à la fois la corrélation spatiale entre pixels d'une même image et leur corrélation temporelle durant toute la séquence. Dans [53], une méthode de segmentation semi-automatique utilisant un modèle de Markov temporel des courbes temps-intensité des différents voxels et la minimisation d'une énergie définie sur un ensemble de variables est présentée. La séquence est d'abord recalée, puis la segmentation est réalisée en trois étapes :

- extraction du rein complet,
- séparation de la médullaire d'une part, du cortex et des cavités d'autre part,
- séparation du cortex et des cavités.

Cette méthode nécessite l'intervention d'un opérateur, en particulier une initialisation par « semences », et le processus de segmentation dure environ une minute par étape. Une image de segmentation résultante est présentée mais aucune mesure de similarité avec d'autres résultats n'est donnée.

Certains auteurs utilisent directement les courbes temps-intensité des voxels rénaux pour les classer suivant leur profil fonctionnel, soit par un algorithme de K-moyennes [42, 54, 55], soit par analyse en composantes indépendantes (ACI) [56]. Une segmentation des structures rénales peut être alors déduite de cette classification. Les méthodes utilisées sont cependant décrites de manière très succincte, sauf dans [42], publié après la publication de nos résultats. Remarquons cependant que l'ACI telle qu'elle est utilisée dans [56] ne conduit pas à une segmentation où chaque voxel est attribué à un compartiment, mais plutôt à des zones fonctionnelles, un voxel donné pouvant appartenir à 0, 1 ou plusieurs de ces zones. L'objectif est ainsi assez différent du nôtre et les résultats ne peuvent pas être réellement comparés à une segmentation anatomique.

Dans [42], les reins sont d'abord extraits par un algorithme de K-moyennes à $K = 3$ classes (distance cosinus)[5] appliqué aux courbes temps-intensité d'une série d'images autour du pic cortical et suivi d'opérations morphologiques. Un second algorithme de K-moyennes avec $K = 5$ à 7 classes est alors appliqué aux voxels rénaux pour aboutir à une

[5] L'algorithme des K-moyennes est détaillé au paragraphe 3.1.3 page 65, le choix d'une distance est discuté au paragraphe 5.1.6 page 164

segmentation en 3 dimensions. Nous reviendrons sur cette méthode, qui est très proche de la nôtre, mais qui a été développée de manière indépendante, au chapitre 4.

Avantages potentiels Les avantages potentiels de ces méthodes sont :
- une plus grande robustesse en raison de l'utilisation de la totalité de la séquence et non de seulement deux images
- une meilleure reproductibilité.

Validation

Il existe très peu d'études incluant une validation sur des données réelles, rendue difficile par l'absence de vérité-terrain. Dans [57], une erreur de segmentation avec une segmentation manuelle est évaluée en additionnant les volumes des voxels « faux positifs » et « faux négatifs », correspondant respectivement à une sur-segmentation et à une sous-segmentation de la zone d'intérêt (cf. paragraphe 4.4.1 pour la définition exacte). Dans [42], un critère de similarité classique, à savoir le coefficient de Tanimoto (cf. paragraphe 3.1.6 page 91) entre les segmentations obtenues et des segmentations manuelles réalisées par des experts, est évalué pour quatre examens, dont deux du même patient. En général, la validation consiste en une évaluation qualitative de la cohérence avec des segmentations effectuées par des radiologues, ou en une comparaison entre les volumes des compartiments correspondants ou entre les rénogrammes obtenus [55].

2.5.3 Conclusion

Notre objectif est d'automatiser en partie les étapes de recalage et de segmentation nécessaires à l'exploitation des séquences d'IRM de perfusion rénale. Un certain nombre de méthodes de segmentation spécifiquement appliquées à ce domaine mais s'appuyant sur des techniques classiques sont présentées dans la littérature. Certaines n'utilisent qu'une ou deux images ; leurs performances sont souvent limitées par le faible rapport signal à bruit et le flou des acquisitions. D'autres sont basées sur l'évolution temporelle des voxels et utilisent une grande partie, voire la totalité des images d'une série, ce qui peut représenter entre 80 et 450 images. C'est cette deuxième option, dont nous avons choisi d'explorer les possibilités, qui semble prometteuse. Nous détaillons la méthode que nous proposons dans les chapitres 4 et 5, alors que les fondements mathématiques sont développés dans le chapitre 3.

La question de la validation des résultats sur les données réelles demeure ouverte. Il s'agit finalement de savoir si la segmentation fonctionnelle obtenue peut se rapprocher d'une segmentation anatomique. La validation reste souvent qualitative, et le choix de critères quantitatifs adéquats reste à discuter.

Ainsi, plusieurs difficultés restent à surmonter. Le recalage, indispensable pour obtenir une évolution temporelle cohérente pour chaque voxel et appliquer convenablement les modèles rénaux du paragraphe 2.3, est rendu particulièrement délicat par les changements de contraste au cours de la perfusion et la qualité des images. L'absence de vérité-terrain rend nécessaire la construction de données simulées réalistes pour tester les méthodes que nous proposons. Il nous faudra définir une méthode de validation des résultats sur des

données réelles ; ce point sera abordé dans les paragraphes 3.1.5 et 4.4.1 respectivement dans le cadre général de la comparaison de *clusterings*, puis dans le cas particulier de l'IRM de perfusion rénale.

Chapitre 3

Modèles statistiques

L'objet de ce chapitre est une présentation du cadre mathématique de notre méthode de segmentation des structures rénales internes. Cependant, nous commençons par introduire les concepts de manière intuitive, et nous donnons déjà quelques liens avec l'application envisagée afin que le lecteur puisse dès à présent percevoir l'intérêt que présentent ces notions dans notre cas. La méthode de segmentation sera exposée en détails dans le chapitre 4. Nous cherchons à segmenter les structures internes d'un rein en trois dimensions sur une séquence d'IRM de perfusion. Nous procéderons en deux étapes : nous commencerons par segmenter la coupe rénale principale, à savoir celle qui contient le plus grand volume rénal, puis nous nous servirons des résultats pour segmenter les autres coupes rénales. La première partie du chapitre est consacrée aux modèles utiles pour la segmentation de la coupe rénale principale : nous ferons une classification des voxels en utilisant des techniques de quantification vectorielle qui sont développées dans la première partie du chapitre. L'extension de la classification pour obtenir une segmentation en trois dimensions s'appuiera sur la théorie exposée dans sa seconde partie.

3.1 Quantification vectorielle appliquée au *clustering*

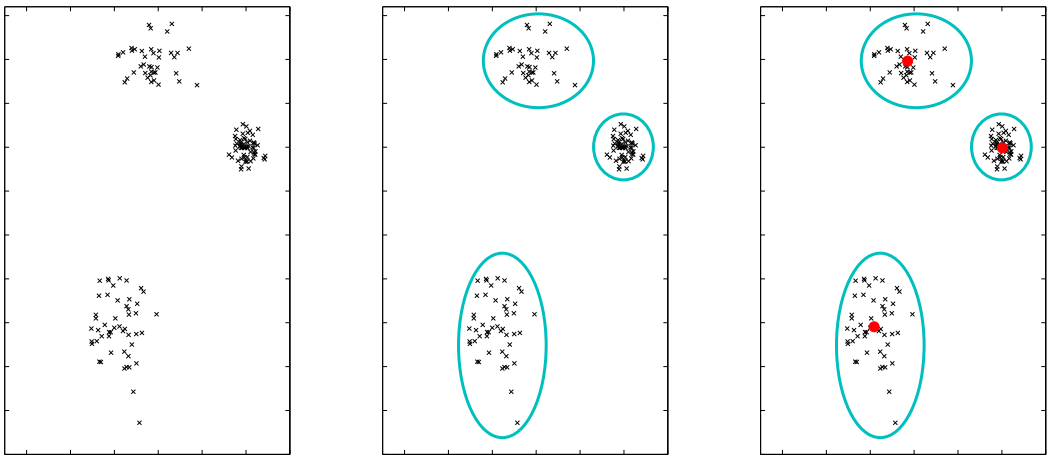
D'après le paragraphe 2.2.3, les trois compartiments anatomiques recherchés (cortex, médullaire et cavités) ont des évolutions temporelles de contraste différentes sur la séquence d'images acquise. Nous allons donc étudier la possibilité de classer les voxels rénaux selon leurs courbes temps-intensité par des méthodes d'apprentissage non supervisé afin d'aboutir à une segmentation du rein dite fonctionnelle.

L'objectif général de l'apprentissage non supervisé est de faire apparaître certaines structures sous-jacentes dans les données, par exemple en les classant en différents ensembles, ce qui correspond bien à nos objectifs. Il se différencie de l'apprentissage supervisé par le fait qu'on n'a pas d'exemples de donnée étiquetée avec la réponse attendue et sur lesquels on pourrait se baser pour approximer la sortie désirée : on ne peut utiliser que les données elles-mêmes. Il s'avère intéressant dans notre cas puisque nous ne disposons pas de vérité-terrain.

La quantification vectorielle fait partie des méthodes d'apprentissage non supervisé pouvant aboutir à une partition des données, au même titre que la classification hiérarchique, l'analyse en composantes indépendantes, les machines à vecteurs supports à une

classe (*one-class SVM*), certains arbres de décision, ... Elle trouve des applications dans des domaines aussi variés que la compression de données, la reconnaissance de formes et l'analyse par regroupement ou *clustering*. C'est cette dernière possibilité que nous nous proposons d'exploiter dans la première étape du traitement de nos données recalées.

Notons que le terme *clustering* est souvent traduit par partition, mais cela ne correspond pas tout-à-fait à l'acception courante de ce mot : s'il s'agit dans les deux cas de grouper les éléments d'un ensemble de données en différents sous-ensembles, appelés *classes* ou *clusters*, il est plus ou moins sous-entendu que, pour un *clustering*, chaque sous-ensemble contient des données qui sont homogènes, au sens où elles possèdent certains attributs similaires, ce qui n'est pas le cas en général pour une partition. On peut par exemple envisager de regrouper des objets d'après leur couleur. Un autre exemple, plus proche de celui que nous allons traiter, est le suivant : supposons qu'on ait un ensemble de points dans un plan, dont on connaît les coordonnées, et qui sont représentés sur la figure 3.1(a). De manière assez naturelle, on peut regrouper ces points d'après leur position en trois sous-ensembles (figure 3.1(b)) et attribuer à chacun un représentant (les points rouges sur la figure 3.1(c), situés « au centre » de chaque ensemble, pourraient convenir).



(a) (b) (c)

FIG. 3.1 – Exemple de *clustering* de points d'après leur position dans un plan : données (a), regroupement en trois sous-ensembles « naturels » (b) et choix d'un représentant (c).

Les attributs seront représentés dans la suite par la variable ξ. Si les attributs sont des points de $\mathbb{R}^p$, la similarité entre deux éléments peut être mesurée par la distance euclidienne entre leurs attributs respectifs. D'autres distances peuvent être utilisées, comme celle de Hamming pour des vecteurs binaires. A chaque sous-ensemble obtenu, on peut associer un élément particulier, ou prototype, qui fait ou non partie des données classées, et qui est supposé représenter convenablement tous les éléments de ce sous-ensemble. Réciproquement, connaissant le représentant d'un des sous-ensembles, il est parfois possible de classer de nouvelles données, en les mettant dans le sous-ensemble du prototype qui leur ressemble le plus. Remarquons qu'il est fréquent que deux données soient plus proches l'une de l'autre que de leur représentant respectif, quand les classes sont mal séparées. Certaines méthodes de quantification vectorielle permettent de classer des données même

dans ce cas.

Pour l'IRM rénale, on suppose que les voxels appartenant à un même compartiment rénal ont des courbes temps-intensité similaires : on va essayer de les classer d'après cet attribut. Pour un rein comportant N voxels $\{x_j\}_{1\leq N}$ dans sa coupe principale, on aura une collection de vecteurs $\{\xi_1, \ldots, \xi_N\}$, chaque vecteur ayant autant de composantes que d'instants d'acquisition. La i-ème composante du j-ème vecteur est liée au niveau de gris du j-ème voxel au i-ème instant. Le représentant de chaque *cluster* aura une courbe temps-intensité ressemblant à celle de tous les voxels qui en font partie.

Parmi les méthodes d'apprentissage non supervisé, nous n'avons pas retenu l'analyse en composantes indépendantes parce que nous ne voyons pas de raison justifiant l'indépendance statistique des signaux que l'on cherche à distinguer. La quantification vectorielle semble mieux adaptée que le *clustering* hiérarchique, qui vise plutôt à relier les données qu'à les séparer. Son objectif premier est de trouver des représentants pour un ensemble de données en minimisant une distorsion, et est très utilisée en codage. Les *clusterings* des données qui peuvent en être déduits dépendent en particulier des attributs associés aux données à classer et des paramètres employés pour les algorithmes de quantification vectorielle, comme par exemple le nombre de classes pour les K-moyennes, la distance utilisée dans l'espace des attributs ou bien la distorsion à atteindre. Ces *clusterings* ne possèdent pas nécessairement les propriétés souhaitées si ces paramètres ou les attributs sont mal adaptés. La quantification vectorielle peut néanmoins donner directement de bons résultats même si les classes sont mal séparées, du moins pour les *clusterings* déduits des K-moyennes.

Dans cette première partie de chapitre, après avoir rappelé les principes généraux de la quantification vectorielle, nous développerons deux manières de l'utiliser pour obtenir un *clustering*, selon que la quantification est associée ou non à une carte préservant la topologie dans l'espace des paramètres. Nous nous interrogerons alors sur la manière de valider les résultats obtenus, sachant que nous n'avons pas de vérité-terrain, mais deux segmentations manuelles réalisées par des radiologues.

3.1.1 Quantification vectorielle

Principes généraux

Soit $\mathcal{X} = \{\xi_1, \ldots, \xi_N\}$ un ensemble fini de points de $\mathbb{R}^p$. Nous supposons que c'est la réalisation d'une séquence de variables aléatoires vectorielles $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ identiquement distribuées de même loi qu'un vecteur aléatoire continu $\mathbf{X} : \Omega \longrightarrow V \subset \mathbb{R}^p$, où $(\Omega, \mathcal{E}, P)$ est un espace probabilisé et V une sous-variété[1] de $\mathbb{R}^p$. Nous l'appellerons *échantillon* $\mathcal{X}$. Notons $\mathcal{B}_{(\mathbb{R}^p)}$ la tribu des boréliens de $\mathbb{R}^p$, $P_{\mathbf{X}}$ la loi de $\mathbf{X}$ sur $(\mathbb{R}^p, \mathcal{B}_{(\mathbb{R}^p)})$ et E l'espérance mathématique. En pratique, ni V, ni $P_{\mathbf{X}}$ en général ne sont connus ; seul l'échantillon $\mathcal{X}$ est donné. L'objectif des techniques de quantification vectorielle est de déterminer, grâce aux ξ_i, un ensemble fini $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ de vecteurs de référence, ou *prototypes* $\mathbf{w}_j \in \mathbb{R}^p$, $j = 1, \ldots, K$ qui représente « au mieux » l'échantillon $\mathcal{X}$ [58, 59]. Un vecteur donné $\xi \in V$ est décrit par le prototype $w(\xi)$ qui lui correspond le mieux, dans un sens à définir.

[1] On trouvera au paragraphe 3.1.2 page 60 des notions concernant les variétés et leur topologie.

Définition d'un quantificateur vectoriel

Dans ce paragraphe, nous définissons formellement un quantificateur vectoriel et introduisons les notions de codeur et de décodeur, qui nous seront en particulier utiles pour définir par la suite la notion de carte préservant la topologie.

Définition 3.1 (Quantificateur vectoriel). *Soit $V \subset \mathbb{R}^p$ une sous-variété de $\mathbb{R}^p$, $\mathcal{I} = \{1, 2, \ldots, K\}$ un ensemble d'indices et $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de vecteurs $\mathbf{w}_j \in \mathbb{R}^p$, $j = 1, \ldots, K$. Un quantificateur vectoriel w de domaine V, de dimension p et de taille K est une application de V dans W qui à tout point ξ de V associe l'un des vecteurs $\mathbf{w}_j$, noté $w(\xi)$.*

$$\begin{aligned} w : V &\longrightarrow W \\ \xi &\longmapsto w(\xi). \end{aligned} \tag{3.1}$$

On le notera (V, W, w).

En compression de données, l'ensemble W est appelé ensemble des mots du code, mais dans le cadre du *clustering*, nous parlerons plutôt d'ensemble de prototypes.

A tout quantificateur vectoriel de taille K, on peut associer une partition de V, voire de $\mathbb{R}^p$ quand $V = \mathbb{R}^p$, en K sous-régions ou cellules V_j, $j = 1, \ldots, K$. La $j^{\text{ième}}$ cellule est définie par :

$$V_j = \{\xi \in V : w(\xi) = \mathbf{w}_j\} = w^{-1}(\mathbf{w}_j). \tag{3.2}$$

Définition 3.2 (Quantificateur vectoriel régulier). *Un quantificateur vectoriel (V, W, w) est dit régulier quand*

1. *chaque cellule V_j est un ensemble convexe, c'est-à-dire*

$$\xi, \eta \in V_j \Rightarrow \alpha\xi + (1 - \alpha)\eta \in V_j, \ (\forall \alpha \in]0, 1[), \tag{3.3}$$

2. *chaque prototype $\mathbf{w}_j$ appartient à la cellule V_j.*

Définition 3.3 (Quantificateur vectoriel borné). *Un quantificateur vectoriel (V, W, w) est dit borné quand son domaine V est borné.*

Un quantificateur vectoriel (V, W, w) peut se décomposer en deux opérateurs : l'encodeur $\mathcal{E}$ et le décodeur $\mathcal{D}$. L'encodeur est l'application de V dans l'ensemble d'indices $\mathcal{I}$, et le décodeur l'application de $\mathcal{I}$ dans W définis par :

$$\begin{aligned} \mathcal{E} : V &\longrightarrow \mathcal{I} \\ \xi &\longmapsto j \text{ si } \xi \in V_j, \end{aligned} \qquad \begin{aligned} \mathcal{D} : \mathcal{I} &\longrightarrow W \\ j &\longmapsto \mathbf{w}_j. \end{aligned} \tag{3.4}$$

Mesures de performance d'un quantificateur vectoriel

Pour analyser les performances d'un quantificateur vectoriel sur l'échantillon $\mathcal{X}$, nous utiliserons l'erreur quadratique moyenne, qui est l'espérance du carré de la distance euclidienne entre une donnée d'entrée et son représentant dans W. L'erreur quadratique moyenne est un cas particulier de distorsion moyenne. Nous pourrons alors définir la notion de quantificateur vectoriel optimal.

Définition 3.4 (Distorsion moyenne). *Soit d une fonction coût définie sur l'espace $\mathbb{R}^p \times \mathbb{R}^p$ à valeurs dans $\mathbb{R}$ ou $\mathbb{R}_+$. Soit un quantificateur vectoriel (V, W, w). On note $\mathbf{W}$ la matrice dont les colonnes sont les vecteurs prototypes $\mathbf{w}_j$, c'est-à-dire que $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$. La performance du quantificateur vectoriel (V, W, w) appliqué à l'échantillon $\mathcal{X}$ peut être mesurée par la distorsion moyenne*

$$\Psi(\mathbf{W}) = \mathrm{E}_{\mathbf{X}}[d(\mathbf{X}, w(\mathbf{X}))] = \int_V d(\xi, w(\xi)) \mathrm{d}P_{\mathbf{X}}(\xi), \tag{3.5}$$

ou encore

$$\Psi(\mathbf{W}) = \sum_{j=1}^{K} \Psi_j, \tag{3.6}$$

où

$$\Psi_j = \int_{V_j} d(\xi, w_j) \mathrm{d}P_{\mathbf{X}}(\xi). \tag{3.7}$$

La fonction d la plus utilisée correspond à l'erreur quadratique, c'est-à-dire au carré de la distance euclidienne entre deux vecteurs :

$$d(\xi, \eta) = \|\xi - \eta\|^2 = (\xi - \eta)^t (\xi - \eta). \tag{3.8}$$

C'est celle que nous choisirons pour toute la suite. D'autres possibilités pour le choix de d sont présentées dans [59].

Quantificateurs au plus proche voisin

Les quantificateurs au plus proche voisin (*nearest neighbor quantizer*) sont une catégorie particulièrement importante puisque tous les quantificateurs optimaux (cf. définition 3.8) en font partie. Ce type de quantificateur permet de construire une partition de $\mathbb{R}^p$ en un nombre fini de polyhèdres.

Définition 3.5 (Polyhèdre de Voronoï). *Soit $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de vecteurs $\mathbf{w}_j \in \mathbb{R}^p$, $j = 1, \ldots, K$. Le polyhèdre de Voronoï de $\mathbf{w}_j$, noté $V_j^{(p)}(W)$, est l'ensemble des points de $\mathbb{R}^p$ qui sont plus proches de $\mathbf{w}_j$ que de n'importe quel autre $\mathbf{w}_i$:*

$$V_j^{(p)}(W) = \{\xi \in \mathbb{R}^p : d(\xi, \mathbf{w}_j) \leq d(\xi, \mathbf{w}_i) \ \forall i = 1, \ldots, K\}. \tag{3.9}$$

Dans le cas où plusieurs prototypes sont à la même distance d'un vecteur $\xi \in \mathbb{R}^p$, ξ n'appartient qu'à un seul polyhèdre de Voronoï, celui correspondant au prototype de plus petit indice. L'ensemble des polyhèdres de Voronoï $\{V_j^{(p)}(W)\}_{1 \leq j \leq K}$ forme une partition de $\mathbb{R}^p$.

Quand il n'y aura pas d'ambiguïté sur W, nous noterons simplement $V_j^{(p)}$ le polyhèdre de Voronoï de $\mathbf{w}_j$.

Un exemple de polyhèdres de Voronoï d'un ensemble de vecteurs $\mathbf{w}_j$ est représenté figure 3.2.

Définition 3.6 (Quantificateur vectoriel au plus proche voisin). *Un quantificateur vectoriel (V, W, w) au plus proche voisin, dit aussi quantificateur de Voronoï, est tel que les cellules V_j sont définies par :*

$$V_j = \{\xi \in V : w(\xi) = \mathbf{w}_j\} = V \cap V_j^{(p)}(W). \tag{3.10}$$

Définition 3.7 (Cellule de Voronoï). *Dans ce cas, les $V_j, 1 \leq j \leq K$ sont appelées cellules de Voronoï de W dans V : chaque V_j est constituée de tous les points de V qui sont plus proches de $\mathbf{w}_j$ que de n'importe quel autre $\mathbf{w}_i$.*

Tout vecteur ξ appartenant à la cellule V_j est ainsi représenté par $w(\xi) = \mathbf{w}_j$, qui est le vecteur de l'ensemble W pour lequel l'erreur quadratique est minimale, i.e.

$$w(\xi) = \text{argmin}_{\mathbf{w}_j}\{d(\xi, \mathbf{w}_j)\}. \tag{3.11}$$

Un exemple de cellules de Voronoï sur une variété V est représenté figure 3.3.

Pour distinguer la quantification vectorielle sur une variété V de celle dans $\mathbb{R}^p$, on parlera respectivement de la cellule de Voronoï V_j ou du polyhèdre de Voronoï $V_j^{(p)}$ d'un prototype $\mathbf{w}_j$.

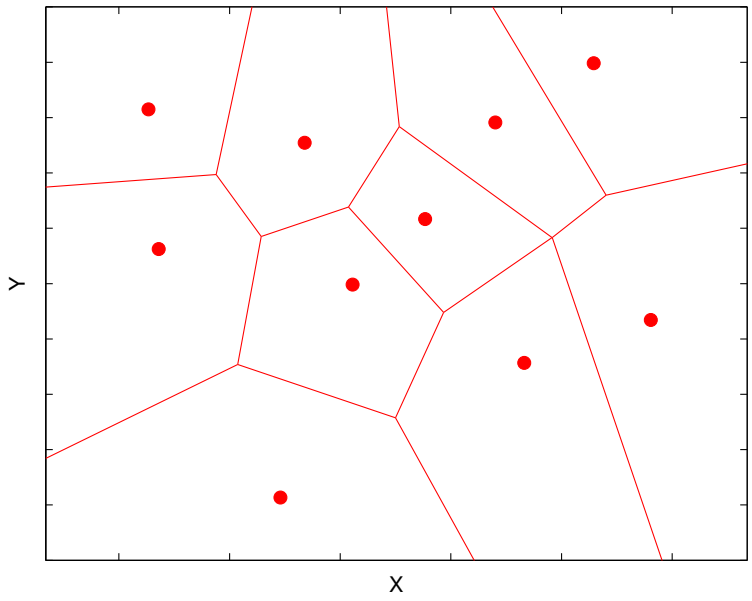


FIG. 3.2 – Pohyhèdres de Voronoï d'un ensemble de prototypes dans $\mathbb{R}^2$.

Conditions d'optimalité pour la quantification vectorielle

Après avoir défini la notion de quantificateur vectoriel optimal, nous rappelons trois conditions nécessaires que doit vérifier un tel quantificateur, sachant qu'elles sont insuffisantes pour impliquer l'optimalité, même locale.

Définition 3.8 (Quantificateur vectoriel optimal). *Soit un quantificateur vectoriel (V, W, w) dont la taille K est supposée fixée. Ce quantificateur est optimal pour l'échantillon $\mathcal{X}$ et le coût quadratique d s'il minimise la distorsion moyenne $\Psi(\mathbf{W})$, où $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ est la matrice dont les colonnes sont les vecteurs prototypes $\mathbf{w}_j$.*

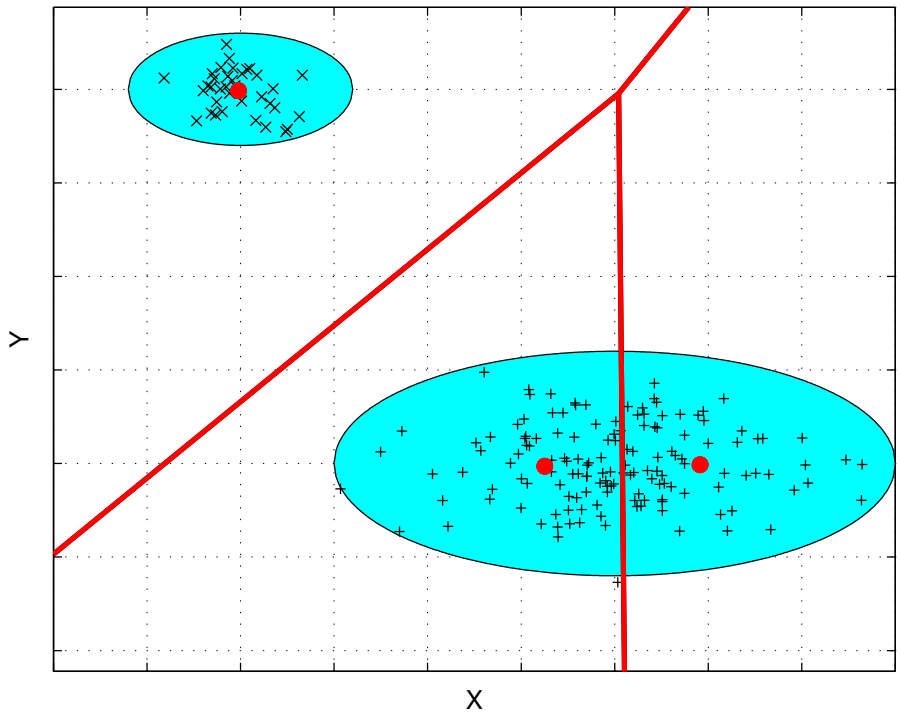


FIG. 3.3 – Cellules de Voronoï d'un ensemble de trois prototypes sur une variété $V \subset \mathbb{R}^2$: sur les polyhèdres de Voronoï (trait continu), la cellule de Voronoï de chaque prototype (gros point) sur la variété V (zone colorée) correspond à l'intersection de son polyhèdre de Voronoï et de la variété V.

Pour qu'un quantificateur vectoriel soit optimal, il doit vérifier les conditions suivantes[2] ([59] pages 349 et suivantes) :

1. condition du plus proche voisin : pour un ensemble donné W de prototypes, les cellules de la partition optimale satisfont

$$V_j \subset V_j^{(p)}(W), \tag{3.12}$$

2. condition du centroïde : étant donnée une partition $\{V_j\}_{1 \leq j \leq K}$ de V, les vecteurs du code optimal pour l'échantillon $\mathcal{X}$ sont les centroïdes des V_j, c'est-à-dire qu'ils vérifient, pour un coût quadratique :

$$\mathbf{w}_j = \mathrm{E}_{\mathbf{X}}[\mathbf{X}|\mathbf{X} \in V_j], 1 \leq j \leq K. \tag{3.13}$$

Définition 3.9 (Optimalité locale, optimalité globale). *Un quantificateur vectoriel (V, W, w) appliqué à l'échantillon $\mathcal{X}$ est dit :*

- *localement optimal si aucune petite perturbation des vecteurs prototypes $\mathbf{w}_j$ ou des cellules V_j (K et V étant fixés) ne fait décroître la distorsion moyenne,*
- *globalement optimal si, pour tout ensemble $W' \subset V$ de K prototypes et toute application w' de V dans W', on a*

$$\mathrm{E}_{\mathbf{X}}[d(\mathbf{X}, w(\mathbf{X}))] \leq \mathrm{E}_{\mathbf{X}}[d(\mathbf{X}, w'(\mathbf{X}))]. \tag{3.14}$$

Un quantificateur vectoriel vérifiant les deux conditions nécessaires d'optimalité n'est cependant pas nécessairement optimal, même localement [59]. Toutefois, ces conditions suggèrent une méthode itérative d'amélioration d'un quantificateur, qui est utilisée dans l'algorithme de Lloyd généralisé (cf. paragraphe 3.1.3 page 65).

3.1.2 Cartes préservant la topologie

Parallèlement à la quantification vectorielle, dans le cas où le support de la loi de $\mathbf{X}$ est une variété de $\mathbb{R}^p$, on peut construire un graphe topologique qui est une approximation de la triangulation de Delaunay des prototypes induite par V (on parlera plus brièvement de la variété encodée), et qui forme une carte préservant la topologie de cette variété [60]. L'ensemble peut être utilisé pour obtenir un *clustering* des données, en supposant qu'une classe correspond à une composante connexe de la variété (voir paragraphe 3.1.4 page 74).

Dans ce paragraphe, après avoir rappelé la définition d'une variété et montré l'insuffisance de la quantification vectorielle pour représenter sa topologie, nous définirons rigoureusement la notion de carte préservant la topologie.

Variété à composantes connexes

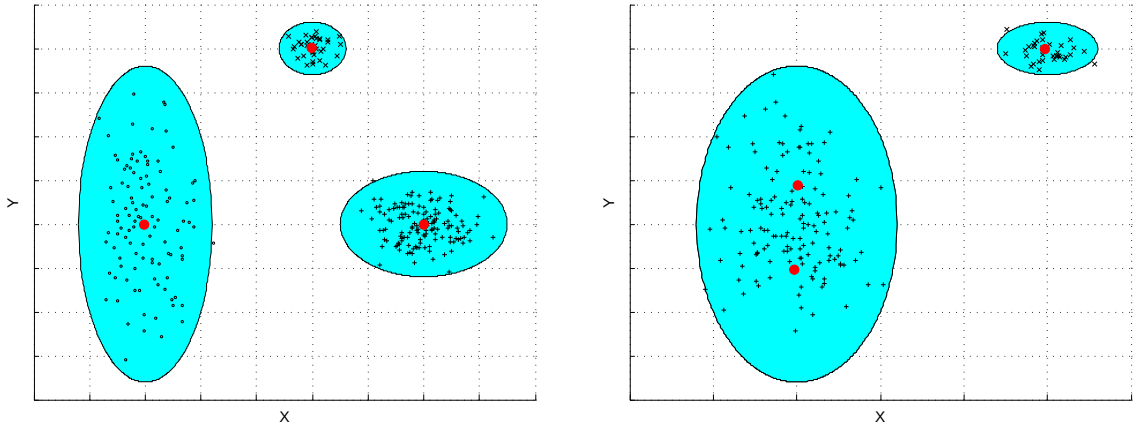
Définition 3.10 (Variété topologique). *Une variété topologique de dimension m ou m-variété, est un espace topologique séparé[3] localement homéomorphe à l'espace euclidien $\mathbb{R}^m$.*

[2] Dans le cas où $\mathbf{X}$ n'est pas continue, une troisième condition est nécessaire : les points des frontières des cellules de Voronoï doivent avoir une probabilité nulle.

[3] Un espace séparé est un espace topologique dans lequel deux points distincts quelconques admettent toujours des voisinages disjoints.

La dimension de la variété est confondue avec celle de l'espace euclidien auquel on la compare : les lignes sont de dimension 1, les plans de dimension 2,... La ressemblance peut n'être que locale, la structure de la variété complète peut être beaucoup plus compliquée (un exemple classique est celui d'une sphère, qui peut être considérée comme localement plane).

Les variétés peuvent être constituées d'un seul morceau ou bien d'une union de composantes connexes de dimensions éventuellement différentes. Quand la variété $V \subset \mathbb{R}^p$ est inconnue ou mal connue, la quantification vectorielle seule peut être insuffisante pour en séparer les différentes composantes connexes sauf dans le cas où chaque composante est incluse dans la cellule de Voronoï d'un seul prototype $\mathbf{w}_j$. Cependant, elle demeure en général insuffisante pour représenter la topologie de la variété. Une solution est d'adjoindre à l'ensemble de prototypes W une carte préservant la topologie de V, ce qui permettra de déduire de l'ensemble un *clustering* des données.



(a) (b)

FIG. 3.4 – Exemple de quantification vectorielle : sur la figure (a), les prototypes donnent une certaine idée de la topologie de la variété encodée puisque chaque prototype correspond à une composante connexe différente ; sur la figure (b), en revanche, il n'apparaît pas que les deux prototypes de gauche représentent des points appartenant à une même composante connexe si l'on ne dispose que des résultats de la quantification vectorielle.

Définition 3.11 (Points adjacents sur une variété). *Soit $V \subseteq \mathbb{R}^p$ une variété, $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes $\mathbf{w}_j \in V$. Deux prototypes $\mathbf{w}_i$, $\mathbf{w}_j \in V$ sont adjacents sur V si leurs cellules de Voronoï V_i et V_j sont adjacentes, c'est-à-dire si*

$$\overline{V_i} \cap \overline{V_j} \neq \emptyset. \tag{3.15}$$

Il apparaît intuitivement que, dans le graphe que nous souhaitons adjoindre à W, une relation doit exister entre deux prototypes si et seulement s'ils sont adjacents sur la variété. Dans le paragraphe suivant, avant de définir un graphe vérifiant cette propriété, nous rappelons la notion de triangulation de Delaunay dans $\mathbb{R}^p$: ce graphe met en relation les prototypes qui sont adjacents dans $\mathbb{R}^p$, même s'ils ne sont pas adjacents sur la variété V.

Triangulation de Delaunay

A un ensemble de prototypes, on peut associer le graphe qui met en relation chaque prototype et ses voisins « naturels » dans $\mathbb{R}^p$. Ce graphe, qui est la structure duale du diagramme de Voronoï, est appelé *triangulation de Delaunay*. Nous le nommerons aussi *triangulation de Delaunay complète* afin de le distinguer de la *triangulation de Delaunay induite* qui sera définie par la suite.

Définition et exemples Dans la définition de l'encodeur ou du décodeur, on peut remplacer l'ensemble des indices $\mathcal{I}$ par un graphe G de sommets, encore appelés *neurones* $i \in \mathcal{I}$ et dont les arcs (connections entre sommets ou neurones) sont décrits par une matrice d'adjacence $\mathbf{A}$ telle que

$$A_{ij} = \begin{cases} 1 & \text{si les neurones } i \text{ et } j \text{ sont connectés,} \\ 0 & \text{sinon.} \end{cases} \tag{3.16}$$

La matrice $\mathbf{A}$ est symétrique car les arcs ne sont pas orientés. On notera le graphe $G = (\mathcal{I}, \mathbf{A})$.

La triangulation de Delaunay complète peut être définie comme un graphe particulier qui relie les neurones dont les représentants ont des polyhèdres de Voronoï adjacents [60]. Par abus de langage, nous dirons parfois que le graphe relie les prototypes au lieu des neurones qu'ils représentent.

Définition 3.12 (Triangulation de Delaunay complète). *La triangulation de Delaunay complète d'un ensemble $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ de points de $\mathbb{R}^p$ est le graphe $G = (\mathcal{I}, \mathbf{A})$ tel que*

$$A_{ij} = \begin{cases} 1 & si \quad \overline{V_i^{(p)}(W)} \cap \overline{V_j^{(p)}(W)} \neq \emptyset, \\ 0 & sinon. \end{cases} \tag{3.17}$$

Ainsi, deux neurones i et j sont reliés par un arc si et seulement si les polyhèdres de Voronoï des prototypes associés $\mathbf{w}_i$ et $\mathbf{w}_j$ sont adjacents.

Un exemple de triangulation de Delaunay complète et du diagramme de Voronoï est représenté figure 3.5. Cette triangulation est définie indépendamment de la variété encodée : tous les prototypes adjacents dans $\mathbb{R}^p$ sont reliés, même s'ils ne sont pas adjacents sur la variété. Généralement, la triangulation de Delaunay complète ne respecte donc pas la topologie de V, mais intuitivement, le graphe représentant la topologie de V est inclus dans cette triangulation.

Triangulation de Delaunay induite et cartes préservant la topologie

Dans ce paragraphe, nous définissons rigoureusement la notion de carte préservant la topologie. Les exemples choisis visent à montrer comment on peut retrouver les différentes composantes connexes d'une variété, en vue du *clustering*[4].

[4]Des exemples de cartes préservant ou non la topologie sans faire intervenir la notion de composantes connexes sont présentés dans [60]

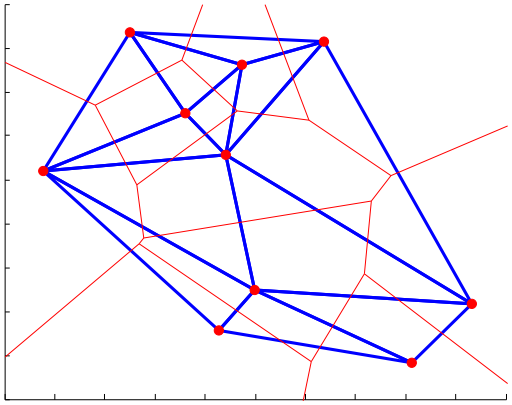


FIG. 3.5 – Triangulation de Delaunay (trait épais) et diagramme de Voronoï associé (traits fins) d'un ensemble de prototypes (gros points) dans $\mathbb{R}^2$.

Définition 3.13 (Encodeur préservant les voisinages). *Soit $V \subseteq \mathbb{R}^p$ une variété et $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes $\mathbf{w}_j \in V$ représentant les neurones j d'un graphe $G = (\mathcal{I}, \mathbf{A})$ dont les arcs sont définis par la matrice d'adjacence $\mathbf{A}$. L'application $\mathcal{E}_V$ de V vers G définie par :*

$$\begin{aligned} \mathcal{E}_V : V &\longrightarrow G \\ \xi &\longmapsto j^*(\xi), \end{aligned} \tag{3.18}$$

où $j^*(\xi)$ est déterminé par $\xi \in V^{(p)}_{j^*(\xi)}(W)$, préserve les voisinages quand les prototypes $\mathbf{w}_i$ et $\mathbf{w}_j$ qui sont adjacents sur la variété V sont les représentants de neurones i et j connectés dans le graphe G (i.e. si $A_{ij} = 1$).

Définition 3.14 (Décodeur préservant des voisinages). *Soit $V \subseteq \mathbb{R}^p$ une variété et $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes $\mathbf{w}_j \in V$ représentant les neurones j d'un graphe $G = (\mathcal{I}, \mathbf{A})$ dont les arcs sont définis par la matrice d'adjacence $\mathbf{A}$. L'application $\mathcal{D}_V$ de G vers V définie par :*

$$\begin{aligned} \mathcal{D}_V : G &\longrightarrow V \\ j &\longmapsto \mathbf{w}_j \in \mathbf{W} \end{aligned} \tag{3.19}$$

préserve les voisinages quand les neurones i et j connectés dans G ont des représentants $\mathbf{w}_i$ et $\mathbf{w}_j$ adjacents sur V.

Définition 3.15 (Carte préservant la topologie). *Soit $V \subseteq \mathbb{R}^p$ une variété et $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes $\mathbf{w}_j \in V$ représentant les neurones j d'un graphe $G = (\mathcal{I}, \mathbf{A})$ dont les arcs sont définis par la matrice d'adjacence $\mathbf{A}$. Le graphe G forme une carte préservant la topologie quand l'application $\mathcal{E}_X$ de V vers G (encodeur) et l'application réciproque $\mathcal{D}_X$ de G vers V (décodeur) préservent les voisinages.*

Dans le cas où la variété est formée de plusieurs composantes connexes, la triangulation de Delaunay complète des prototypes ne tient pas compte de cette particularité (cf.

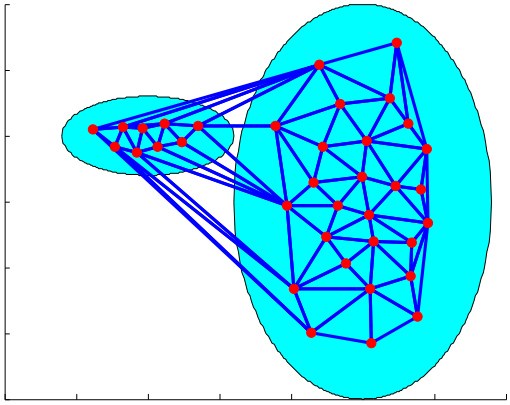


FIG. 3.6 – Non conservation de la topologie : Triangulation de Delaunay (trait épais) d'un ensemble de prototypes $\mathbf{w}$ (gros points) encodant une variété X (zone colorée) ; la topologie de la variété n'est pas préservée par la triangulation complète puisque des points appartenant à des composantes connexes différentes sont reliés par des arcs.

figure 3.6). La topologie de la distribution est alors mieux représentée par la *triangulation de Delaunay de W induite par V* (cf. figure 3.7).

Définition 3.16 (Triangulation de Delaunay induite). *Soit $V \subseteq \mathbb{R}^p$ une variété et $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes $\mathbf{w}_j \in V$. La triangulation de Delaunay de W induite par V, notée $\mathcal{D}_w^{(V)}$, est déterminée par le graphe qui connecte deux prototypes $\mathbf{w}_i$ et $\mathbf{w}_j$ si et seulement si leurs cellules de Voronoï V_i et V_j sont adjacentes.*

Ainsi, la triangulation de Delaunay de W induite par V (appelée plus brièvement *triangulation de Delaunay induite* s'il n'y a pas d'ambiguïté sur W et V) forme un sous-graphe de la triangulation de Delaunay complète de W.

Théorème 3.1. *Soit $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes répartis sur une variété $V \subseteq \mathbb{R}^p$ et $G = (\mathcal{I}, \mathbf{A})$ un graphe de sommets j représentés par les prototypes w_j et dont les arcs sont définis par la matrice d'adjacence $\mathbf{A}$. Le graphe G forme une carte préservant parfaitement la topologie de V si et seulement si G est la triangulation de Delaunay de W induite par V.*

La démonstration de ce théorème se trouve dans [60].

Maintenant que sont définies les notions de quantificateur optimal pour un échantillon et de cartes préservant la topologie, nous présentons des algorithmes permettant de les construire, ou au moins de s'en approcher.

3.1.3 Algorithmes de quantification vectorielle et construction de cartes préservant la topologie

Nous utiliserons deux algorithmes de quantification vectorielle :

– l'algorithme des K-moyennes, dit aussi de Lloyd généralisé [59],

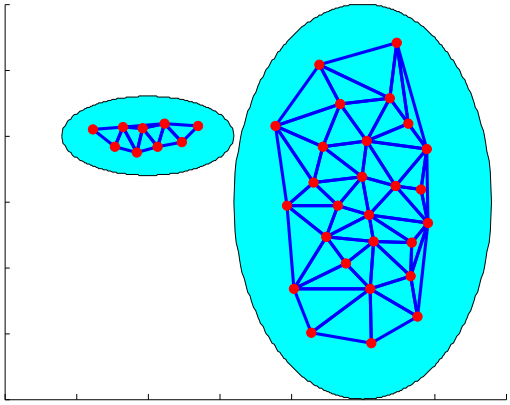


FIG. 3.7 – Conservation de la topologie par la triangulation de Delaunay induite (trait épais) d'un ensemble de prototypes (croix) sur une variété (zone colorée) dans $\mathbb{R}^2$.

– l'algorithme du *Growing Neural Gas with Targeting* (GNG-T) [61], variante de l'algorithme de *Growing Neural Gas* (GNG) [62].

La philosophie pour déduire le *clustering* des données selon que l'on utilise l'un ou l'autre de ces algorithmes sera quelque peu différente et détaillée dans le paragraphe 3.1.4. Cependant, l'un comme l'autre visent à minimiser la distorsion donnée dans l'équation (3.5) par l'ajustement itératif de la position des prototypes $\mathbf{w}_j$ en utilisant seulement un nombre fini de données ξ_i, supposées identiquement distribuées selon une loi de probabilité inconnue.

Algorithme des K-moyennes ou de Lloyd généralisé

Il existe une ambiguïté sur le terme K-moyennes, algorithme pour lequel il est parfois imposé que les prototypes fassent partie des données, alors que l'algorithme de Lloyd généralisé ne respecte pas cette contrainte. Dans la suite, nous emploierons indifféremment l'un ou l'autre terme, au sens de l'algorithme de Lloyd.

Cet algorithme recherche les K prototypes $\{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ qui minimisent la fonction de coût :

$$\widehat{\Psi}(\mathbf{W}) = \frac{1}{N} \sum_{j=1}^{K} \sum_{\xi_i \in V_j} \|\xi_i - \mathbf{w}_j\|^2, \tag{3.20}$$

où $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ est la matrice dont les colonnes sont les vecteurs prototypes $\mathbf{w}_j$ et où V_j est le polyhèdre de Voronoï de $\mathbf{w}_j$. Remarquons que, d'après la loi des grands nombres, quand les observations $\mathbf{X}_1, \ldots, \mathbf{X}_N$ sont indépendantes et N assez grand, $\widehat{\Psi}(\mathbf{W})$ vaut approximativement $\Psi(\mathbf{W})$ définie en (3.5) pour un coût quadratique.

Processus de minimisation Il s'agit d'un algorithme itératif, qui utilise à chaque itération l'ensemble des données ξ_i disponibles (algorithme dit *batch* ou hors ligne). Le cardinal K de W est prédéfini.

Initialisation $m = 0$;
$\widehat{\mathbf{W}}_m = [\xi_{\sigma(1)}, \ldots, \xi_{\sigma(K)}]$, où σ est une permutation de $[\![1, N]\!]$.

$\widehat{\Psi}_1 = \Psi(\widehat{\mathbf{W}}_m)$;

Etape 1 Etant donné l'ensemble de prototypes $\{\widehat{\mathbf{w}}_j^m\}_{1\leq j\leq K}$, générer l'ensemble de prototypes $\{\widehat{\mathbf{w}}_j^{m+1}\}_{1\leq j\leq K}$ de la manière suivante :

(a) trouver la partition optimale C des ξ_i telle que le *cluster* C_k soit l'ensemble de tous les ξ_i appartenant au polyhèdre de Voronoï $V_k^{(p)}(\widehat{W}_m)$ de $\widehat{\mathbf{w}}_k$,

(b) calculer les nouveaux prototypes $\{\widehat{\mathbf{w}}_j^{m+1}\}_{1\leq j\leq K}$:

$$\widehat{\mathbf{w}}_k^{m+1} = \frac{1}{N_k} \sum_{\xi_i \in C_k} \xi_i \tag{3.21}$$

avec

$$N_k = \mathrm{card}(C_k). \tag{3.22}$$

(c) $\widehat{\Psi}_0 = \widehat{\Psi}_1$;
$m = m + 1$;
$\widehat{\mathbf{W}}_m = [\widehat{\mathbf{w}}_1^m, \ldots, \widehat{\mathbf{w}}_K^m]$;
$\widehat{\Psi}_1 = \Psi(\widehat{W}_m)$.

Etape 2 Si $\widehat{\Psi}_0 - \widehat{\Psi}_1 > \epsilon \geq 0$, aller à l'étape 1, sinon retourner $\widehat{\Psi}_1$ et $\widehat{\mathbf{W}}_m$.

Remarquons que l'étape 1(a) correspond à la condition du plus proche voisin et l'étape 1(b) à la règle du centroïde, qui sont des conditions nécessaires pour qu'un quantificateur soit optimal : l'équation (3.21) donne en effet une estimation du centroïde de la cellule de Voronoï $V_k(\widehat{W}_m)$ si N est assez grand et les $\xi_1, \ldots, \xi_N$ suffisamment indépendants.

$$\widehat{\mathbf{w}}_k^{m+1} \simeq \mathrm{E}[\mathbf{X}|\mathbf{X} \in V_k(\widehat{W}_m)]. \tag{3.23}$$

Une étude de la convergence de cet algorithme est présentée dans [63], où il est en particulier démontré (cf. annexe B) que ce processus, applique un algorithme d'optimisation de Newton avec la règle d'adaptation (3.21).

Le principal défaut de l'algorithme des K-moyennes est que le résultat obtenu est très dépendant des conditions initiales et peut correspondre à un minimum local et non global de la distorsion [59]. Par ailleurs, le nombre de prototypes K doit être défini à l'avance, ce qui peut être particulièrement délicat pour des distributions de forme complexe dans des espaces de grande dimension (cf. paragraphe 3.1.4 page 77).

Algorithme GNG-T

L'algorithme GNG-T est une variante du GNG classique [62], lui-même dérivé de l'algorithme de *Neural Gas* associé à une compétition selon la règle de Hebb [58]. Il n'existe pas, à notre connaissance, de preuve de convergence de cet algorithme, et il n'est pas prouvé non plus qu'il minimise la distorsion de l'équation (3.5). Des simulations peuvent cependant être trouvées dans [62]. Les résultats de cet algorithme consistent à la fois en un ensemble de prototypes encodant une variété et en un graphe reliant les neurones associés et qui forme une carte préservant la topologie de cette variété.

Dans ce paragraphe, nous commencerons par présenter en quelques mots l'algorithme de *Neural Gas*, pour lequel les preuves de convergence sont établies, en insistant particulièrement sur la règle d'adaptation des prototypes puis nous rappellerons la règle d'apprentissage compétitif de Hebb qui permet, sous certaines conditions, de construire la triangulation de Delaunay des prototypes induite par la variété encodée. Nous expliquerons alors comment ces règles sont en quelque sorte combinées pour aboutir au GNG-T.

Neural Gas Bien qu'une version *batch* de cet algorithme existe [64], c'est nécessairement la version en ligne qui est utilisée conjointement à la compétition selon la règle de Hebb : il y a ajustement des prototypes après la présentation de chaque donnée ξ_i et pas seulement après la présentation de l'ensemble de ces données. Les résultats exposés ci-dessous et la convergence de cet algorithme sont détaillés dans [58].

Règle d'adaptation pour l'algorithme de *Neural Gas* Pour $\mathcal{I} = \{1, \ldots, K\}$, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$ et $\xi \in \mathbb{R}^p$ donnés, on ordonne les prototypes selon leur distance à ξ. Soit la permutation $\sigma \in \mathrm{Perm}(\mathcal{I})$ telle que

$$\|\xi - \mathbf{w}_{\sigma(1)}\| \leq \|\xi - \mathbf{w}_{\sigma(2)}\| \leq \ldots \leq \|\xi - \mathbf{w}_{\sigma(K)}\|. \tag{3.24}$$

Dans le cas d'égalité dans les inégalités (3.24), la permutation σ est choisie de telle sorte que, si $\|\xi - \mathbf{w}_{\sigma(i)}\| = \|\xi - \mathbf{w}_{\sigma(i+1)}\| = \ldots = \|\xi - \mathbf{w}_{\sigma(i+k)}\|$, alors $\sigma(i) < \sigma(i+1) < \ldots < \sigma(i+k)$.

Pour $j \in \mathcal{I}$, on pose

$$k_j(\xi, \mathbf{W}) = \sigma^{-1}(j) - 1. \tag{3.25}$$

Ainsi, pour $j, k \in \mathcal{I}$, on a $k_j(\xi, \mathbf{W}) = k - 1$ quand $\mathbf{w}_j$ est le k-ième prototype de W le plus proche de ξ avec la convention ci-dessus dans le cas où plusieurs prototypes sont à égale distance de ξ. Le prototype $\mathbf{w}_j$ est appelé *k-ième voisin de ξ*.

La règle d'adaptation est alors la suivante :

$$\widehat{\mathbf{w}}_j(t+1) = \widehat{\mathbf{w}}_j(t) + \epsilon \mathrm{h}_\lambda[k_j(\xi, \widehat{\mathbf{W}}(\mathbf{t}))](\xi - \widehat{\mathbf{w}}_j(t)), \tag{3.26}$$

$\epsilon \in [0, 1]$ étant un paramètre d'ajustement, h_λ désignant une fonction telle que

$$\mathrm{h}_\lambda[k_i(\xi, \mathbf{W})] = 1 \text{ pour } k_i = 0 \tag{3.27}$$

et décroissant vers 0 quand k_i augmente, avec une constante caractéristique λ.

Interprétation Soit la fonction $\Psi_{\mathrm{NG}} : V \longrightarrow \mathbf{R}^+$ définie par :

$$\Psi_{\mathrm{NG}}(\mathbf{W}) = \mathrm{E}_{\mathbf{X}} \left[\sum_{j=1}^{K} \mathrm{h}_\lambda[k_j(\mathbf{X}, \mathbf{W})] d(\mathbf{X}, \mathbf{w}_j) \right]. \tag{3.28}$$

La règle d'adaptation de l'équation (3.26) correspond à une descente stochastique de gradient sur la fonction Ψ_{NG} (voir annexe C.2).

Remarquons que, lorsque $\mathrm{h}_\lambda(1) = 0$, la fonction minimisée est la même que celle minimisée par l'algorithme des K-moyennes.

Conclusion Si on compare cet algorithme à la version en ligne des K-moyennes, où seul le prototype le plus proche de la donnée courante ξ_i est modifié [63], l'ensemble complet des prototypes est ajusté à chaque itération, l'ampleur de la modification étant plus petite pour les $\mathbf{w}_j$ les plus éloignés de ξ_i. Ceci diminue le risque d'aboutir à un minimum local de la fonction minimisée Ψ_{NG}, laquelle tend, sous certaines conditions, vers la distorsion de l'équation (3.5).

Construction d'une carte préservant la topologie Conjointement à la construction de l'ensemble de prototypes W, il est possible, sous certaines conditions, de construire une carte préservant la topologie de la variété encodée. Pour d'autres algorithmes comme les cartes de Kohonen [65], une structure de graphe doit être choisie à l'avance, alors qu'on n'est pas sûr qu'elle convienne aux données traitées. Ici, aucune connaissance *a priori* sur la topologie de la distribution n'est nécessaire.

Règle d'adaptation de Hebb La règle d'adaptation de Hebb, qui fait partie des modes d'apprentissage dits compétitifs, consiste à relier, pour chaque entrée ξ_i, les deux prototypes les plus proches au sens de la distance euclidienne.

Définition 3.17 (Polyhèdre de Voronoï du second ordre). *Soit $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ un ensemble de vecteurs $\mathbf{w}_j \in \mathbb{R}^p$, $j = 1, \ldots, K$. Le polyhèdre de Voronoï du second ordre $V_{ik}^{(p)}(W)$ des prototypes $\mathbf{w}_i$ et $\mathbf{w}_k$ est l'ensemble des points $\xi \in \mathbb{R}^p$ dont $\mathbf{w}_i$ et $\mathbf{w}_k$ sont les deux plus proches voisins dans W, c'est-à-dire :*

$$V_{ik}^{(p)}(W) = \{\xi \in \mathbb{R}^p : d(\xi, w_i) \leq d(\xi, w_j) \text{ et } d(\xi, w_k) \leq d(\xi, w_j), \forall j \neq i, k\}. \tag{3.29}$$

Elle permet de construire la triangulation de Delaunay complète de l'ensemble W seulement si chaque polyhèdre de Voronoï du second ordre $V_{ik}^{(p)}(W)$ est, au moins partiellement, recouvert par la variété V. Dans le cas contraire, moyennant les conditions décrites dans le théorème 3.2, on en obtient seulement une partie, appelée triangulation de Delaunay induite de W sur V, qui constitue une carte préservant parfaitement la topologie de V.

Théorème 3.2. *Soit $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ un ensemble de prototypes répartis sur une variété $V \subseteq \mathbb{R}^p$. Si l'ensemble W est tel que, pour tout point $\xi \in V$, les segments $[\xi, \mathbf{w}_i]$ et $[\xi, \mathbf{w}_k]$ formés par ξ et ses deux plus proches voisins $\mathbf{w}_i$ et $\mathbf{w}_k$ dans W sont entièrement inclus dans V, alors le graphe G formé par la règle d'adaptation de Hebb est la triangulation de Delaunay induite $\mathcal{D}_w^{(V)}$ de V.*

Eléments de démonstration. D'après la règle de Hebb rappelée au début du paragraphe, deux prototypes $\mathbf{w}_i$ et $\mathbf{w}_k$ sont reliés si et seulement s'il existe $\xi \in V$ dont $\mathbf{w}_i$ et $\mathbf{w}_k$ sont les deux plus proches voisins. La variété V étant le support de la loi de $\mathbf{X}$, $\mathbf{w}_i$ et $\mathbf{w}_k$ sont donc reliés si et seulement si

$$V_{ik}^{(p)}(W) \cap V \neq \emptyset. \tag{3.30}$$

Montrons alors que la condition 3.30 est équivalente à

$$V_i \cap V_k \neq \emptyset. \tag{3.31}$$

Si $V_i \cap V_k \neq \emptyset$, il existe $\xi_0 \in V$ tel que

$$d(\xi_0, \mathbf{w}_i) = d(\xi_0, \mathbf{w}_k) \text{ et } d(\xi_0, \mathbf{w}_i) \leq d(\xi_0 - \mathbf{w}_j), \forall j \in [\![1, \ldots, K]\!], \tag{3.32}$$

donc $\xi_0 \in V_{ik}^{(p)}(W)$. Si $V_{ik}^{(p)}(W) \cap V \neq \emptyset$, il existe $\xi^\star \in V_{ik}^{(p)}(W)$ dont $\mathbf{w}_i$ et $\mathbf{w}_k$ sont les deux plus proches voisins, et tels que $[\xi^\star, \mathbf{w}_i] \subseteq V$ et $[\xi^\star, \mathbf{w}_k] \subseteq V$. En supposant sans perte de généralité que $\mathbf{w}_i$ est le plus proche voisin de $\xi^\star$, on montre qu'il existe un point $\xi' \in [\xi^\star, \mathbf{w}_k]$ tel que $d(\xi', \mathbf{w}_i) = d(\xi', \mathbf{w}_k)$, donc que $V_i \cap V_k \neq \emptyset$. □

La démonstration complète de ce théorème peut être trouvée dans [60]. Une de ses conséquences, d'après le théorème 3.1, est que le graphe G ainsi construit forme une carte préservant parfaitement la topologie de V.

GNG-T GNG-T [61] est un algorithme au fil de l'eau (*on line* ou en ligne) qui traite des ensembles de données de taille N (échantillons) successifs, correspondant à différentes époques (*epochs*). Pendant une époque, N réalisations ξ_i sont présentées en tant qu'entrées. A chaque donnée ξ_i présentée, le $\mathbf{w}_j$ le plus proche (appelé prototype gagnant, ou simplement gagnant) et tous ses voisins topologiques sont modifiés. Alors que dans *Neural Gas*, l'amplitude de l'ajustement d'un prototype dépend uniquement de sa distance à ξ_i, cette amplitude est liée à sa place dans le graphe topologique construit conjointement selon la règle de Hebb. Une variable d'accumulation e_j est associée à chaque nœud $\mathbf{w}_j$: elle est égale à la somme des carrés des écarts entre la donnée courante et le prototype $\mathbf{w}_j$ à chaque fois que ce dernier gagne, c'est-à-dire qu'après le traitement de ξ_n :

$$e_j = \sum_{k \in S_j} d(\xi_k, \mathbf{w}_j) \text{ avec } S_j = \{k \in [\![1, \ldots, n]\!] : w(\xi_k) = \mathbf{w}_j\} \tag{3.33}$$

Lorsque la fonction coût définie par l'équation (3.6) est minimale, asymptotiquement, quand le nombre de prototypes tend vers l'infini, tous les Ψ_j tendent vers la même valeur [66, 67] ; cela reste vrai dans une certaine mesure quand le nombre de prototypes est fini [68].

A la fin d'une époque donnée, Ψ_j, définie dans l'équation (3.8) peut être estimée par :

$$\Psi_j \approx \frac{e_j}{N}. \tag{3.34}$$

Cette valeur commune des e_j, que nous noterons T', permet de contrôler la précision de la quantification vectorielle et peut être comparée à une valeur cible désirée T : elle sert à adapter le nombre de nœuds à la fin de chaque époque. Si $T' > T$, la précision est insuffisante, donc de nouveaux nœuds seront ajoutés. Au contraire, si $T' < T$, la quantification est trop précise, il faut donc éliminer certains nœuds. Le nombre de nœuds est donc adapté itérativement pendant l'exécution de l'algorithme. Le paramètre T ne change pas au cours de l'algorithme, il est donc défini avant son commencement. Il est également nécessaire de supprimer les arcs qui ne font plus partie de la triangulation de Delaunay induite en raison du mouvement des prototypes qu'ils relient. Un âge est ainsi attribué à chaque arc : cet âge prend la valeur 0 lors de la création de l'arc, augmente de 1 à chaque itération, et est remis à 0 à chaque fois que les deux prototypes qu'il relie sont

les deux plus proches voisins de la donnée courante. Tous les arcs dont l'âge dépasse la valeur $\mathcal{A}$ sont supprimés.

L'implantation est la suivante.

Initialisation Commencement d'une nouvelle époque. Vérifier que le graphe comprend bien deux nœuds. Si ce n'est pas le cas et que des exemples sont disponibles, choisir deux prototypes $\mathbf{w}_1$ et $\mathbf{w}_2$ selon la loi $P_{\mathbf{X}}$.
$n \leftarrow 0$.

Etape 1 Mettre e_j à 0 pour tous les neurones j.

Etape 2 Si $n = N$, l'époque est terminée, aller à l'étape 8. Sinon, aller à l'étape suivante.

Etape 3 $n \leftarrow n + 1$, prendre l'entrée ξ_n et déterminer le prototype gagnant $\widehat{\mathbf{w}}_{j_1}$ et le second voisin $\widehat{\mathbf{w}}_{j_2}$ parmi les $\widehat{\mathbf{w}}_j$.

Etape 4 Si $\widehat{\mathbf{w}}_{j_1}$ et $\widehat{\mathbf{w}}_{j_2}$ sont déjà connectés par un arc, mettre son âge à 0. Sinon, créer un arc entre $\widehat{\mathbf{w}}_{j_1}$ et $\widehat{\mathbf{w}}_{j_2}$, d'âge 0. Incrémenter l'âge de tous les arcs issus de $\widehat{\mathbf{w}}_{j_1}$ et détruire ceux dont l'âge dépasse $\mathcal{A}$.

Etape 5 Mettre à jour l'accumulateur : $e_{j_1} \leftarrow e_{j_1} + d(\xi, \widehat{\mathbf{w}}_{j_1})$.

Etape 6 Mettre à jour les prototypes de la manière suivante, l'indice i dénotant les voisins de $\widehat{\mathbf{w}}_{j_1}$ dans le graphe :

$$\widehat{\mathbf{w}}_{j_1} \leftarrow \widehat{\mathbf{w}}_{j_1} + \alpha_1(\xi - \widehat{\mathbf{w}}_{j_1}),\ \forall i\, \widehat{\mathbf{w}}_i \leftarrow \widehat{\mathbf{w}}_i + \alpha_2(\xi - \widehat{\mathbf{w}}_i). \tag{3.35}$$

Etape 7 Aller à l'étape 2.

Etape 8 Eliminer les nœuds qui n'ont pas gagné pendant l'époque.

Etape 9 Calculer $\overline{e} = \frac{1}{K}\sum_{j=1}^{K} e_j$

Etape 10 Si $T < \overline{e}$, aller à l'étape 11, sinon aller à l'étape 12.

Etape 11 La quantification n'est pas assez précise. Déterminer $\mathbf{w}_a$, le prototype $\mathbf{w}_j$ avec la plus forte distorsion e_j et trouver parmi ses voisins topologiques le prototype avec la plus forte distorsion e_i, que nous appellerons $\mathbf{w}_b$. Détruire les arcs entre les neurones a et b, ajouter un neurone c de prototype $\mathbf{w}_c = 0,5 \times (\mathbf{w}_a + \mathbf{w}_b)$, ajouter un nouvel arc d'âge 0 entre a et c, et entre a et b. Aller à l'étape 13.

Etape 12 La quantification est trop précise. Eliminer le neurone j qui a la plus petite distorsion e_j.

Etape 13 Fin de l'époque.

Si les résultats du *Neural Gas* et du GNG (avec ou sans cible) sont comparables, GNG est considéré comme ayant une meilleure convergence que le *Neural Gas*, au sens où les étapes intermédiaires semblent plus significatives (cf. [62], en particulier la figure 3 : du fait de la croissance, au cours des itérations, du nombre de prototypes que relie le réseau, la carte topologique de GNG semble être une meilleure approximation de la topologie de la distribution que celle fournie par le *Neural Gas*, et les prototypes sont mieux positionnés parce que chaque prototype supplémentaire est introduit dans la zone où la distorsion locale est la plus élevée).

Exemples comparatifs

De nombreux exemples peuvent être trouvés dans la littérature, en particulier dans [61] pour GNG-T. Sur les figures 3.8 et 3.9, un exemple de quantification vectorielle d'un même ensemble de données par un algorithme de K-moyennes ou de GNG-T est présenté, les paramètres étant réglés dans l'optique d'en déduire un *clustering*. Dans le paragraphe 3.1.4, nous exposerons d'autres résultats de quantification vectorielle avec les *clusterings* induits pour des données en dimension 2, afin d'illustrer ce qui peut être obtenu pour les données en grande dimension que nous utiliserons par la suite.

3.1.4 *Clustering* par quantification vectorielle

Dans ce paragraphe, nous montrons comment utiliser les résultats d'une quantification vectorielle pour obtenir un *clustering* des données, la philosophie étant différente suivant que l'un ou l'autre algorithme aura été utilisé pour effectuer la quantification.

Clustering et fonction de *clustering*

Soit une variété $V \subseteq \mathbb{R}^p$ [69]. Un ensemble $\mathcal{X} = \{\xi_1, \ldots, \xi_N\}$ est un ensemble fini d'éléments de V, que l'on supposera être la réalisation de N vecteurs aléatoires continus $\mathbf{X}_i : \Omega \longrightarrow V$ identiquement distribués de loi $P_{\mathbf{X}}$. La définition habituellement adoptée pour le *clustering* d'une variété est la suivante.

Définition 3.18 (*Clustering* d'une variété V). *Un clustering C d'une variété V est une partition finie de V. Les sous-ensembles $C_1, C_2, \ldots, C_K$ qui la forment sont appelés clusters.*

$$C = \{C_1, C_2, \ldots, C_K\} \text{ tel que } \forall k \neq l, C_k \cap C_l = \emptyset \text{ et } \cup_{k=1}^{K} C_k = V. \tag{3.36}$$

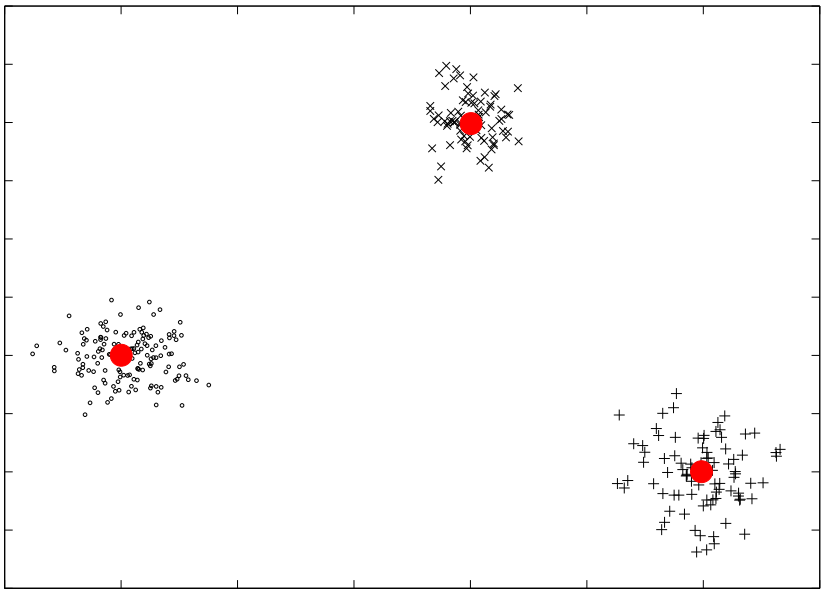
Le clustering peut aussi être considéré comme une relation d'équivalence sur V avec un nombre fini de classes d'équivalences appelées clusters.

La même définition est valable pour le *clustering* d'un ensemble dénombrable X. L'ensemble des partitions de l'ensemble X est noté $\mathcal{C}(X)$, ou plus simplement $\mathcal{C}$ s'il n'y a pas d'ambiguïté sur X. Remarquons cependant que, dans ces définitions, la connotation relative à l'homogénéité des données au sein d'un *cluster* (cf. paragraphe 3.1 page 53), liée à la proximité dans l'espace des attributs, n'apparaît pas. La quantification vectorielle est une méthode de *clustering* qui prend cette hypothèse en compte.

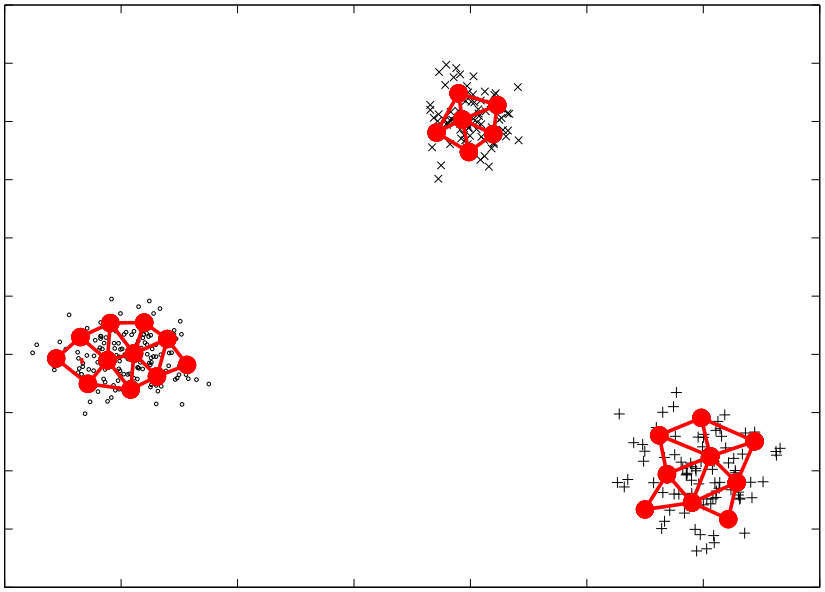
Définition 3.19 (*Clustering* uniformément centré (*center-based*)). *Un clustering $C = \{C_1, C_2, \ldots, C_K\}$ d'une variété V est dit uniformément centré s'il existe des points $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K \in \mathbb{R}^p$ tels que*

$$\forall i, \forall \xi \in C_i, \forall j \neq i, \|\xi - \mathbf{w}_i\| \leq \|\xi - \mathbf{w}_j\|. \tag{3.37}$$

Pour $1 \leq i \leq K$, $\mathbf{w}_i$ est alors le barycentre de C_i.

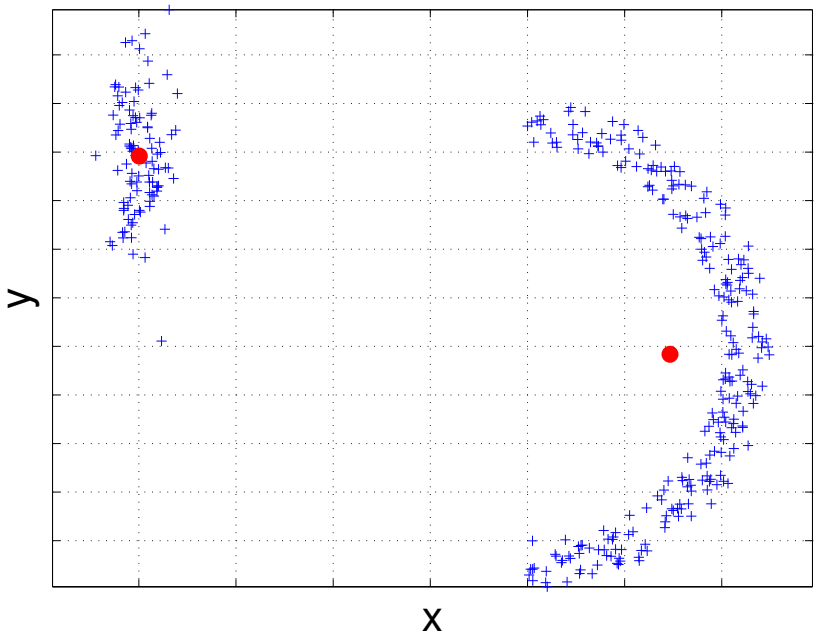


(a)

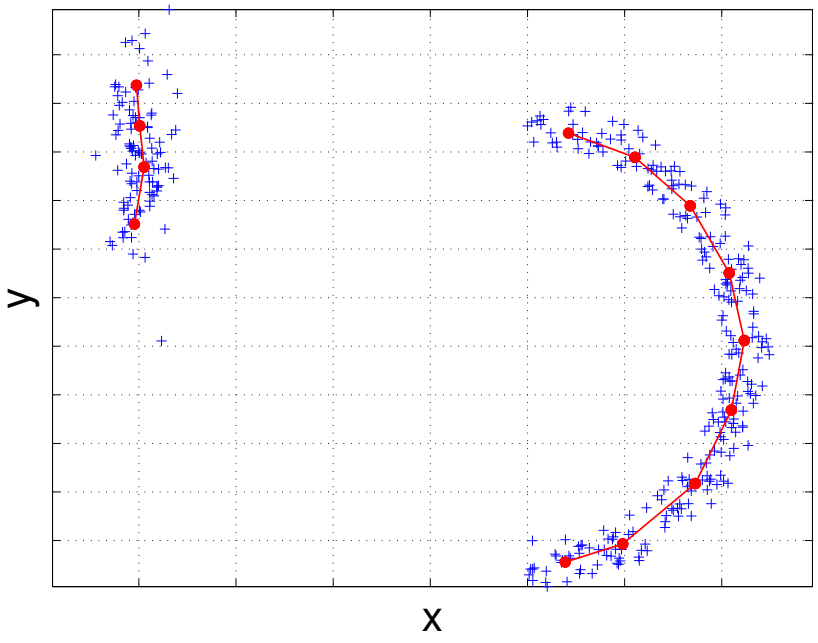


(b)

FIG. 3.8 – Exemples de quantification vectorielle d'un même ensemble de données par un algorithme de K-moyennes (a) ou de GNG-T (b) : les croix, les signes plus et les ronds représentent les échantillons de la distribution, les gros points sont les prototypes après quantification vectorielle, reliés par des arcs pour GNG-T.



(a)



(b)

FIG. 3.9 – Exemples de quantification vectorielle d'un même ensemble de données par un algorithme de K-moyennes (a) ou de GNG-T (b) : les croix représentent les échantillons de la distribution, les gros points sont les prototypes après quantification vectorielle, reliés par des arcs pour GNG-T.

Clustering induit par une quantification vectorielle

Que l'on utilise un algorithme de K-moyennes ou un GNG-T, on obtient à la fois une partition d'une variété, voire de l'espace $\mathbb{R}^p$ entier, et de l'ensemble des observations $\mathcal{X}$. Dans ce paragraphe, nous montrons comment nous utilisons les résultats de la quantification vectorielle pour obtenir le *clustering* qui nous intéresse.

Cas des K-moyennes Comme cela est fait de manière habituelle, nous supposerons que les données appartenant à une même cellule de Voronoï forment une classe de la partition (cf. figure 3.10a). Les *clusterings* ainsi obtenus sont uniformément centrés. Cette hypothèse peut se révéler inadéquate, par exemple dans le cas où la valeur de K n'est pas adaptée à la structure des données (cf. figure 3.12(b)) ou dans le cas d'un mauvais choix d'attributs pour les données à classifier.

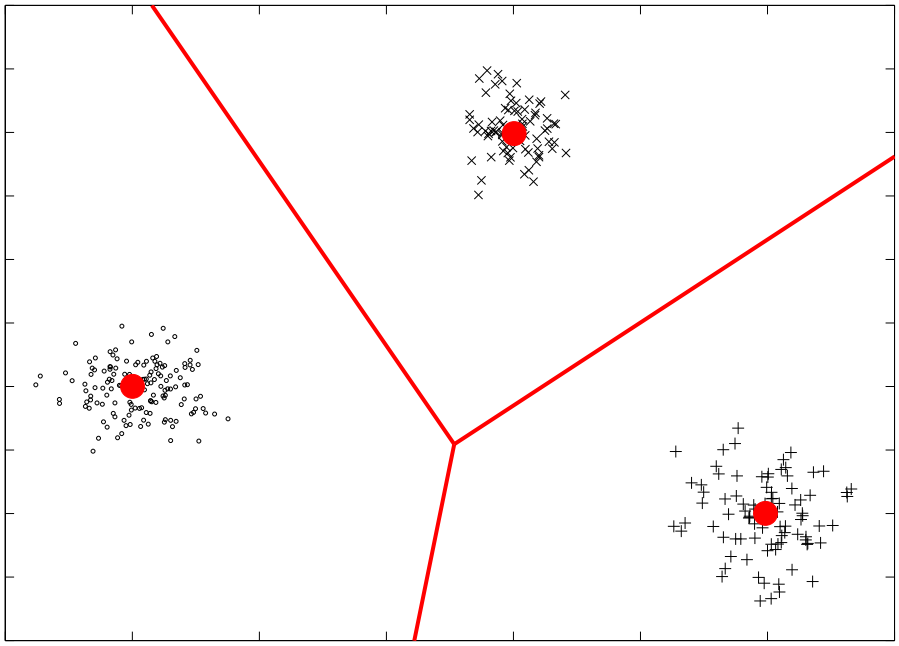
Cas du GNG-T Si l'on ne tient compte que des prototypes et non de la topologie du graphe, on peut procéder de la même manière que pour les K-moyennes. Cependant, nos hypothèses en impliqueront une autre utilisation, qui tient compte de la triangulation de Delaunay induite obtenue en appliquant la règle de Hebb.

Soit une variété V comportant H composantes connexes, la triangulation de Delaunay induite est alors formée de H sous-graphes $G_h, h = 1, \ldots, H$ de neurones connectés (cf. l'exemple de la figure 3.11, où cette triangulation comprend trois sous-graphes). Supposons que chaque composante connexe puisse être considérée comme une classe, tous les prototypes appartenant à une classe donnée sont alors les représentants de neurones éléments d'un des sous-graphes ; cette hypothèse prend en compte à la fois la proximité dans l'espace des attributs et la séparation des clusters. La partition finale est alors composée d'autant de *clusters* que de sous-graphes. Chaque *cluster* est constitué

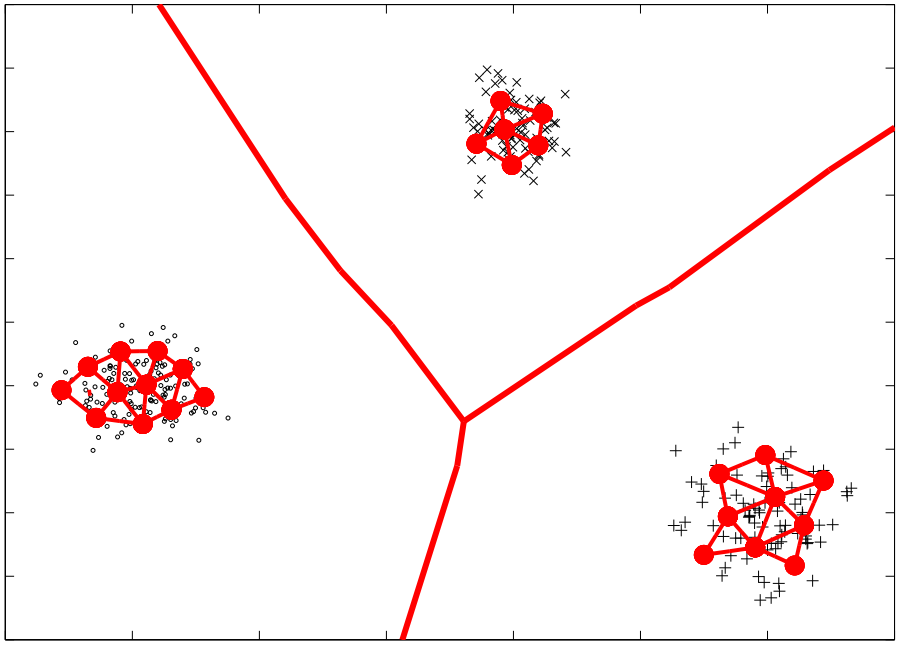
- de l'association des polyhèdres de Voronoï de tous les prototypes représentant les neurones d'un sous-graphe donné (pour une partition de $\mathbb{R}^p$).
- de l'association des cellules de Voronoï de tous les prototypes représentant les neurones d'un sous-graphe donné (pour une partition de V).
- par les points de l'échantillon $\mathcal{X}$ appartenant à l'association des cellules de Voronoï de tous les prototypes représentant les neurones d'un sous-graphe donné (pour une partition de $\mathcal{X}$).

Notons que de tels *clusterings* ne sont généralement pas uniformément centrés.

Résultats Des exemples où les deux méthodes de quantification vectorielle conduisent à des *clusterings* semblables sont présentés sur la figure 3.10. Remarquons qu'en imposant le même nombre de prototypes pour les K-moyennes que celui obtenu par le GNG-T, nous aboutirions à des résultats similaires mais le *clustering* ne pourrait se faire car nous n'aurions pas la carte topologique. D'autres exemples où l'un ou l'autre donne de meilleurs résultats sont présentés dans le paragraphe suivant. Bien que nous travaillerons en IRM rénale dans un espace de dimension 256, pour illustrer les difficultés rencontrées, nous nous sommes placés dans $\mathbb{R}^2$.

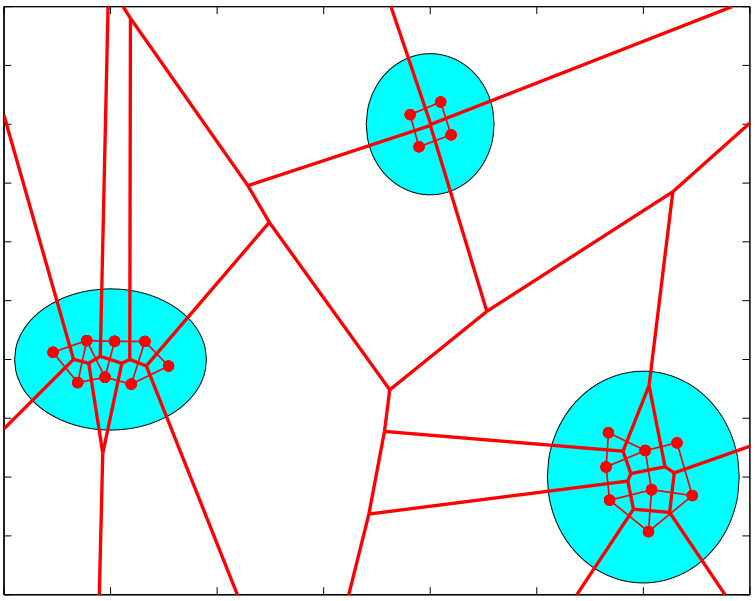


(a)

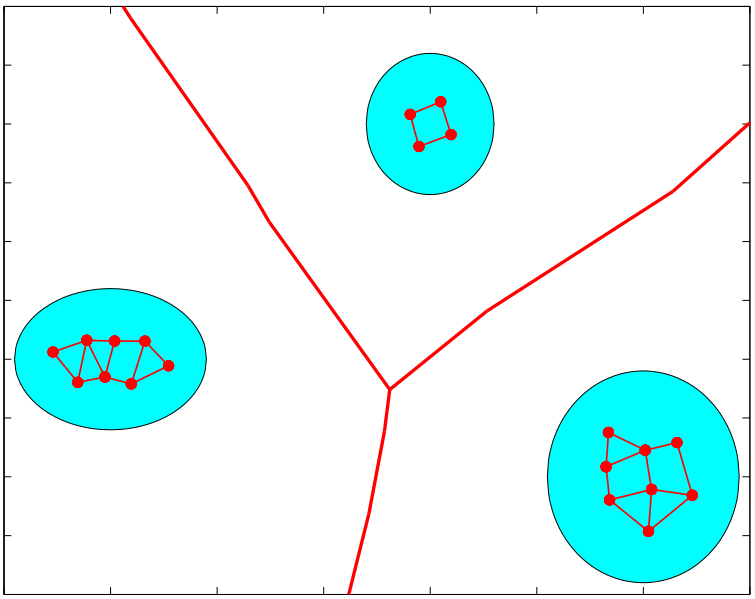


(b)

FIG. 3.10 – *Clustering* des mêmes données induit par un algorithme de K-moyennes (a) ou de GNG-T (b) : les deux *clusterings* obtenus pour l'échantillon $\mathcal{X}$ sont les mêmes, bien que les frontières de la partition soient légèrement différentes (les petits points, les croix et les signes plus représentent les observations ξ_i de l'échantillon $\mathcal{X}$, les gros points sont les prototypes après quantification vectorielle, reliés par des arcs pour GNG-T).



(a)



(b)

FIG. 3.11 – Formation d'une partition dans $\mathbb{R}^2$ à partir de prototypes et d'une approximation de la triangulation de Delaunay induite sur une variété V formée de trois composantes connexes (en bleu) : tous les polyhèdres de Voronoï (frontières en lignes rouges épaisses) des prototypes (points rouges) sont représentés sur (a) ; la partition finale (frontières en lignes rouges épaisses) apparaît sur (b), chaque *cluster* étant l'union des polyhèdres de Voronoï des prototypes représentant des neurones appartenant à un même sous-graphe.

Quelques problèmes pouvant être rencontrés

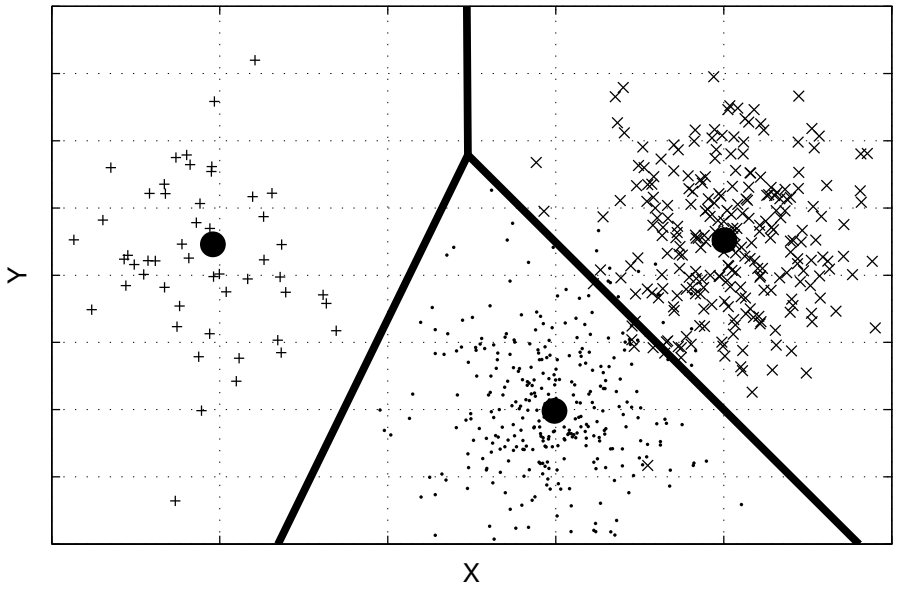
Choix a priori du nombre de classes non adapté à la distribution pour les K-moyennes Dans l'algorithme des K-moyennes, le nombre de *clusters* K doit être prédéfini et un *cluster* est traditionnellement associée à chaque prototype. Dans l'exemple de la figure 3.12, nous pouvons considérer que l'échantillon est bien décrit par trois prototypes qui conduisent à une partition convenable des données, alors que la quantification vectorielle avec quatre prototypes aboutit à des résultats moins satisfaisants : des parties importantes des deux *clusters* de droite sont mélangées dans la cellule de Voronoï du prototype ajouté. Pour des données réelles de loi de probabilité inconnue, il peut être difficile de prédéfinir une valeur de K adéquate. Bien que des méthodes basées sur la stabilité des résultats soient proposées dans cette optique dans la littérature, elles peuvent être difficiles à mettre en œuvre et des résultats récents peuvent mettre en doute leur validité [69].

Déséquilibre des représentants des classes Supposons que l'un des *clusters* recherchés soit représenté par un pourcentage relativement faible de l'ensemble des données (ce qui, en IRM de perfusion, pourra être le cas pour les cavités, voir paragraphe 5.1.2). S'il est relativement proche des autres, dans une classification par K-moyennes, ce *cluster* risque d'être associé de manière erronée à l'un des autres *clusters* en raison du faible poids des erreurs associées à ses représentants dans la distorsion globale. Il n'apparaîtrait donc pas dans la classification finale, même pour une valeur de K adaptée au problème. Un premier exemple est présenté figure 3.13 : le *cluster* central est mélangé à celui de gauche, alors que celui de droite est représenté par 2 prototypes. Au contraire, avec GNG-T, qui s'adapte à la densité locale, on obtient trois sous-graphes indépendants permettant d'établir une partition séparant convenablement les trois *clusters*. Une autre manière de procéder est d'ajouter un quatrième prototype pour la quantification vectorielle par K-moyennes : sur la figure 3.14, en regroupant les cellules de Voronoï des deux prototypes de droite obtenus pour $K = 4$, on aboutit à une partition convenable, ce qui n'était pas le cas avec $K = 3$. Il faut cependant trouver une règle pour effectuer le regroupement.

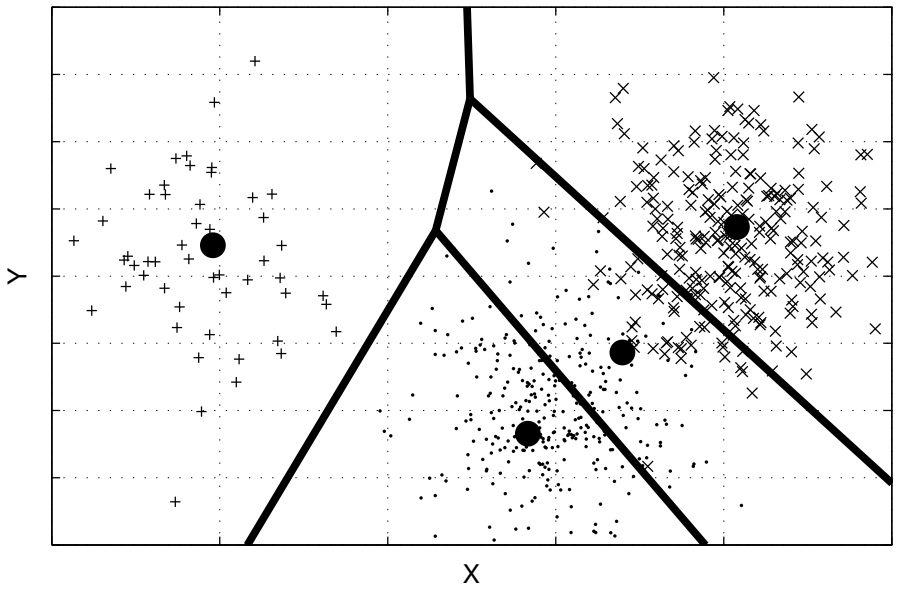
Un second exemple similaire est présenté sur la figure 3.15, où le déséquilibre entre classes est associé à un étalement encore plus marqué de l'un des nuages de points. La situation est encore plus défavorable parce que ce dernier est réparti sur les trois *clusters* finaux. Le problème peut être résolu de la même manière que précédemment.

Bruit entraînant des connections parasites dans la triangulation de Delaunay induite Dans le cas où les classes ne sont pas séparables de manière très évidente, c'est-à-dire si du bruit apparaît dans des zones où la loi $P_{\mathbf{X}}$ devrait être nulle (ce qui correspond à un élargissement du support de la variété idéale), il arrive qu'avec GNG-T, on obtienne non pas un ensemble de sous-graphes indépendants, mais un seul graphe, les prototypes étant tous connectés. Un tel exemple est présenté figure 3.16(a) pour une variété à trois composantes connexes dans un espace de dimension 2.

Pour aboutir à une approximation du *clustering* désiré, il est nécessaire de briser certains arcs non significatifs (figure 3.16(b)). Les critères utilisés pour cela dépendent de l'application [61] ; en IRM rénale, nous injecterons de la connaissance par le biais d'un opérateur, tout en utilisant à la fois certains critères physiologiques équivalents à une



(a)



(b)

FIG. 3.12 – Exemple de quantification vectorielle pour un mélange gaussien à trois composantes : résultats de l'algorithme des K-moyennes pour $K = 3$ (a) et $K = 4$ (b) avec les polyhèdres de Voronoï associés (lignes épaisses) ; les petits points (.), les croix ($\times$) et les signes plus ($+$) représentent l'échantillon $\mathcal{X}$, les gros points sont les prototypes.

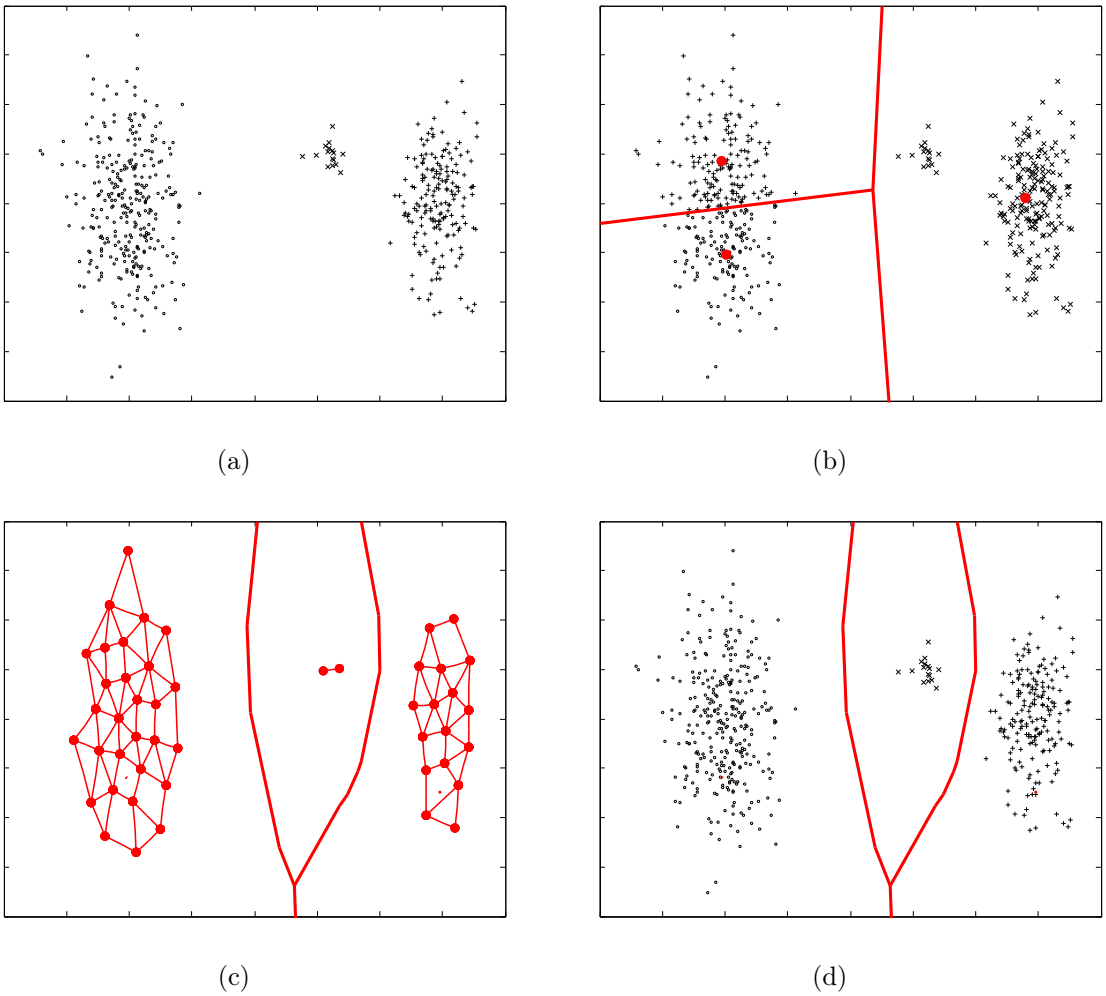


FIG. 3.13 – *Clustering* induit par quantification vectorielle dans le cas d'un déséquilibre des classes : l'échantillon $\mathcal{X}$ est représenté sur (a), les prototypes obtenus par K-moyennes et la partition induite le sont sur (b) ; le graphe obtenu par GNG-T et les frontières de la partition induite apparaissent sur (c), le *clustering* de $\mathcal{X}$ correspondant sur (d).

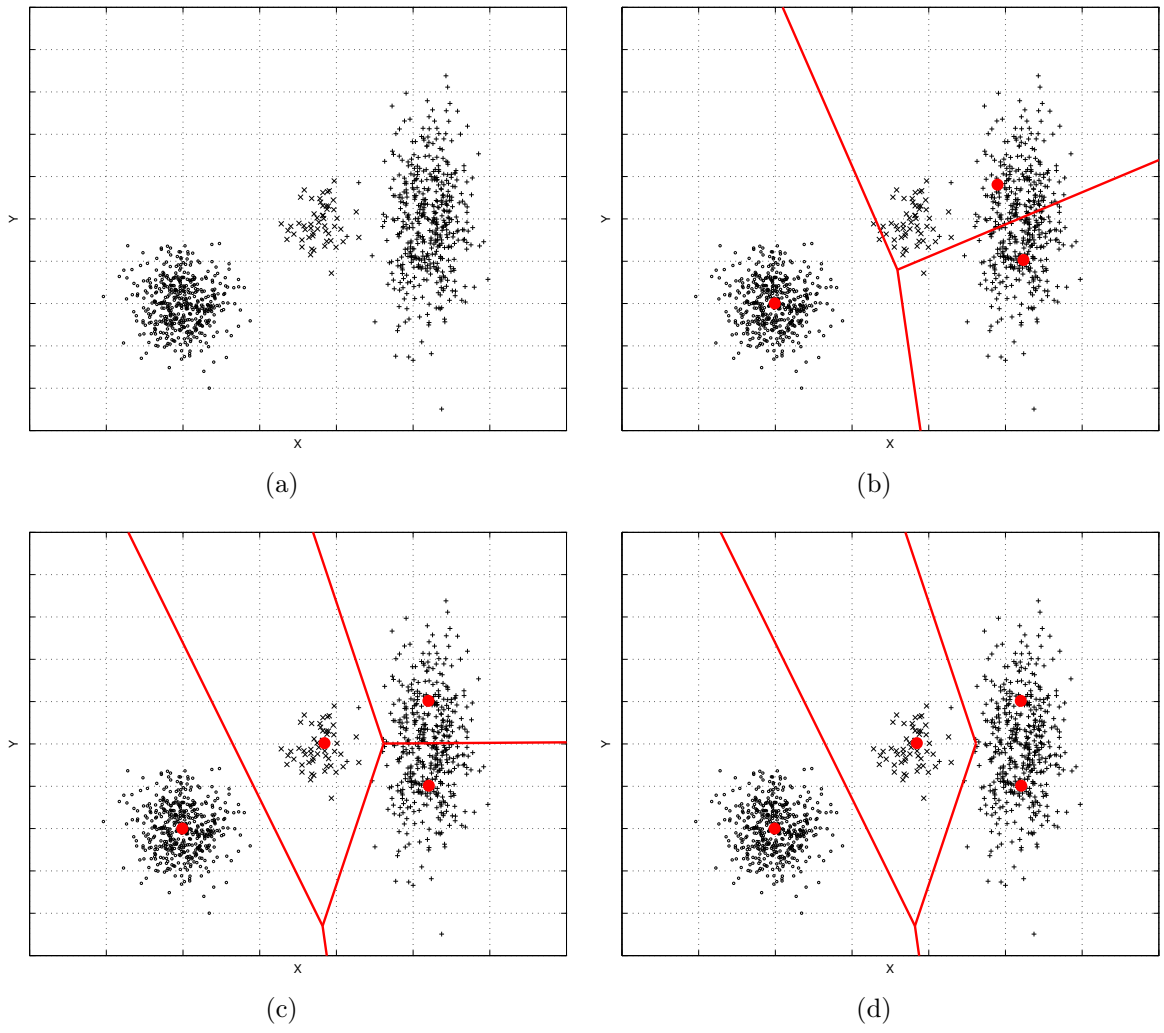


FIG. 3.14 – *Clustering* induit par quantification vectorielle dans le cas d'un déséquilibre des classes : l'échantillon $\mathcal{X}$ est représenté sur (a), les prototypes obtenus par K-moyennes et les polyhèdres de Voronoï le sont sur (b) pour $K = 3$ et sur (c) pour $K = 4$; sur (d), les cellules des deux prototypes de droite ont été regroupées.

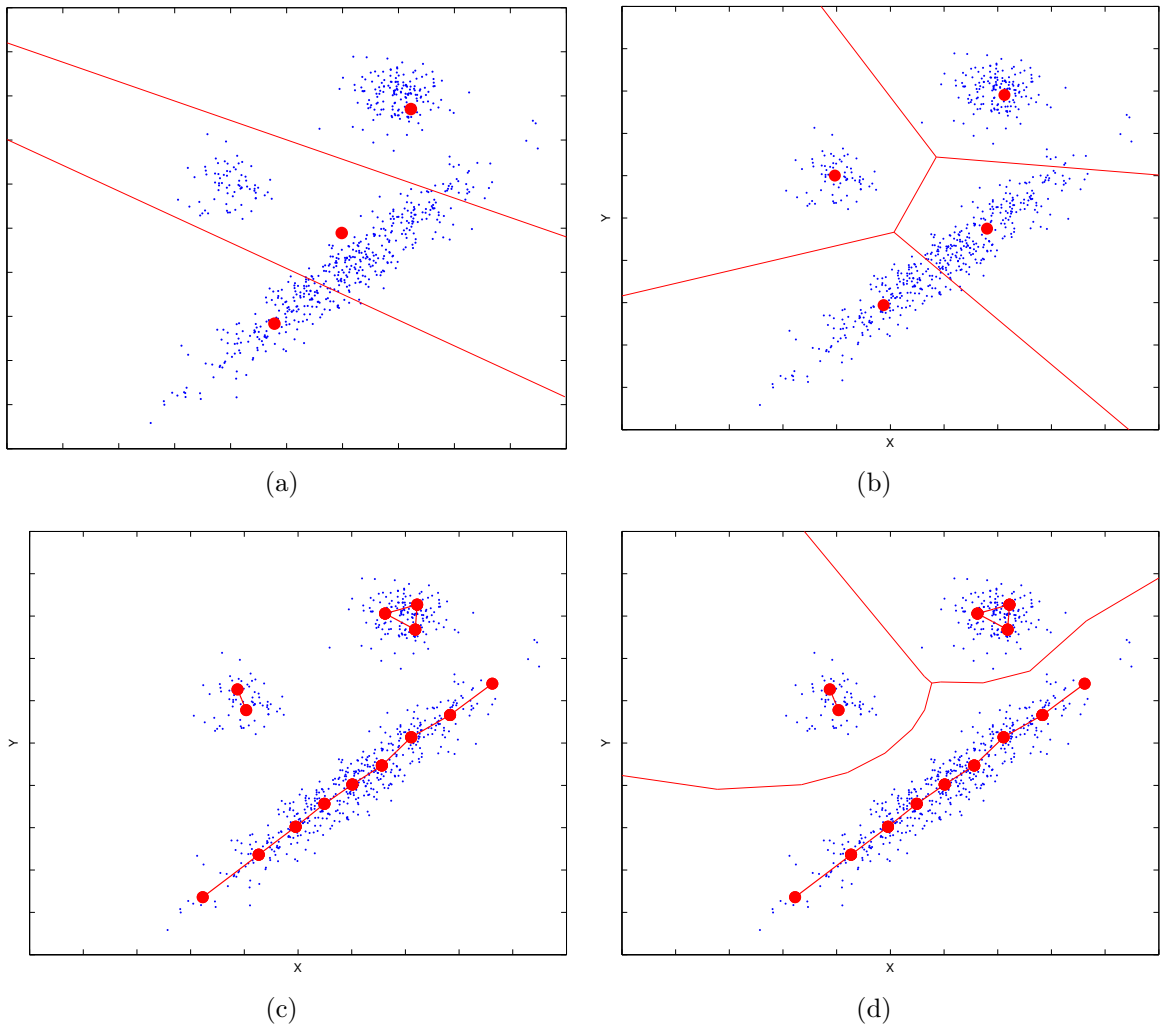
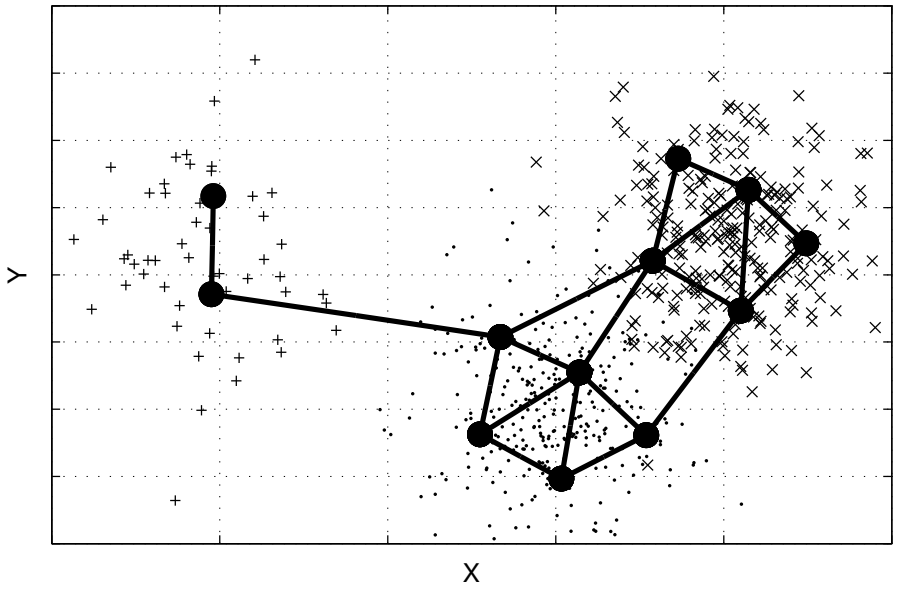
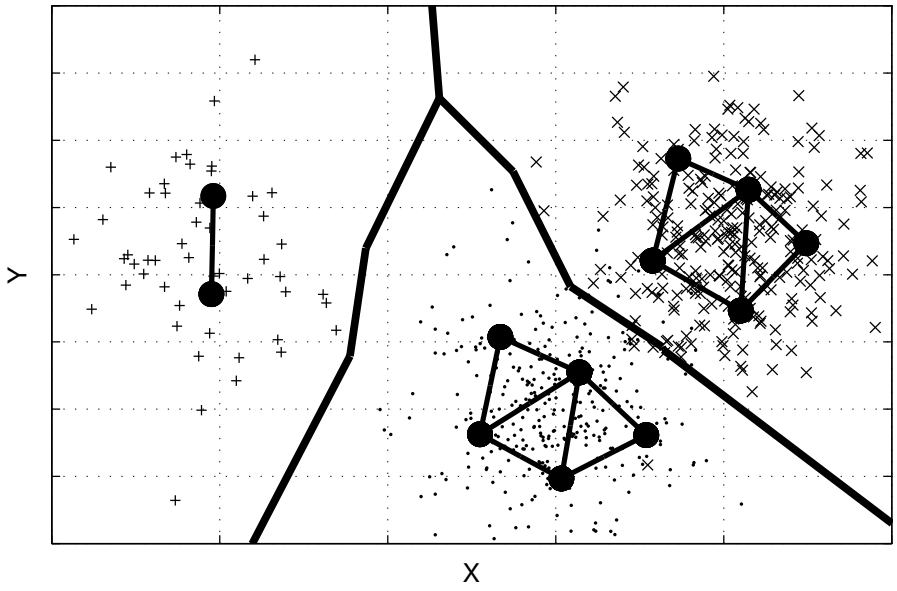


FIG. 3.15 – *Clustering* induit par quantification vectorielle dans le cas d'un déséquilibre des classes : l'échantillon $\mathcal{X}$ est représenté par des points ; sur les prototypes obtenus par K-moyennes et les polyhèdres de Voronoï apparaissent sur (a) pour $K = 3$ et sur (b) pour $K = 4$; les résultats du GNG-T sont représentés sur (c), les frontières induites sur (d).



(a)



(b)

FIG. 3.16 – Exemple de quantification vectorielle par GNG-T pour un mélange gaussien à trois composantes : résultats obtenus avec le graphe complet (a) et avec la partition finale (lignes épaisses) après cassure des arcs non significatifs (b) ; les petits points (.), les croix ($\times$) et les signes plus (+) représentent l'échantillon $\mathcal{X}$, les gros points sont les prototypes reliés par des arcs.

séparation par hyperplan dans l'espace des attributs et la carte topologique obtenue par le GNG-T (cf. paragraphe 5.1.5 page 161) : une interprétation d'une telle séparation en dimension 2 est présentée ci-dessous. En tout cas, chaque classe correspond encore à une association de polyhèdres de Voronoï. Il se peut que, dans certains de ces cas, l'algorithme des K-moyennes donne directement des résultats satisfaisants : sur la figure 3.13(a), des données identiques à celles de la figure 3.16 sont classifiées en 3 *clusters* par un algorithme de K-moyennes, aboutissant à un *clustering* des données similaire à celui du GNG-T après cassure des arcs parasites.

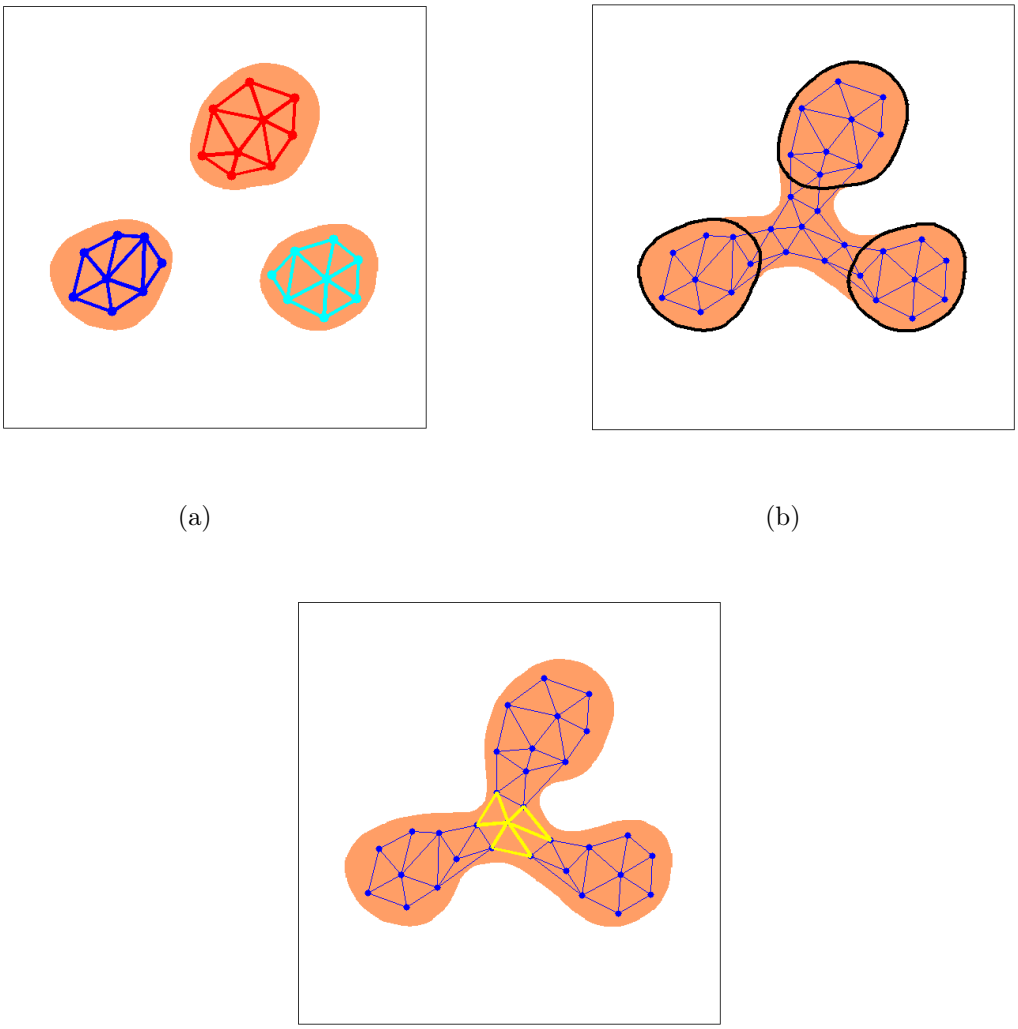
Séparation des *clusters* par injection de connaissance : intérêt de la carte topologique Supposons par exemple que l'on sache que la variété « idéale »(figure 3.17(a)) comprend trois composantes connexes correspondant à trois *clusters*, et que l'un d'entre eux contient les points d'abscisses les plus élevées (*cluster* inférieur droit, de couleur cyan). Suite à l'élargissement du support de cette variété en présence de bruit, un seul graphe entièrement connecté est obtenu par un algorithme de GNG-T (figure 3.17(b)), au lieu des trois sous-graphes attendus (figure 3.17(a)). Comment briser les arcs parasites, représentés en trait gras jaune sur la figure 3.17(c), pour obtenir un *clustering* qui serait une approximation convenable de celui obtenu sur la figure 3.17(a) ? Une première idée serait d'attribuer à un même *cluster* tous les prototypes dont l'abscisse est supérieure à un seuil donné, matérialisé ici par la droite verte sur la figure 3.18(a) : ceci revient à réaliser une séparation de l'espace des attributs par un hyperplan. Ce critère est cependant insuffisant pour obtenir un *clustering* approchant l'idéal. En effet, si nous représentons en cyan les prototypes conservés pour différents niveaux décroissants de ce seuil (figure 3.18, première colonne), nous pouvons observer (figure 3.18(e)) qu'une grande partie du *cluster* supérieur est sélectionnée avant que l'inférieur droit ne soit convenablement séparé des autres.

En revanche, prenons maintenant aussi en compte la topologie indiquée par le graphe. Appelons semence le prototype qui vérifie « au mieux »le critère retenu, à savoir ici le point d'abscisse la plus élevée (cf. figure 3.19). Seuls les prototypes qui vérifient le critère de seuil et et qui sont reliés à la semence par un chemin ne passant que par des prototypes vérifiant eux-mêmes le critère sont alors considérés comme appartenant au *cluster* cyan (figure 3.18, seconde colonne). Pour la valeur de seuil matérialisée sur la figure 3.18(f), les prototypes jaunes sont cette fois-ci exclus du *cluster* cyan. Il est ainsi possible, en fixant une bonne valeur de seuil, et en brisant les arcs considérés comme parasites, représentés ici en brun (figure 3.20(a)), de trouver une approximation convenable du *cluster* cyan d'origine (figure 3.20(b)).

En procédant de manière similaire pour séparer les deux autres *clusters*, on peut ainsi aboutir, à condition d'injecter la bonne connaissance, à un *clustering* beaucoup plus satisfaisant (figure 3.20(c)) que si l'on ne tient pas compte de la topologie de la variété.

Conclusion

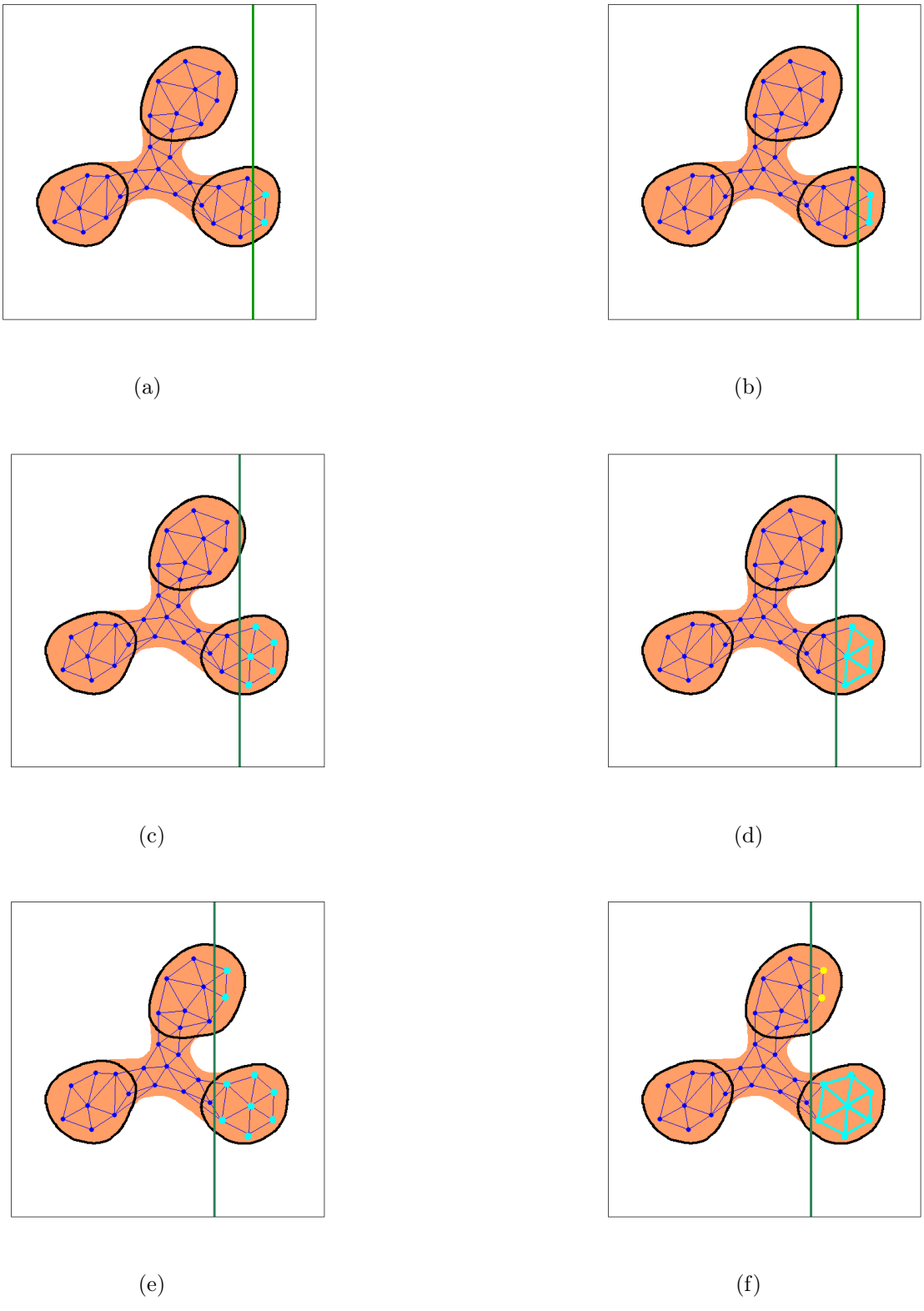
Deux méthodes pour déduire de la quantification vectorielle une fonction de *clustering* des données ou de l'espace entier ont été présentées, correspondant chacune à une manière différente d'effectuer la quantification vectorielle. L'algorithme des K-moyennes est utilisé en supposant que chaque classe correspond à la cellule de Voronoï d'un des



(a) (b)

(c)

FIG. 3.17 – Effet du bruit : quantification vectorielle et carte topologique obtenue par GNG-T sur une variété de support orange en l'absence de bruit (a) et en cas d'élargissement du support dû au bruit (b). Les arcs parasites à briser sont représentés en trait gras jaune sur (c).



(a) (b) (c) (d) (e) (f)

FIG. 3.18 – Séparation des *clusters* selon la méthode exposée dans le paragraphe 3.1.4 page 83 : coupure par hyperplan sans prise en compte de la topologie de la variété pour des seuils décroissants ((a), (c) et (e)) puis en tenant compte de la carte respectant la topologie de la variété ((b), (d) et (f)).

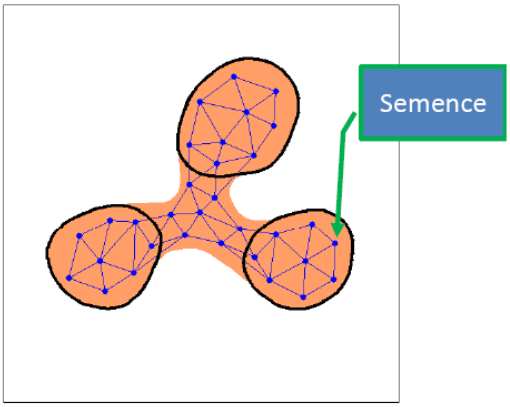


FIG. 3.19 – Prototype semence pour la méthode de séparation des *clusters* exposée dans le paragraphe 3.1.4 page 83.

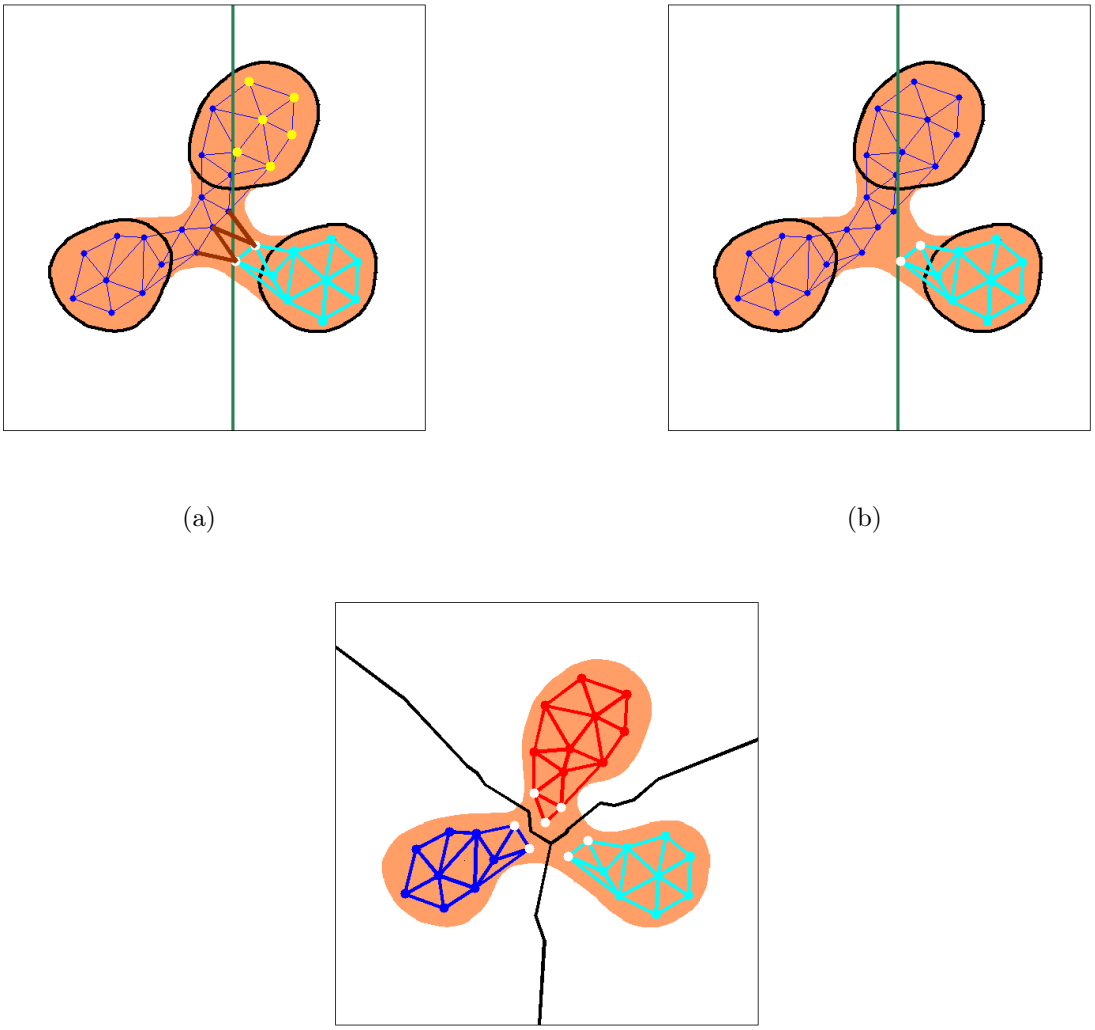
prototypes obtenus, ce qui suppose de définir préalablement le nombre K adéquat grâce à des connaissances a priori. En revanche, pour celui du GNG-T, une classe correspond à l'association des cellules de Voronoï de chaque ensemble de prototypes appartenant à un même sous-graphe de la carte respectant la topologie. La validation quantitative des résultats est abordée au paragraphe suivant.

3.1.5 Analyse des résultats : comparaisons de *clusterings*

Dans les paragraphes précédents, nous avons exposé deux méthodes de *clustering* par quantification vectorielle. Comment pouvons-nous qualifier et quantifier la validité des *clusterings* obtenus, avant d'éventuellement étendre les résultats à d'autres données tirées suivant la même loi de probabilité ? Nous soulignerons d'abord l'insuffisance de la mesure de performance du quantificateur vectoriel pour estimer la qualité des résultats. Nous examinerons ensuite les différentes approches proposées dans la littérature et verrons si elles sont utilisables dans le cadre de notre étude. Notons que nous ne nous préoccuperons pas ici des aspects algorithmiques.

Insuffisance de la mesure de performance d'un quantificateur vectoriel pour estimer la qualité d'un *clustering*

Rappelons que la mesure de performance d'un quantificateur est directement la distorsion globale (cf. paragraphe 3.1.1). L'estimateur $\frac{1}{N}\sum_{i=1}^{N} d(\xi_i, w(\xi_i))$ tend vers la distorsion lorque N tend vers l'infini, d'après la loi des grands nombres. Cet estimateur est directement la fonction minimisée par l'algorithme des K-moyennes. Cette mesure est difficile à interpréter dans l'absolu pour juger de la qualité d'un *clustering*. En particulier, elle n'est pas invariante par changement d'échelle, et ne permet pas de comparer des *clusterings* d'un même ensemble de données comportant un nombre de *clusters* différent [70].



(a) (b)

(c)

FIG. 3.20 – Séparation des *clusters* par la méthode exposée dans le paragraphe 3.1.4 page 83 : les arcs parasites à briser pour séparer le *cluster* cyan sont représentés en brun (a), les prototypes jaunes ont une abscisse supérieure au seuil retenu (droite verte) mais ne sont pas reliés à la semence par un chemin ne passant que par des prototypes vérifiant eux-mêmes le critère ; sous-graphes obtenus après séparation du cluster cyan (b) ; *clustering* final (c).

Par ailleurs, pour l'algorithme de GNG-T, la formule ne tient pas compte du regroupement correct ou non entre polyhèdres de Voronoï, alors que, pour certaines composantes connexes, le centroïde obtenu par K-moyennes n'est pas nécessairement très représentatif du *cluster* (cf. figure 3.9). Il nous faudra donc définir d'autres mesures de qualité.

Deux grandes tendances se dégagent dans la littérature :

1. pour une application donnée, des comparaisons sont effectuées avec une vérité-terrain. C'est sans conteste la méthode de validation la plus utilisée dans le domaine médical.
2. des mesures de la qualité du *clustering*, valables en soi sans comparaison avec une vérité-terrain, peuvent être définies.

Dans les paragraphes suivants, nous développerons dans cet ordre ces différentes notions.

Comparaison de deux *clusterings*, indépendamment de l'espace des paramètres

Nous souhaitons en pratique comparer un *clustering* obtenu par nos méthodes de quantification vectorielle et une « pseudo vérité-terrain » établie par un radiologue (voir paragraphe 4.4.3). En d'autres termes, en reprenant les notations du paragraphe 3.1.4, il s'agit de mesurer la ressemblance entre deux partitions finies d'un ensemble $\mathcal{X}$ d'observations, et non d'une variété $V \in \mathbb{R}^p$ ou de $\mathbb{R}^p$ entier car la classification par un radiologue n'y donne pas accès, dans la mesure où elle n'est pas uniformément centrée. La manière dont les *clusters* ont été générés n'est pas prise en compte : nous ne nous préoccuperons pas de savoir si l'algorithme utilisé aboutit à un optimum ou non.

Dans [71], l'auteur compare deux *clusterings* C et C' **du même ensemble de données** $\mathcal{X}$ de cardinal N, mais ayant éventuellement des nombres de classes K et K' différents. On définit une distance dans l'espace des partitions de cet ensemble $\mathcal{X}$, et non dans l'espace des paramètres ayant permis la formation du *clustering*. Tous les critères cités peuvent être évalués à partir de la matrice de confusion $\mathbf{M}$ de C et C', de taille $K \times K'$. L'élément M_{ij} est le nombre de points communs au *cluster* C_i de C et au *cluster* C'_j de C' :

$$M_{ij} = \text{card}(C_i \cap C'_j). \tag{3.38}$$

Comptage de paires Une première série de critères se base sur le comptage des paires de points sur lesquelles deux *clusterings* sont en accord ou non. Toute paire de points de X entre dans l'une des quatre catégories suivantes :

- {paires de points qui sont dans le même *cluster* tant dans C que dans C'}, dont le cardinal vaut N_{11}. Cet ensemble est aussi appelé *ensemble des bonnes paires*.
- {paires de points qui sont dans le même *cluster* dans C mais pas dans C'}, dont le cardinal vaut N_{10}.
- {paires de points qui sont dans le même *cluster* dans C' mais pas dans C}, dont le cardinal vaut N_{01}. Son union avec l'ensemble précédent forme l'*ensemble des mauvaises paires*.
- {paires de points qui sont dans des *clusters* différents tant dans C que dans C'}, dont le cardinal vaut N_{00}. Cet ensemble est aussi appelé *ensemble des paires neutres* parce

que les paires qu'il contient n'indiquent pas vraiment un accord ou un désaccord entre les deux *clusterings*, par opposition aux trois autres ensembles.

Remarquons que

$$N_{11} + N_{00} + N_{10} + N_{01} = \frac{N(N-1)}{2}. \tag{3.39}$$

N_{11} et N_{00} mesurent les classements cohérents, alors que N_{10} et N_{01} mesurent les désaccords [72]. Les N_{ij} peuvent être calculés à partir de la matrice de confusion. Par exemple [73],

$$2N_{11} = \sum_{i,j} M_{ij}^2 - N. \tag{3.40}$$

Divers coefficients faisant intervenir ces classements peuvent être définis. Quelques exemples sont donnés ci-dessous.

Indice de Rand L'indice de Rand, noté $R(C, C')$ est défini par [72] :

$$R(C, C') = \frac{N_{11} + N_{00}}{N_{11} + N_{00} + N_{10} + N_{01}} = \frac{N_{11} + N_{00}}{N(N-1)/2}. \tag{3.41}$$

L'indice de Rand représente la proportion de paires qui ne sont pas mauvaises (c'est-à-dire qui sont bonnes ou neutres) par rapport au nombre total de paires. Il varie entre 0 et 1. Il vaut 0 si et seulement s'il n'y a aucune cohérence entre les deux segmentations ($N_{11} = N_{00} = 0$). Il vaut 1 si et seulement si les deux *clusterings* sont les mêmes ($N_{10} = N_{01} = 0$). Il s'agit d'un coefficient symétrique, il n'est pas nécessaire de préciser si C ou C' est le *clustering* de référence. Cependant, son espérance n'est pas nulle lorsque l'on compare des partitions aléatoires. En pratique, sa valeur varie sur $[1 - \epsilon, 1]$ avec ϵ petit [71].

Coefficient de Jaccard Le coefficient de Jaccard, noté $J(C, C')$, est défini par

$$J(C, C') = \frac{N_{11}}{N_{11} + N_{10} + N_{01}}. \tag{3.42}$$

Ce coefficient varie entre 0 et 1. Il peut être considéré comme la proportion de bonnes paires par rapport à la somme des paires non neutres. Contrairement à l'indice de Rand, il ne tient pas compte des paires neutres.

Coefficients asymétriques de Wallace Les coefficients asymétriques de Wallace, notés $W_1(C, C')$ et $W_2(C, C')$, sont définis de la manière suivante :

$$W_1(C, C') = \frac{N_{11}}{\sum_i N_i(N_i - 1)/2} \tag{3.43}$$

et

$$W_2(C, C') = \frac{N_{11}}{\sum'_i N'_i(N'_i - 1)/2}, \tag{3.44}$$

où N_i est le cardinal du *cluster* C_i de C et N'_i celui du *cluster* C'_i de C'. Ils représentent la probabilité qu'une paire de points qui sont dans le même *cluster* dans C (respectivement dans C') soient aussi dans le même *cluster* dans C' (respectivement dans C).

Coefficient de Fowlkes et Mallows Le coefficient de Fowlkes et Mallows, noté $F(C, C')$ est la moyenne géométrique des coefficients de Wallace [73].

$$F(C, C') = \sqrt{W_1(C, C')W_2(C, C')}. \tag{3.45}$$

Ce coefficient représente un produit scalaire et peut être considéré comme une modification non linéaire du coefficient de Jaccard.

Les performances de ces coefficients sont peu aisées à comparer (un exemple particulier est cité dans [72]). En plus, il est difficile de supposer une linéarité entre la valeur du coefficient et la qualité de la partition dans l'intervalle utile de ces coefficients. D'après [71], ces coefficients ne dépendent que relativement peu de N pour des faibles valeurs de N, alors qu'asymtotiquement, d'après la loi des grands nombres, ils en sont indépendants quand N devient grand.

Correspondance d'ensembles (*set matching*) Dans [71], des mesures de similarité entrant dans cette catégorie sont décrites, comme les coefficients de Larsen ou de Van Dongen, ou la variation d'information. Cependant, ces coefficients souffrent pour la plupart des mêmes inconvénients que ceux basés sur le comptage de paires. Un critère particulièrement intéressant est l'erreur de classification D : à chaque *cluster* de C, est associé le *cluster* qui lui correspond le mieux dans C'. La grandeur D est alors la masse de probabilité totale qui ne correspond pas dans la matrice de confusion.

$$D(C, C') = 1 - \frac{1}{N} \max_{\pi \in \Pi} \sum_{k=1}^{K} M_{k,\pi(k)}, \tag{3.46}$$

où Π est l'ensemble des applications injectives de $\{1, 2, \ldots, K\}$ dans $\{1, 2, \ldots, K'\}$. Si $K = K'$, Π est l'ensemble des permutations de $\{1, 2, \ldots, K\}$. En lien avec le paragraphe 4.4.1 du chapitre 4, où nous comparons également des segmentations, supposons que la permutation qui fait se recouvrir au mieux les *clusters* corresponde aux mêmes étiquettes que dans les segmentations (ou classifications) comparées. Nous pouvons remarquer que dans le cas où $K = 2$, on montre que $D(C, C') = 1 - SC$, où SC est le coefficient de correspondance simple (*Simple matching coefficient*). Dans les mêmes conditions, pour $K = 3$,

$$D(C, C') = 1 - \mathrm{WCP}, \tag{3.47}$$

où WCP est la proportion d'échantillons bien classés, et représente ainsi le nombre d'échantillons mal classés. Ces coefficients sont donc simples à interpréter. De plus, ce critère peut s'interpréter comme un risque empirique dans le cadre de la théorie de la généralisation de la classification (cf. chapitre 3.2).

Peut-on estimer la qualité d'un *clustering*, sans comparaison avec une vérité terrain ?

Si la notion de *clustering* apparaît comme assez intuitive, il est cependant difficile de la définir dans un cadre axiomatique, qui serait pertinent indépendamment de l'application envisagée et des algorithmes utilisés [74]. Il paraît d'autant plus ardu de définir

dans l'absolu la notion de « *clustering* convenable ». Dans [70], une mesure de qualité de *clustering* est définie comme étant une fonction qui, à un ensemble d'observations $\mathcal{X}$ et à une partition finie de $\mathcal{X}$, associe un réel positif d'autant plus élevé que le *clustering* paraît concluant ; par exemple, la marge relative tient compte à la fois du caractère groupé des éléments d'un même *cluster* et de l'éloignement entre les différents *clusters*.

Il est même proposé d'estimer la capacité d'un ensemble de données à être plus ou moins bien divisé en *clusters* (notion de *clusterability*). Dans [75], il est démontré que les différentes notions proposées dans la littérature et ayant trait à ce sujet sont incohérentes entre elles, bien qu'elles essaient de traduire les mêmes propriétés intuitives.

Remarquons que, pour les mêmes raisons que la distorsion globale, ces mesures de qualité sont surtout adaptées au *clustering* dit uniformément centré et guère, en général, au *clustering* par composante connexe d'une variété. En plus, la quantification vectorielle peut permettre de réaliser des partitions sur des données qui ne sont pas nécessairement très bien séparées, alors qu'un certain nombre de mesures de qualité sont sensibles aux points aberrants. Seules des comparaisons avec une partition de référence nous permettront de valider nos résultats.

3.1.6 Analyse des résultats : comparaisons de segmentations

Outre les comparaisons de *clusterings*, il est possible d'utiliser des critères de comparaison de segmentations, qui sont peut-être plus faciles à interpréter. En effet, dans le cas traité, on a accès indifféremment à l'un ou l'autre. Cependant, on comparera des *clusterings* ayant un nombre de *clusters* quelconque. En revanche, pour les segmentations, on cherchera à comparer, pour chaque compartiment, deux segmentations obtenues par des méthodes différentes, ce qui revient aussi à comparer deux images binaires.

Un état de l'art sur les critères couramment utilisés dans le cas d'images médicales est dressé dans [76]. Les deux segmentations à comparer peuvent être considérées comme deux images binaires R (référence) et T (test), avec une étiquette égale à 1 à l'intérieur de la ROI, à 0 à l'extérieur. Quatre types de pixels peuvent alors être définis, d'après leurs étiquettes dans les deux images 3.21 :

Type de pixel	Etiquette dans R	Etiquette dans T
Vrai Positif (VP)	1	1
Faux Négatif (FN)	1	0
Faux Positif (FP)	0	1
Vrai Négatif (VN)	0	0

Les mesures de ressemblance suivantes sont fréquemment utilisées.

Pourcentage de recouvrement

Le pourcentage de recouvrement est défini par :

$$\mathrm{PR} = 100 \times \frac{\mathrm{VP}}{(\mathrm{VP} + \mathrm{FN})}. \tag{3.48}$$

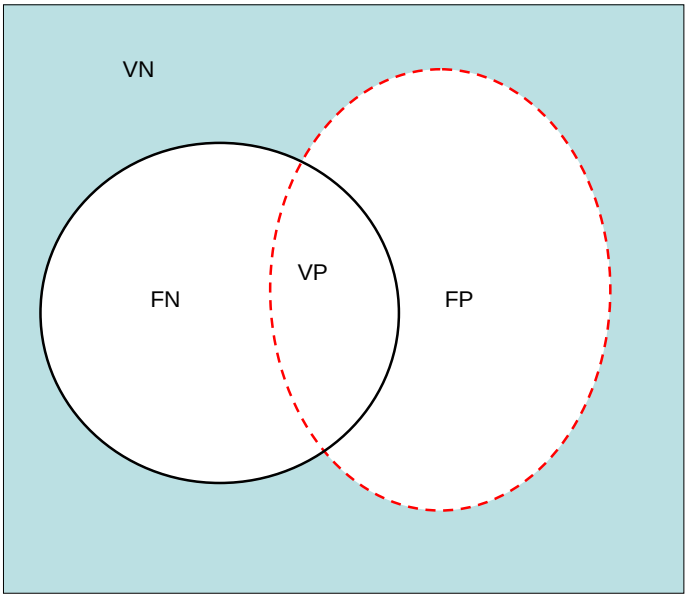


FIG. 3.21 – Les différents types de pixels intervenant dans les critères de similarité pour les comparaisons de segmentations : les pixels valant 1 dans la segmentation de référence R (respectivement test T) sont à l'intérieur de l'ellipse de gauche en trait continu noir (respectivement l'ellipse de droite en trait pointillé rouge).

Il s'agit du pourcentage de pixels de la ROI de référence qui sont également dans la ROI test.

Pourcentage de dépassement

Le pourcentage de dépassement est défini par :

$$\mathrm{PD} = 100 \times \frac{\mathrm{FP}}{(\mathrm{VP} + \mathrm{FN})}. \tag{3.49}$$

Il s'agit du nombre de pixels se trouvant dans la ROI test mais hors de la ROI de référence, divisé par le nombre de pixels dans la ROI de référence. Une segmentation parfaite conduirait à un PR de 100 % et à un PD de 0%. Des valeurs élevées pour à la fois PR et PD pour une segmentation donnée peuvent traduire une sur-segmentation de la zone correspondante, au sens où cette zone dans la segmentation test recouvrirait quasiment celle de la segmentation de référence tout en étant beaucoup plus étendue. Une valeur élevée de PD associée à une faible valeur de PR peut indiquer une position erronée de la zone.

Distance moyenne entre les contours

Il s'agit de la distance moyenne DM entre les pixels de bordure de la zone test et le pixel le plus proche de la zone de référence. Pour une segmentation idéale, DM = 0.

Distance de Hausdorff

La distance de Hausdorff DH est la distance maximale entre chaque pixel de bordure de la zone de test par rapport à la zone de référence ou vice-versa. Pour une segmentation idéale, DH = 0. Notons que cette distance est très sensible aux pixels aberrants, puisqu'il suffit d'un seul pixel éloigné de la bordure pour que DH prenne une valeur élevée.

Coefficient de correspondance simple

Le coefficient de correspondance simple est défini par :

$$\mathrm{SC} = \frac{\mathrm{VP} + \mathrm{VN}}{\mathrm{VP} + \mathrm{VN} + \mathrm{FP} + \mathrm{FN}}. \tag{3.50}$$

SC représente la proportion de pixels bien classés, et vaut 1 en cas de segmentation parfaite.

Indice de similarité

L'indice de similarité, ou indice de Dice, est défini par :

$$\mathrm{SI} = \frac{2 \times \mathrm{VP}}{2 \times \mathrm{VP} + \mathrm{FN} + \mathrm{FP}}. \tag{3.51}$$

Dans le cas d'une segmentation parfaite, SI vaudrait 1.

Coefficient de Tanimoto

Le coefficient de Tanimoto est défini par :

$$\mathrm{CT} = \frac{\mathrm{VP} + \mathrm{VN}}{\mathrm{VP} + \mathrm{VN} + 2 \times (\mathrm{FP} + \mathrm{FN})}. \tag{3.52}$$

CT vaut 1 pour une segmentation parfaite.

Comparaison des coefficients de correspondance simple, de Tanimoto, et de l'indice de similarité

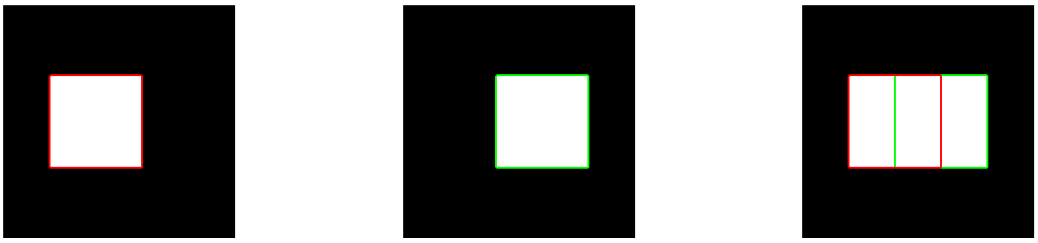
Pour comparer les trois derniers coefficients, qui sont sensibles à la fois aux différences en taille et en position [77], nous proposons de les évaluer pour les segmentations des figures 3.22, 3.23, 3.24 et 3.25. Les résultats apparaissent dans le tableau 3.1, où la taille de la ROI de référence est exprimée en pourcentage du nombre total de pixels. Les segmentations des figures 3.22 et 3.23 sont telles que les ROI de R et T sont de taille identique et partagent la moitié de leurs pixels, mais la ROI de R représente 32% du nombre total de pixels dans le premier cas et 16% dans le second. Dans les segmentations des figures 3.24 et 3.25, la ROI de T couvre complètement celle de R, qui est deux fois plus petite qu'elle. Dans le premier cas, cette dernière représente 24% de l'ensemble des pixels, et 12% dans le second. Notons que l'indice SI prend les mêmes valeurs si on échange le rôle de R et de T, ce qui n'est pas le cas pour les deux autres coefficients. Remarquons aussi que, pour

ces exemples auxquels on ajoute le cas où R et T sont identiques, la plage de variation de SC est nettement plus restreinte que celles de CT et SI (respectivement entre $0,68$ et 1, entre $0,52$ et 1, et entre $0,50$ et 1). Par ailleurs, si nous comparons les colonnes d'un même tableau, il apparaît que SC et CT donnent des résultats plus favorables pour les ROI de petite taille, en raison de l'influence des pixels étiquetés VP, alors que SI reste le même. Remarquons également que nous ne pouvons pas deviner le défaut que présente la segmentation uniquement grâce à la valeur de ces coefficients.

Ces remarques, ainsi que des exemples de mesures de ressemblance pour quelques segmentations rénales (données au paragraphe 4.4 page 130) nous guideront pour choisir les critères les mieux adaptés à notre problème.

(a) (b) (c)

FIG. 3.22 – Segmentation de référence R (a), segmentation test T (b) et comparaison des deux (c).

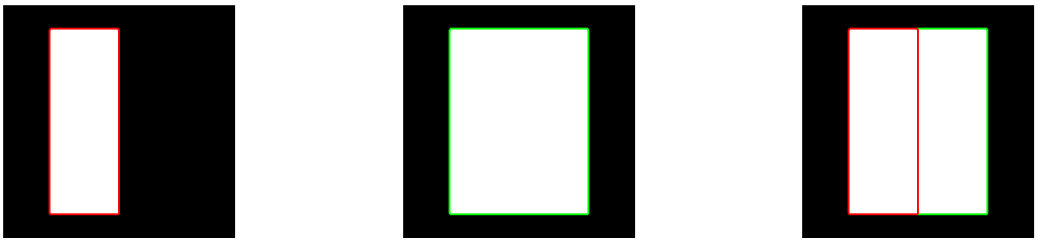


(a) (b) (c)

FIG. 3.23 – Segmentation de référence R (a), segmentation test T (b) et comparaison des deux (c).

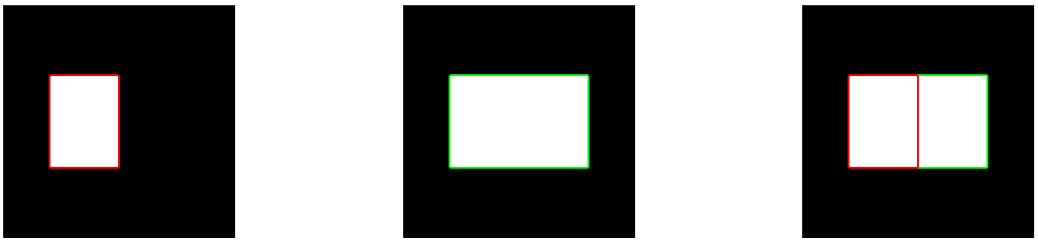
3.1.7 Conclusion

Deux méthodes de *clustering* par quantification vectorielle destinées à être testées sur des données d'IRM de perfusion rénale simulées ou réelles ont été proposées : l'une utilise un algorithme de K-moyennes, l'autre un GNG-T. Des exemples comparatifs de résultats en dimension 2 montrent qualitativement que, selon les cas, l'une ou l'autre peut être mieux adaptée. Des mesures de distance entre *clusterings* permettent de comparer les



(a) (b) (c)

FIG. 3.24 – Segmentation de référence R (a), segmentation test T (b) et comparaison des deux (c).



(a) (b) (c)

FIG. 3.25 – Segmentation de référence R (a), segmentation test T (b) et comparaison des deux (c).

Taille de la ROI de référence (% du total)	32	16
SC	0,68	0,84
CT	0,52	0,72
SI	0,50	0,50

(a) Segmentations des figures 3.22 et 3.23

Proportion de la ROI de référence (% du total)	24	12
SC	0,76	0,88
CT	0,61	0,79
SI	0,67	0,67

(b) Segmentations des figures 3.24 et 3.25

TAB. 3.1 – Exemples de valeurs pour les critères de similarité SC, CT et SI.

résultats obtenus à des pseudo vérités-terrain. Le choix des critères retenus sera discuté au paragraphe 4.4.1, d'après les spécificités de notre application.

Pour celle-ci, à ce stade, on aura construit plusieurs classificateurs permettant à la fois de classer les voxels de la coupe rénale principale en trois compartiments anatomiques à partir de leurs courbes temps-intensité. On aura pu tester leurs résultats pour chaque rein de la petite base de données dont nous disposons actuellement, par exemple en évaluant l'erreur de classement commise par rapport à la segmentation d'un radiologue sur la coupe rénale principale. Il reste cependant à segmenter les autres coupes ; ce que l'on peut faire avec n'importe lequel des classificateurs construits, puisqu'ils permettent tous d'attribuer un compartiment à tout voxel dont on connaîtrait la courbe temps-intensité. L'étude théorique de cette généralisation fait l'objet de la seconde partie de ce chapitre.

3.2 Théorie de la généralisation

Tout comme dans la première partie de ce chapitre, nous commençons par donner une explication intuitive de la démarche proposée, pour introduire l'exposé du cadre mathématique. Les correspondances précises avec notre problème seront exposées en détail au chapitre 4.

Les acquisitions IRM dont nous disposons sont réalisées simultanément pour plusieurs coupes (on acquiert en réalité un volume, mais les voxels sont 5 à 10 fois plus allongés dans la direction perpendiculaire au plan de coupe). En d'autres termes, à chaque instant d'acquisition, une image de chacune de ces coupes est enregistrée. Les méthodes dont le cadre mathématique est défini dans le chapitre 3.1 sont utilisées pour segmenter la coupe rénale principale, qui est celle contenant le plus grand volume rénal, c'est-à-dire celle où le rein est représenté par le plus grand nombre de voxels. Elles permettent de construire des classificateurs qui, à chaque voxel de cette coupe, représenté par sa courbe temps-intensité, associent un compartiment rénal. Il reste alors à segmenter les autres coupes du même rein.

Pour faire cette segmentation, nous pouvons a priori utiliser n'importe lequel des classificateurs précédemment construits. En effet, à chaque fois, une partition de $\mathbb{R}^p$ a

été réalisée, donc nous pouvons attribuer une classe à tout vecteur de $\mathbb{R}^p$. Il est alors possible de classer les voxels des autres coupes à partir de leur courbe temps-intensité en généralisant les résultats obtenus pour la première. Notons cependant que, pour la suite, la manière dont ces classificateurs a été obtenue n'entre pas en considération.

Le cadre mathématique dans lequel nous nous plaçons est celui de la théorie de la généralisation de Vapnik [78], qui se base sur la minimisation du risque empirique. Notre application est très différente de celle envisagée par Vapnik, mais les résultats peuvent s'appliquer à notre cas, comme nous le verrons au chapitre 5. D'autres points de vue sont possibles.

Considérons le système de la figure 3.26. Pour les reins de notre base de tests, nous disposons, pour la coupe de référence, d'une segmentation faite par un radiologue[5]. A chaque voxel, dont l'attribut est sa courbe temps-intensité, on peut donc associer une classe correspondant à un compartiment anatomique en utilisant :

- soit la segmentation faite par le radiologue, considérée comme la référence ; la classe résultat est représentée par la variable y.
- soit la réponse d'un des M classificateurs construits précédemment, indicés par la variable α, et dont la classe résultat, en sortie de l'entité LM, est représentée par la fonction $f(\xi, \alpha)$.

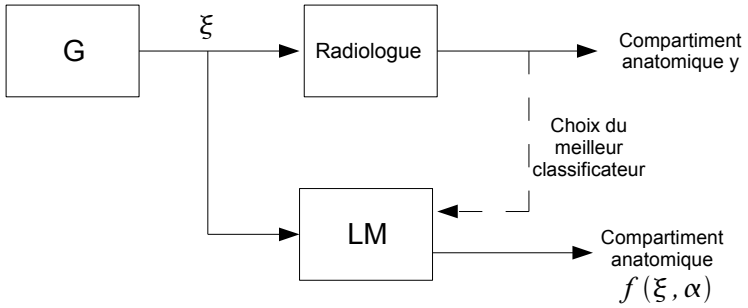


FIG. 3.26 – Généralisation : soit ξ le vecteur temps-intensité d'un voxel x et y le compartiment anatomique que lui associe un radiologue ; l'entité LM permet de choisir, parmi M classificateurs, le meilleur de ceux-ci pour les paires (ξ, y) de la base d'exemples tirés de la coupe rénale principale, c'est-à-dire celui qui minimise l'erreur de classement pour ces paires. Ce choix définitif étant effectué après la présentation de N exemples, la machine doit, pour toute entrée ξ, donner un compartiment anatomique $\hat{y} = f(\xi, \alpha)$. L'objectif est de retourner le plus souvent possible le même compartiment que le radiologue.

L'objectif est que LM retourne le plus souvent possible le même compartiment $\widehat{y} = f(\xi, \alpha)$ que le radiologue. Supposons que l'on présente les vecteurs ξ les uns après les autres. Au $\ell^{\text{ième}}$ vecteur ξ présenté, pour chacun des M classificateurs, il est possible de calculer l'erreur de classement commise sur les ℓ paires (ξ, y). A cette étape, le classificateur que l'on peut considérer comme étant le meilleur est celui dont l'erreur de classement

[5]Pour l'application finale, sur les cas autres que ceux de la base de test, en réalité, nous n'aurons pas une telle segmentation, mais ce cas « fictif » permet de mettre en place les concepts.

est minimale ; l'entité LM rend alors le résultat $f(\xi, \alpha)$ correspondant. A chaque ξ supplémentaire, le classificateur optimum peut changer.

A priori, nous allons garder comme classificateur celui qui donne la plus petite erreur de classification par rapport à la segmentation du radiologue pour la coupe de référence, une fois présentés les N vecteurs ξ des voxels de cette coupe. On peut alors utiliser ce classificateur pour classer les voxels des autres coupes.

Cependant, deux questions se posent : le classificateur qui donne les meilleurs résultats pour la coupe de référence est-il vraiment le classificateur optimum, c'est-à-dire celui qui minimise la probabilité d'erreur de classification pour des voxels provenant d'autres coupes ? Si le classificateur choisi n'est pas le classificateur optimum, leurs performances sont-elles très différentes ? Par ailleurs, mêmes si elles sont bonnes sur la coupe de référence, restent-elles acceptables si on généralise la classification aux autres coupes ? En effet, on peut par exemple imaginer qu'un classificateur donnant de bons résultats sur un « faible » nombre de voxels, qui ne seraient pas suffisamment représentatifs, voie ses performances se dégrader largement quand on l'essaie sur d'autres échantillons.

Ceci peut être décrit dans le cadre de la théorie de la généralisation des résultats obtenus à partir des exemples de la première coupe pour un problème dit « de reconnaissance de forme », à condition que nous puissions considérer que les courbes temps-intensité des voxels des autres coupes soient des réalisations d'un vecteur aléatoire de même loi que pour la coupe de référence. Cette hypothèse est plausible dans la mesure où elles correspondent au passage du même traceur dans des tissus de nature identique. Il est alors possible de borner la probabilité d'erreur de classement lors de l'extension de la classification.

Dans ce chapitre, nous présentons des résultats de la théorie de la généralisation valables dans un cadre général, et d'autres s'appliquant uniquement dans le cadre de la reconnaissance de forme (définie au paragraphe 3.2.2). Nous définirons d'abord un estimateur de risque moyen, qui correspondra dans notre cas à la probabilité d'erreur de classement, montrerons à quelles conditions il est consistant, puis étudierons sa vitesse de convergence afin d'évaluer si le nombre d'exemples utilisés pour choisir le classificateur optimum est suffisant pour obtenir des résultats convenables lors de la généralisation.

3.2.1 Théorie de la généralisation

Dans le schéma de la figure 3.27 [78], G est un générateur de vecteurs aléatoires identiquement distribués tirés suivant une loi de probabilité $P_{\mathbf{X}}(\xi)$ inconnue[6]. Un oracle, ou superviseur S, renvoie, pour le vecteur d'entrée $\mathbf{X}$, une valeur de sortie $Y : \Omega \longrightarrow \mathcal{I}$, où $\mathcal{I} = \{1, 2, \ldots, K\}$ d'après une loi de probabilité conditionnelle $F(y|\xi)$ fixe mais inconnue (cela inclut le cas où le superviseur utilise une fonction $g(\xi)$). On appelle machine (*learning machine*) et on note LM une entité capable d'implanter un ensemble de fonctions $\{f(\xi, \alpha), \alpha \in \Lambda\}$, où Λ est un ensemble de paramètres. Le paramètre α peut être totalement abstrait, de manière qu'on puisse considérer n'importe quel ensemble de fonctions. Il s'agit de choisir, parmi l'ensemble de fonctions $\{f(\xi, \alpha), \alpha \in \Lambda\}$, celle qui est la meilleure estimation de la réponse du superviseur. La sélection de la fonction désirée se

[6]Le problème est donc différent de celui de l'estimation pure puisque les vraisemblances ne sont pas connues.

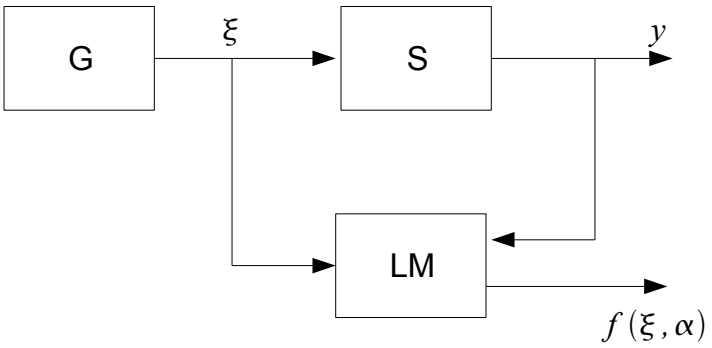


FIG. 3.27 – Modèle d'apprentissage à partir d'exemples : pendant la phase d'apprentissage, la machine LM (*learning machine*) « observe » les paires (ξ, y) de la base d'apprentissage. Cette phase étant achevée, la machine doit, pour toute entrée ξ, retourner une sortie $\hat{y} = f(\xi, \alpha)$. L'objectif est de retourner une valeur proche de celle de la réponse y du superviseur S.

fait en utilisant une base d'apprentissage de N observations indépendantes et identiquement distribuées tirées selon la loi de probabilité jointe $P_{\mathbf{X},Y}(\xi, y) = P_{\mathbf{X}}(\xi)F(y|\xi)$. On peut mesurer la perte (*loss*) en définissant $L[y, f(\xi, \alpha)]$ qui quantifie le coût de la décision issue de LM « la réponse est $f(\xi, \alpha)$ » alors que sa vraie valeur, c'est-à-dire la réponse du superviseur, est y. La perte moyenne est donnée par la fonction risque moyen :

$$R(\alpha) = \int L[y, f(\xi, \alpha)] \mathrm{d}P_{\mathbf{X},Y}(\xi, y). \tag{3.53}$$

Il s'agit de minimiser cette fonction dans le cas où la loi de probabilité jointe $P_{\mathbf{X},Y}(\xi, y)$ est inconnue et où toute l'information disponible est contenue dans la base d'apprentissage.

Pour alléger les notations, on pose $\mathbf{Z} = (\mathbf{X}, Y)$, $\mathbf{z} = (\xi, y)$, $P_{\mathbf{Z}}(\mathbf{z}) = P_{\mathbf{X},Y}(\xi, y)$ et $Q(\mathbf{z}, \alpha) = L[y, f(\xi, \alpha)]$. L'objectif devient la minimisation de

$$R(\alpha) = \int Q(\mathbf{z}, \alpha) \mathrm{d}P_{\mathbf{Z}}(\mathbf{z}), \tag{3.54}$$

où la loi $P_{\mathbf{Z}}(\mathbf{z})$ est inconnue mais où un échantillon d'observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$ est donné.

3.2.2 Problème de la reconnaissance de forme

Le problème de la reconnaissance de forme correspond au cas particulier où le superviseur associe à toute valeur d'entrée ξ la classe à laquelle elle appartient, selon la loi de probabilité conditionnelle $F(y|\xi)$, avec $y \in \{1, 2, \ldots, K\}$: $y = k$ traduit le fait que le superviseur attribue la $k^{\text{ième}}$ classe à ξ. L'ensemble des fonctions $\{f(\xi, \alpha), \alpha \in \Lambda\}$ ne contient que des fonctions à valeurs dans $\{1, 2, \ldots, K\}$. Une fonction coût très simple est un coût d'erreur constant, qui ne prend que les valeurs 0 ou 1 :

$$L(y, f) = \begin{cases} 0 & \text{si } y = f, \\ 1 & \text{si } y \neq f. \end{cases} \tag{3.55}$$

La fonction risque moyen $R(\alpha)$ représente alors la probabilité d'erreur de classification et les fonctions $\{Q(\mathbf{z}, \alpha), \alpha \in \Lambda\}$ sont des fonctions indicatrices (c'est-à-dire qui prennent les valeurs 0 ou 1 uniquement). Le processus d'extension de la segmentation d'une coupe aux coupes voisines entre dans ce cadre (cf. paragraphe 5.2 page 186).

3.2.3 Principe inductif de la minimisation du risque empirique moyen (MRE)

Une fois qu'on aura déterminé le paramètre $\widehat{\alpha}_l$ qui minimise le risque empirique moyen défini ci-dessous, on considérera la fonction $Q(\mathbf{z}, \widehat{\alpha}_l)$ comme une estimation de la fonction $Q(\mathbf{z}, \alpha_0)$ qui minimise le risque moyen $R(\alpha)$: on parlera de processus d'apprentissage basé sur la minimisation du risque empirique. Dans ce paragraphe, nous verrons comment caractériser la qualité de cette estimation et quelles sont les conditions nécessaires et suffisantes pour que le processus d'apprentissage converge. Nous définirons aussi le taux de convergence du processus et nous verrons comment le contrôler.

Consistance d'un processus d'apprentissage

Soit $\mathcal{X} = \{\xi_1, \ldots, \xi_N\}$ un échantillon ; à chaque ξ_i, on associe une classe $y_i \in \{1, \ldots, K\}$, et on pose $\mathbf{z}_i = (y_i, \xi_i)$. Soit $Q(\mathbf{z}, \widehat{\alpha}_\ell)$ une fonction qui minimise le risque empirique moyen d'ordre ℓ, appelé plus simplement risque empirique moyen, défini par :

$$\widehat{R}_\ell(\alpha) = \frac{1}{\ell} \sum_{i=1}^{\ell} Q(\mathbf{z}_i, \alpha). \tag{3.56}$$

$\widehat{\alpha}_\ell(\xi_1, \ldots, \xi_N) = \widehat{\alpha}_\ell = \arg_\alpha \min \widehat{R}_\ell(\alpha)$ est l'estimateur du minimum de risque empirique (MRE).

Définition 3.20 (Consistance de l'estimateur MRE). *L'estimateur MRE est dit consistant pour l'ensemble des fonctions $\{Q(z, \alpha), \alpha \in \Lambda\}$ et pour la loi de probabilité $P_{\mathbf{Z}}(\mathbf{z})$ si les deux suites suivantes convergent en probabilité vers la même limite quand les $\mathbf{z}_i$ sont indépendants (cf. [79] page 36) :*

$$\lim_{\ell \to +\infty} R(\widehat{\alpha}_\ell) = \inf_{\alpha \in \Lambda} R(\alpha), \tag{3.57}$$

$$\lim_{\ell \to +\infty} \widehat{R}_\ell(\widehat{\alpha}_\ell) = \inf_{\alpha \in \Lambda} R(\alpha). \tag{3.58}$$

En d'autres termes, l'estimateur est consistant s'il fournit une suite de paramètres $\widehat{\alpha}_\ell, \ell = 1, 2, \ldots$ pour laquelle le risque moyen et le risque empirique moyen convergent tous les deux vers la valeur minimale du risque moyen. L'équation (3.57) prouve que les valeurs des risques moyens obtenus convergent vers la meilleure valeur possible. L'équation (3.58) montre que l'on peut estimer la valeur minimale du risque sur la base des valeurs du risque empirique moyen.

Pour exclure le cas de consistance triviale (que l'on rencontre quand l'ensemble des fonctions $\{Q(\mathbf{z}, \alpha), \alpha \in \Lambda\}$ contient une fonction qui minore toutes les autres), on définit la notion de consistance non triviale.

Définition 3.21 (Consistance non triviale de l'estimateur MRE). *L'estimateur MRE est dit non trivialement consistant pour l'ensemble des fonctions* $\{Q(\mathbf{z},\alpha), \alpha \in \Lambda\}$ *et pour la loi de probabilité* $P_{\mathbf{Z}}(\mathbf{z})$ *si pour tout sous-ensemble non vide* $\Lambda(c)$ *de* Λ*, où* $c > 0$*, tel que*

$$\Lambda(c) = \{\alpha : \int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) \geq c, \alpha \in \Lambda\}, \tag{3.59}$$

la convergence en probabilité suivante est vérifiée :

$$\lim_{\ell\to+\infty} \inf_{\alpha\in\Lambda(c)} \widehat{R}_\ell(\alpha) = \inf_{\alpha\in\Lambda(c)} R(\alpha). \tag{3.60}$$

Il est démontré dans [78] page 83 que la condition de l'équation (3.57) est alors également satisfaite. Un contre-exemple prouvant que la convergence (3.57) n'implique pas nécessairement la convergence (3.58) y est également présenté.

Théorème clé de la théorie de l'apprentissage

Théorème 3.3. *Soit* $\{Q(\mathbf{z},\alpha), \alpha \in \Lambda\}$ *un ensemble de fonctions vérifiant la condition*

$$\exists A, B > 0 \text{ tels que } \forall\alpha \in \Lambda,\ A \leq \int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) \leq B, \tag{3.61}$$

c'est-à-dire

$$\exists A, B > 0 \text{ tels que } \forall\alpha \in \Lambda,\ A \leq R(\alpha) \leq B. \tag{3.62}$$

Pour que l'estimateur MRE soit consistant, il faut et il suffit que le risque empirique moyen $\widehat{R}_\ell(\alpha)$ *converge uniformément de manière unilatérale vers le risque moyen* $R(\alpha)$ *sur l'ensemble* $\{Q(\mathbf{z},\alpha), \alpha \in \Lambda\}$*, c'est-à-dire que :*

$$\lim_{\ell\to+\infty} P\{\sup_{\alpha\in\Lambda}(R(\alpha) - \widehat{R}_\ell(\alpha)) > \epsilon\} = 0, \forall\epsilon > 0. \tag{3.63}$$

Une preuve de ce théorème se trouve dans [78] page 82. La convergence est dite unilatérale par opposition à la convergence bilatérale définie par :

$$\lim_{\ell\to+\infty} P\{\sup_{\alpha\in\Lambda}|R(\alpha) - \widehat{R}_\ell(\alpha)| > \epsilon\} = 0, \forall\epsilon > 0. \tag{3.64}$$

Les conditions de convergence de l'estimateur sont déterminées par la fonction Q qui est « la pire » dans l'ensemble choisi.

Convergence uniforme bilatérale et unilatérale dans le cas d'un ensemble fini de fonctions indicatrices

On définit dans le cas général deux processus stochastiques empiriques : l'un unilatéral, l'autre bilatéral. D'après le théorème clé, la convergence uniforme unilatérale est une condition nécessaire et suffisante pour la convergence de l'estimateur MRE. Cependant, les conditions pour qu'il y ait convergence uniforme bilatérale jouent un rôle important dans la construction des conditions pour la convergence uniforme unilatérale. Les conditions

de convergence sont données ici uniquement dans le cas du modèle le plus simple, où l'ensemble des fonctions indicatrices est fini.

Soit la suite de fonctions aléatoires

$$\mu^\ell(\mathbf{z}_1,\ldots,\mathbf{z}_\ell) = \mu^\ell = \sup_{\alpha\in\Lambda}\left|\int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha)\right|,\ \ell = 1,2,\ldots \tag{3.65}$$

Quand la suite de valeurs $(\mathbf{z}_1,\ldots,\mathbf{z}_\ell,\ldots)$ est remplacée par la suite de variables aléatoires $(\mathbf{Z}_1,\ldots,\mathbf{Z}_\ell,\ldots)$, la suite μ^ℓ obtenue, qui dépend à la fois de $P_{\mathbf{Z}}(\mathbf{z})$ et de l'ensemble des fonctions $\{Q(\mathbf{z},\alpha),\alpha\in\Lambda\}$, est un processus empirique bilatéral. Sa convergence en probabilité vers 0 signifie que l'égalité suivante est vérifiée :

$$\lim_{\ell\to+\infty} P\left\{\sup_{\alpha\in\Lambda}\left|\int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha)\right| > \epsilon\right\} = 0,\ \forall\epsilon > 0. \tag{3.66}$$

De la même façon, on définit le processus empirique unilatéral

$$\mu^\ell_+(\mathbf{z}_1,\ldots,\mathbf{z}_\ell) = \mu^\ell_+ = \sup_{\alpha\in\Lambda}\left(\int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha)\right)_+,\ \ell = 1,2,\ldots \tag{3.67}$$

avec

$$(u)_+ = \begin{cases} u & \text{si } u > 0, \\ 0 & \text{sinon.} \end{cases} \tag{3.68}$$

La convergence en probabilité de ce processus signifie que l'égalité suivante est vérifiée :

$$\lim_{\ell\to+\infty} P\left\{\sup_{\alpha\in\Lambda}\left(\int Q(\mathbf{z},\alpha)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha)\right) > \epsilon\right\} = 0,\ \forall\epsilon > 0. \tag{3.69}$$

Conditions de convergence uniforme bilatérale dans le cas du modèle le plus simple Dans le modèle le plus simple, l'ensemble de fonctions indicatrices $Q(\mathbf{z},\alpha_k), k = 1,\ldots,M$ est fini, de cardinal M. A k fixé, les $\mathbf{z}_i, i = 1,\ldots,\ell$ étant des réalisations indépendantes et identiquement distribuées, les $Q(\mathbf{z}_i,\alpha_k), i = 1,\ldots,\ell$ peuvent être considérés comme ℓ réalisations indépendantes d'un processus de Bernoulli. Les risques moyens $R(\alpha_k), k = 1,\ldots,M$ sont les probabilités $P\{Q(\mathbf{z},\alpha_k) > 0\}$ des événements $A_k = \{\omega\in\Omega : Q(Z(\omega),\alpha_k) > 0\}, k = 1,\ldots,M$, et les risques empiriques $\widehat{R}_\ell(\alpha_k), k = 1,\ldots,M$ sont les fréquences $\nu_\ell\{Q(z_i,\alpha_k) > 0\}$ de ces événements obtenues sur les observations $z_1,\ldots,z_\ell$. L'équation (3.66) peut s'écrire :

$$\lim_{\ell\to+\infty} P\left\{\max_{1\le k\le M}|P\{Q(\mathbf{z},\alpha_k) > 0\} - \nu_\ell\{Q(\mathbf{z}_i,\alpha_k) > 0\}| > \epsilon\right\} = 0,\ \forall\epsilon > 0. \tag{3.70}$$

D'après la loi des grands nombres, pour tout événement fixé A_{k_0}, les fréquences convergent vers la probabilité quand le nombre d'observations tend vers l'infini et quand on suppose les $(\mathbf{Z}_i), i \ge 1$ indépendants en plus d'être identiquement distribués. Le taux de convergence est donné par l'inégalité de Chernoff ([80] page 190) :

$$P\left\{|P\{Q(\mathbf{z},\alpha_{k_0}) > 0\} - \nu_\ell\{Q(\mathbf{z}_i,\alpha_{k_0}) > 0\}| > \epsilon\right\} \le 2\exp(-2\epsilon^2\ell). \tag{3.71}$$

Alors,

$$P\left\{\max_{1\leq k\leq M}|P\{Q(\mathbf{z},\alpha_k)>0\}-\nu_\ell\{Q(\mathbf{z}_i,\alpha_k)>0\}|>\epsilon\right\}$$
$$\leq\sum_{k=1}^{M}P\{|P\{Q(\mathbf{z},\alpha_k)>0\}-\nu_\ell\{Q(\mathbf{z}_i,\alpha_k)>0\}|>\epsilon\}$$
$$\leq 2M\exp(-2\epsilon^2\ell), \quad (3.72)$$

ou encore

$$P\left\{\max_{1\leq k\leq M}|P\{Q(\mathbf{z},\alpha_k)>0\}-\nu_\ell\{Q(\mathbf{z}_i,\alpha_k)>0\}|>\epsilon\right\}\leq 2\exp\left\{\left(\frac{\ln M}{\ell}-2\epsilon^2\right)\ell\right\}, \quad (3.73)$$

ce qui prouve qu'il y a convergence uniforme bilatérale selon l'équation (3.66).

Conditions de convergence uniforme unilatérale dans le cas du modèle le plus simple La convergence bilatérale peut être décrite de la manière suivante :

$$\lim_{\ell\to+\infty}P\left\{\left[\max_{1\leq k\leq M}(R(\alpha_k)-\widehat{R}_\ell(\alpha_k))>\epsilon\right]\text{ ou }\left[\max_{1\leq k\leq M}(\widehat{R}_\ell(\alpha_k)-R(\alpha_k))>\epsilon\right]\right\}=0. \quad (3.74)$$

Cette condition, qui inclut la condition de convergence unilatérale, est donc suffisante pour assurer la convergence de l'estimateur MRE.

3.2.4 Bornes sur le taux de convergence d'un processus d'apprentissage dans le cas d'un ensemble fini de fonctions indicatrices

Il s'agit maintenant de caractériser la rapidité de la convergence du processus d'apprentissage. Supposons que le minimum de la fonction risque moyen de l'équation (3.54) soit atteint pour la fonction $Q(z,\alpha_0)$, et le minimum du risque empirique moyen pour la fonction $Q(\mathbf{z},\widehat{\alpha}_\ell)$. On souhaite estimer :

- $R(\widehat{\alpha}_\ell)$, c'est-à-dire le risque moyen obtenu avec la fonction $Q(\mathbf{z},\widehat{\alpha}_\ell)$,
- $R(\widehat{\alpha}_\ell)-R(\alpha_0)$, c'est-à-dire la différence entre le risque moyen minimal qu'on pourrait atteindre avec la « meilleure » fonction $Q(\mathbf{z},\alpha_0)$ et le risque moyen obtenu avec la fonction $Q(\mathbf{z},\widehat{\alpha}_\ell)$ qui minimise le risque empirique moyen.

Pour cela, on est amené à étudier le taux de convergence uniforme en probabilité vers 0 de

$$\sup_{1\leq k\leq M}\left(\int Q(z,\alpha_k)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z})-\frac{1}{\ell}\sum_{i=1}^{\ell}Q(z_i,\alpha_k)\right).$$

Rappelons que l'ensemble des fonctions indicatrices contient un nombre fini M d'éléments $Q(z,\alpha), k=1,2,\ldots,M$.

Cas pessimiste

Les bornes de Chernoff additives ([80] page 190) sont valables pour les ℓ réalisations indépendantes $\mathbf{z}_1 \ldots, \mathbf{z}_\ell$:

$$P\{p - \nu_\ell > \epsilon\} < \exp\{-2\epsilon^2\ell\}, \tag{3.75}$$

$$P\{\nu_\ell - p > \epsilon\} < \exp\{-2\epsilon^2\ell\}, \tag{3.76}$$

où p et ν_l représentent respectivement les probabilités et les fréquences. Alors,

$$\begin{aligned} P\left\{\sup_{1\le k\le M}\left(\int Q(\mathbf{z},\alpha_k)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha_k)\right) > \epsilon\right\} \\ \le \sum_{k=1}^{M} P\left\{\left(\int Q(\mathbf{z},\alpha_k)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha_k)\right) > \epsilon\right\} \\ \le M\exp\{-2\epsilon^2\ell\}. \end{aligned} \tag{3.77}$$

Posons $\eta = M\exp\{-2\epsilon^2\ell\}$, alors, pour ℓ assez grand, $0 < \eta \le 1$. Résolvons cette équation par rapport à ϵ :

$$\epsilon = \sqrt{\frac{\ln M - \ln\eta}{2\ell}}. \tag{3.78}$$

L'inégalité (3.77) devient alors :

$$P\left\{\sup_{1\le k\le M}\left(\int Q(\mathbf{z},\alpha_k)\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(\mathbf{z}_i,\alpha_k)\right) > \sqrt{\frac{\ln M - \ln\eta}{2\ell}}\right\} \le \eta. \tag{3.79}$$

Autrement dit, avec une probabilité $1-\eta$, simultanément pour l'ensemble des M fonctions $Q(\mathbf{z},\alpha), k = 1, 2, \ldots, M$, on a :

$$\int Q(z,\alpha_k)\mathrm{d}F(z) - \frac{1}{\ell}\sum_{i=1}^{\ell} Q(z_i,\alpha_k) \le \sqrt{\frac{\ln M - \ln\eta}{2\ell}}. \tag{3.80}$$

C'est en particulier vrai pour la fonction caractérisée par $\widehat{\alpha}_\ell$ qui minimise le risque empirique moyen, ce que l'on peut écrire ainsi :

$$R(\widehat{\alpha}_\ell) \le \widehat{R}_\ell(\widehat{\alpha}_\ell) + \sqrt{\frac{\ln M - \ln\eta}{2\ell}}, \tag{3.81}$$

ce qui permet d'estimer le risque moyen obtenu avec le paramètre $\widehat{\alpha}_\ell$.

Par ailleurs, en appliquant l'inégalité de Chernoff (3.76) à la fonction $Q(\mathbf{z},\alpha_0)$ caractérisée par le paramètre α_0 qui minimise la fonction risque moyen (3.53), on obtient :

$$P\left\{\widehat{R}_\ell(\alpha_0) - R(\alpha_0) > \epsilon\right\} \le \exp\{-2\epsilon^2\ell\}. \tag{3.82}$$

En d'autres termes, avec une probabilité $1-\eta$,

$$R(\alpha_0) \ge \widehat{R}_\ell(\alpha_0) - \sqrt{\frac{-\ln\eta}{2\ell}}. \tag{3.83}$$

Comme $\widehat{\alpha}_\ell$ minimise le risque empirique moyen, on a :

$$\widehat{R}_\ell(\alpha_0) - \widehat{R}_\ell(\widehat{\alpha}_\ell) \geq 0, \tag{3.84}$$

donc

$$R(\widehat{\alpha}_\ell) - R(\alpha_0) \leq \sqrt{\frac{\ln M - \ln \eta}{2\ell}} + \sqrt{\frac{-\ln \eta}{2\ell}}. \tag{3.85}$$

On a ainsi réussi à borner la différence entre le risque moyen minimum et le risque moyen obtenu avec le paramètre qui minimise le risque empirique moyen.

Cas optimiste

Plaçons-nous dans le cas où l'ensemble de fonctions contient au moins une « bonne fonction », c'est-à-dire une fonction qui donne une probabilité d'erreur égale à zéro. Supposons qu'il existe au moins une telle fonction dans l'ensemble des fonctions indicatrices considéré (contenant au minimum deux fonctions). S'il y en a plusieurs, choisissons n'importe laquelle d'entre elles.

Il est démontré dans [78] page 125 qu'avec une probabilité $1 - \eta$, simultanément pour l'ensemble des paramètres $\widehat{\alpha}_\ell$ qui annulent le risque empirique moyen, on a :

$$R(\widehat{\alpha}_\ell) \leq \frac{\ln(M-1) - \ln \eta}{\ell}, \; M > 1. \tag{3.86}$$

Il est aussi démontré qu'avec une probabilité au moins égale à $1 - \eta$,

$$R(\widehat{\alpha}_\ell) - R(\alpha_0) < \frac{\ln(M-1) - \ln \eta}{\ell}. \tag{3.87}$$

Notons l'amélioration due au fait que les bornes sont ici proportionnelles à $1/\ell$, au lieu de $1/\sqrt{\ell}$ dans le cas pessimiste.

Cas général

En utilisant les mêmes notations que précédemment, d'après les bornes de Chernoff multiplicatives ([80] page 190), on a :

$$P\left\{\frac{p - \nu_\ell}{\sqrt{p}} > \epsilon\right\} < \exp\left(\frac{-\epsilon^2 \ell}{2}\right), \tag{3.88}$$

$$P\left\{\frac{\nu_\ell - p}{\sqrt{p}} > \epsilon\right\} < \exp\left(\frac{-\epsilon^2 \ell}{3}\right). \tag{3.89}$$

On peut en déduire de la même manière que précédemment que :

$$P\left\{\sup_{1 \leq k \leq M} \frac{R(\alpha_k) - \widehat{R}_l(\alpha_k)}{\sqrt{R(\alpha_k)}} > \epsilon\right\} < M \exp\left(\frac{-\epsilon^2 \ell}{2}\right). \tag{3.90}$$

Posons $\eta = M \exp\left(\frac{-\epsilon^2 \ell}{2}\right)$, alors, pour ℓ assez grand, $0 < \eta \leq 1$. Résolvons cette équation par rapport à ϵ :

$$\epsilon = \sqrt{2\frac{\ln M - \ln \eta}{\ell}}. \tag{3.91}$$

L'inégalité (3.90) s'écrit alors :

$$P\left\{ \sup_{1 \leq k \leq M} \frac{R(\alpha_k) - \widehat{R}_\ell(\alpha_k)}{\sqrt{R(\alpha_k)}} > \sqrt{2\frac{\ln M - \ln \eta}{\ell}} \right\} < \eta, \tag{3.92}$$

donc la probabilité pour que

$$\frac{R(\alpha_k) - \widehat{R}_l(\alpha_k)}{\sqrt{R(\alpha_k)}} \leq \epsilon \tag{3.93}$$

est au moins égale à $1 - \eta$. En résolvant l'inégalité (3.93) par rapport à $R(\alpha_k)$ après avoir remplacé ϵ par sa valeur donnée par (3.91), on obtient que l'inégalité suivante est vérifiée simultanément pour les M fonctions $\{Q(\mathbf{z}, \alpha_k), k = 1, \ldots, M\}$ avec une probabilité de $1 - \eta$:

$$R(\alpha_k) < \widehat{R}_\ell(\alpha_k) + \frac{\ln M - \ln \eta}{\ell} \left(1 + \sqrt{1 + 2\frac{\widehat{R}_\ell(\alpha_k)\ell}{\ln M - \ln \eta}} \right). \tag{3.94}$$

Ceci est en particulier vérifié pour la fonction qui minimise le risque empirique moyen, caractérisée par le paramètre $\widehat{\alpha}_\ell$, ce qui permet d'estimer le risque moyen obtenu avec cette fonction, $R(\widehat{\alpha}_\ell)$:

$$R(\widehat{\alpha}_\ell) < \widehat{R}_\ell(\widehat{\alpha}_\ell) + \frac{\ln M - \ln \eta}{\ell} \left(1 + \sqrt{1 + 2\frac{\widehat{R}_\ell(\widehat{\alpha}_\ell)\ell}{\ln M - \ln \eta}} \right). \tag{3.95}$$

Par ailleurs, l'inégalité (3.83) est vérifiée. En la combinant avec l'inégalité précédente, on peut borner la différence entre le risque moyen minimum et le risque moyen obtenu avec la fonction qui minimise le risque empirique moyen : avec une probabilité au moins égale à $1 - 2\eta$, on a

$$R(\widehat{\alpha}_\ell) - R(\alpha_0) < \sqrt{\frac{-\ln \eta}{2\ell}} + \frac{\ln M - \ln \eta}{\ell} \left(1 + \sqrt{1 + 2\frac{\widehat{R}_\ell(\widehat{\alpha}_\ell)\ell}{\ln M - \ln \eta}} \right). \tag{3.96}$$

3.2.5 Conclusion

Dans la seconde partie de ce chapitre, nous avons développé le cadre mathématique dans lequel nous étudierons l'extension de la segmentation d'une coupe donnée à toutes les autres coupes rénales. Dans un ensemble de classificateurs, on choisit celui qui minimise la fréquence d'erreur sur une suite finie donnée d'exemples (encore appelée risque empirique moyen). Il est alors possible de calculer des bornes limitant d'une part le risque moyen obtenu avec ce classificateur, d'autre part la différence entre celui-ci et le risque moyen minimum qui pourrait être obtenu avec le meilleur classificateur de l'ensemble.

Son application à nos données d'IRM sera développée dans le chapitre 4. Remarquons que l'hypothèse d'indépendance des $(\mathbf{z}_i)_{i\geq 1}$ est importante dans la preuve des inégalités de Chernoff, sur lesquelles sont fondés les théorèmes de convergence.

Notons également que le cadre proposé diffère de ce que l'on nomme « extension d'un *clustering* ». Dans [81], il est démontré que dans ce cas, il n'existe pas l'équivalent des bornes de la théorie de la généralisation sauf dans des cas très particuliers, comme les K-moyennes, mais elles concernent uniquement la fonction de risque minimisée par l'algorithme lui-même, qui serait la distorsion dans notre cas, et cela ne présente pas d'intérêt pour l'application envisagée.

Chapitre 4

Segmentations fonctionnelles : données, pré-traitement et méthodes de validation des résultats

Dans ce chapitre, nous abordons l'ensemble des problèmes liés à l'absence de vérité-terrain pour les données réelles. Comme nous l'avons expliqué dans l'introduction du chapitre 3.1, notre objectif est de classer les voxels rénaux selon leurs courbes temps-intensité afin d'aboutir à une segmentation du rein dite fonctionnelle ; différentes méthodes seront testées, avec validation des résultats. Nous préciserons d'abord la nature des données traitées, qu'il s'agisse de simulations ou de données réelles. Nous avons été amenés à construire un modèle rénal le plus réaliste possible pour réaliser les premiers tests de notre méthode ; à notre connaissance, aucun modèle semblable n'a été proposé auparavant. Nous exposerons alors les différentes stratégies envisagées pour réaliser une segmentation volumique. En ce qui concerne la validation, il s'agira de savoir si la segmentation fonctionnelle obtenue est semblable à la segmentation anatomique des mêmes données. Pour des images de synthèse, nous verrons que la segmentation anatomique est inhérente au modèle utilisé pour générer les données. Pour des données réelles, les segmentations anatomiques seront réalisées par des radiologues.

4.1 Données

Les méthodes proposées seront testées d'abord sur des données simulées puis sur des données réelles. Dans ce paragraphe, nous présenterons le processus de génération des données de synthèse en détaillant les différents effets pris en compte et la manière dont nous les modélisons. Nous décrirons ensuite les données réelles dont nous disposons.

4.1.1 Données simulées : reins sains

Il s'agit de générer des séquences d'images de synthèse correspondant à l'acquisition d'une coupe rénale en IRM dynamique de perfusion avec rehaussement de contraste. Il n'est pas question ici de simuler l'ensemble du processus d'acquisition d'un modèle rénal anatomique tridimensionnel avec reconstruction des images à partir des données brutes,

mais de reconstituer une série d'images à partir de ce modèle et de courbes d'évolution temporelle du contraste typiques pour chaque compartiment. Le bruit d'acquisition ainsi que l'effet de volume partiel (voir ci-dessous) seront pris en compte.

Génération d'un modèle de rein anatomique à trois compartiments

Rappelons que nous considérons que le rein est composé de trois compartiments anatomiques : le cortex, la médullaire et les cavités (cf. paragraphe 2.1 page 29). Deux modèles de complexités différentes sont proposés. Pour le premier, un voxel de l'acquisition IRM correspond entièrement à un compartiment donné. Pour le second, un voxel est un mélange de tissus appartenant à différents compartiments.

Modèle le plus simple Un rein est modélisé par un tableau M_{rein} à 3 dimensions correspondant à une segmentation de cet organe en trois compartiments à la même résolution spatiale que celle de l'acquisition IRM à simuler. Un élément de la matrice vaut 0 si le voxel est à l'extérieur du rein, 1 s'il correspond à des cavités, 2 à de la médullaire, 3 à du cortex. C'est cette segmentation qu'il nous faudrait idéalement retrouver en faisant une classification fonctionnelle des voxels rénaux. Un exemple en est présenté figure 4.1.

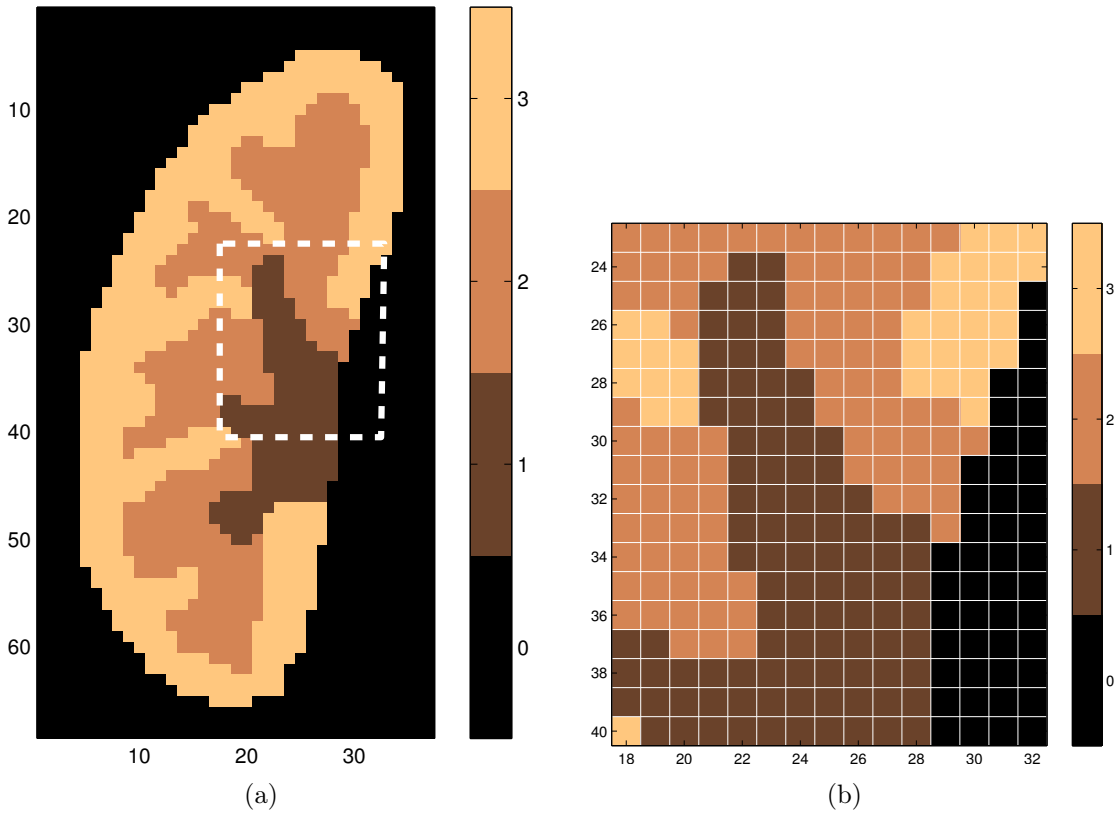


FIG. 4.1 – Exemple de coupe d'un modèle rénal anatomique simple à trois compartiments : coupe complète (a) et zoom (b) faisant apparaître les voxels étiquetés.

Modèle tenant compte de l'effet de volume partiel Vu la finesse des structures rénales et la résolution des acquisitions, chaque voxel de la séquence acquise correspond en réalité non pas à un seul mais à un mélange de compartiments et d'organes autres que le rein, ce qui crée un effet de volume partiel, dont la modélisation est détaillée page 116.

Un rein est alors décrit par un tableau M_{rein} à 3 dimensions (représentation volumique), dont chaque élément est un quadruplet donnant les proportions de chaque compartiment et de l'extérieur du rein dans la composition du voxel correspondant. Un exemple de description globale d'une coupe est donné sur la figure 4.2, et le détail de la partie entourée se trouve figure 4.3.

Détermination des proportions des différents compartiments

Solution 1 Si on disposait d'un modèle spatial 3D semblable au modèle le plus simple mais avec une résolution plus fine que celle de l'acquisition IRM, on pourrait facilement générer le modèle proposé en regroupant plusieurs voxels du modèle le plus fin et en attribuant au voxel généré la composition correspondante. Nous ne possédons cependant pas un tel modèle, donc nous n'avons pas retenu cette solution.

Solution 2 Ce modèle est généré à partir du modèle le plus simple, décrit ci-dessus au paragraphe 4.1.1 qui représente alors le compartiment « dominant » associé à chaque voxel. On module ensuite sa composition en y incluant un pourcentage des compartiments dominants de ses 24 plus proches voisins en dimension 2. Ceci se justifie dans la mesure où ce sont les voxels situés aux frontières des compartiments du premier modèle qui sont soumis à l'effet de volume partiel : ils sont ainsi composés d'un mélange de ces compartiments anatomiques.

Remarquons que chaque voxel sera en définitive attribué à un seul compartiment, qui sera considéré comme dominant. Ici apparaît une ambiguïté dans la définition du terme « dominant ». Du point de vue anatomique, le compartiment dominant est de manière évidente celui dont le volume est le plus important dans la composition du voxel. En revanche, du point de vue fonctionnel, nous devons définir plus précisément ce terme pour un mélange de comportements : nous y reviendrons dans la partie du paragraphe 4.1.1 consacrée à la simulation du volume partiel.

Quelques ordres de grandeur pour les valeurs numériques du modèle Typiquement, un modèle correspond à un rein de volume 25 cm^3 acquis à une résolution de 1,25 mm $\times$ 1,25 mm dans le plan de coupe, chaque coupe étant épaisse de 1 cm, ce qui correspond effectivement à la résolution des données réelles utilisées par la suite. Le rein est alors représenté par environ 5000 voxels répartis dans quatre coupes, dont 1600 dans la coupe principale (qui est celle contenant le volume de rein le plus grand). Pour un bébé, les dimensions sont bien sûr réduites : par exemple, le rein est représenté par environ 800 voxels dans la coupe principale.

Modèle de comportement temporel de chaque compartiment anatomique

Il s'agit d'associer à chaque compartiment anatomique un modèle d'évolution temporelle du contraste qui sera utilisé pour créer les signaux associés aux différents voxels.

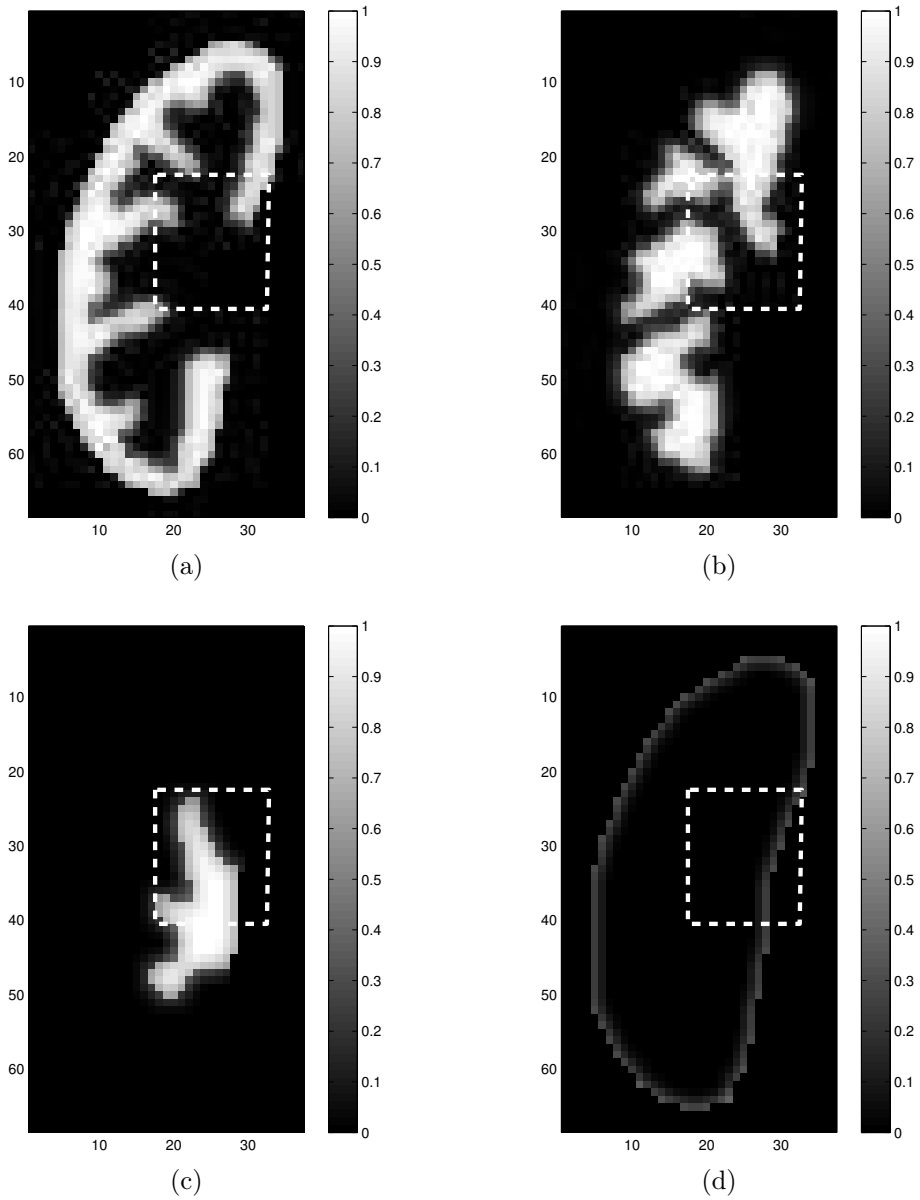


FIG. 4.2 – Exemple de coupe d'un modèle rénal anatomique simple à trois compartiments avec, pour chaque voxel, les proportions de cortex (a), médullaire (b), cavités (c) et organes extérieurs (d).

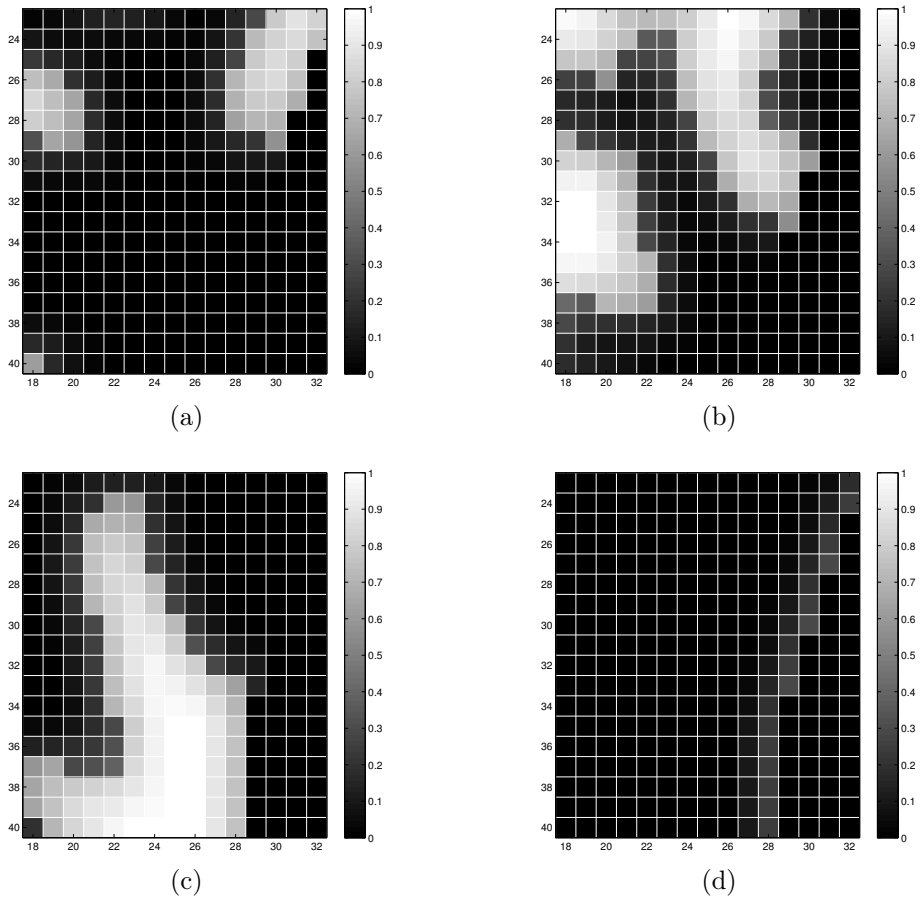


FIG. 4.3 – Exemple de coupe d'un modèle rénal anatomique à trois compartiments avec, pour chaque voxel, les proportions de cortex (a), médullaire (b), cavités (c) et organes extérieurs (d) : zoom sur la partie indiquée sur la figure 4.2.

Quatre courbes temps-intensité typiques obtenues à partir d'acquisitions réelles, filtrées passe-bas pour les lisser, sont utilisées (cf. figure 4.4). Les trois premières sont celles des trois compartiments rénaux ; la quatrième est celle des organes voisins pour la simulation du volume partiel dans les voxels situés en périphérie du rein[1]. Nous pouvons cependant remarquer que, même si certaines caractéristiques principales des courbes temps-intensité restent en principe identiques d'un rein à l'autre (cf. paragraphe 2.2.3), les valeurs absolues et les formes de ces courbes varient, et dépendent également des conditions d'injection du produit de contraste et de la machine utilisée pour l'acquisition. Nous ne pourrons pas conclure que les simulations recouvrent l'ensemble des cas rencontrés dans la réalité. Nous espérons être en mesure de mettre en évidence et d'identifier l'origine de certains phénomènes fréquemment rencontrés, ce qui à notre connaissance, n'a jamais été publié. Nous ne pouvons cependant pas vraiment espérer les quantifier de manière valable.

[1]Ces voxels sont en effet formés d'un mélange de tissus rénaux et d'organes extérieurs au rein. Notons bien que les organes extérieurs ne seront bien sûr jamais un compartiment majoritaire, et que nous recherchons trois, et non quatre compartiments.

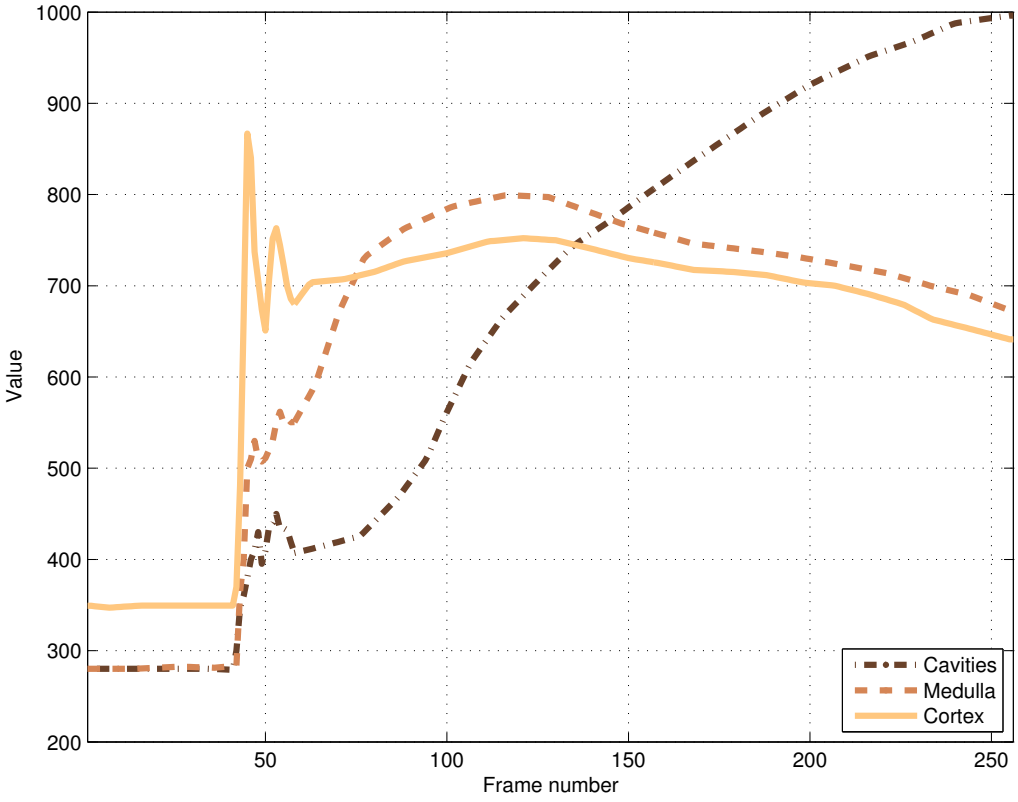


FIG. 4.4 – Courbes temps-intensité typiques non bruitées pour les trois compartiments rénaux.

Simulation du bruit d'acquisition

Le bruit en IRM est essentiellement d'origine thermique et peut être considéré comme un bruit blanc gaussien complexe centré $N(\mu,\nu)$ qui s'ajoute au signal $S(\mu,\nu)$ dans les données brutes de l'espace K [82], dans lequel les coordonnées sont représentées par les variables μ et ν.

$$Y(\mu,\nu) = S(\mu,\nu) + N(\mu,\nu). \tag{4.1}$$

L'image reconstruite $x[m,n]$ est le plus souvent le module de la transformée de Fourier discrète inverse des données brutes Y (cf. paragraphe 2.2.1) et peut s'écrire de la manière suivante :

$$x[m,n] = \|y[m,n]\| = \big[(s[m,n]\cos\theta[m,n] + n_{Re}[m,n])^2 + (s[m,n]\sin\theta[m,n] + n_{Im}[m,n])^2\big]^{\frac{1}{2}}, \tag{4.2}$$

où s est la transformée de Fourier inverse de S, donc le signal d'intérêt, où $\theta[m,n]$ représente l'erreur de phase pour le voxel $[m,n]$ et où n_{Re} et n_{Im} sont deux bruits blancs gaussiens de variance σ^2.

Les données après reconstruction sont modélisées par une distribution de Rice. Le bruit de Rice n'est pas un simple bruit additif mais dépend du signal lui-même. Son comportement varie spatialement suivant l'intensité locale. La densité de probabilité est la suivante [83] :

$$p(x|s,\sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2+s^2}{2\sigma^2}\right) I_0\left(\frac{sx}{\sigma^2}\right) u(x), \tag{4.3}$$

où I_0 est la fonction de Bessel modifiée d'ordre zéro de première espèce et u la fonction d'Heaviside. Pour des rapports signal à bruit élevés, on peut assimiler cette densité de probabilité à une distribution gaussienne, alors que, pour des rapports faibles, elle tend vers une densité de Rayleigh. Pour des rapports intermédiaires, ni l'une ni l'autre ne sont des approximations convenables. Un exemple comparatif est donné figure 4.5.

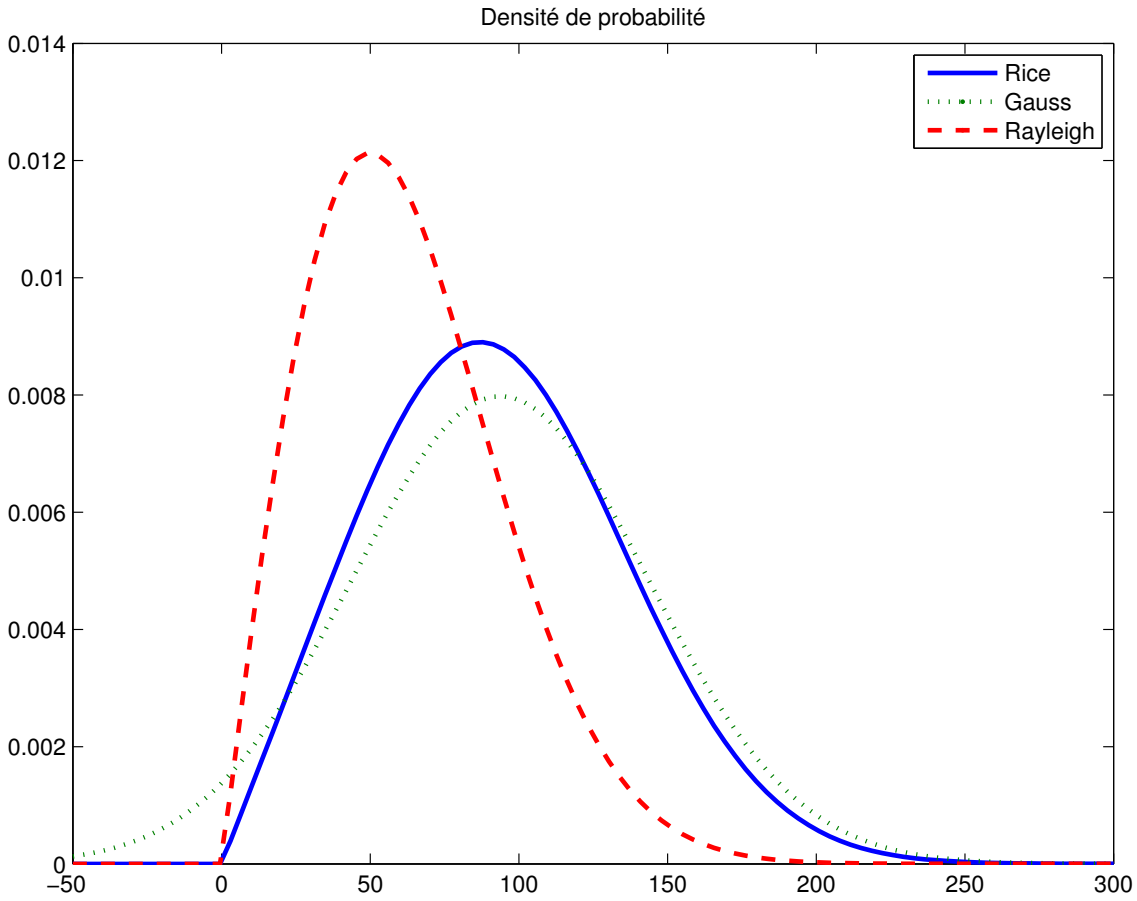


FIG. 4.5 – Loi de probabilité de Rice pour $s = 75$ et $\sigma = 50$, comparée à une loi de Gauss de même moyenne et de même écart-type σ et à une loi de Rayleigh de paramètre σ correspondant aux valeurs extrêmes du signal (respectivement $s \to \infty$ et $s = 0$).

Cependant, le post-traitement appliqué par le constructeur sur les données IRM dont nous disposons est inconnu. Il est ainsi difficile d'estimer le rapport signal à bruit par des méthodes classiques [84] : en effet, les zones hors du patient ont été préalablement filtrées de sorte que le signal y est uniformément nul. Nous avons donc généré des images sous

l'hypothèse que les données ont une distribution de Rice pour différentes valeurs σ de l'écart-type du bruit gaussien complexe. Il apparaît qu'aux niveaux d'intensité envisagés pour obtenir des images « réalistes », le bruit est pratiquement gaussien. On ne tient ici pas compte des éventuelles inhomogénéités d'illumination ou des scintillements.

Effet de volume partiel

L'intensité du signal pour un voxel est caractéristique du type de tissu qu'il contient [85]. Si un voxel donné contient des tissus appartenant à plusieurs compartiments différents, le signal est une combinaison linéaire des contributions de chaque compartiment : ce phénomène est appelé effet de volume partiel. Pour un modèle à C compartiments, l'intensité I d'un voxel quelconque peut s'écrire :

$$I = \sum_{i=1}^{C} \alpha_i I_i \text{ avec } \sum_{i=1}^{C} \alpha_i = 1 \text{ et } 0 \leq \alpha_i \leq 1, \tag{4.4}$$

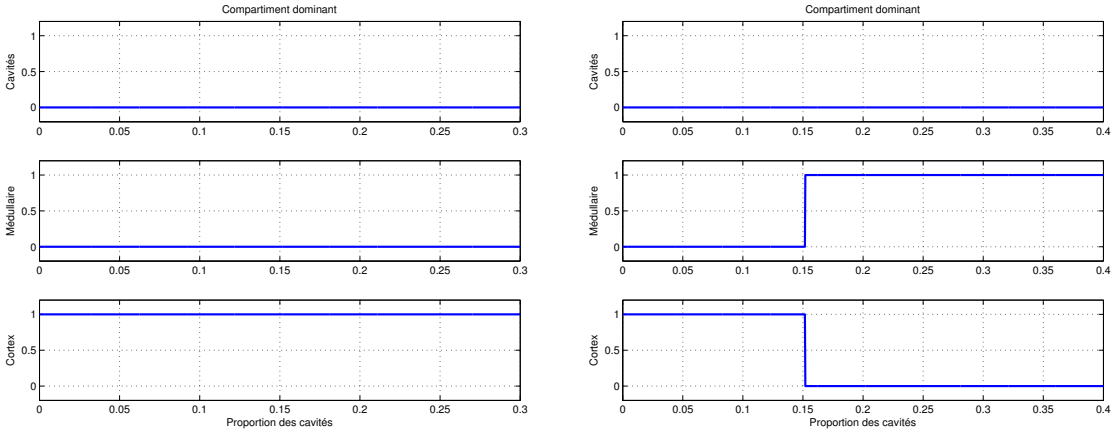
I_i étant tiré suivant la loi de probabilité pour le compartiment i. Le second modèle anatomique proposé ci-dessus permet de tenir compte de cet effet.

Compartiment fonctionnel dominant Pour le modèle de simulation proposé, on dispose de trois évolutions temporelles typiques ; on peut donc mesurer la distance (dans un sens à définir) entre la courbe temps-intensité d'un voxel donné et chacune des trois courbes types, la distance la plus faible indiquant le comportement et donc le compartiment dominants. Cependant, il n'est pas certain que le compartiment ainsi défini soit bien le compartiment dominant au sens anatomique.

Nous avons réalisé quelques simulations de la manière suivante : la proportion d'un compartiment donné étant fixée, nous faisons varier les proportions relatives des deux autres compartiments ; nous comparons alors le compartiment anatomique au compartiment fonctionnel obtenu, sur des données non bruitées. Un exemple de résultat est présenté sur la figure 4.6 : pour chaque compartiment, la fonction représentée vaut 1 s'il s'agit du compartiment fonctionnel dominant, 0 sinon. Pour une proportion de cortex fixée à 0,6 (le cortex est alors le compartiment anatomique majoritaire), nous pouvons remarquer sur la figure 4.6(b) que c'est la médullaire, et non le cortex, qui a le comportement fonctionnel dominant lorsque la proportion de cavités est supérieure à 0,15. Ce phénomène s'atténue pour des mélanges où la proportion de compartiment anatomique majoritaire augmente dans la composition du voxel : pour une proportion de cortex fixée à $0,7$, celui-ci reste le comportement fonctionnel dominant quelles que soient les proportions des deux autres compartiments. Pour le modèle de volume partiel utilisé en simulation suivant la solution 2 page 111, seuls environ 1% des voxels ont une courbe temps-intensité se rapprochant davantage de celle d'un des compartiments minoritaires entrant dans sa composition volumique. Ainsi, sur les données de synthèse, en l'absence de bruit, les segmentations anatomiques et fonctionnelles ne coïncident pas tout à fait.

Il est cependant difficile, en raison du manque de vérité-terrain, de valider les proportions obtenues sur les données simulées pour des données réelles. Le phénomène dépend en particulier des valeurs relatives des courbes, qui changent pour chaque rein. Sur certaines acquisitions, il peut s'accentuer, en particulier sur les reins de petite taille et sur

les coupes extrêmes. Par ailleurs, notons qu'il influence aussi la segmentation manuelle par un radiologue en particulier pour les voxels à la frontière des compartiments, dans la mesure où les compartiments ne sont pas segmentés sur la même image, mais sur deux images différentes (cf. paragraphe 4.4.3 page 139) : par exemple, certains voxels dont le comportement se rapproche davantage de celui des cavités en phase tardive seront identifiés comme tels, alors qu'ils se rapprocheraient davantage de la médullaire près du pic cortical.



(a) (b)

FIG. 4.6 – Compartiment fonctionnel dominant d'un voxel en fonction de la proportion de cavités pour une proportion de cortex fixée de 0,7 (a) et de 0,6 (b).

Post-traitement constructeur

Comme nous l'avons déjà évoqué précédemment dans le paragraphe concernant le bruit, les données brutes sont filtrées lors de la reconstruction, mais le traitement n'est pas explicitement connu. Cependant, l'observation des transformées de Fourier des images laisse supposer qu'un filtre de type moyenneur de fréquence de coupure réduite $\nu = 1/4$ a été appliqué. Aussi, nous ajoutons ce post-traitement sur nos images de synthèse pour en étudier l'influence.

Exemples d'images obtenues

Des exemples d'images réelles et simulées avec bruit, volume partiel et filtrage sont présentés figures 4.7 et 4.8 : leur comparaison visuelle semble montrer que nous parvenons à générer des images de synthèse relativement réalistes. Nous pouvons cependant noter une différence entre données simulées et données réelles que notre modèle ne prend pas en compte. Sur les « vidéos » des séquences réelles, on peut souvent observer une zone qui reste plus sombre que le reste du rein, et qui a donc un comportement en phase tardive qui ne correspond à aucun de ceux envisagés dans le modèle. Ceci peut augmenter la difficulté pour classer les voxels autour de la frontière entre la médullaire et les cavités.

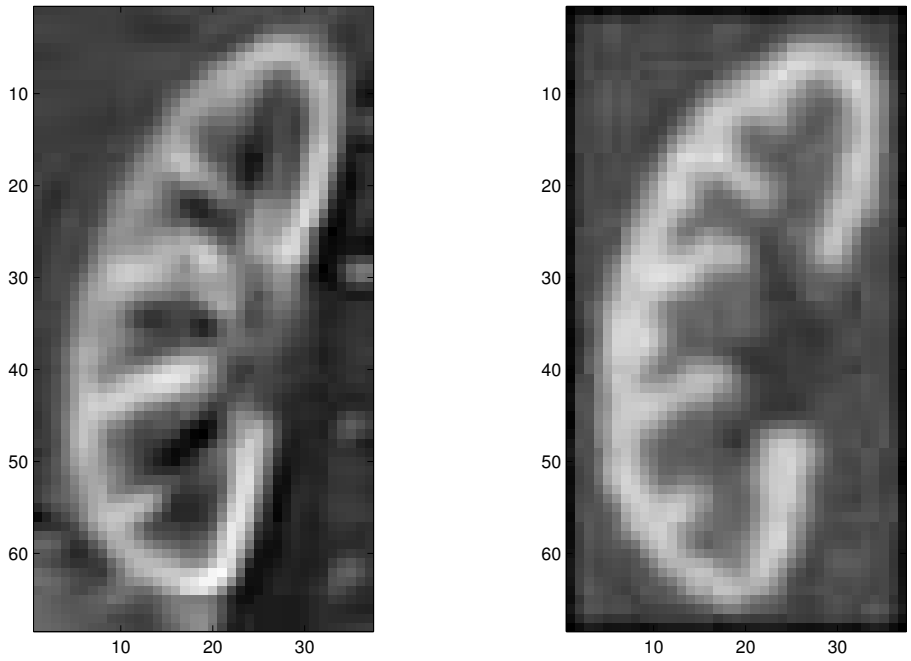


FIG. 4.7 – Exemple d'images réelle (a) et simulée (b) au pic cortical.

Effet d'erreurs de recalage

Vu les difficultés rencontrées lors de l'étape de recalage (cf. chapitre 2.4), il est souhaitable de pouvoir simuler un recalage défectueux. Pour ce faire, nous introduisons des translations aléatoires horizontales et verticales indépendantes sur chaque image de la série, à l'exception de l'image de référence. Ces translations ont une moyenne nulle et une déviation standard égale à 0,5 pixel. Un histogramme typique est représenté figure 4.9 pour 256×2 translations, ce qui nous paraît assez représentatif des résultats obtenus sur les données réelles.

Effet du filtre moyenneur et des erreurs de recalage sur le comportement dominant

Lorsque nous ajoutons, en plus du bruit et du volume partiel, un filtre moyenneur sur l'image, nous constatons que les compartiments dominants anatomiques et fonctionnels ne coïncident plus que pour environ 92% des voxels. L'ajout des erreurs de recalage ne modifie cette proportion que d'environ 1%.

Conclusion

Le modèle proposé pour générer des séquences d'images de synthèse tient compte du bruit d'acquisition, de l'effet de volume partiel, et des erreurs de recalage. En revanche, les inhomogénéités dues au caractère non uniforme de la sensibilité des antennes sont igno-

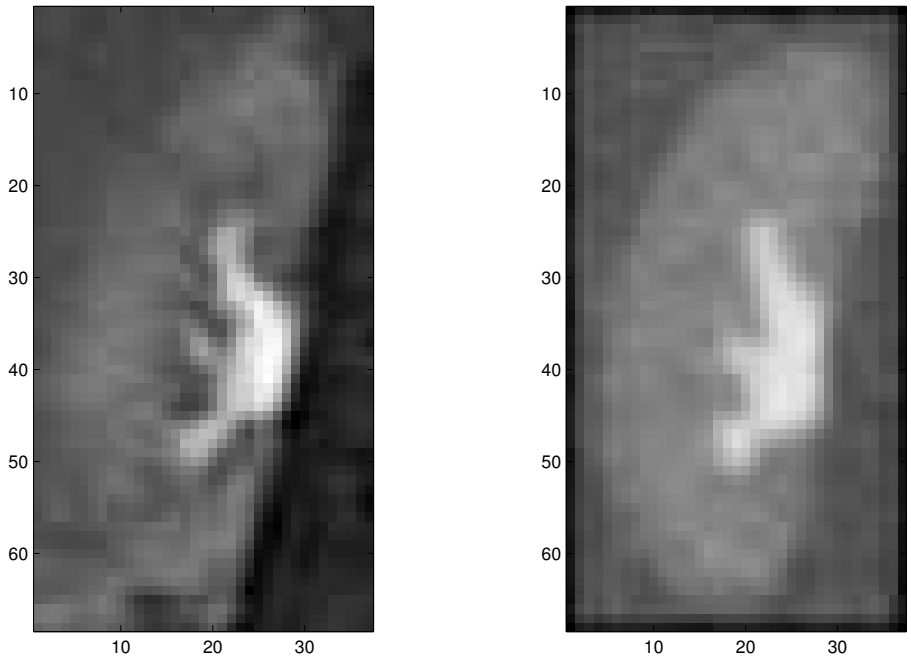


FIG. 4.8 – Exemple d'images réelle (a) et simulée (b) en phase tardive.

rées. Par ailleurs, l'anatomie du modèle tout comme les courbes temps-intensité typiques utilisées correspondent à un rein sain. La méthode devant à terme servir aussi pour des reins pathologiques, nous proposons de modifier le modèle pour l'adapter aussi à ce cas.

4.1.2 Données simulées : reins pathologiques

En raison de la variété des évolutions temporelles pour les reins pathologiques, en particulier pour le remplissage des cavités, il nous est impossible de recouvrir tous les cas. Nous avons choisi un modèle anatomique tenant compte des particularités généralement présentes et que nous avons associé à des évolutions temporelles typiques de deux cas fréquemment rencontrés. Le premier correspond à un remplissage tardif des cavités, le second à des cavités restant vides. Le reste du procédé de simulation reste identique à celui utilisé pour les reins sains.

Modèle anatomique

En général, le cortex et la médullaire sont beaucoup plus minces pour un rein pathologique que pour un rein sain, et il est parfois difficile de les différencier ; les cavités, en revanche, sont la plupart du temps dilatées (cf. figure 4.10). Pour générer les images de synthèse simulant l'acquisition, nous procédons de la même manière que pour le rein sain mais en utilisant le modèle anatomique simple représenté figure 4.11, qui tient compte des particularités mentionnées ci-dessus. Le modèle à volume partiel correspondant se trouve

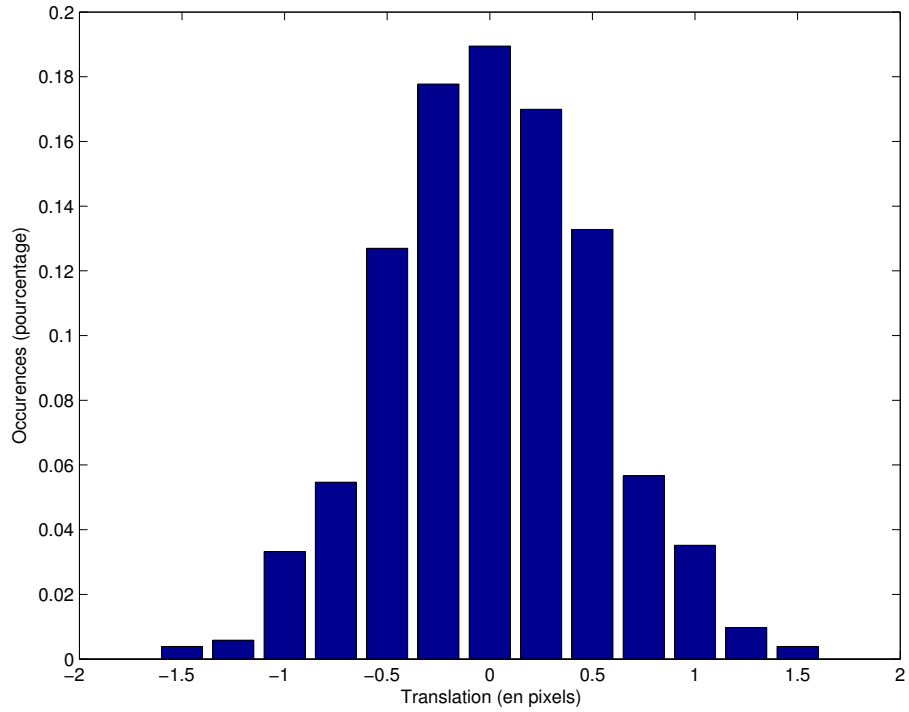


FIG. 4.9 – Histogramme des translations ajoutées aux images.

figure 4.12. On remarquera qu'aucun voxel n'est entièrement constitué de médullaire. Pour simuler un remplissage inhomogène dans les cavités, nous serons amenés à diviser ce compartiment en deux parties (voir ci-dessous).

Modèle de comportement temporel de chaque compartiment anatomique

Dans tous les cas, les courbes pour le cortex et la médullaire restent les mêmes que pour les reins sains. On comparera l'évolution temporelle pour les cavités à celle de la figure 4.4 page 114.

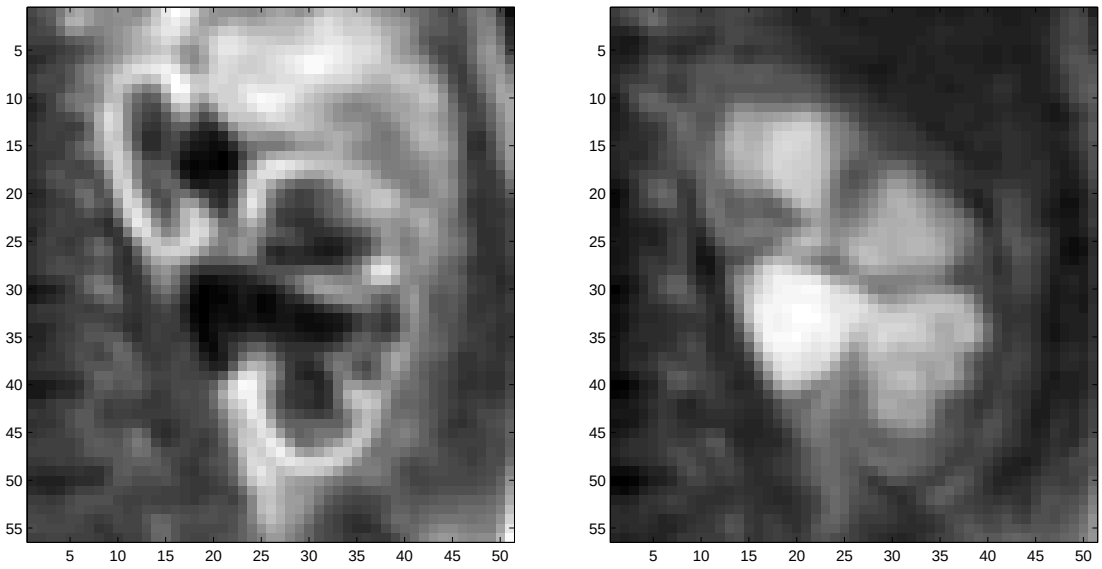
Cas de cavités se remplissant très tardivement Sur la figure 4.13, la courbe temporelle des cavités prend des valeurs élevées seulement en fin de la phase de perfusion tardive.

Cas de cavités ne se remplissant pas du tout Sur la figure 4.14, la courbe temps-intensité des cavités garde des valeurs très faibles tout au long de la perfusion.

Compartiment dominant

Tout comme pour les reins sains, nous étudions la correspondance entre compartiments anatomique et fonctionnel dominants.

En l'absence de bruit, pour le modèle à volume partiel proposé dans la solution 2 page 111, les segmentations anatomique et fonctionnelle sont les mêmes pour plus de



(a) (b)

FIG. 4.10 – Exemple d'images extraites d'une acquisition sur un rein pathologique près du pic cortical (a) et en phase tardive (b), avec cavités dilatées.

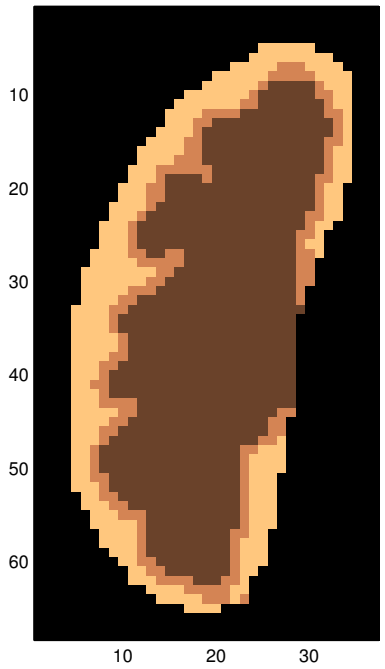


FIG. 4.11 – Modèle anatomique de rein pathologique.

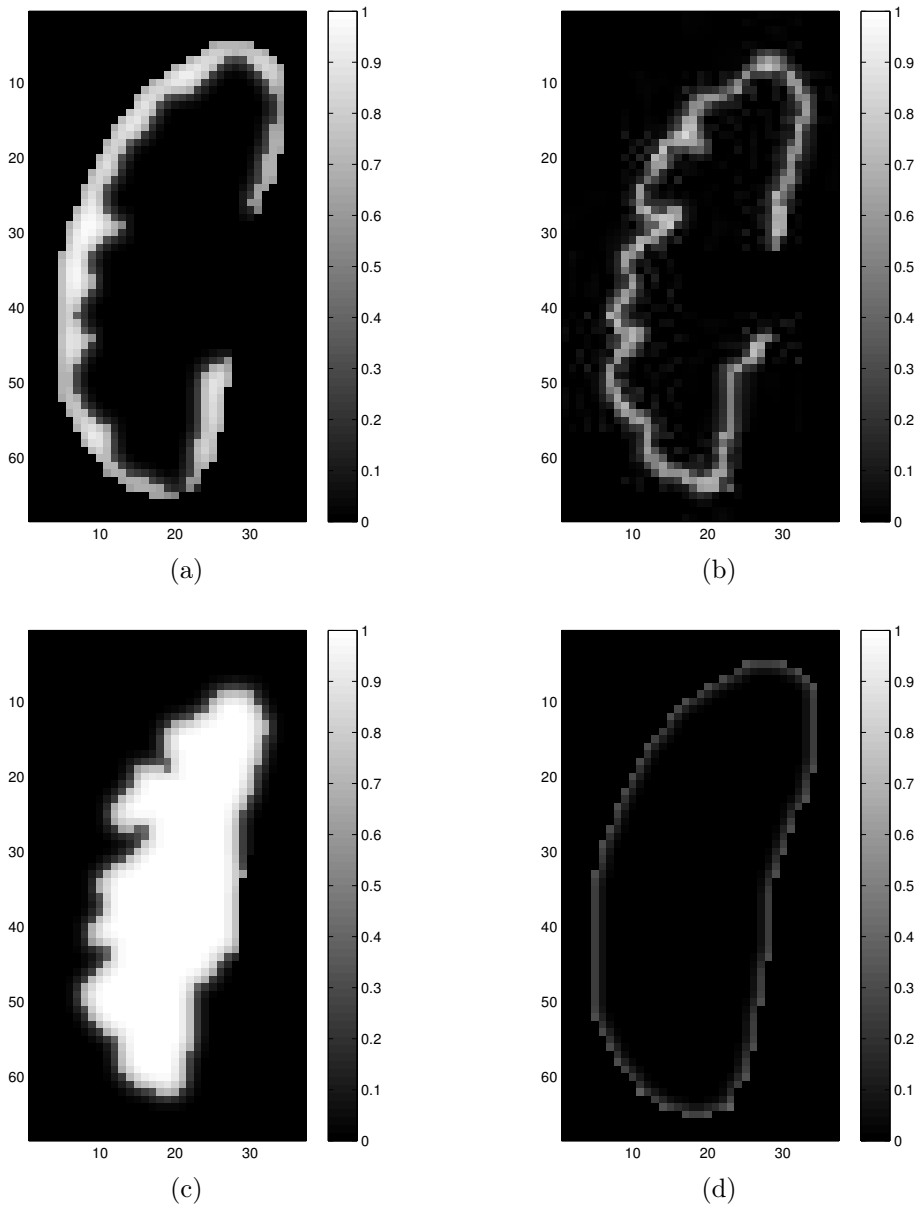


FIG. 4.12 – Exemple de coupe d'un modèle rénal anatomique pathologique à trois compartiments avec, pour chaque voxel, les proportions de cortex (a), médullaire (b), cavités (c) et organes extérieurs (d).

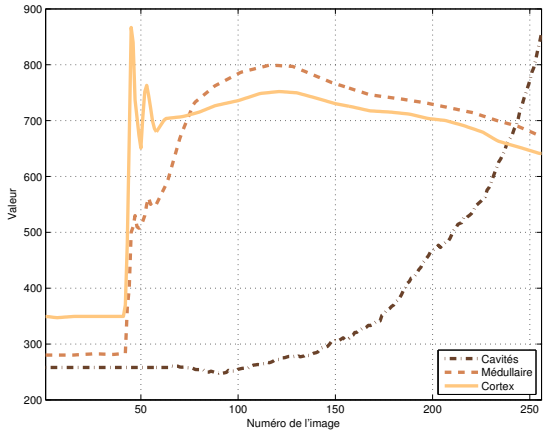


FIG. 4.13 – Courbes temps-intensité typiques non bruitées pour les trois compartiments rénaux pour un rein pathologique avec remplissage tardif des cavités.

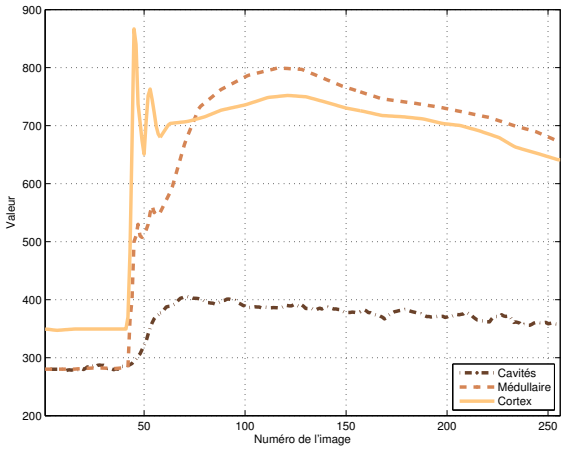


FIG. 4.14 – Courbes temps-intensité typiques non bruitées pour les trois compartiments rénaux pour un rein pathologique sans aucun remplissage des cavités.

99,5% des pixels. En revanche, en ajoutant du bruit ($\sigma = 40$), le pourcentage n'est plus que de 88% : en raison de la faible épaisseur de la médullaire, seuls 25% des pixels de ce compartiment sont identifiés, les autres étant attribués au cortex ou aux cavités. Cette confusion est accentuée par les erreurs de recalage : si le pourcentage de pixels pour lequel les segmentations anatomique et fonctionnelle coïncident est de 85%, ceci n'est vrai que pour 5% des pixels de la médullaire. Ces phénomènes sont illustrés sur les segmentations de la figure 4.15, à comparer avec la segmentation anatomique de la figure 4.11.

En conclusion, sur le modèle utilisé, alors que la non-correspondance entre compartiments anatomique et fonctionnel dominants n'est globalement guère plus élevée pour les reins pathologiques que pour les reins sains, elle est en revanche très élevée pour la médullaire, particulièrement touchée par le volume partiel à cause de son faible volume et de sa position intermédiaire entre le cortex et les cavités. Nous reviendrons sur ce point délicat au paragraphe 5.1.7.

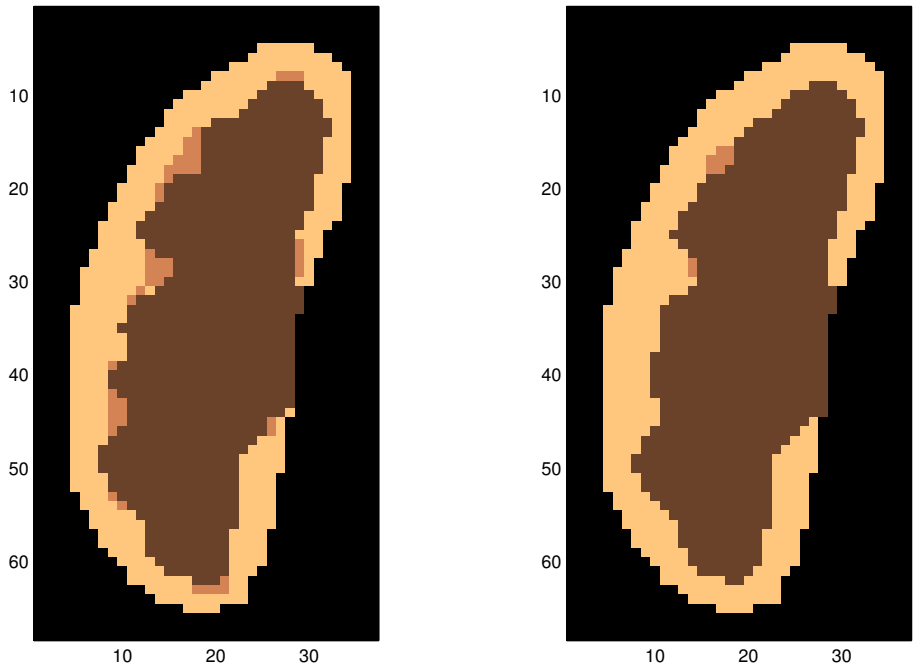


FIG. 4.15 – Segmentations fonctionnelles obtenues avec le modèle complet et un recalage parfait (a) et avec erreurs de recalage résiduelles (b).

4.1.3 Données réelles 2D ou 3D

Des éléments sur le processus d'acquisition en IRM dynamique à rehaussement de contraste sont donnés dans le paragraphe 2.2.1 page 32.

Huit séquences d'IRM de perfusion rénale à rehaussement de contraste pour des reins normaux sont à notre disposition. Ces acquisitions ont été réalisées sur huit patients avec un rein sain et un rein pathologique. Seul le rein sain a été retenu pour cette étude. Les examens ont été réalisés sur un appareil General Electric Healthcare à 1,5 T (corps

entier). Des séquences LAVA (écho de gradient 3D ultra-rapide) ont été utilisées, avec les paramètres suivants :

- taille de la matrice : 256×256 pixels ;
- taille des voxels : entre $1,172$ et $1,875$ mm dans le plan de coupe, 10 mm d'épaisseur de coupe
- angle de bascule : $15°$;
- T_R : $2,3$ ms ;
- T_E : $1,1$ ms.

L'examen dure en moyenne une dizaine de minutes. Un rectangle recouvrant la totalité du rein est tout d'abord sélectionné, la taille de la matrice correspondante variant entre 47×35 et 84×59 pixels dans la coupe contenant la plus grande proportion de tissu rénal. Des segmentations seront réalisées par deux radiologues suivant le protocole décrit au paragraphe 4.4.3 page 139.

Remarque sur les données réelles 3D

Pour l'instant, nous disposons de deux jeux de données 3D mais pas des segmentations manuelles associées.

4.2 Recalage

Nous avons souhaité tester une méthode qui ne nécessiterait pas le tracé d'un masque au préalable, pour se garder la possibilité de l'extraire automatiquement. Nous l'avons testée sur des données de synthèse et des données réelles.

4.2.1 Méthode choisie

Comme suggéré au paragraphe 2.4.3, pour éviter les risques de dérives, nous choisissons l'image enregistrée au pic cortical comme référence : toutes les images de la série seront recalées sur cette image. La mesure de similarité choisie est l'information mutuelle, en raison des changements de contraste pendant la perfusion et du fait que nous sommes amenés à recaler des images temporellement éloignées. Comme il nous semble impossible de distinguer localement les changements de contraste dus aux mouvements ou à la diffusion de l'agent de contraste, nous préférons nous limiter à des transformations rigides ; non seulement les transformations non rigides n'améliorent pas nécessairement le recalage, qui est d'ailleurs difficile à valider, mais le risque d'obtenir des résultats non pertinents n'est pas négligeable. L'examen des variances par compartiments après *clustering* [42] ne nous semble pas apporter nécessairement une preuve de la cohérence du recalage. Il demeure difficile de fixer *a priori* des contraintes réalistes sur les paramètres de la transformation afin d'aboutir à des résultats concluants.

Il est demandé à l'opérateur d'extraire un rectangle contenant le rein et le dépassant de quelques pixels à chaque extrémité. En effet, les transformations ne peuvent être considérées comme uniformes que localement, et le rein doit occuper la majorité du rectangle, pour que ce soient les pixels qui le représentent, et non ceux des organes voisins, qui influencent le plus la mesure de similarité.

L'information mutuelle est estimée à l'aide de fenêtres de Parzen, selon la méthode exposée dans [86]. Après un recalage au pixel près, un recalage subpixel avec rotation est réalisé. Cette méthode a été également choisie mais avec des transformations élastiques dans [42], de manière indépendante, les résultats ayant été publiés après que nous l'avons utilisée.

Notons que notre objectif final est de segmenter les structures internes du rein et que nous souhaitons donc surtout obtenir des résultats suffisants pour aboutir par la suite à une segmentation convenable, ce qui sera étudié dans le chapitre 5. Ainsi, pour valider les résultats du recalage avant segmentation, nous avons choisi une méthode qualitative, essentiellement pour des questions de temps et en raison des difficultés évoquées ci-dessus pour les transformations autres que les translations. Nous avons tracé manuellement les contours du rein sur l'image de référence, puis nous les avons reportés sur les autres images de la série, comme sur la figure 2.11. La « vidéo » de la série est alors visualisée.

Les résultats sont aussi comparés à ceux obtenus par la méthode des gradients de [43].

Des planches de résultats se trouvent dans l'annexe A page 205 et suivantes, tant pour des données de synthèse que pour des données réelles.

4.2.2 Résultats

Données de synthèse

Lorsque l'on recale une image et sa transformée par une translation connue, comme sur la figure A.1, on retrouve la transformation au vingtième de pixel près. Les différences observées sur la figure (d) viennent essentiellement de la méthode d'interpolation, qui utilise des splines d'ordre 3. Des résultats pour une rotation simple, puis pour une translation combinée à une rotation sont données respectivement sur les figures A.2 et A.3.

Les mêmes transformations sont appliquées à une image proche de l'image de référence dans la séquence temporelle (figures A.4 à A.6) puis à une image en phase de perfusion tardive (figures A.7 à A.9). Nous ne retrouvons pas exactement l'image avant transformation, à la fois à cause de l'interpolation et parce que la transformation obtenue par optimisation n'est pas tout à fait identique à la transformation réellement utilisée. Nous pouvons cependant remarquer que le rein est toujours très bien cadré dans le masque global.

Données réelles

En ce qui concerne les données réelles, deux exemples sont présentés sur les figures A.10 et A.11 : dans le premier cas, les images sont relativement proches dans la séquence (pic cortical et filtration) et plus éloignées dans le second (l'image à recaler se situant dans la phase tardive). Sur ces images apparaît la difficulté de valider le résultat en raison du manque de netteté des contours et de l'apparition ou de la disparition de certaines structures internes selon la phase de perfusion. On constate cependant qu'après recalage, le rein paraît beaucoup mieux cadré dans le masque global.

Conclusion

Les tests ont été réalisés sur huit reins (cf. paragraphe 4.1.3 pour le détail des acquisitions). Rappelons que les mouvements hors plan de coupe ne peuvent pas être corrigés, même en utilisant les coupes voisines, en raison de la faible résolution spatiale dans la direction perpendiculaire aux coupes sur nos séquences. Par ailleurs, l'acquisition demeure très bruitée, ce qui est typique de ce genre de séquences très rapides. Nous pouvons remarquer que, sur certaines images particulièrement dégradées, aucune des deux méthodes testées ne donne des résultats convenables, alors qu'ils sont similaires sur les autres images. Il est donc à noter que le recalage des données réelles ne sera pas parfait pour l'intégralité des images de la série, mais on peut considérer qu'il est correct à une fraction de pixel près sur la plupart d'entre elles, que l'erreur peut atteindre parfois 1 à deux pixels sur quelques images (10%), et qu'il subsiste sur certaines séries quelques mouvements non corrigibles (environ 1%).

Nous verrons dans le chapitre 5 si la correction de mouvement ainsi obtenue est suffisante pour ne pas gêner la segmentation fonctionnelle des compartiments rénaux internes.

4.3 Stratégie proposée

Notre objectif est d'effectuer, à partir d'une séquence de DCE-MRI 3D recalée (qui contient donc plusieurs coupes), la segmentation fonctionnelle d'un rein en classant les voxels rénaux en 3 compartiments (cortex, médullaire et cavités) d'après leur évolution temporelle, c'est-à-dire en utilisant l'ensemble des vecteurs temps-intensité $\{\xi_i\}_{1\leq i\leq N'}$, de dimension p égale au nombre d'instants d'acquisition, N' étant le nombre total de voxels rénaux sur l'ensemble des coupes, ξ_i le vecteur dont la j-ème composante est liée à l'intensité du i-ème voxel au j-ème temps d'acquisition. Ces vecteurs, selon les hypothèses du paragraphe 3.1.1 page 55 sont considérés comme la réalisation d'une séquence de variables aléatoires vectorielles continues $\{\mathbf{X}_1, \ldots, \mathbf{X}'_N\}$ identiquement distribuées de même loi $P_{\mathbf{X}}$. Différentes stratégies mettant en jeu des méthodes d'apprentissage sont envisageables. Le choix final a été effectué en tenant compte de la vraisemblance des hypothèses et des préférences des radiologues à propos du côté pratique de la méthode.

4.3.1 Apprentissage supervisé

Une méthode possible serait la suivante :

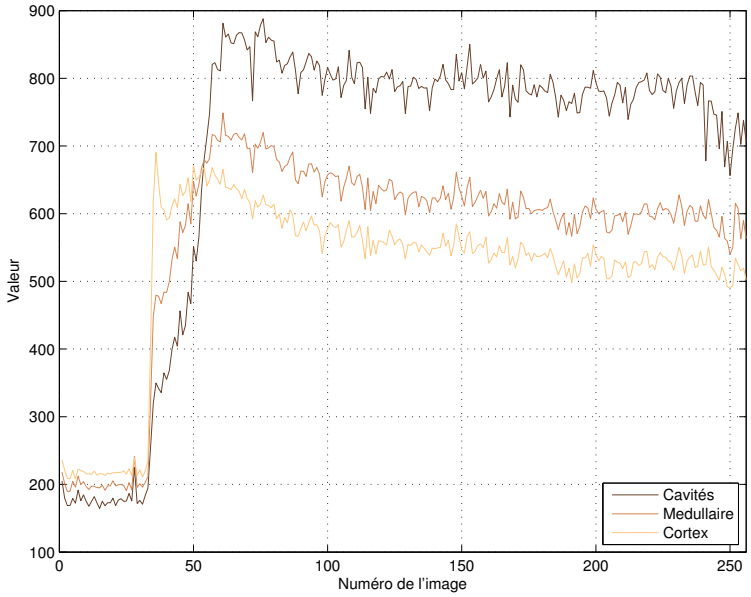
Etape 1 Créer une base de données constituée de reins segmentés par des radiologues,

Etape 3 Réaliser un apprentissage (supervisé) sur cette base,

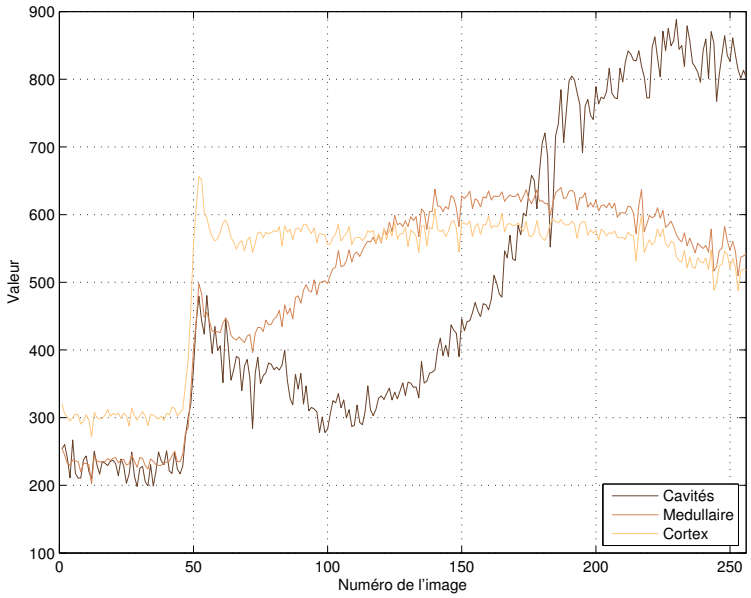
Etape 3 Utiliser le classificateur obtenu sur les autres reins.

Cependant, nous ne disposons pas actuellement de données en nombre suffisant pour constituer une telle base d'apprentissage. Par ailleurs, comme un rein comporte trois compartiments, nous pouvons nous attendre à ce que la distribution de probabilité des vecteurs temps-intensité ait l'équivalent de trois modes en dimension p. Les conditions d'acquisition peuvent néanmoins modifier assez considérablement l'évolution temporelle apparente d'un rein à l'autre (cf. figure 4.16) : comme nous l'avons déjà souligné page 111, la forme

des courbes temps-intensité dépend de plusieurs facteurs, dont la vitesse d'injection du produit de contraste, de la fonction cardiaque, de la dispersion et de la recirculation [26]. Ainsi, en prenant les données issues de plusieurs reins, les trois modes risquent de ne plus être présents, en raison du mélange des distributions, et l'hypothèse de base selon laquelle toutes les données sont tirées suivant une densité de probabilité fixée permettant de distinguer les trois compartiments n'est plus valide. C'est pourquoi nous avons éliminé cette solution.



(a)



(b)

FIG. 4.16 – Exemple de courbes temps intensité par compartiment, issues de reins de deux patients différents.

4.3.2 Apprentissage non supervisé par quantification vectorielle sur l'ensemble des voxels rénaux

Nous proposons de classifier les voxels d'un rein donné en construisant de manière semi-automatique ou entièrement automatique un classificateur adapté à ce rein particulier. Il est possible de faire directement la classification sur l'ensemble des voxels rénaux, de manière à obtenir immédiatement une segmentation tridimensionnelle. Cependant, plusieurs inconvénients apparaissent :

- les effets de volume partiel sur les coupes extrêmes risquent de détériorer l'ensemble de la segmentation pour certaines acquisitions,
- le risque de déséquilibre entre classes est augmenté, pouvant mener à la perte d'un des compartiments,
- l'éventuelle intervention de l'opérateur pour les méthodes semi-automatiques est moins aisée : par exemple, les seuils que l'opérateur doit fixer lors de l'étape de fusion (cf. paragraphe 5.1.5 page 159) sont plus difficiles à régler.

Nous n'avons pas retenu cette méthode car il est ressorti de nos discussions avec des radiologues que, pour des raisons d'ergonomie essentiellement, ces derniers préfèrent participer à la réalisation de la segmentation de la meilleure coupe disponible, en ajustant les seuils pour la méthode utilisant le GNG-T ou en testant différents nombres de classes pour l'algorithme des K-moyennes. Il est en effet moins aisé d'observer en même temps 4 à 6 coupes, surtout si les structures internes ne sont pas toujours bien visibles sur celles-ci. Il s'agit alors d'utiliser le meilleur classificateur parmi ceux qui auront été construits pour classer les voxels des autres coupes. Cette méthode, appelée méthode mixte, est décrite plus en détails dans le paragraphe suivant.

4.3.3 Apprentissage non supervisé sur une coupe de référence et généralisation

La méthode que nous proposons pour réaliser la segmentation d'un rein donné en dimension 3 est la suivante :

Etape 1 réaliser une segmentation semi-automatique sur la coupe contenant le volume rénal le plus important, dite coupe de référence, sur laquelle figure une partie de chacun des trois compartiments à segmenter,

Etape 2 généraliser les résultats aux autres coupes.

La première étape consiste à choisir, parmi une série de classificateurs disponibles, celui qui donne la segmentation de la coupe de référence qui se rapprocherait le plus de celle qu'aurait faite le radiologue, dans un sens à définir (cf. paragraphe 5.2). Les voxels des autres coupes sont alors soumis à ce même classificateur pour avoir la segmentation complète en dimension 3. En procédant ainsi, on limite le risque de détérioration de l'ensemble de la classification par les voxels des coupes extrêmes, ce qui est préférable même si certains voxels de ces coupes peuvent être au final mal classés. Par ailleurs, l'intervention de l'opérateur est plus aisée dans ce cas. Le cadre mathématique choisi pour la modélisation de cette démarche est décrit au paragraphe 5.2, en lien avec le chapitre 3.2.

Une fois la stratégie choisie, il reste à déterminer une méthode de validation des résultats, ce qui fait l'objet du paragraphe suivant.

4.4 Validation des résultats

Les résultats peuvent être validés qualitativement et quantitativement.

Pour une validation qualitative, nous pouvons soumettre la segmentation obtenue à un radiologue, en superposant ses contours à certaines images issues de la série, comme sur la figure 5.31, et recueillir son avis d'expert quant à la cohérence des structures obtenues par rapport à l'anatomie rénale. Nous verrons que cette première étape est indispensable car nous n'avons pas trouvé de critère quantitatif susceptible de traduire véritablement cette qualité.

Pour valider quantitativement les résultats, qu'il s'agisse de données simulées ou de données réelles, nous serons amenés à comparer deux *clusterings* ou deux segmentations. La méthode de validation diffère quelque peu pour les données simulées et les données réelles, du fait que nous disposons d'une vérité-terrain pour les premières mais seulement d'une « pseudo vérité-terrain » pour les secondes.

4.4.1 Critères quantitatifs retenus pour les comparaisons

Nous souhaitons conserver seulement un nombre limité de critères suffisamment significatifs et facilement interprétables parmi ceux présentés dans les paragraphes 3.1.5 et 3.1.6. Rappelons qu'il est possible de comparer soit des *clusterings*, soit des segmentations, puisque dans notre application, nous avons accès à l'un et l'autre. Dans la littérature concernant l'IRM rénale, ce sont des segmentations qui sont comparées, ce qui peut se comprendre dans la mesure où un radiologue peut donner son avis sur une segmentation et non sur un *clustering*. Nous conservons tout de même un critère de comparaison de clusterings, qui permet de juger le résultat dans son ensemble, sans le décomposer par compartiment comme pour les segmentations. Vu les inconvénients des critères liés au comptage de paires (cf. paragraphe 3.1.5), nous avons accordé notre préférence au pourcentage de pixels bien classés WCP (cf. paragraphe 3.1.5 et équation (3.47)).

Une segmentation fonctionnelle peut être qualifiée de satisfaisante si elle ressemble suffisamment à une segmentation anatomique, c'est-à-dire si on retrouve bien des structures cohérentes proches de celles définies manuellement. Il faut aussi que les trois compartiments soient convenablement représentés, donc trouver des critères qui soient représentatifs pour des régions d'intérêt (*regions of interest* ou ROI) de tailles différentes. Remarquons qu'il existe une certaine subjectivité et des *a priori* sur l'anatomie rénale qui ont une influence lors de l'établissement des segmentations manuelles, ainsi que sur la satisfaction « visuelle » qui en découle et qui est extrêmement difficile à quantifier.

Pour illustrer ce point, nous avons représenté différentes segmentations sur les figures 4.17 à 4.19, à comparer avec la segmentation de référence de la figure 4.17(a). Nous avons évalué quelques critères de similarité, les résultats étant présentés dans les tableaux 4.1 à 4.3.

Parmi les critères permettant de comparer deux segmentations, et qui sont présentés dans l'état de l'art du paragraphe 3.1.6 page 91, il nous a semblé intéressant de conserver :

- les pourcentages de recouvrement et de dépassement, faciles à interpréter, et qui doivent toujours être examinés ensemble ;
- un indice parmi les coefficients de correspondance simple, de Tanimoto et de Dice ;
- une des distances entre frontières.

Ceci permet d'avoir des critères basés sur différentes propriétés, sans en multiplier le nombre à outrance.

Nous avons éliminé les coefficients de Tanimoto et de correspondance simple, parce que nous souhaitons segmenter au mieux trois compartiments de tailles différentes, les cavités étant souvent sensiblement plus petites que les deux autres. Or, nous avons vu que les deux premiers donnent des résultats plus favorables pour les compartiments de petite taille en cas de sous-segmentation, à cause de l'importance des pixels étiquetés VP, même si c'est moins vrai pour le coefficient de Tanimoto que pour l'indice de correspondance simple (cf. tableau 3.1 page 96). L'indice de Dice SI semble mieux adapté dans notre cas. Il permet de savoir d'emblée si un compartiment est négligé au profit des deux autres, ce qui plus difficile à voir avec les pourcentages de recouvrement et de dépassement ; ces derniers peuvent servir à expliquer l'origine d'un bon ou d'un mauvais SI. Ainsi, sur la figure 4.19, la médullaire a été dilatée de manière homogène, conduisant à une érosion comparable pour les deux autres compartiments ; le SI est approximativement le même pour les trois ROIs, et les pourcentages de recouvrement traduisent bien la dilatation et les érosions décrites.

Nous n'avons pas retenu la distance de Hausdorff car elle peut prendre des valeurs très élevées même pour une segmentation qui, au final, paraît satisfaisante : par exemple, cela se produit lorsque quelques pixels en périphérie du rein sont mal classés par des méthodes fonctionnelles à cause d'un masque global un peu déficient, et qui n'ont pas une grande importance pour ce que nous voulons faire de la segmentation par la suite. Ce cas est illustré sur la figure 4.18(c), les résultats numériques apparaissant dans le tableau 4.2 : la distance de Hausdorff vaut entre $6,1$ et $19,6$ selon le compartiment, ce qui n'est pas du tout représentatif du résultat. Ainsi, seule la distance moyenne apparaîtra dans les autres tableaux de comparaison.

Remarquons que nous évaluons à la fois un critère global (WCP) tenant compte de la classification complète en trois compartiments, et des critères par compartiment. Cela évite que les cavités, par exemple, ne soient trop négligées parce que, leur volume étant moindre par rapport à celui des autres régions, elles ont moins de poids dans les critères globaux : cela permet une analyse plus fine des résultats.

Examen de quelques segmentations

Il est évident que les segmentations de la figure 4.17(b), (c) et (d) ne sont pas acceptables, et qu'elles seraient rejetées par un simple examen visuel. Si, pour les deux premières, les critères de similarité ont des valeurs basses, pour la troisième, les résultats ne sont pas tellement éloignés de ceux obtenus pour des segmentations paraissant plus satisfaisantes, hormis pour la distance moyenne des cavités ; le WCP est par exemple le même que celui des segmentations (a) et (c) de la figure 4.19. D'où l'importance d'exa-

miner d'abord visuellement la cohérence de la segmentation. Dans nos tests, que ce soit sur données de synthèse ou sur données réelles, nous n'avons cependant jamais abouti à de tels résultats.

La segmentation de la figure 4.18(a) est meilleure que la (b), ce qui se retrouve dans tous les critères de comparaison. La segmentation (d) est à peu près comparable à la (b) pour le cortex et la médullaire, mais semble visuellement moins bonne pour les cavités : la petite partie détachée est plus éloignée, et la forme de la partie principale est dégradée ; les critères de similarité ont des valeurs nettement moins bonnes, par exemple le SI baisse de $0,87$ à $0,78$.

Quant aux segmentations de la figure 4.19, sur (b) et (d), la médullaire a été dilatée de deux façons différentes, alors qu'elle a été érodée au profit du cortex et des cavités sur (c) : cette situation se rencontre fréquemment car les frontières entre compartiments sont parfois mal définies en raison du flou des images et de la difficulté à interpréter les niveaux de gris pour les segmentations manuelles (cf. par exemple la figure 5.31 page 197). Sur (a), l'ensemble est décalé de deux pixels vers le bas et d'un pixel vers la droite, ce qui peut correspondre à des choix d'images différents avec un recalage déficient. La qualité de ces segmentations paraît à peu près semblable, et les critères de ressemblance du tableau 4.3 ont des valeurs similaires.

En conclusion, il n'existe pas véritablement de critère qui traduise de manière satisfaisante la qualité d'une segmentation par comparaison à une référence, sachant qu'il est également difficile de l'exprimer qualitativement. Les critères que nous avons choisis permettent de savoir si deux segmentations sont relativement proches l'une de l'autre. Si, pour les exemples choisis, nous atteignons un WCP supérieur à 80%, un SI supérieur à $0,75$ et une distance moyenne entre frontière inférieure à $0,7$ pixel, donner une valeur limite à chacun de ces critères pour décider si une segmentation fonctionnelle ressemble suffisamment à celle d'un radiologue pour être acceptable reste cependant sujet à caution, d'autant plus que les segmentations manuelles ne constituent pas une vérité-terrain à proprement parler et peuvent varier sensiblement d'un radiologue à l'autre. Ceci devra être pris en compte lors du choix de la méthode de validation, surtout pour les données réelles.

4.4.2 Méthode de validation pour les données simulées

Pour valider les résultats, nous pourrions évaluer des critères de similarité en prenant comme référence la segmentation ayant servi à générer les images et la segmentation fonctionnelle. Il en est de même pour les *clusterings*. Il nous a cependant paru suffisant, au vu des résultats obtenus, de ne donner que le WCP et de décrire les différences entre la référence et la segmentation test ; ceci évite la multiplication de tableaux de résultats.

4.4.3 Méthode de validation pour les données réelles

Pour les données réelles, nous ne disposons pas de vérité-terrain à proprement parler. Nous pouvons comparer nos résultats à des segmentations réalisées par des radiologues, démarche fréquemment adoptée dans la littérature.

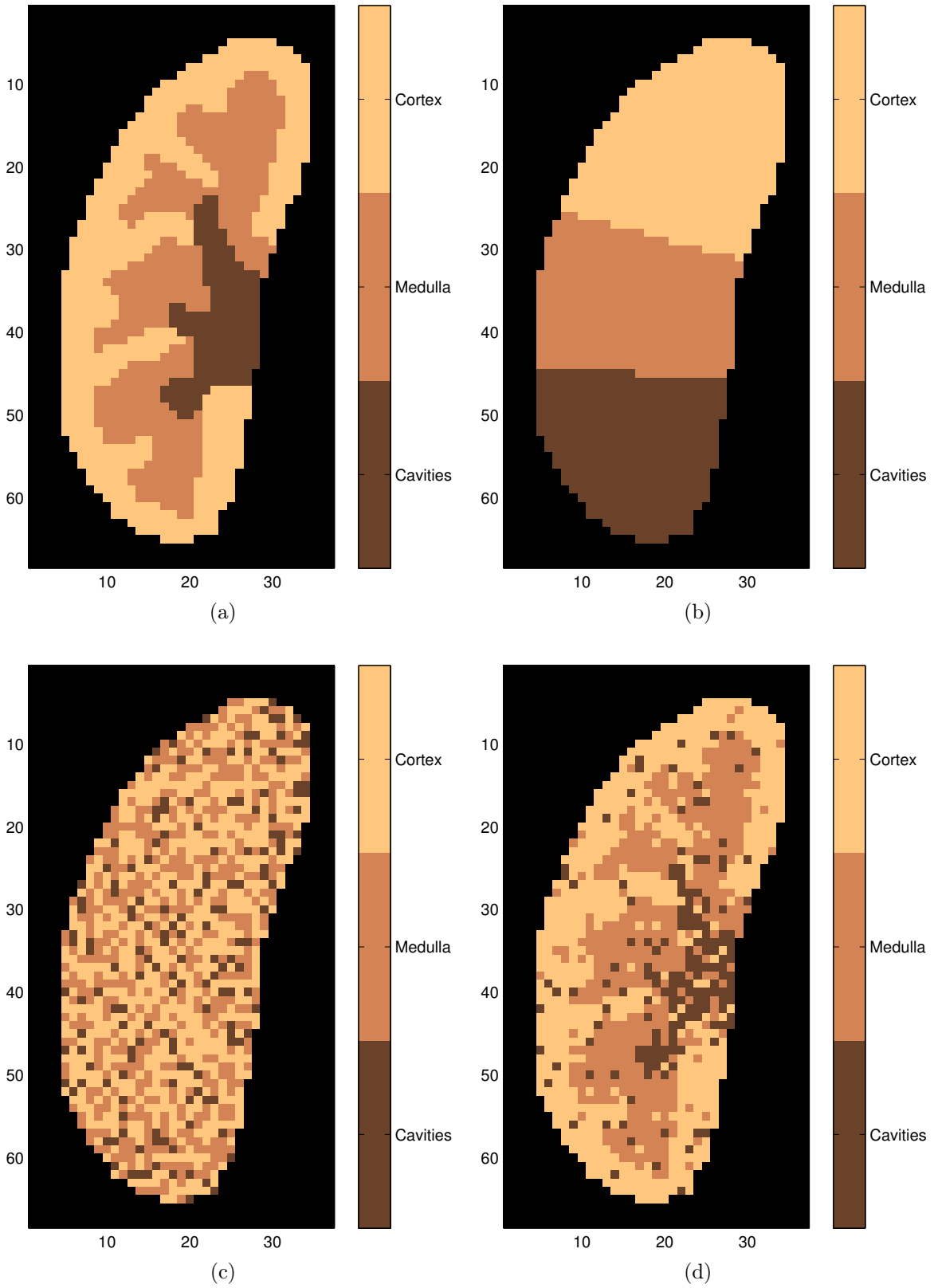


FIG. 4.17 – Comparaison de segmentations : segmentation de référence (a) et segmentations à tester (b), (c) et (d).

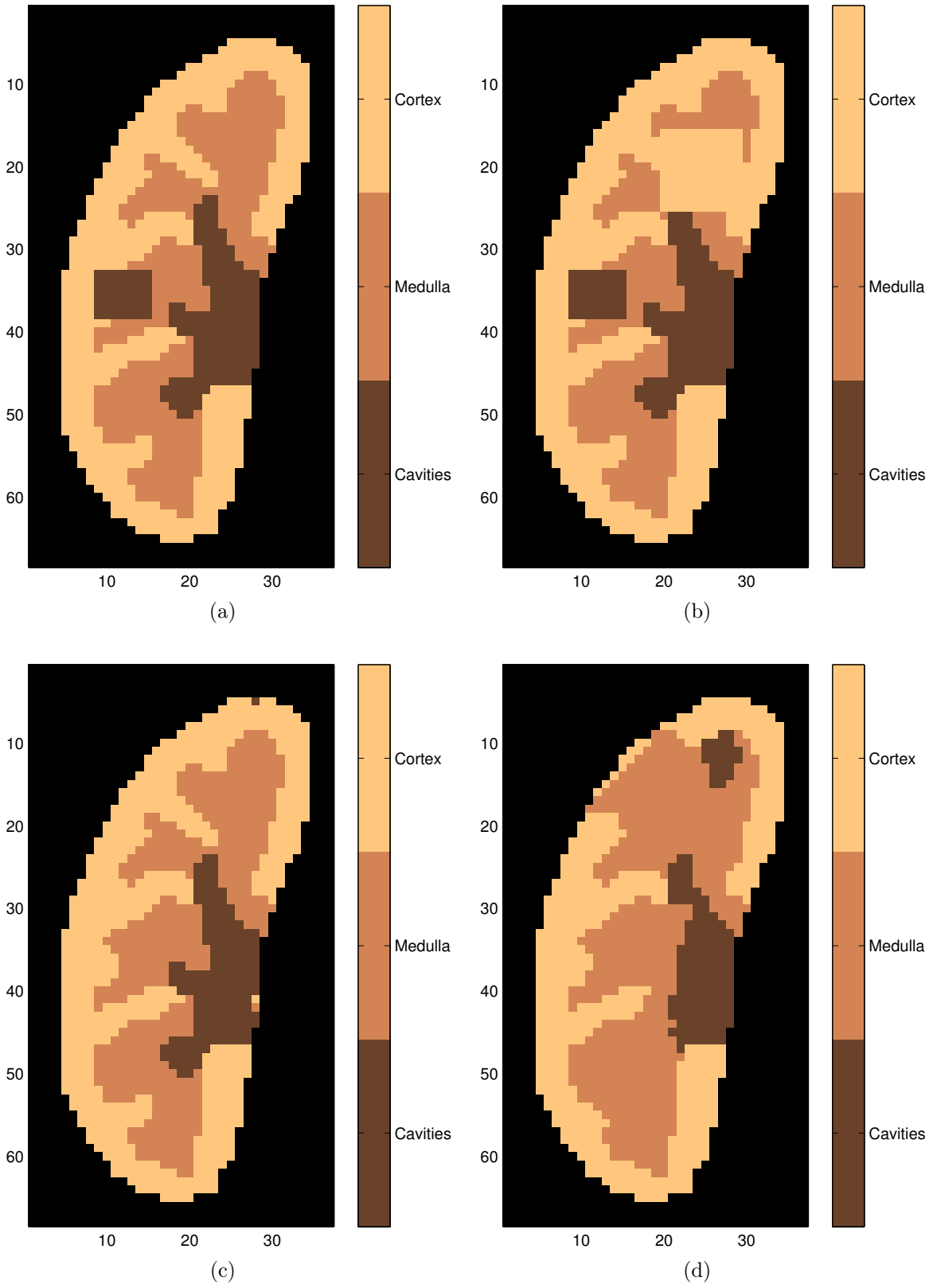


FIG. 4.18 – Comparaison de segmentations : segmentation de référence (a) et segmentations à tester (b), (c) et (d).

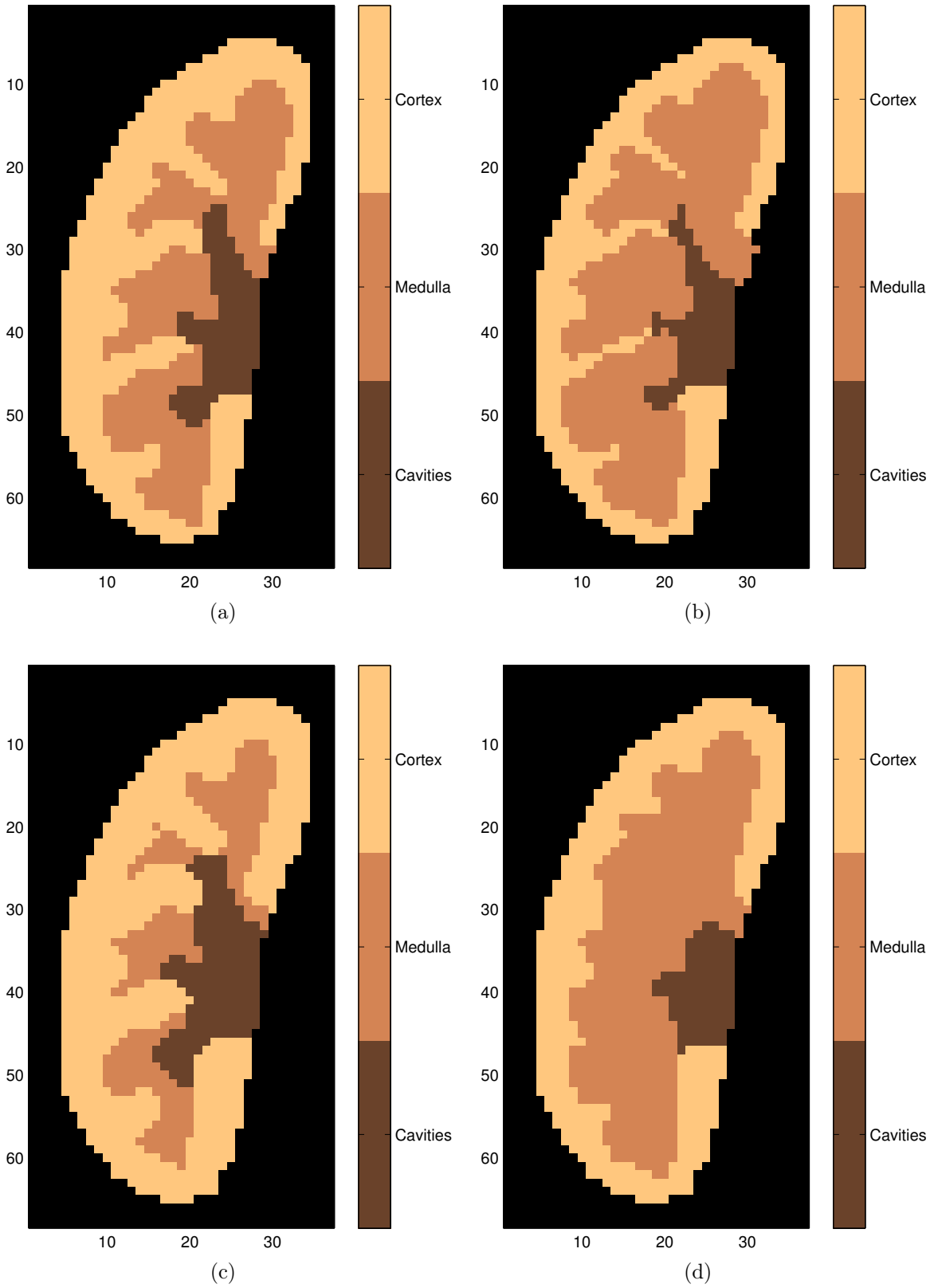


FIG. 4.19 – Comparaison de segmentations : segmentation de référence (a) et segmentations à tester (b), (c) et (d).

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	100	100	100
Dépassement (%)	0	0	0
Indice de similarité	1	1	1
Distance moyenne	0	0	0

(a) Segmentation identique à la segmentation de référence, WCP = 100%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	16	28	61
Dépassement (%)	212	57	0
Indice de similarité	0,1	0,30	0,75
Distance moyenne	9,2	1,15	0,0

(b) Segmentation de la figure 4.17(b), WCP = 33%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	18	37	48
Dépassement (%)	82	63	52
Indice de similarité	0,18	0,36	0,48
Distance moyenne	7,6	1,8	1,6

(c) Segmentation de la figure 4.17(c), WCP = 40%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	61	86	90
Dépassement (%)	40	13	10
Indice de similarité	0,61	0,86	0,90
Distance moyenne	3,4	0,9	0,57

(d) Segmentation de la figure 4.17(d), WCP = 84%

TAB. 4.1 – Mesures de similarité pour les segmentations des trois ROIs sur les segmentations de la figure 4.17.

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	100	94	98
Dépassement (%)	25	0	0
Indice de similarité	0,88	0,97	0,99
Distance moyenne	1,3	0,0	0,0

(a) Segmentation de la figure 4.18(a), WCP = 97%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	97	77	98
Dépassement (%)	26	0	14
Indice de similarité	0,87	0,87	0,92
Distance moyenne	1,4	0,1	0,2

(b) Segmentation de la figure 4.18(b), WCP = 90%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	99	100	100
Dépassement (%)	0	0	0
Indice de similarité	0,99	1,00	1,00
Distance moyenne	0,3	0,0	0,0
Distance de Hausdorff	19,6	7,8	6,1

(c) Segmentation de la figure 4.18(c), WCP = 100%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	75	94	85
Dépassement (%)	18	28	0
Indice de similarité	0,78	0,84	0,91
Distance moyenne	3,2	0,6	0,0

(d) Segmentation de la figure 4.18(d), WCP = 79%

TAB. 4.2 – Mesures de similarité pour les segmentations des trois ROIs sur les segmentations de la figure 4.18.

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	80	77	88
Dépassement (%)	11	20	15
Indice de similarité	0,84	0,79	0,88
Distance moyenne	0,5	0,7	0,4

(a) Segmentation de la figure 4.19(a), WCP = 84%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	73	100	76
Dépassement (%)	0	42	0
Indice de similarité	0,84	0,83	0,86
Distance moyenne	0,4	0,7	0,3

(b) Segmentation de la figure 4.19(b), WCP = 90%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	92	61	100
Dépassement (%)	32	0	45
Indice de similarité	0,82	0,75	0,81
Distance moyenne	0,7	0,7	0,7

(c) Segmentation de la figure 4.19(c), WCP = 84%

Critère	Cavités	Médullaire	Cortex
Recouvrement (%)	63	100	85
Dépassement (%)	0	32	0
Indice de similarité	0,78	0,86	0,92
Distance moyenne	0,2	0,3	0,0

(d) Segmentation de la figure 4.19(d), WCP = 88%

TAB. 4.3 – Mesures de similarité pour les segmentations des trois ROIs sur les segmentations de la figure 4.19.

Segmentations manuelles anatomiques de référence pour les données réelles

Deux radiologues expérimentés ont examiné les séquences pour délimiter d'abord un masque du rein entier, puis les structures internes : cortex, médullaire et cavités. La procédure retenue est la suivante :

- visualisation de la séquence complète,
- sélection d'une image en phase tardive, sur laquelle les cavités apparaissent nettement, détermination du masque global puis segmentation des cavités,
- sélection du pic cortical, segmentation du cortex, puis de la médullaire (par différence entre le cortex et les cavités précédemment délimitées).

Le masque global ainsi défini peut légèrement différer d'un radiologue à l'autre en raison du flou des images et parce que les images choisies pour réaliser la segmentation peuvent être différentes. Le masque global finalement retenu est l'intersection des masques globaux des deux radiologues. Les trois ROIs définies par la suite sont incluses dans ce masque global, qui est également utilisé par la suite pour la segmentation fonctionnelle. Un exemple de segmentation manuelle est présenté figure 5.31(a) et (b). Notons que cette manière de procéder peut « favoriser » les cavités au détriment de la médullaire : en raison du volume partiel, certains voxels composés d'un mélange de médullaire et de cavités ont un comportement dominant de cavités en phase tardive et seront considérés comme tels parce que ce compartiment est extrait en premier, alors qu'il ressemble davantage à de la médullaire aux alentours du pic cortical.

Méthode de validation

Les segmentations manuelles précédemment décrites sont cependant sujettes à une variabilité intra- et inter-opérateur importante et comportent bien sûr une part de subjectivité : l'opérateur a en particulier un *a priori* sur la structure rénale qui l'influence sans doute lors de la délimitation des zones où la frontière est peu nette. Un choix d'images différent dans la séquence peut également entraîner des différences dans les segmentations, surtout en raison de mouvements hors plan de coupe. Il nous paraît donc difficile de considérer ces segmentations comme des vérités-terrain au sens strict du terme puisqu'elles peuvent varier notablement pour un même rein et une même séquence d'acquisition.

Nous proposons donc de comparer quantitativement les segmentations (semi-)automatiques avec des segmentations manuelles en évaluant les quelques critères de ressemblance retenus au paragraphe 4.4.1 et, comme point de référence, de calculer ces mêmes critères entre deux segmentations manuelles.

Parmi ces critères, comme nous l'avons déjà remarqué au paragraphe 3.1.6, seul l'indice de similarité garde la même valeur lorsque l'on échange les rôles de la référence et de la segmentation test, mais les mêmes valeurs apparaissent deux fois dans les tableaux de résultats pour faciliter les comparaisons. Pour les données réelles, ces tableaux contiennent les valeurs moyennes sur les huit reins des critères retenus.

En ce qui concerne le pourcentage de pixels bien classés, pour le rein complet, le coefficient WCP est la somme de tous les voxels VP sur les huit reins testés, quel que soit le compartiment, divisé par la somme de tous les voxels rénaux. Pour la segmentation d'un compartiment donné, WCP est la somme sur les huit reins des voxels VP divisé par le nombre total de voxels pour cette ROI. Le coefficient WCP est un peu différent

du pourcentage de recouvrement moyen, dans la mesure où les reins de petites taille l'influencent moins.

Représentations graphiques des résultats

Afin de faciliter l'analyse des résultats, nous proposons aussi quelques représentations graphiques pour l'indice de similarité et les pourcentages de recouvrement et de dépassement.

Indice de similarité Nous suggérons de représenter les résultats pour chaque compartiment sous forme de boîte à moustache, dont un exemple est représenté sur la figure 4.20 : le trait rouge représente la valeur médiane de l'indice de similarité, la boîte bleue le premier et le troisième quartile, les segments noirs les valeurs extrêmes sur l'ensemble des reins testés.

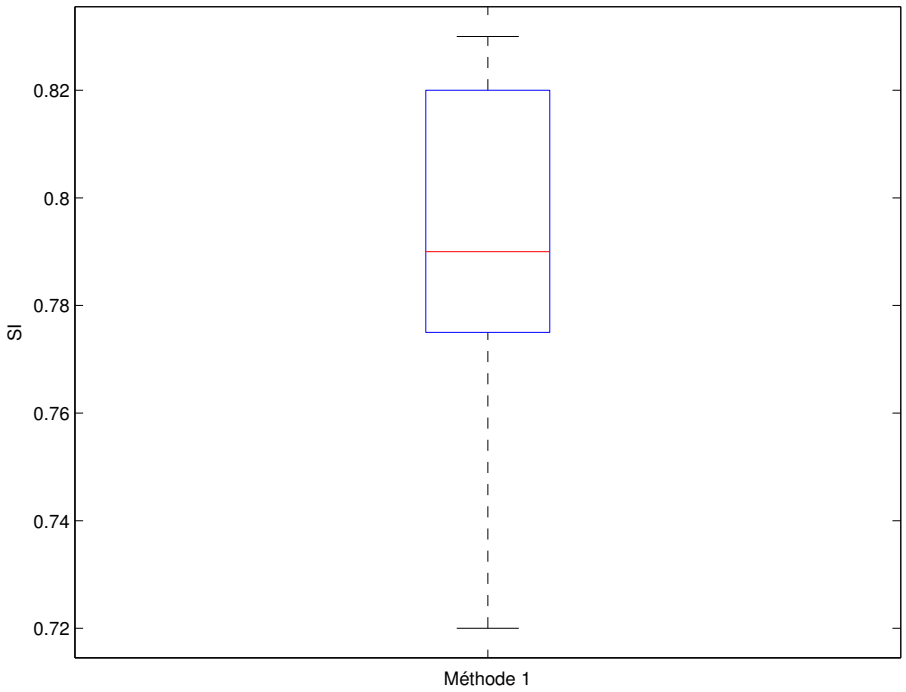


FIG. 4.20 – Représentation des indices de similarité obtenus pour une série de reins sous forme de boîte à moustache : pour un compartiment donné, le trait rouge représente la valeur médiane, la boîte bleue le premier et le troisième quartile, les segments noirs les valeurs extrêmes sur l'ensemble des reins testés.

Pourcentages de recouvrement et de dépassement Les pourcentages de recouvrement et de dépassement doivent être examinés conjointement : une véritable amélioration de la segmentation se traduit à la fois par un recouvrement plus élevé et un dépassement moindre. Sur la figure 4.21, supposons que l'on ait représenté par un carré rouge la valeur moyenne du dépassement sur l'ensemble des reins traités en fonction de celle du recouvrement par rapport à la segmentation de référence pour la seconde segmentation manuelle :

le plan est alors divisé en quatre secteurs A, B, C et D. Représentons ces mêmes valeurs pour une segmentation semi-automatique par un losange bleu. Selon le secteur où il est positionné, nous pouvons tirer les conclusions suivantes :

- sur (a), la segmentation semi-automatique ressemble davantage à la pseudo vérité-terrain que la segmentation de référence, puisque le pourcentage de recouvrement est supérieur et le pourcentage de dépassement inférieur (secteur cyan A),
- sur (b) la segmentation semi-automatique ressemble moins à la pseudo vérité-terrain que la segmentation de référence, puisque le pourcentage de recouvrement est inférieur et le pourcentage de dépassement supérieur (secteur blanc B),
- sur (c), la segmentation semi-automatique donne un meilleur recouvrement mais davantage de dépassement que la segmentation de référence (secteur jaune C),
- sur (d) la segmentation semi-automatique donne moins de dépassement mais aussi un recouvrement plus faible que la segmentation de référence (secteur vert D).

Si on peut conclure sans ambiguïté dans les deux premiers cas, aucune des deux segmentations n'est « meilleure » que l'autre dans les deux derniers, sauf si l'un des critères s'éloigne beaucoup plus que l'autre des valeurs de référence. Le recouvrement et le dépassement peuvent faciliter l'analyse de l'indice de similarité, en précisant l'origine des différences entre les segmentations.

4.5 Conclusion

Dans ce chapitre, nous avons étudié un certain nombre de difficultés liées essentiellement à l'absence de vérité-terrain pour les données réelles et à leur nature particulière.

Pour réaliser une première série de tests de nos méthodes de recalage et de segmentation, nous avons construit un modèle de rein suffisamment réaliste, que ce soit anatomiquement ou du point de vue de l'évolution temporelle des voxels. Le bruit, l'effet de volume partiel, les erreurs résiduelles de recalage ont été pris en compte.

Nous avons ensuite défini une méthode de recalage adaptée aux forts changements de contraste, ainsi qu'à l'apparition et à la disparition de certaines structures rénales internes au cours de la perfusion. Toutes les images de chaque série ont été recalée sur une image de référence prise au pic cortical, en n'autorisant que des transformations rigides (translations et rotation) et en mesurant la ressemblance entre images par l'information mutuelle.

Nous avons enfin discuté des possibilités pour valider au mieux les résultats sur des données réelles. Nous avons choisi d'évaluer quelques critères de similarité entre une segmentation établie par un radiologue, et des segmentations réalisées de manière semi-automatique. Il est cependant difficile de fixer une limite sur les critères pour savoir si une segmentation peut être considérée comme acceptable ou non. C'est pourquoi nous évaluerons ces mêmes critères entre la segmentation de référence et une segmentation manuelle faite par un second expert, afin d'avoir un point de comparaison, sachant que les deux segmentations effectuées par les radiologues sont considérées comme satisfaisantes.

Dans le chapitre suivant, nous proposerons différentes solutions pour construire des classificateurs à partir de la coupe de référence, et nous évaluerons leurs performances. Nous étudierons leur généralisation dans le cadre mathématique défini dans le para-

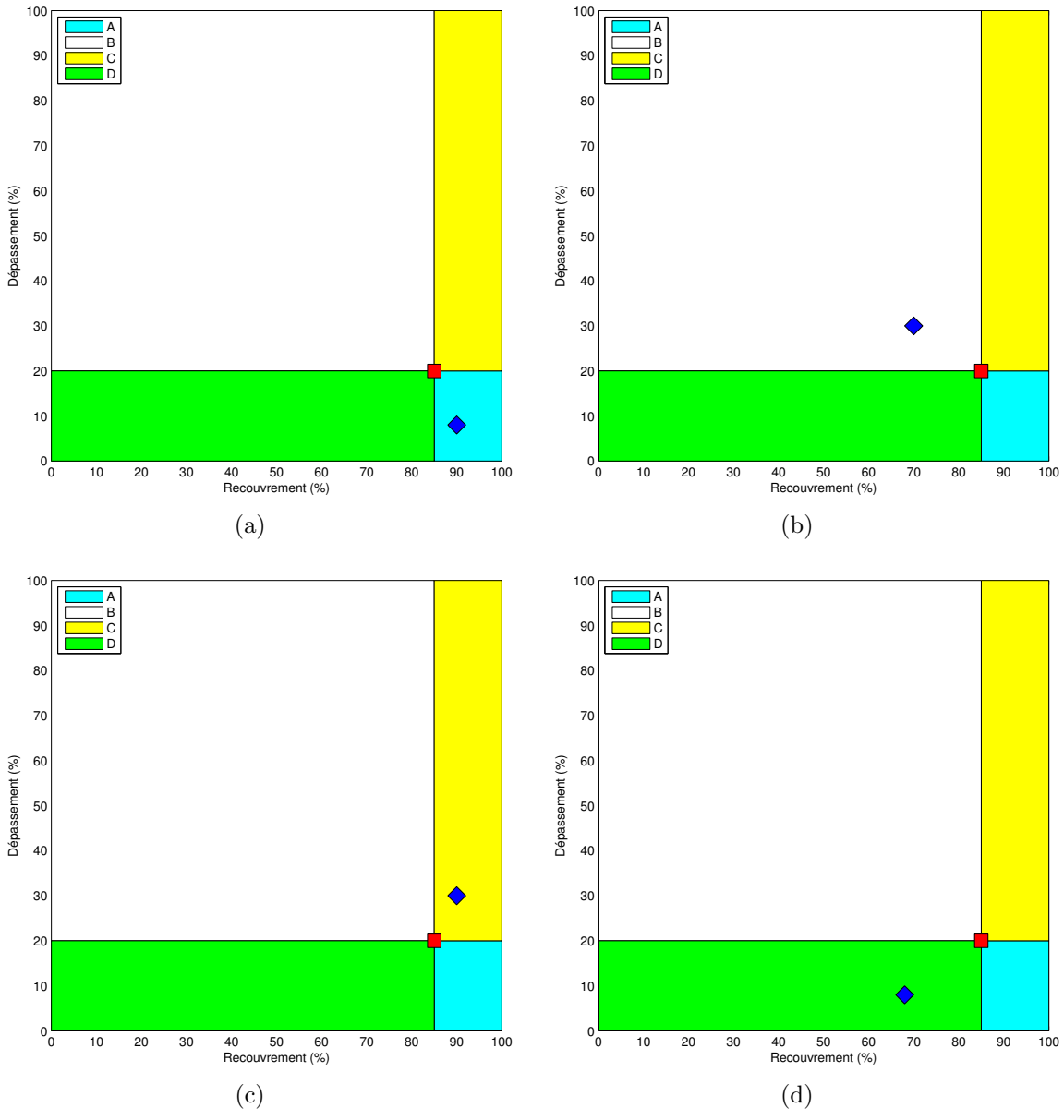


FIG. 4.21 – Comparaison de segmentations par les critères de recouvrement et de dépassement : en (a), la segmentation automatique (losange bleu) est meilleure que la segmentation manuelle (rectangle rouge), alors que c'est le contraire en (b) ; pour le cas (c) (respectivement (d)), la segmentation automatique a un meilleur recouvrement (respectivement dépassement) mais un moins bon dépassement (respectivement recouvrement) que la segmentation manuelle.

graphe 3.2.

Chapitre 5

Segmentations fonctionnelles : méthodes et résultats

Dans ce chapitre, nous construisons puis testons différents classificateurs pour les voxels rénaux de la coupe principale. Grâce aux données simulées, nous donnons certaines explications sur l'origine des qualités et des défauts de chacun d'entre eux, et nous proposons des améliorations. Nous évaluons également leurs performances sur des données réelles. Nous étudions ensuite la généralisation de la classification aux autres coupes rénales.

5.1 Construction des classificateurs grâce à la coupe de référence

La construction des classificateurs se fait grâce à un *clustering* par quantification vectorielle des courbes temps-intensité (éventuellement normalisées, cf. paragraphe 5.1.6 page 164) des voxels de la coupe de référence, selon les principes exposés au chapitre 3.1. Différentes variantes, s'appuyant néanmoins sur un ensemble d'hypothèses communes, seront étudiées. En effet, si certaines peuvent donner satisfaction sur les données simulées, leurs performances peuvent s'avérer insuffisantes sur des données réelles. Il s'agit toujours de déduire d'une quantification vectorielle de leurs courbes temps-intensité, un *clustering*, puis une classification des voxels rénaux. Notre objectif étant non pas d'extraire le rein, mais d'en segmenter les structures internes (cortex, médullaire, cavités), un masque global à N voxels x_i, $1 \leq i \leq N$, est créé sur la coupe de référence avant la segmentation fonctionnelle (cf. paragraphe 4.4.3). Seules les courbes temps-intensité des voxels de ce masque, obtenues à partir des p images de la séquence IRM, servent à la quantification vectorielle.

Nous commencerons par présenter les hypothèses communes aux différentes variantes, qui permettent d'appliquer une quantification vectorielle dans le cadre mathématique présenté dans le paragraphe 3.1.

5.1.1 Hypothèses communes en lien avec la quantification vectorielle

La quantification vectorielle (cf. paragraphe 3.1.1) est la première étape de la segmentation fonctionnelle rénale. Un *clustering* des données et de l'espace des courbes temps-intensité en est déduit, selon les principes décrits au paragraphe 3.1.4. Ce *clustering* dépend de l'algorithme utilisé et des hypothèses issues de la physiologie du rein.

Hypothèses sur les données communes aux différentes méthodes

Soit $\mathbf{I}_i = (I_{i1}, \ldots, I_{ip})$ le vecteur à p composantes associé au voxel x_i, où $I_{i\tau}$ est le contraste du voxel x_i au temps τ (on l'appellera également courbes temps-intensité du voxel x_i). A partir de ce vecteur, par normalisation (cf. paragraphes 5.1.4 page 152 et 5.1.6 page 164), on construit le vecteur $\xi_i = (\xi_{i1}, \ldots, \xi_{ip})$ à p composantes, où $\xi_{i\tau}$ est le contraste normalisé du voxel x_i au temps τ.

Les N vecteurs ξ_i peuvent être considérés comme une réalisation d'une séquence de variables aléatoires $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ identiquement distribuées de même loi qu'un vecteur aléatoire continu $\mathbf{X} : \Omega \longrightarrow V \subset \mathbb{R}^p$, où $(\Omega, \mathcal{E}, P)$ est un espace probabilisé et V une sous-variété de $\mathbb{R}^p$. Ni V, ni la loi $P_{\mathbf{X}}$ ne sont connues ; seul l'échantillon $\mathcal{X} = \{\xi_1, \ldots, \xi_N\}$ est donné[1]. La quantification vectorielle permet de déterminer, grâce aux ξ_i, un ensemble fini $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ de prototypes $\mathbf{w}_j \in \mathbb{R}^p$, $j = 1, \ldots, K$ qui représente au mieux l'échantillon $\mathcal{X}$. Un vecteur donné $\xi \in V$ est décrit par le prototype $w(\xi)$ qui lui est le plus proche, au sens de la distance euclidienne.

Quelques propriétés des quantificateurs utilisés

Quel que soit l'algorithme choisi, le quantificateur vectoriel utilisé est borné :
- si les vecteurs de données ne sont pas normalisés : les niveaux d'intensité sont bornés par le système d'acquisition ;
- si les vecteurs de données sont tous normés à 1 avant quantification (cf. paragraphe 5.1.6), ils se situent tous sur une hypersphère de rayon 1.

Il s'agit de quantificateurs dits au plus proche voisin (cf. paragraphe 3.1.1) ou encore de Voronoï : chaque donnée est finalement attribuée au prototype qui lui est le plus proche au sens de la distance utilisée.

5.1.2 K-moyennes à trois classes

Supposons qu'en accord avec le paragraphe 2.2.3 page 37, de manière qualitative, les comportements des voxels d'un même compartiment anatomique soient homogènes (ou tout au moins présentent une certaine homogénéité) et que les différences d'évolution temporelle entre des compartiments distincts soient assez marquées pour que les trois compartiments recherchés s'obtiennent directement par une quantification vectorielle sur trois prototypes. Un algorithme de K-moyennes avec $K = 3$ est utilisé pour réaliser une

[1] Remarquons qu'on pourrait aussi considérer que chaque compartiment correspond à une distribution de loi différente, mais dans le cadre de la quantification vectorielle, on suppose une loi de probabilité commune permettant de distinguer les trois compartiments, par exemple en ayant trois modes.

quantification vectorielle des ξ_i. Un *clustering* en est alors déduit : tous les voxels associés à des vecteurs ξ_i appartenant à la cellule de Voronoï d'un même prototype sont regroupés dans la même classe. On aboutit ainsi à une segmentation de la coupe de référence.

Tests sur données simulées

Effet du bruit seul Avec le modèle de rein sans effet de volume partiel ni filtre moyenneur, les segmentations anatomiques et fonctionnelles coïncident à 100 % pour des niveaux de bruit variant de $\sigma = 5$ à 60.

Bruit et effet de volume partiel Qualitativement, au lieu d'avoir une répartition gaussienne (ou de Rice) des intensités autour de la courbe moyenne, comme pour le bruit seul, il y a une sorte d'étirement de chacun des trois *clusters* vers les *clusters* voisins.

Les résultats obtenus restent très convenables (WCP $> 99\%$). Quelques pixels contenant majoritairement du cortex et situés en périphérie des trois compartiments sont attribués à la médullaire. On peut comparer ce qui vient d'une différence entre segmentations anatomique et fonctionnelle, et ce qui est dû au bruit : quelques dixièmes de pourcent pour chacune de ces causes à des niveaux de bruit qui paraissent « réalistes ». En revanche pour des mélanges plus marqués et un rapport signal à bruit très faible (qui ne semblent plus réalistes), nous avons un mélange de compartiments, mais pas aussi important que sur les données réelles.

Bruit, effet de volume partiel et filtre moyenneur avec recalage parfait Lorsque l'on rajoute le filtrage moyenneur, on observe une augmentation du nombre de pixels mal classés en périphérie des compartiments, qui fait diminuer le WCP à 92% : on retrouve dans ce cas exactement les compartiments fonctionnels dominants, les 8% restants correspondant aux voxels dont les compartiments anatomique et fonctionnel dominants diffèrent (cf. paragraphe 4.1.1).

Bruit, effet de volume partiel et filtre moyenneur avec erreurs de recalage Nous ajoutons enfin des erreurs de recalage de la manière décrite au paragraphe 4.1.1. Nous observons sur la figure 5.1 une confusion assez importante sur les voxels en bordure des compartiments, qui fait diminuer le pourcentage de voxels bien classés vers 80% : on ne retrouve pas tous les compartiments fonctionnels convenables comme avec un recalage parfait.

Tests sur données réelles

Les segmentations obtenues ne sont pas vraiment satisfaisantes. Pour trois des reins testés, les cavités sont « perdues » car entièrement incluses dans la médullaire, alors que le cortex est plutôt divisé en deux *clusters*. Pour les autres reins, on observe, pour les voxels situés à la frontière des compartiments, un mélange entre médullaire et cortex ou entre médullaire et cavités très semblable à celui de la figure 5.1.

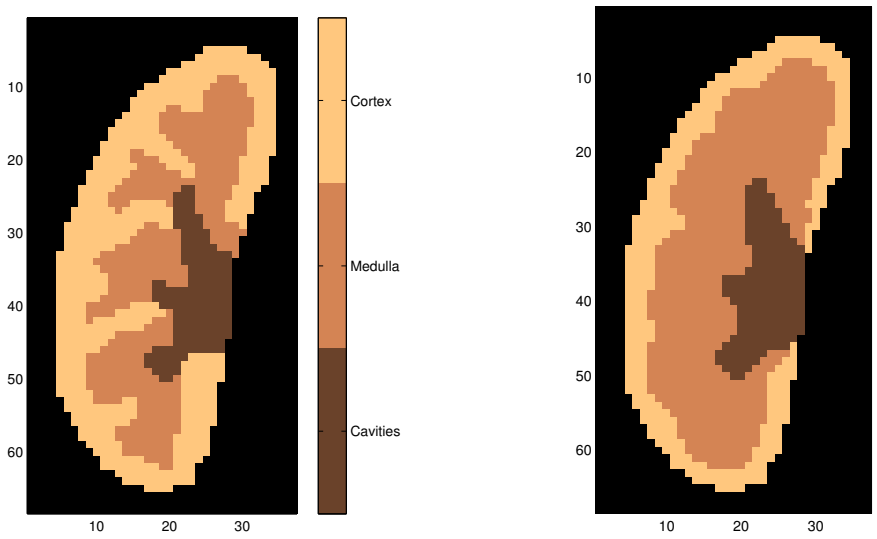


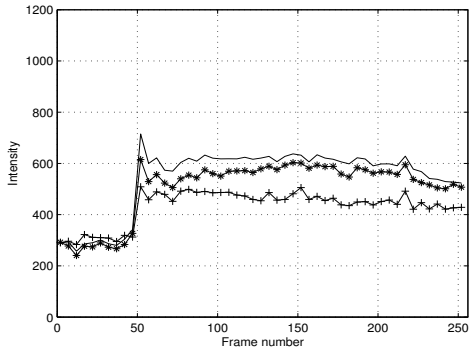
FIG. 5.1 – Segmentation par K-moyennes sur des vecteurs non normalisés avec le modèle le plus complet avec $K = 3$ classes.

Conclusion

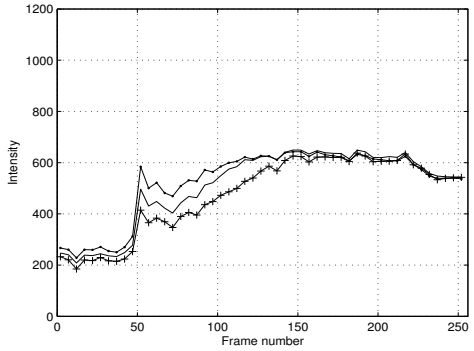
Les résultats sont très satisfaisants sur les données de synthèse, mais assez peu sur les données réelles. Si le bruit est en soi bien supporté par la quantification vectorielle, l'effet de volume partiel et du filtre moyenneur ne sont pas très gênants aux niveaux considérés puisqu'on retrouve le bon compartiment fonctionnel dominant, mais il n'y a pas correspondance entre les compartiments fonctionnel et anatomique dominants pour environ 10% des voxels. En revanche, les erreurs de recalage résiduelles dégradent nettement la classification. On obtient des comportements hybrides pour les voxels situés sur les frontières des compartiments, puisque selon les instants, l'intensité du voxel sera plutôt celle de l'un ou l'autre des compartiments limitrophes. Les vecteurs temps-intensité présenteront ainsi un caractère hétérogène pour des voxels ayant le même compartiment dominant. Il se peut qu'il y ait une déformation du nuage de points par rapport à la situation idéale, de sorte que le placement des prototypes ne permet plus d'aboutir à un *clustering* convenable comme sur la figure 3.15 page 81. Ceci peut expliquer le mélange entre compartiments (essentiellement cortex et médullaire). Pour les cavités, la conjugaison de ces deux phénomènes associés à une sous-représentation de ce compartiment, qui correspondrait à un nuage de points situé à une faible distance des autres *clusters* peut-être plus étendus (cf. paragraphe 3.1.4, en particulier la figure 3.13 page 79).

Afin d'améliorer les résultats, plusieurs stratégies ont été envisagées et testées parallèlement. Une première solution est d'utiliser l'algorithme des K-moyennes avec plus que trois classes dans le but de retrouver tous les compartiments, quitte à ajouter une étape de fusion de classes. Dans la même optique, nous nous proposons également de faire la quantification vectorielle avec un autre algorithme, le *Growing Neural Gas* et d'utiliser le graphe topologique associé comme cela a été évoqué au paragraphe 3.1.4 page 74. Une alternative est enfin de normaliser les courbes pour essayer de diminuer l'influence des

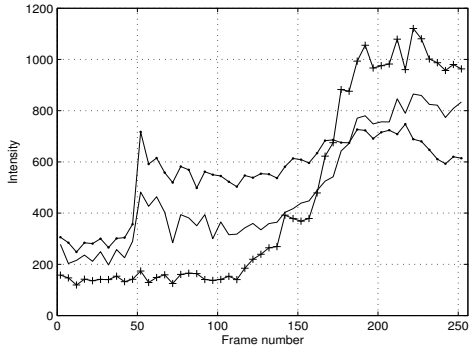
compartiments hybrides en tenant davantage compte de la forme des courbes que de leurs valeurs absolues.



(a) Prototypes pour le cortex



(b) Prototypes pour la médullaire



(c) Prototypes pour les cavités

FIG. 5.2 – Quelques exemples de courbes temps-intensité de prototypes attribués au cortex, à la médullaire et aux cavités d'un rein donné.

5.1.3 K-moyennes à au moins 4 classes et fusion manuelle

Pour les raisons invoquées précédemment (déséquilibre entre classes, étalement des *clusters*), il est envisageable qu'au moins quatre prototypes soient nécessaires pour encoder convenablement la variété, comme nous l'avons exposé au paragraphe 3.1.4 (cf. en particulier la figure 3.14 pour une variété dans $\mathbb{R}^2$). L'algorithme des K-moyennes est ici utilisé avec de faibles valeurs de K ($K \leq 7$). La quantification vectorielle aboutit donc à un faible nombre K de courbes prototypes. Dans le *clustering* induit, une classe de voxels est associée à chacun de ces prototypes. Afin d'obtenir les trois compartiments finaux, un opérateur peut facilement regrouper ces classes manuellement.

Tests sur données simulées

On observe pour le modèle avec recalage parfait une légère dégradation des résultats, ce qui n'est pas étonnant. Ceci est dû au fait que la quantification vectorielle avec $K = 3$ donne des résultats satisfaisants, alors que $K > 3$ n'est pas adapté à la structure des données (cf. paragraphe 3.1.4) : il y a création d'une ou plusieurs classes « parasites » entre médullaire et cavités, ou entre cortex et médullaire (voxels présentant le plus fort volume partiel) mais avec peu de représentants. En revanche, on constate une amélioration pour le modèle avec erreur de recalage : pour $K = 4$ (cf. figure 5.3a), le pourcentage de voxels bien classés varie entre 90 et 92%, soit celui obtenu par K-moyennes à trois classes avec un recalage parfait. On classe ainsi correctement les voxels qui ont les mêmes compartiments anatomique et fonctionnel dominants. L'augmentation du nombre de classes n'améliore pas sensiblement ce résultat (cf. figure 5.3b) : certains *clusters* sont seulement coupés en deux sans que cela ne modifie la répartition des autres voxels hormis pour quelques voxels en périphérie de compartiments.

Tests sur données réelles

On constate une amélioration par rapport à la méthode précédente. Les cavités sont retrouvées dans tous les cas dès que $K \geq 5$ (les résultats pour $K = 4$ n'apparaîtront donc pas dans les tableaux). Cependant, qualitativement, le découpage entre cortex et médullaire ne satisfait pas toujours les radiologues.

Les résultats quantitatifs, publiés dans [87], sont présentés dans les tableaux 5.1 et 5.2. Les valeurs des critères de similarité apparaissent en gras pour les segmentations fonctionnelles si les résultats sont meilleurs ou équivalents à ceux obtenus pour la segmentation manuelle test (gras italique). Les figures 5.4 à 5.7 complètent ces tableaux : les boîtes à moustache de référence (comparaison des deux segmentations manuelles) sont sur la gauche des figures 5.4 et 5.6 ; les valeurs de référence sont matérialisées par des carrés rouges sur les figures 5.5 et 5.7.

Les résultats obtenus pour les comparaisons quantitatives sont relativement convenables, dans la mesure où on atteint pratiquement les valeurs de critères obtenus pour des comparaisons entre deux segmentations faites par des radiologues. La variabilité inter-opérateur est relativement élevée puisque le pourcentage de pixels bien classés est seulement de 74,9%, ce qui peut s'expliquer par une certaine subjectivité du tracé des frontières sur des acquisitions assez floues et par un choix différent des images sur lesquelles sont

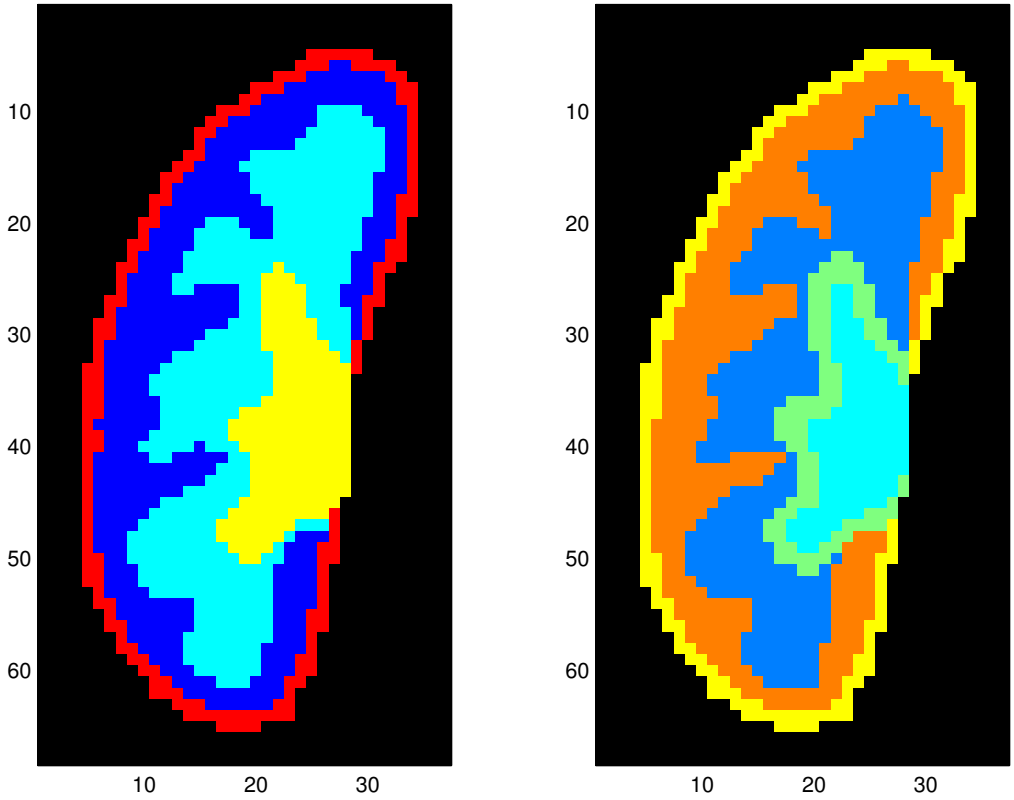


FIG. 5.3 – Segmentation par K-moyennes avant fusion sur des vecteurs non normalisés avec le modèle le plus complet avec $K = 4$ classes (a) et $K = 5$ classes (b).

réalisées les segmentations manuelles. Ce pourcentage varie entre 72,6 et 73,2 pour les segmentations automatiques suivant la valeur de K, ce qui n'est que très légèrement inférieur aux 74,9% obtenus entre les segmentations manuelles. Pour ces dernières, l'indice de similarité varie entre 0,70 et 0,79 selon les compartiments, la valeur la plus basse étant atteinte pour la médullaire, sa place intermédiaire entre le cortex et les cavités la rendant plus sujette aux variations de positionnement des frontières en fonction de l'opérateur. Ce phénomène se traduit également sur les pourcentages de recouvrement et de dépassement, en particulier dans les parties (b) des tableaux de résultats : le second monte à 56,6% quand OP1 est la référence (tableau 5.1 et figure 5.5), et le premier n'atteint que 60,5% quand OP2 est la référence (tableau 5.2 et figure 5.7). Pour les segmentations automatiques, l'indice de similarité prend des valeurs moyennes comprises entre 0,66 et 0,77, ce qui reste globalement un peu inférieur aux valeurs obtenues pour les segmentations manuelles. Les valeurs extrêmes sont dans la plupart des cas un peu moins bonnes pour les segmentations automatiques (cf. figures 5.4 et 5.6).

Notons que les pourcentages de recouvrement et de dépassement PR et PD doivent être examinés conjointement pour mettre en évidence une amélioration véritable de la segmentation dans son ensemble : la plupart du temps, une augmentation de PR (amé-

lioration) va de pair avec une augmentation de PD (dégradation), et vice-versa (cette remarque restera bien sûr valable pour les autres tableaux de résultats). Par exemple, dans les colonnes 2 et 3 du tableau 5.2 ($K = 5$ ou 6), nous pouvons observer que les valeurs de PD et PR augmentent conjointement pour le cortex, alors qu'elles diminuent pour la médullaire : cette évolution traduit bien une amélioration du recouvrement pour le cortex, mais aussi une sur-segmentation croissante de ce compartiment, au détriment de la médullaire. Cette évolution est visible sur la figure 5.7(b) et (c), où nous pouvons également observer une dispersion des résultats plus importante pour les méthodes automatiques que pour la segmentation manuelle. Sur la figure 5.5, nous constatons que la segmentation par K-moyennes à 5 classes est un peu moins bonne que les autres pour le cortex et la médullaire.

Conclusion

La méthode est rapide et simple d'utilisation, et donne de bons résultats sur les données de synthèse. Les résultats pour les données réelles sont convenables mais perfectibles. Le fait que plusieurs défauts d'acquisition, comme la non uniformité de la sensibilité des antennes, ainsi que certains comportements temporels peu courants ne soient pas modélisés (comme nous l'avons souligné au paragraphe 4.1.1 page 117) peut expliquer que, pour segmenter convenablement les données réelles, un nombre de classes plus élevé que pour les données de synthèse soit nécessaire.

5.1.4 GNG-T automatique

Rappelons que les résultats de l'algorithme (cf. paragraphe 3.1.3, dont nous reprenons les notations) consistent en un ensemble de neurones représentés chacun par une courbe temps-intensité prototype et connectés par un graphe G respectant la topologie de la variété $V \subset \mathbb{R}^p$, support de la loi de $\mathbf{X}$ (le graphe G est une approximation de la triangulation de Delaunay des prototypes induite par la variété V). Dans le *clustering* induit (cf. paragraphe 3.1.4), l'espace est divisé en autant de sous-ensembles que de sous-graphes indépendants dans le graphe G. Chaque sous-ensemble est constitué de l'association des cellules de Voronoï des prototypes dont les neurones forment un de ces sous-graphes et correspond à une classe. Le *clustering* des voxels en découle naturellement.

On suppose ici que la variété V comporte trois composantes connexes, chaque composante correspondant à un compartiment[2]. On s'attend à ce qu'avec une quantification vectorielle plus précise associée à un graphe topologique, dans le cas d'un déséquilibre des classes (cf. paragraphe 3.1.4), les résultats soient meilleurs ; en particulier, on espère ne plus perdre les cavités. Si on obtient un graphe topologique avec trois sous-graphes, le *clustering* des voxels en est déduit par la méthode décrite au paragraphe 3.1.4 page 74. Cependant, comme c'est très souvent le cas avec des données réelles, on peut tomber sur

[2]Remarquons que la connexité dans l'espace des attributs n'implique en aucune manière une connexité des pixels de chaque compartiment dans les images. Nous obtenons au contraire dans la plupart des cas plusieurs zones non connexes pour chaque compartiment (cf. l'exemple figure 5.8(b) page 160 pour le cortex et la médullaire). Ceci peut représenter un avantage supplémentaire par rapport à la méthode de segmentation manuelle décrite dans le paragraphe 4.4.3 page 139 : la non-connexité peut correspondre à une réalité anatomique mais alourdit encore davantage la méthode manuelle.

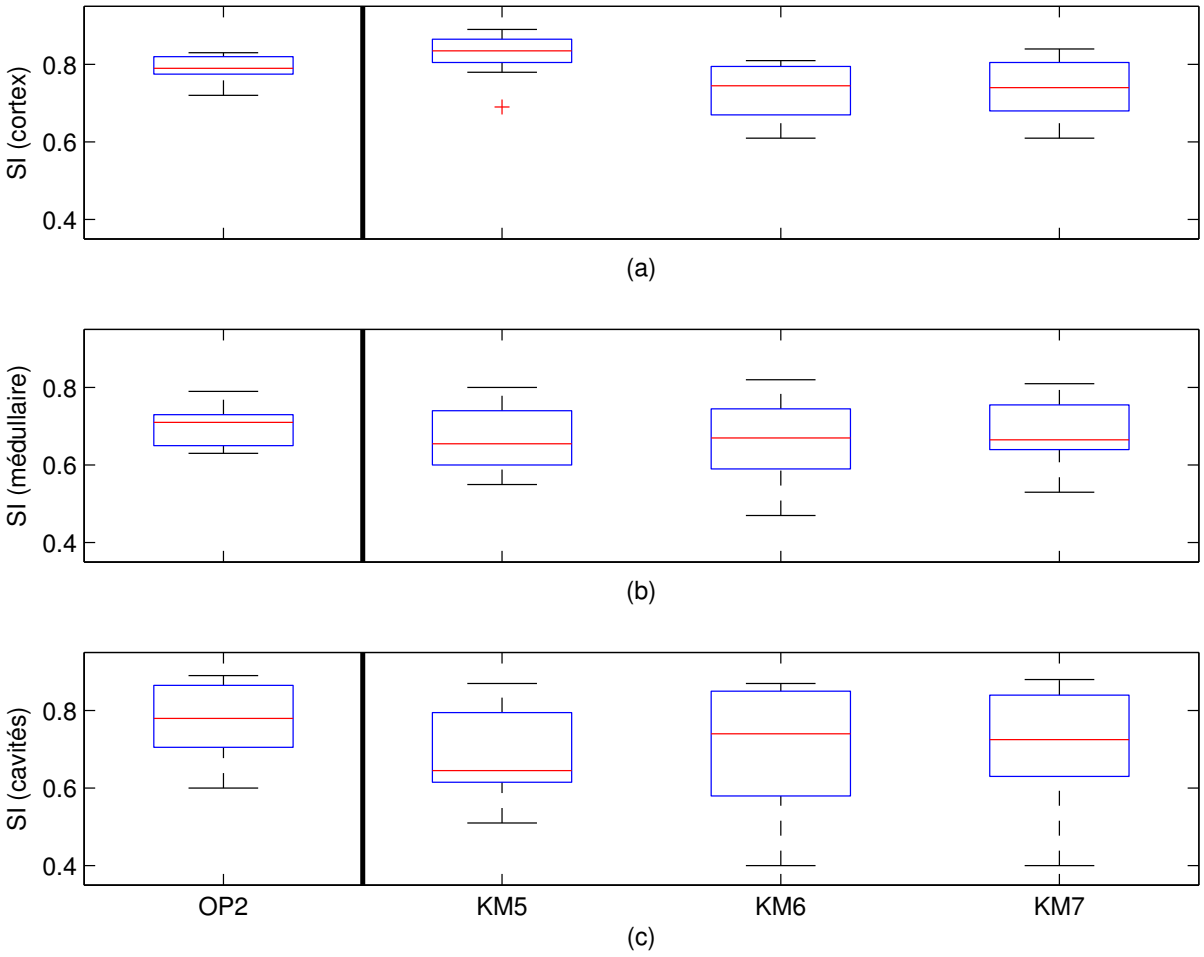


FIG. 5.4 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP1) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP2 et les segmentations automatiques par K-moyennes à 5, 6 et 7 classes (respectivement notées KM5, KM6 et KM7).

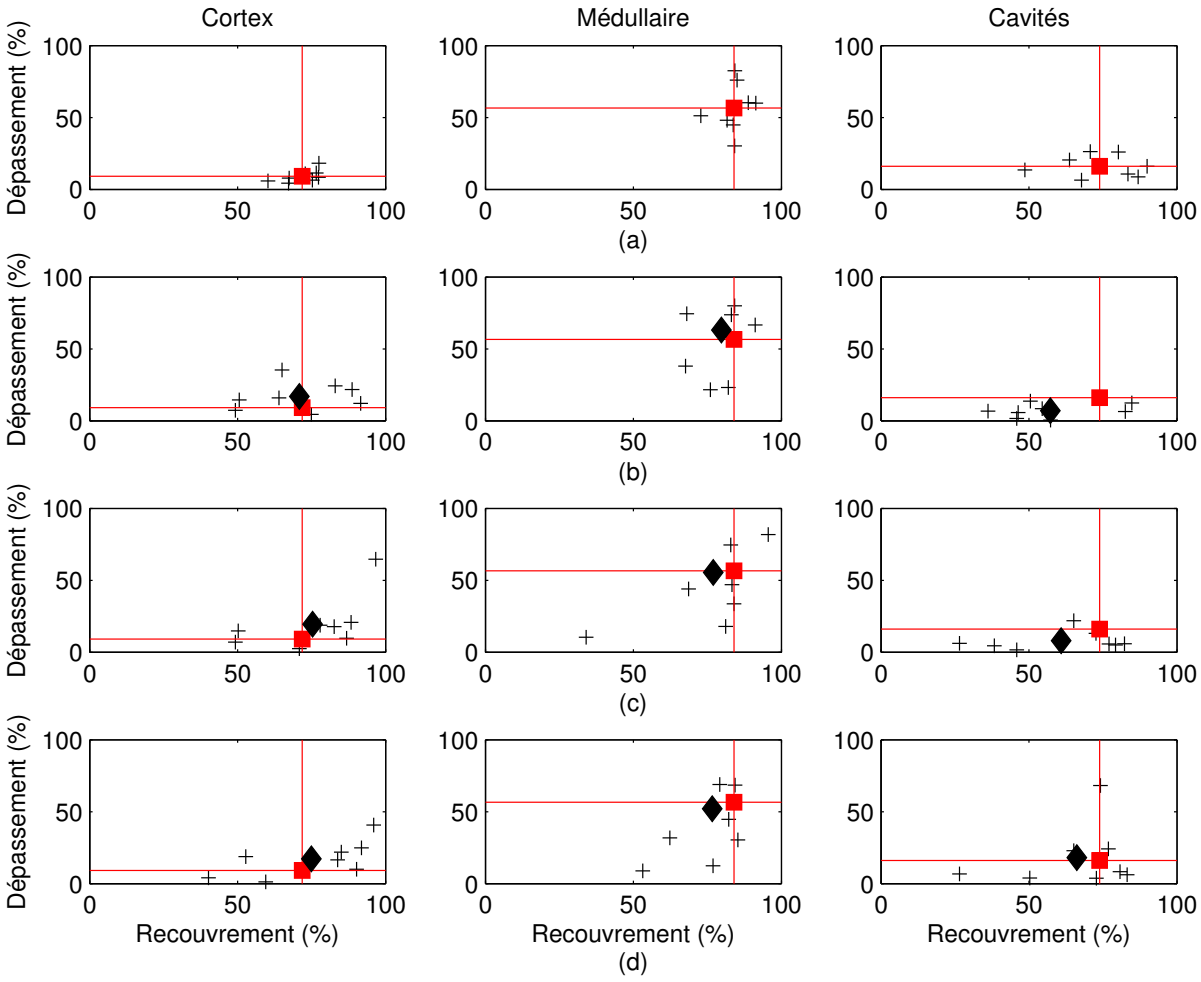


FIG. 5.5 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP1 (référence) pour les différents compartiments pour la segmentation manuelle de OP2 (a) et les segmentations par K-moyennes à 5 classes (b), 6 classes (c) et 7 classes (d).

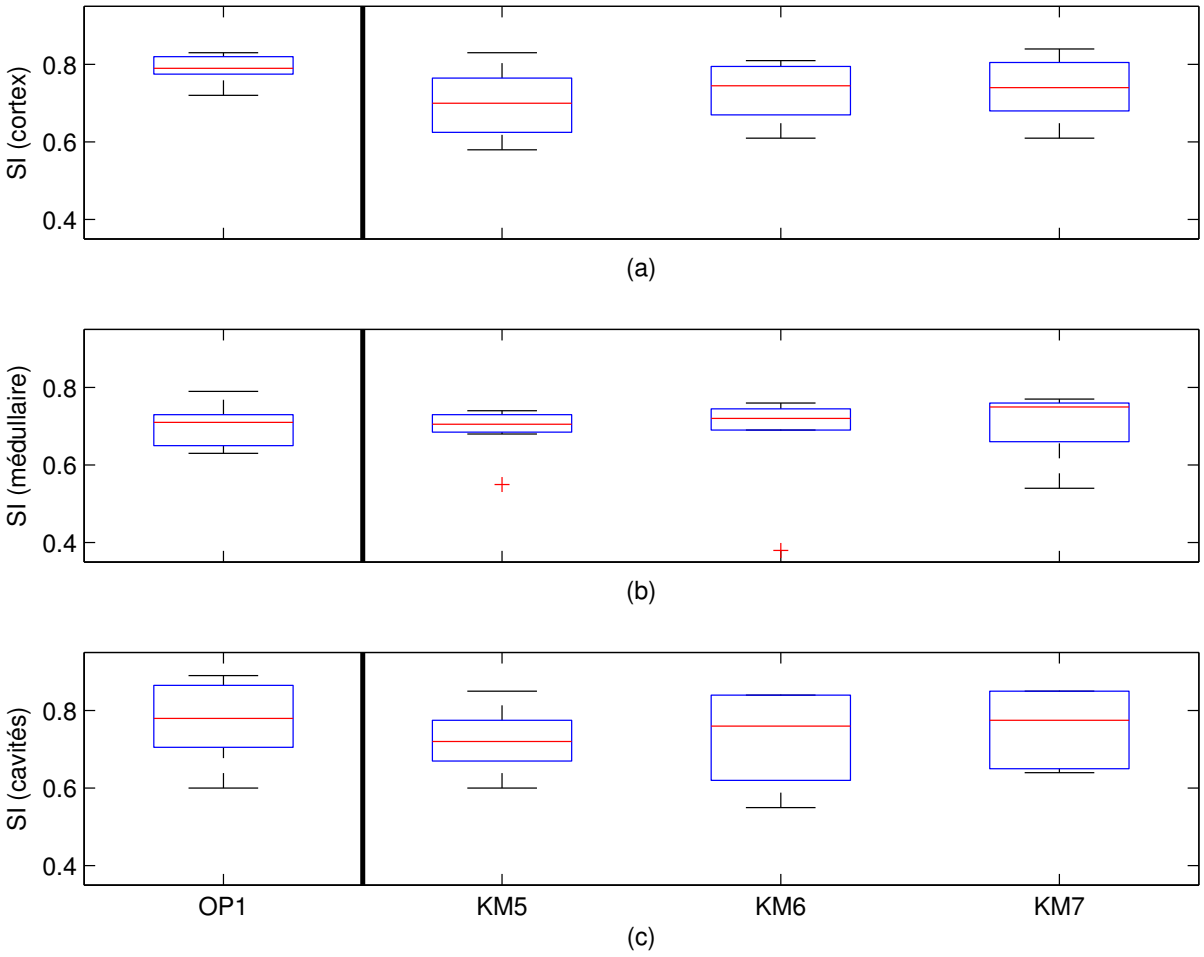


FIG. 5.6 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP2) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP1 et les segmentations automatiques par K-moyennes à 5, 6 et 7 classes (respectivement notées KM5, KM6 et KM7).

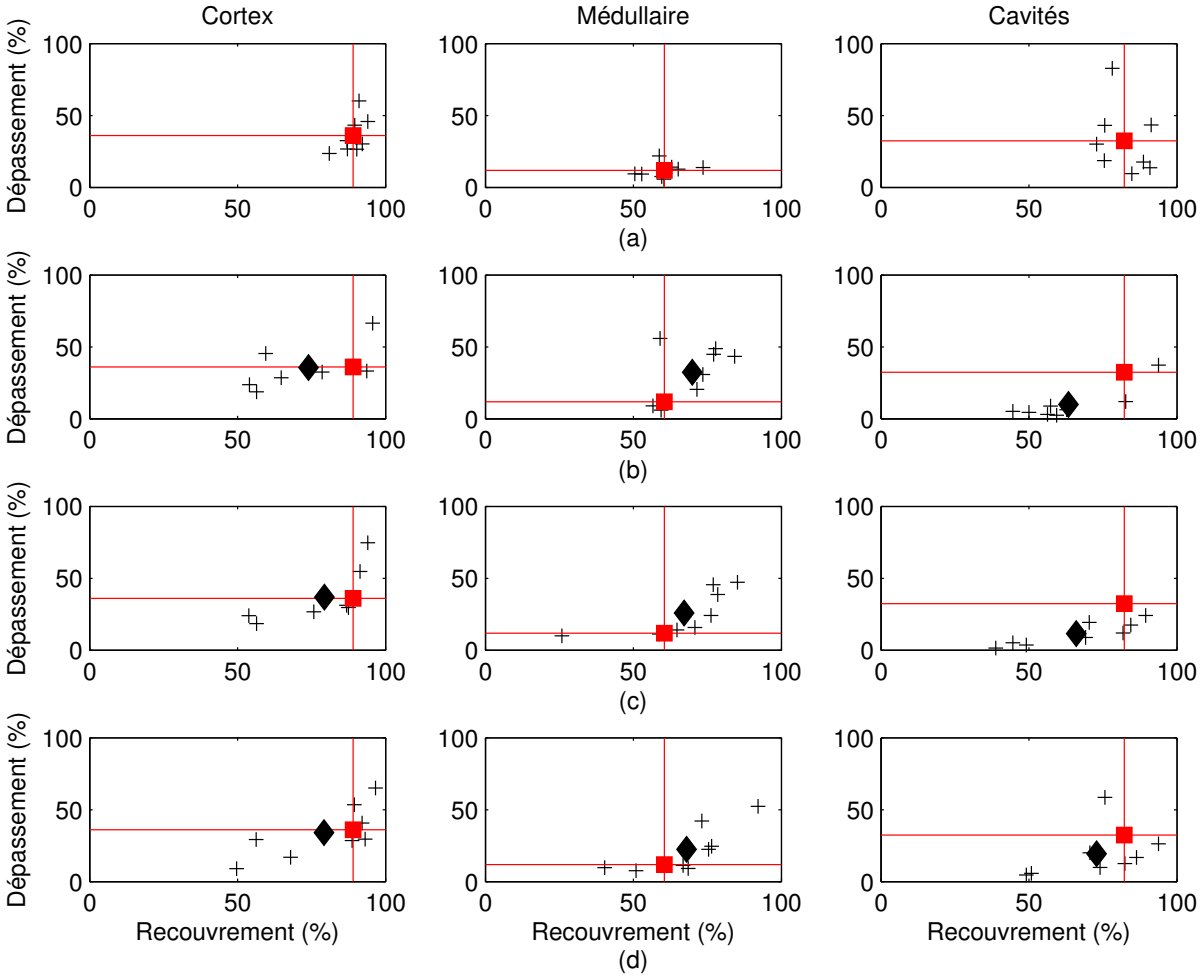


FIG. 5.7 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP2 (référence) pour les différents compartiments pour la segmentation manuelle de OP1 (a) et les segmentations par K-moyennes à 5 classes (b), 6 classes (c) et 7 classes (d).

Test	OP2	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	***69,8***	**74,3**	**75,4**	**75,9**
Recouvrement (%)	***71,8***	70,9	**75,3**	**74,9**
Dépassement (%)	9,2	17,0	19,5	17,3
Indice de similarité	0,79	0,75	0,77	0,77
Distance moyenne	***0,6***	0,7	0,7	**0,6**

(a) Cortex

Test	OP2	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	84,4	79,2	78,4	75,7
Recouvrement (%)	84,0	79,7	77,0	76,7
Dépassement (%)	***56,6***	63,1	55,4	**52,1**
Indice de similarité	0,70	0,67	0,66	0,68
Distance moyenne	1,0	1,3	1,2	1,1

(b) Médullaire

Test	OP2	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	74,9	59,1	63,1	68,2
Recouvrement (%)	73,9	57,2	60,9	66,2
Dépassement (%)	***16,1***	**7,0**	**8,0**	18,0
Indice de similarité	0,77	0,69	0,70	0,71
Distance moyenne	***0,8***	**0,7**	**0,6**	1,1

(c) Cavités

Test	OP2	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	74,9	71,3	72,7	73,5

(d) Pourcentage de pixels bien classés dans le rein entier

TAB. 5.1 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP1 : les résultats apparaissent en gras s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

Test	OP1	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	89,7	78,2	80,9	81,7
Recouvrement (%)	89,0	74,0	79,3	79,2
Dépassement (%)	***36,1***	35,6	36,8	**34,1**
Indice de similarité	0,79	0,70	0,73	0,74
Distance moyenne	0,7	0,9	0,8	0,8

(a) Cortex

Test	OP1	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	***60,3***	**68,2**	**68,0**	**66,4**
Recouvrement (%)	***60,5***	**69,9**	**67,1**	**67,9**
Dépassement (%)	11,8	32,4	25,8	22,4
Indice de similarité	***0,70***	0,69	0,68	**0,71**
Distance moyenne	0,8	1,2	1,1	1,0

(b) Médullaire

Test	OP1	Auto	Auto	Auto
K	3	5	6	7
Pixels bien classés (%)	81,8	63,6	66,9	74,1
Recouvrement (%)	82,2	63,3	65,9	72,8
Dépassement (%)	***32,4***	**9,9**	**11,5**	**19,3**
Indice de similarité	0,77	0,72	0,73	0,76
Distance moyenne	***0,9***	**0,6**	**0,6**	**0,9**

(c) Cavités

Test	OP1	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	74,9	72,6	73,5	73,8

(d) Pourcentage de pixels bien classés dans le rein entier

TAB. 5.2 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP2 : les résultats apparaissent en gras s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

le problème dû aux arcs parasites dans la triangulation de Delaunay induite (cf. paragraphe 3.1.4 page 77). Dans ce cas, en raison du nombre relativement élevé de nœuds, un regroupement manuel serait trop fastidieux. Il faudra alors trouver des critères pour briser les arcs parasites.

Tests sur données simulées

Effet du bruit seul Sans effet de volume partiel, pour des niveaux de bruit variant de $\sigma = 5$ à 60, les segmentations anatomiques et fonctionnelles coïncident à 100 %. Le graphe topologique contient bien trois sous-graphes correspondant chacun à une composante connexe de la variété V.

Bruit et effet de volume partiel ; erreurs de recalage L'effet de volume partiel s'avère beaucoup plus gênant, même pour des rapports signal à bruit assez élevés. Pour des mélanges "faibles", on obtient souvent une variété à deux composantes connexes, le cortex et la médullaire étant rassemblés dans la même ; cependant, il suffit généralement de briser un seul arc parasite pour obtenir la classification correcte. Quand le mélange devient plus complexe, le graphe ne comporte qu'une seule composante connexe. Il en est de même pour le modèle complet avec erreurs de recalage. La méthode ne permettant pas d'obtenir une segmentation à trois compartiments comparable à la vérité-terrain, nous ne sommes pas en mesure de donner des résultats numériques.

Tests sur données réelles

Dans la plupart des cas, on obtient un seul graphe où tous les neurones sont connectés indirectement, de sorte que l'on n'est pas en mesure d'en déduire directement un *clustering* des voxels.

Conclusion

La méthode n'est pas utilisable directement sur des données réelles. Cependant, le « découpage anatomique » obtenu après quantification vectorielle (cf. figure 5.8(a)) peut permettre d'aboutir à une segmentation tout à fait satisfaisante si nous trouvons un critère pour briser les arcs « parasites » dans la triangulation de Delaunay induite (figure 5.8(b) à (e)). Diverses possibilités sont examinées dans le paragraphe suivant.

5.1.5 GNG-T semi-automatique

Avec l'algorithme de GNG-T, dans le cas idéal, la triangulation de Delaunay induite comprend trois sous-graphes connectant les neurones associés aux prototypes d'un même compartiment. En réalité, à cause du bruit et du volume partiel, sur données réelles et sur certaines données simulées, on obtient un seul graphe avec un nombre de prototypes relativement élevé (entre 10 et 30 sur les données réelles). Un regroupement manuel n'est donc pas envisageable. Un critère utilisé dans la littérature pour séparer les prototypes est la variance par nœud e_j/n_j, où e_j est une approximation de la distorsion associée au prototype w_j et n_j le nombre de fois où w_j a gagné au cours de l'époque considérée [61] ;

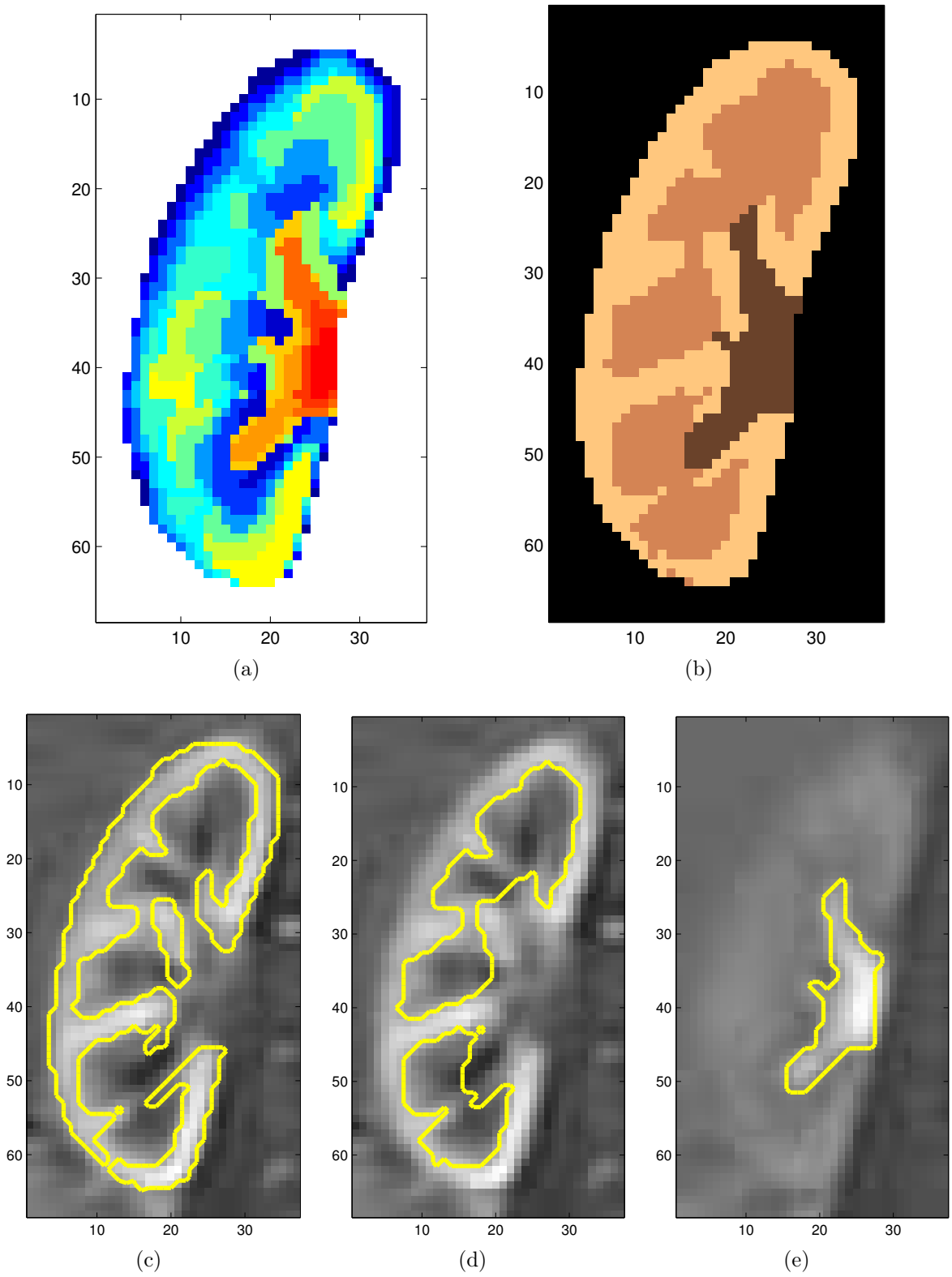


FIG. 5.8 – Classes après quantification vectorielle par GNG-T (a) et segmentation pouvant être obtenue par regroupement correct des classes (b) ; compartiments obtenus superposés à des images de la séquence : cortex (c), médullaire (d) et cavités (e).

ce critère ne donne cependant pas de résultats satisfaisants dans notre cas, sauf peut-être parfois pour séparer les cavités des deux autres compartiments (cela reste cependant assez peu fiable). Nous avons alors choisi de nous baser sur des critères physiologiques communs à une partie des voxels d'un compartiment donné pour briser convenablement certains arcs du graphe G, tout en tenant compte de la topologie de ce dernier. La segmentation est alors réalisée en deux étapes : une quantification vectorielle associée à la construction d'un graphe respectant la topologie, suivie d'une étape de séparation des *clusters* par réglage de deux seuils indépendants, ce qui permet de couper le graphe en trois sous-ensembles pour obtenir la segmentation finale à trois compartiments [88].

Méthode détaillée

La procédure que nous avons utilisée est la suivante.

- Extraction des cavités : les cavités sont le compartiment anatomique contenant les voxels dont le rehaussement est maximal en fin de perfusion. Appelons prototype semence celui dont l'intensité moyenne est maximale en phase tardive. Tous les prototypes dont l'intensité moyenne durant cette phase est supérieure à un premier seuil s_1 et qui sont connectés dans le graphe G au prototype semence par un chemin ne passant que par des prototypes qui eux-mêmes vérifient le critère, sont supposés représenter les cavités.
- Séparation du cortex et de la médullaire : le taux de filtration dépend de la nature des tissus et est utilisé pour séparer cortex et médullaire. La pente des courbes temps-intensité pendant la phase de filtration est évaluée automatiquement par une méthode de régression linéaire standard pour tous les prototypes n'appartenant pas aux cavités. Appelons prototype semence celui dont la pente est maximale durant cette phase. Un nœud est attribué à la médullaire si la pente correspondante est inférieure à un seuil s_2 et s'il est relié au prototype semence par un chemin ne passant que par des prototypes qui eux-mêmes vérifient le critère de seuil. Les prototypes restants sont attribués au cortes.

Les deux seuils s_1 et s_2 sont initialisés de sorte que le cortex représente approximativement 50% du rein et les cavités 20%. La seule intervention manuelle est leur réglage par un opérateur. Comme l'algorithme est très rapide, l'ajustement peut se faire en temps réel. Il est assez aisé car, à chaque changement de niveau des seuils, un nombre important de voxels avec une évolution temporelle semblable est ajouté ou retiré. Remarquons que le second critère ne pourrait être utilisé pour séparer les cavités des autres compartiments. En effet, sur la figure 5.2(c), un pic artériel, non prévu en théorie mais relativement important, peut être observé sur des courbes temps-intensité de voxels attribués aux cavités ; ce pic est dû à l'importante vascularisation du rein et également à l'effet de volume partiel pour certains voxels.

Tests sur données simulées

Seuls des tests sur le modèle à volume partiel sont effectués, puisque les résultats sont satisfaisants pour le premier modèle. Les résultats sont très satisfaisants, puisque les segmentations coïncident à plus de 95%. Avec le modèle complet comprenant le filtre moyenneur et les erreurs de recalage, on classe convenablement tous les voxels ayant

des compartiments anatomiques et fonctionnels dominants identiques, comme avec les K-moyennes à 4 classes.

Afin de tester la reproductibilité liée aux conditions initiales et à l'ordre des données pour un rein donné, on lance 20 fois la segmentation de données identiques, par la méthode GNG-T, en faisant varier ces deux paramètres. Notons que, comme un opérateur intervient, on teste en réalité une reproductibilité qui est également liée à l'opérateur. On définit une segmentation moyenne, qui est créée de la manière suivante : chaque voxel est attribué au compartiment auquel il appartient le plus souvent sur l'ensemble des 20 segmentations. On calcule ensuite le WCP entre cette segmentation moyenne et les 20 segmentations obtenues. On obtient un WCP moyen de 98% avec un écart type de 1,4%.

En ce qui concerne les différents compartiments :

- pour les cavités, le SI moyen est de $0,99$, avec un écart-type de $0,007$;
- pour la médullaire, le SI moyen est de $0,97$, avec un écart-type de $0,020$;
- pour le cortex, le SI moyen est de $0,98$, avec un écart-type de $0,013$.

La reproductibilité est donc très bonne, aussi bien pour la segmentation dans son ensemble que par compartiment.

Tests sur données réelles

Les tests ont été réalisés avec une cible T valant 40 pour les reins d'adultes, et 20 pour les reins d'enfants, le choix de ces valeurs résultant de l'observation d'expériences. Qualitativement, les segmentations obtenues sont très cohérentes avec celles des radiologues et tendent à montrer que la méthode est valable. Les résultats quantitatifs sont présentés dans les tableaux 5.3 et 5.4 ; ils sont très satisfaisants dans l'ensemble, et les valeurs des critères de ressemblance entre les segmentations fonctionnelles et manuelles s'avèrent parfois meilleures qu'entre deux segmentations manuelles. Ainsi, pour les segmentations semi-automatiques, le pourcentage de pixels bien classés vaut 76,7 ou 77,6 selon l'opérateur de référence, contre 74,9 pour les segmentations manuelles. L'indice de similarité est pratiquement toujours supérieur pour la segmentation automatique, quel que soit le compartiment considéré, alors que la distance moyenne est inférieure, sauf pour le cortex où elle est supérieure de 0,2 pixel. Nous commenterons ces résultats plus en détail dans le paragraphe 5.1.6 en même temps que ceux de la dernière méthode utilisée, dans la mesure où ils sont similaires.

Pour les tests de reproductibilité liée à l'initialisation et à l'ordre des données, nous procédons de la même façon qu'avec les données de synthèse. On obtient un WCP moyen de 95% avec un écart type de 2,2%, qui est bien supérieur au WCP de 74,9% entre les segmentations manuelles réalisées par les deux opérateurs. En ce qui concerne les différents compartiments :

- pour les cavités, le SI moyen est de 0,97 avec un écart-type de 0,01, alors qu'il vaut 0,77 pour les segmentations manuelles ;
- pour la médullaire, le SI moyen est de 0,93 avec un écart-type de 0,03, alors qu'il vaut 0,70 pour les segmentations manuelles ; ;
- pour le cortex, le SI moyen est de 0,95 avec un écart-type de 0,02, alors qu'il vaut 0,79 pour les segmentations manuelles ;.

Nous pouvons donc conclure que la variabilité due aux conditions initiales et à l'ordre

Segmentation par	OP2	GNG-T
Pixels bien classés (%)	***69,8***	**83,7**
Recouvrement (%)	***71,8***	**83,2**
Dépassement(%)	9,2	21,9
Indice de similarité	***0,79***	**0,81**
Distance moyenne au contour de référence	0,6	0,8

(a) Cortex

Segmentation par	OP2	GNG-T
Pixels bien classés (%)	84,8	73,0
Recouvrement (%)	84,0	73,0
Dépassement(%)	***56,6***	**33,7**
Indice de similarité	***0,70***	**0,71**
Distance moyenne au contour de référence	***1,0***	**0,8**

(b) Médullaire

Segmentation par	OP2	GNG-T
Pixels bien classés (%)	74,9	68,7
Recouvrement (%)	73,9	69,8
Dépassement(%)	***16,1***	**11,1**
Indice de similarité	***0,77***	**0,77**
Distance moyenne au contour de référence	***0,8***	**0,7**

(c) Cavités

Segmentation par	OP2	GNG-T
Pixels bien classés (%)	***74,9***	**77,6**

(d) Rein entier

TAB. 5.3 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP1 : les résultats apparaissent en gras pour GNG-T s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

Segmentation par	OP1	GNG-T
Pixels bien classés (%)	89,7	85,8
Recouvrement (%)	89,0	85,7
Dépassement(%)	***36,1***	**34,8**
Indice de similarité	0,79	0,78
Distance moyenne au contour de référence	0,7	0,9

(a) Cortex

Segmentation par	OP1	GNG-T
Pixels bien classés (%)	***60,3***	**69,7**
Recouvrement (%)	***60,5***	**69,9**
Dépassement(%)	11,8	16,3
Indice de similarité	***0,70***	**0,75**
Distance moyenne au contour de référence	***1,0***	**0,9**

(b) Médullaire

Segmentation par	OP1	GNG-T
Pixels bien classés (%)	81,8	74,9
Recouvrement (%)	82,2	76,4
Dépassement(%)	***32,4***	**12,6**
Indice de similarité	***0,77***	**0,80**
Distance moyenne au contour de référence	***0,9***	**0,5**

(c) Cavités

Segmentation par	OP1	GNG-T
Pixels bien classés (%)	***74,9***	**76,7**

(d) Rein entier

TAB. 5.4 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP2 : les résultats apparaissent en gras pour GNG-T s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

des données pour un même opérateur est très inférieure à la variabilité entre les deux opérateurs pour les segmentations manuelles, que ce soit pour la segmentation dans son ensemble, ou pour les différents compartiments.

5.1.6 K-moyennes à 3 classes ou plus sur courbes normalisées

Afin de tenir compte davantage de la forme de la courbe temps-intensité, qui semble mieux caractériser chaque compartiment que ses valeurs absolues au vu de la figure 5.2, des tests ont été menés en normalisant les vecteurs temps-intensité ξ_i de sorte que leur norme soit égale à 1, avant de les présenter en tant qu'entrées à l'algorithme des K-moyennes. Cette normalisation peut aussi être considérée comme une modification de la fonction coût d de l'équation (3.5) relative à la quantification vectorielle (remplacement de la distance euclidienne par la « distance cosinus »).

Lien entre la distance euclidienne et la « distance cosinus »

La « distance cosinus » fait partie des distances autres que la distance euclidienne couramment utilisées dans l'algorithme des K-moyennes (d'autres exemples étant la norme L_1, ou la distance de Hamming pour les données binaires...)

Soient deux vecteurs $X, Y \in \mathbb{R}^p$. Notons $d_c(X,Y) = 1 - \cos(X,Y)$ la « distance cosinus » et $d_e(X,Y)$ la distance euclidienne entre X et Y. On montre aisément que :

$$d_e^2(X/\sqrt{X^T X}, Y/\sqrt{Y^T Y}) = \sqrt{2} d_c(X,Y). \tag{5.1}$$

Cependant, la « distance cosinus » n'est pas une distance au sens strict du terme puisque le fait que $d_c(X,Y) = 0$ n'implique pas nécessairement que $X = Y$. Elle est utilisée dans [42] dans le cadre de l'IRM de perfusion rénale.

Modification de l'espace des paramètres

Après leur normalisation à 1, les vecteurs de données sont tous situés sur une hypersphère de rayon 1 dans $\mathbb{R}^p$. On garde le carré de la distance euclidienne comme fonction coût. Tous les résultats présentés dans les chapitres 3.1 et 3.2 restent alors valables.

Remarquons qu'on ne peut pas normaliser de cette manière les courbes pour la méthode semi-automatique utilisant le GNG-T : si une normalisation est utilisée, elle doit respecter l'ordre relatif des intensités pour que les critères physiologiques servant à casser les arcs parasites demeurent valables.

Tests sur données simulées

Avec le modèle complet comprenant les erreurs de recalage, on arrive cette fois-ci à classer convenablement les voxels ayant des compartiments fonctionnels et anatomiques identiques, ce qui n'était pas le cas sans normalisation : on aboutit ainsi à un WCP d'environ 92%. L'avantage est de ne pas avoir d'étape de fusion.

Tests sur données réelles

Les résultats quantitatifs ont été publiés dans [88] et sont présentés dans les tableaux 5.5 et 5.6, conjointement à ceux obtenus par le GNG-T semi-automatique, et que nous commenterons parallèlement. Ces tableaux sont complétés par les figures 5.9 à 5.12.

Pour tous les types de comparaisons, l'indice de similarité est du même ordre de grandeur et la distance moyenne entre les contours demeure inférieure à un voxel. Le pourcentage de voxels bien classés est même un peu plus élevé pour la plupart des segmentations fonctionnelles et varie entre 74,4% et 77,6%, alors qu'il vaut 74,9% pour les segmentations manuelles. Les scores les moins bons pour l'ensemble des critères sont globalement obtenus pour la médullaire : la frontière cortico-médullaire n'est pas très nette sur les acquisitions (cf. figures 5.8(c) et (d) page 160) et ce compartiment de forme complexe reste particulièrement difficile à délimiter, même manuellement. Par ailleurs, les erreurs sur les frontières internes du cortex et des cavités se reportent toutes sur la médullaire en raison de sa position interne dans le rein.

Pour l'algorithme des K-moyennes, des résultats très satisfaisants sont obtenus avec $K = 3$ pour la plupart des reins, et ne sont en moyenne pas nécessairement améliorés lorsque K augmente, les valeurs devenant même un peu plus dispersées. Les nuages de points obtenus sont peut-être semblables à ceux de la figure 3.12 en dimension 2. Cela tend à prouver que les courbes temps-intensité normalisées à 1 sont un attribut discriminant pour distinguer directement trois types de voxels. Par ailleurs, aucune étape de fusion n'est nécessaire, ce qui n'est pas le cas lorsque l'algorithme est appliqué aux vecteurs temps-intensité non normalisés : rappelons que pour $K = 3$ ou 4, les cavités n'apparaissent parfois pas. La normalisation entraîne vraisemblablement une plus grande homogénéité des vecteurs temps-intensité à l'intérieur de chaque compartiment et une meilleure séparabilité, ce qui réduit le problème dû à la sous-représentation des cavités.

Les figures 5.13 à 5.16 permettent de comparer les résultats obtenus pour les K-moyennes à 7 classes, les K-moyennes à 3 classes sur vecteurs normalisés, le GNG-T et la segmentation manuelle. Les performances des K-moyennes à 7 classes sont moins bonnes que les autres : moyennes inférieures, valeurs extrêmes moins bonnes, dispersion plus élevée vers les valeurs basses. Les résultats pour les K-moyennes à $K = 3$ classes et pour GNG-T sont très similaires : les valeurs moyennes de l'indice de similarité sont généralement égales ou supérieures, des valeurs extrêmes assez semblables, les plus basses étant compensées par de meilleures valeurs hautes. Les valeurs obtenues pour le recouvrement et le dépassement sont moins dispersées.

Les résultats sont légèrement meilleurs pour le GNG-T parce que sa flexibilité est un peu mieux adaptée au traitement des cas difficiles, en particulier en cas de comportements temporels plus chaotiques, le nombre de prototypes s'ajustant automatiquement pour représenter les vecteurs temps-intensité avec une distorsion donnée. Chacun des trois compartiments est convenablement identifié ; autrement dit, la segmentation dans son ensemble est satisfaisante.

L'ensemble des tests ont été réalisés sur des reins sains. Or, cette méthode est destinée à être utilisée également sur des reins pathologiques. Ce point est discuté dans le paragraphe suivant.

5.1.7 Simulations sur des reins pathologiques

Dans le paragraphe 4.1.2, nous avons proposé deux modèles de reins pathologiques, sur lesquels nous testons les classificateurs construits précédemment. Ces deux modèles tiennent compte de la finesse du parenchyme, et d'un remplissage tardif ou inexistant des cavités. En revanche, nous ne sommes pas en mesure de donner des résultats pour des reins pathologiques réels car nous ne disposons pas encore de segmentations manuelles sur de tels cas. Rappelons que les deux modèles ne prétendent pas recouvrir l'ensemble des situations rencontrées dans la réalité, mais servent à montrer que la méthode peut être adaptée. Nous proposons également des solutions en cas de remplissage inhomogène des cavités.

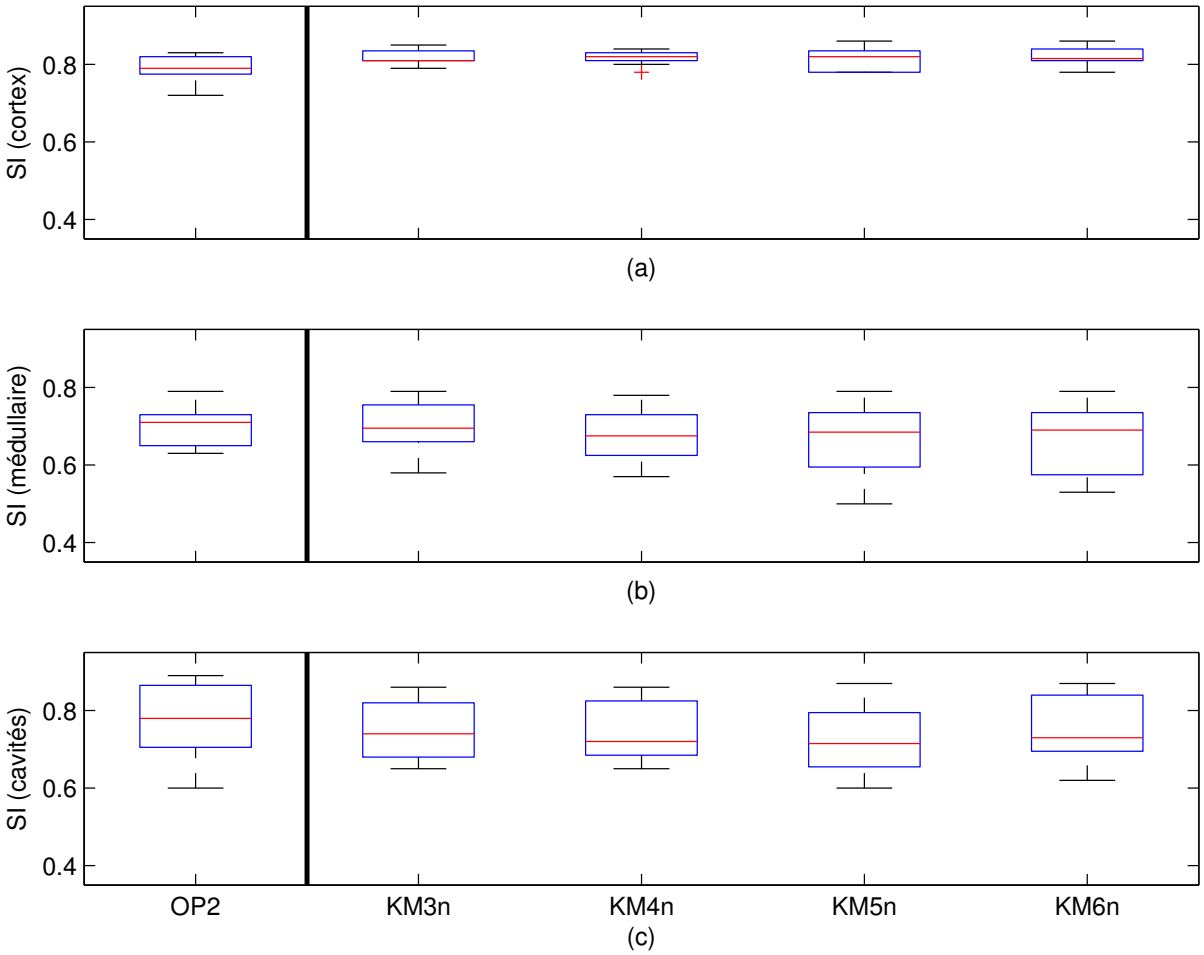


FIG. 5.9 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP1) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP2 et les segmentations automatiques par K-moyennes à 3, 4, 5 et 6 classes sur vecteurs normalisés (respectivement notées KM3n, KM4n et KM5n et KM6n).

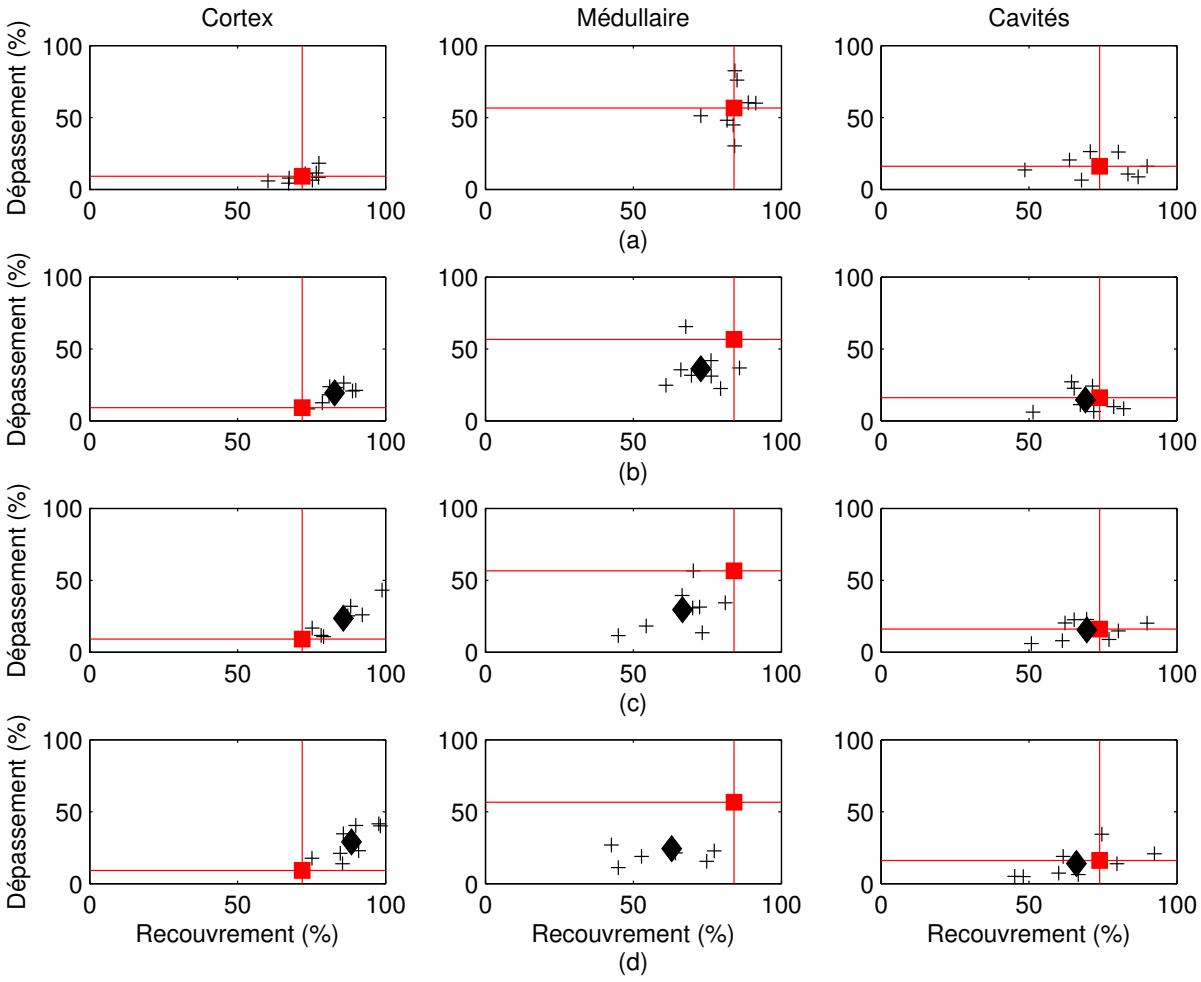


FIG. 5.10 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP1 (référence) pour les différents compartiments pour la segmentation manuelle de OP2 (a) et les segmentations par K-moyennes à 3 classes (b), 4 classes (c) et 5 classes sur vecteurs normalisés (d).

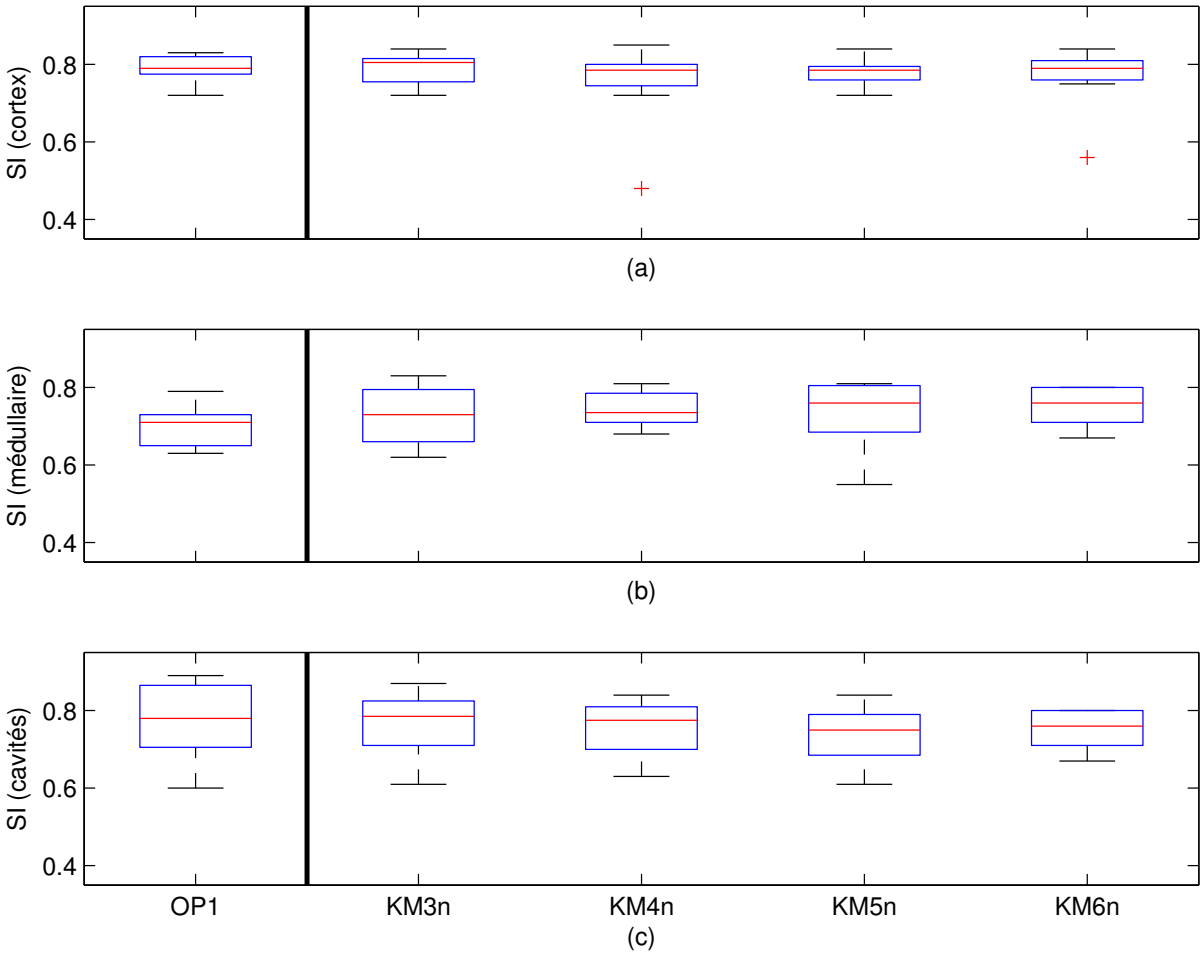


FIG. 5.11 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP2) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP1 et les segmentations automatiques par K-moyennes à 3, 4, 5 et 6 classes sur vecteurs normalisés (respectivement notées KM3n, KM4n et KM5n et KM6n.

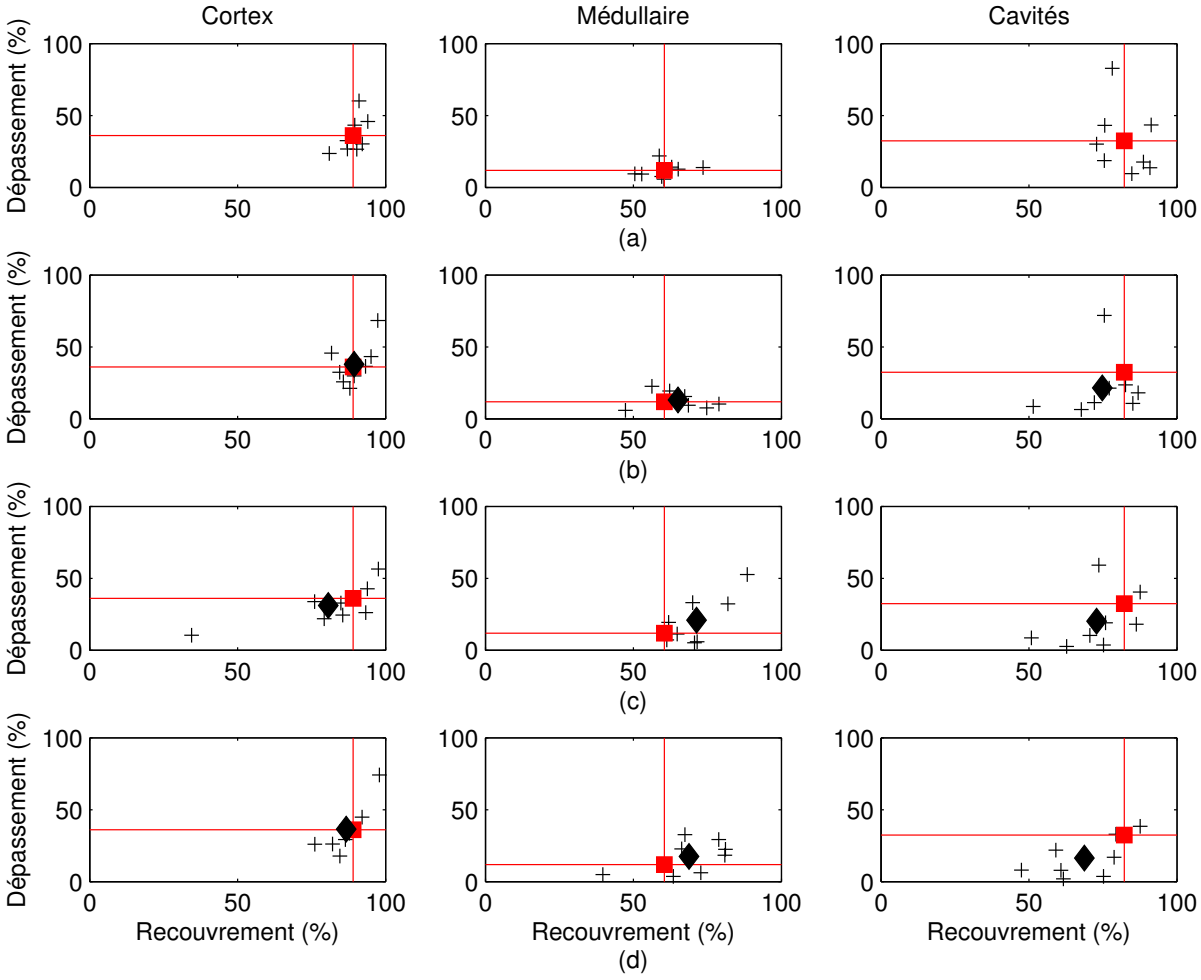


FIG. 5.12 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP2 (référence) pour les différents compartiments pour la segmentation manuelle de OP1 (a) et les segmentations par K-moyennes à 3 classes (b), 4 classes (c) et 5 classes sur vecteurs normalisés (d).

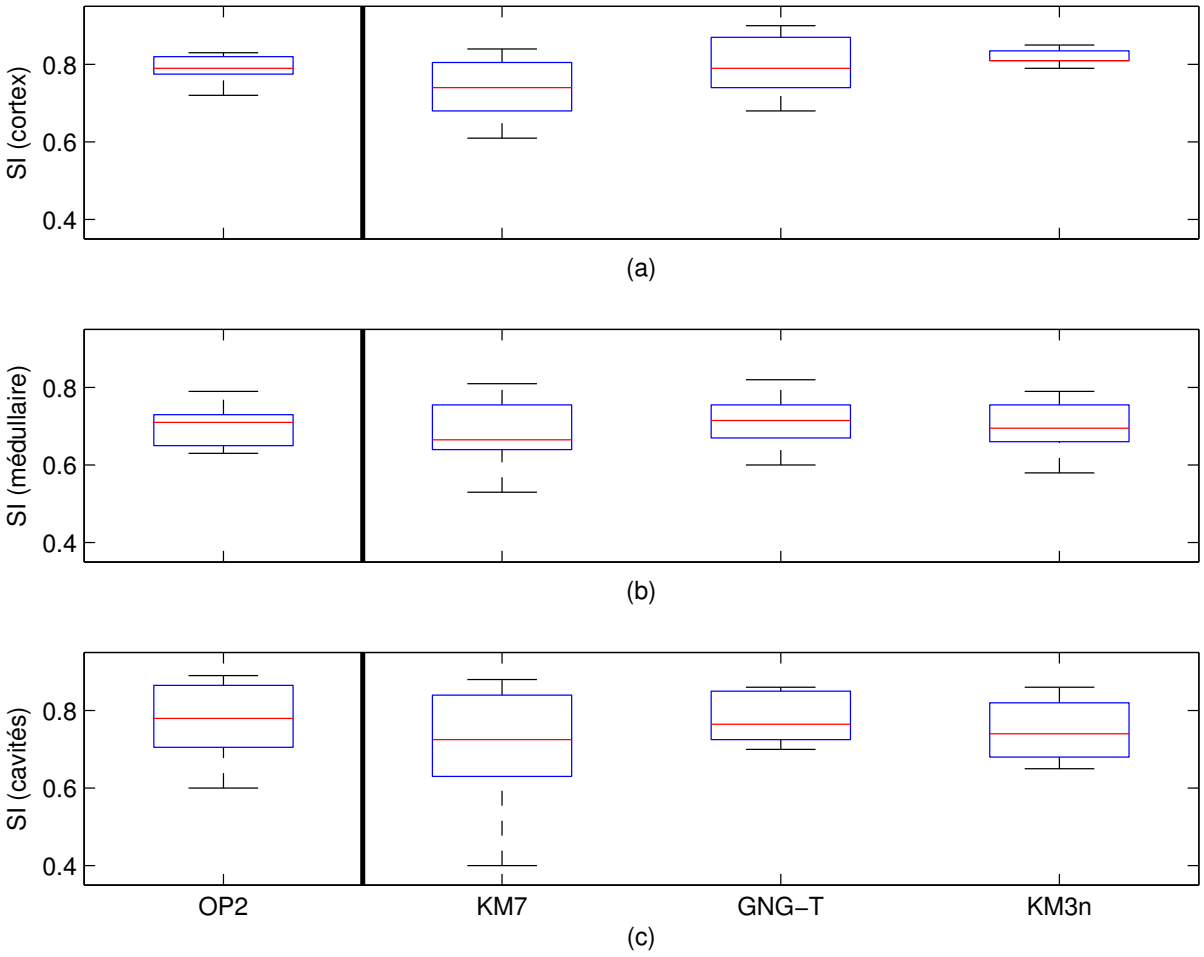


FIG. 5.13 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP1) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP2 et les segmentations automatiques par K-moyennes à 7 classes (KM7), GNG-T et K-moyennes à 3 classes sur vecteurs normalisés (KM3n).

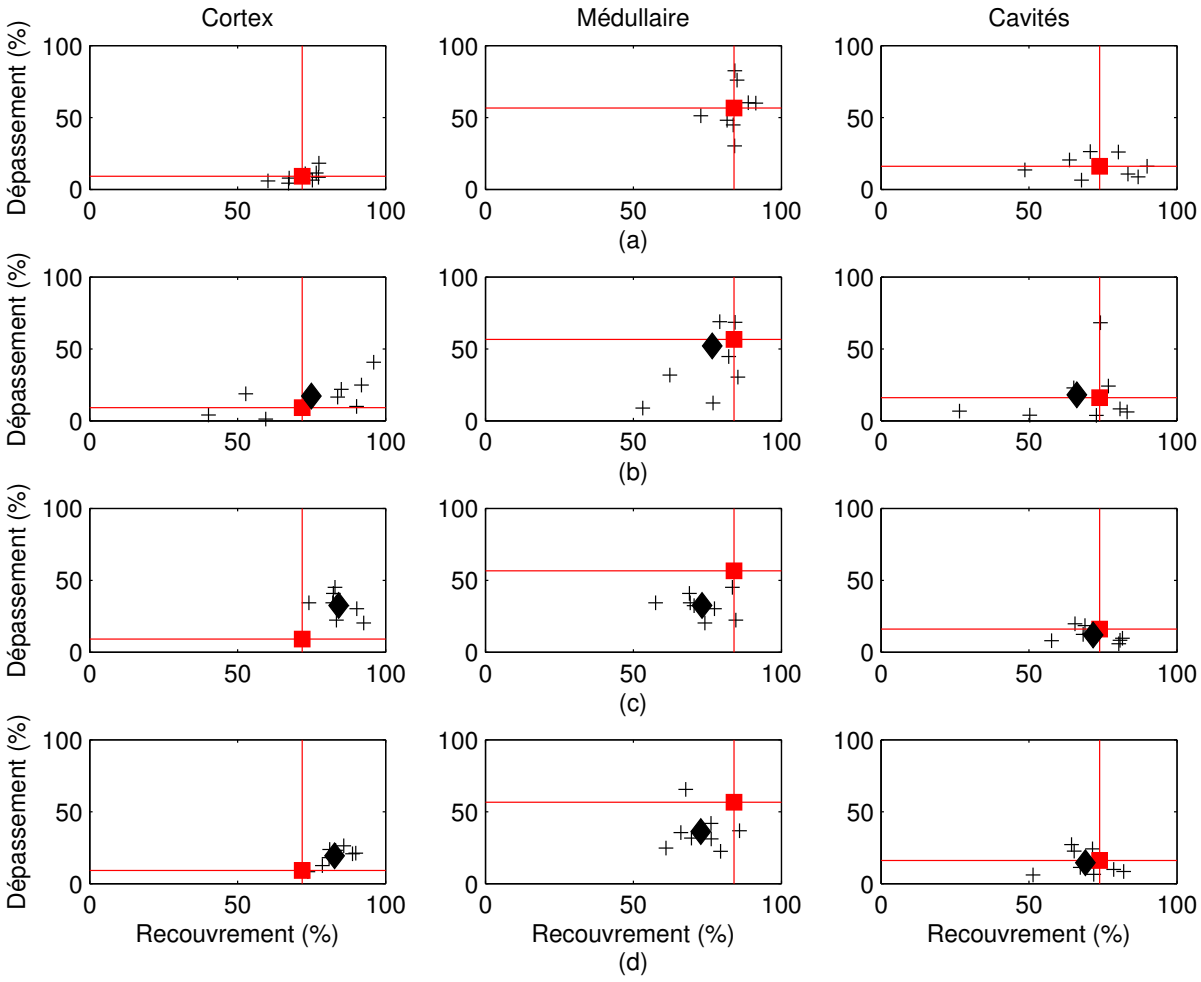


FIG. 5.14 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP1 (référence) pour les différents compartiments pour la segmentation manuelle de OP2 (a) et les segmentations par K-moyennes à 7 classes (b), GNG-T (c) et K-moyennes à 3 classes sur vecteurs normalisés (d).

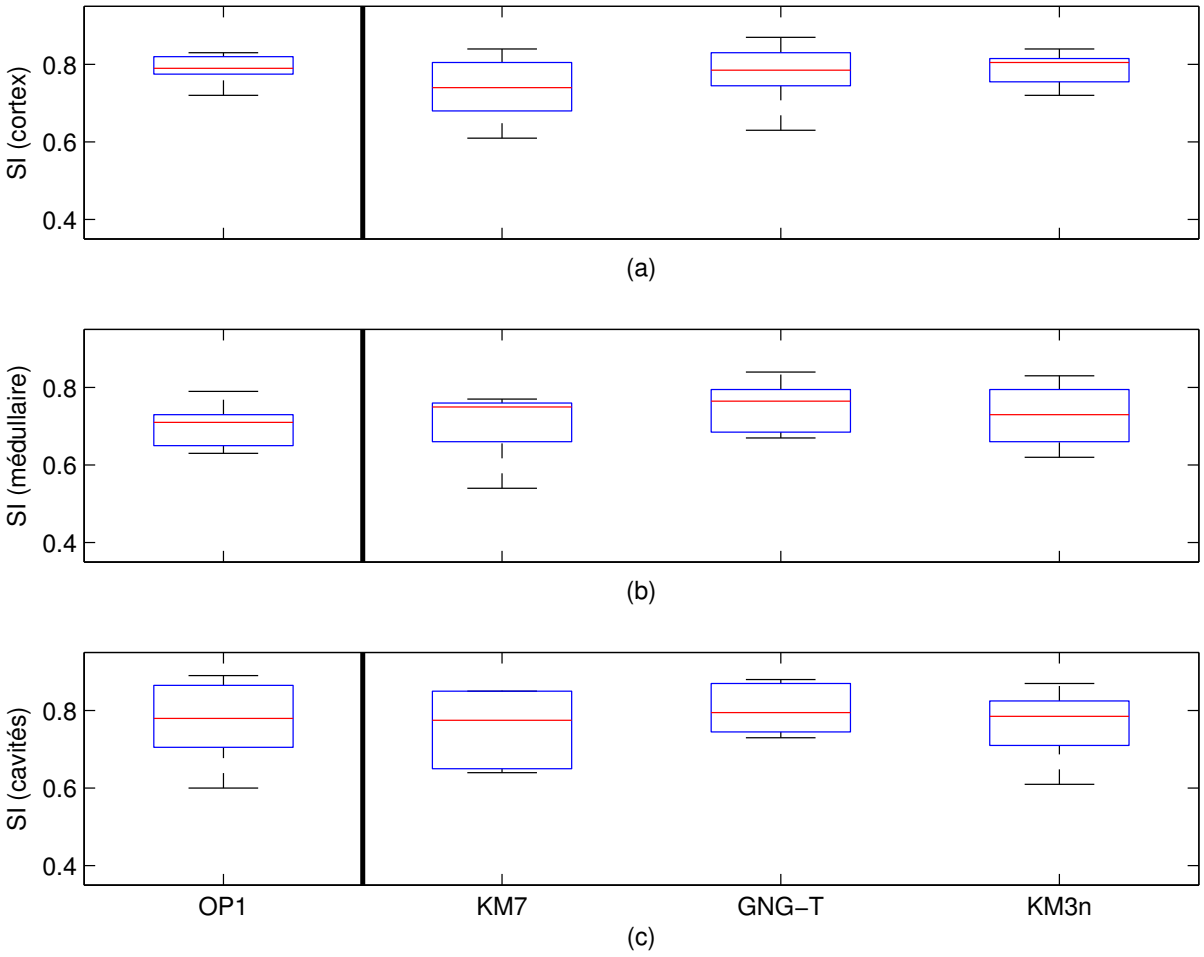


FIG. 5.15 – Comparaisons de segmentations par l'indice de similarité (calculé par rapport à la segmentation de OP2) par boîte à moustache pour le cortex (a), la médullaire (b) et les cavités (c) ; les méthodes utilisées sont la segmentation manuelle par OP1 et les segmentations automatiques par K-moyennes à 7 classes (KM7), GNG-T et K-moyennes à 3 classes sur vecteurs normalisés (KM3n).

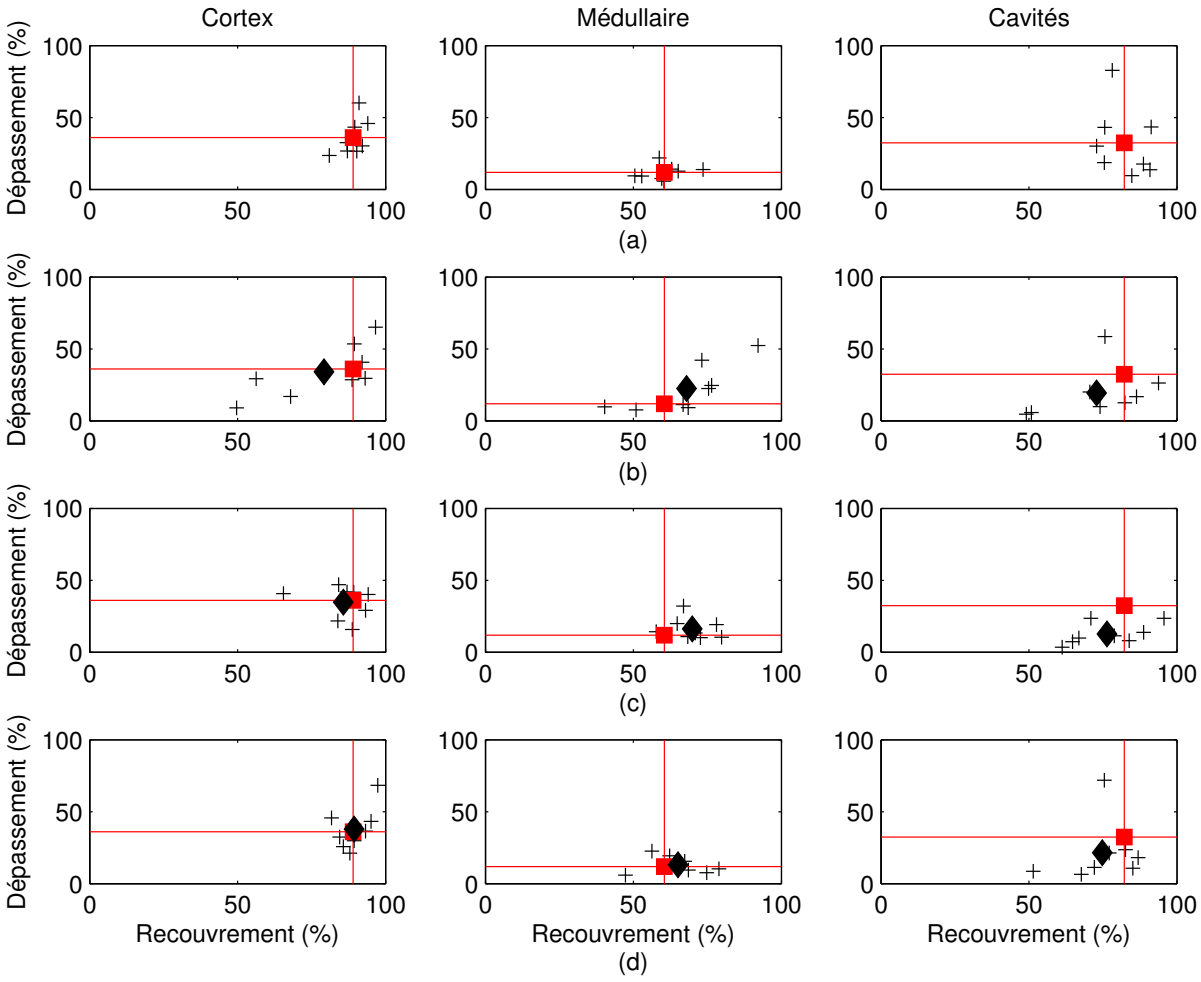


FIG. 5.16 – Comparaisons de segmentations : dépassement (%) en fonction du recouvrement (%) par rapport à la segmentation de OP2 (référence) pour les différents compartiments pour la segmentation manuelle de OP1 (a) et les segmentations par K-moyennes à 7 classes (b), GNG-T (c) et K-moyennes à 3 classes sur vecteurs normalisés (d).

Référence	OP1					
Segmentation test par	OP2	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	***69,8***	**83,6**	86,6	88,8	89,2	**83,7**
Recouvrement (%)	***71,8***	**82,8**	85,2	88,3	88,5	**83,2**
Dépassement (%)	9,2	19,3	23,4	29,0	27,8	21,9
Indice de similarité	***0,79***	**0,82**	0,82	0,81	0,82	**0,81**
Distance moyenne	0,6	0,7	0,8	0,8	0,8	0,8

(a) Cortex

Référence	OP1					
Segmentation test par	OP2	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	84,8	72,2	65,5	62,1	62,2	73,0
Recouvrement (%)	84,0	72,8	66,1	62,7	62,8	73,0
Dépassement (%)	***56,6***	**36,2**	30,3	24,9	24,9	**33,7**
Indice de similarité	***0,70***	**0,70**	0,67	0,66	0,66	**0,71**
Distance moyenne	***1,0***	**0,9**	0,9	0,8	0,8	**0,8**

(b) Médullaire

Référence	OP1					
Segmentation test par	OP2	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	74,9	70,5	71,5	69,1	71,2	68,7
Recouvrement (%)	73,9	69,1	69,8	65,9	69,5	69,8
Dépassement (%)	***16,1***	**14,5**	16,2	13,9	14,2	**11,1**
Indice de similarité	***0,77***	0,75	0,75	0,73	0,75	**0,77**
Distance moyenne	***0,8***	**0,8**	0,8	0,8	0,8	**0,7**

(c) Cavités

Référence	OP1					
Segmentation test par	OP2	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	***74,9***	**77,5**	77,0	76,5	77,1	**77,6**

(d) Rein global

TAB. 5.5 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP1 : les résultats apparaissent en gras s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

Référence	OP2					
Segmentation test par	OP1	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	***89,7***	**90,6**	76,0	75,1	77,5	85,8
Recouvrement (%)	***89,0***	**89,4**	80,3	78,9	80,2	85,7
Dépassement (%)	***36,1***	37,9	30,5	28,0	28,3	**34,8**
Indice de similarité	***0,79***	**0,79**	0,75	0,75	0,76	0,78
Distance moyenne	0,7	0,9	0,9	0,8	0,8	0,9

(a) Cortex

Référence	OP2					
Segmentation test par	OP1	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	***60,3***	**63,2**	73,7	76,4	74,1	**69,7**
Recouvrement (%)	60,5	65,0	71,4	74,6	73,5	69,9
Dépassement (%)	11,8	13,2	21,0	23,3	22,1	16,3
Indice de similarité	***0,70***	**0,73**	0,74	0,76	0,75	**0,75**
Distance moyenne	***1,0***	**0,8**	0,9	0,9	0,9	**0,9**

(b) Médullaire

Référence	OP2					
Segmentation test par	OP1	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	81,8	75,4	72,7	70,4	73,6	74,9
Recouvrement (%)	82,2	74,8	73,0	68,9	73,3	76,4
Dépassement (%)	***32,4***	**21,4**	21,8	16,6	17,4	**12,6**
Indice de similarité	***0,77***	0,76	0,75	0,74	0,77	**0,80**
Distance moyenne	***0,9***	**0,8**	0,8	0,8	0,7	**0,5**

(c) Cavités

Référence	OP2					
Segmentation test par	OP1	K-moyennes				GNG-T
K (pour les K-moyennes)		3	4	5	6	
Pixels bien classés (%)	***74,9***	**75,6**	74,4	74,9	75,3	**76,7**

(d) Rein global

TAB. 5.6 – Mesures de similarité pour les segmentations des trois ROIs, la référence étant OP2 : les résultats apparaissent en gras s'ils sont meilleurs que ceux obtenus avec les segmentations manuelles (gras italique).

Cas de cavités se remplissant très tardivement

Rappelons que nous avons choisi une médullaire extrêmement fine, ce qui est un cas très défavorable. Nous avons vu au paragraphe 4.1.2 que le faible volume de la médullaire expliquait le très fort pourcentage de pixels de ce compartiment anatomique ayant un comportement temporel dominant de cortex ou de cavités pour le modèle avec bruit, volume partiel et filtre moyenneur, ce qui laisse peu d'espoir de retrouver une bonne correspondance entre segmentations anatomique et fonctionnelle pour ce compartiment. Ce point était moins crucial pour le modèle de rein sain, puisque les pixels pour lesquelles les deux segmentations étaient différentes, étaient répartis de manière assez équitable entre les trois compartiments. Il faut cependant souligner que sur des données réelles, les segmentations réalisées par les experts se basent sur des images de type fonctionnel. Au vu des images de synthèse de la figure 5.17, on peut légitimement se demander si ces segmentations ressembleraient davantage à la segmentation anatomique ou à une segmentation dans laquelle chaque pixel serait attribué à son compartiment fonctionnel dominant. Nous gardons cependant comme référence la segmentation anatomique, par souci de cohérence avec les paragraphes précédents.

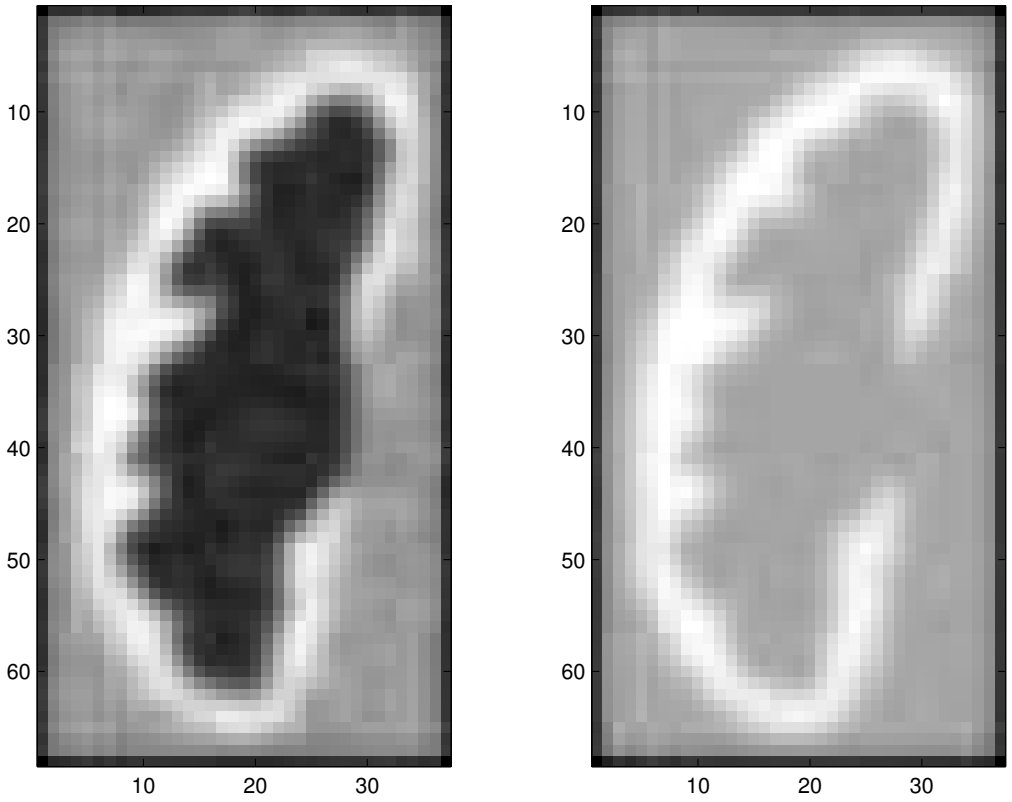


FIG. 5.17 – Images de synthèse pour le modèle de rein pathologique au pic cortical (a) et en phase d'équilibre (b).

K-moyennes à trois classes Pour le modèle avec bruit seul, nous obtenons une parfaite correspondance entre la segmentation idéale et la segmentation automatique pour les mêmes niveaux de bruit que le rein sain.

Lorsque l'on tient compte aussi du volume partiel, on obtient un WCP de 73%. Cortex et médullaire sont regroupés dans une même classe, alors que les cavités sont séparées en deux classes (cf. figure 5.18). Ceci se produit vraisemblablement parce que la médullaire est très peu représentée et toujours mêlée aux autres compartiments (cf. les représentations des compartiments de la figure 4.14). Notons qu'il est difficile même pour les radiologues de séparer cortex et médullaire. Il est dans ce cas envisageable de se limiter à deux compartiments : parenchyme (cortex et médullaire réunis) et cavités. Nous pouvons soit regrouper deux classes sur le résultat obtenu par les K-moyennes avec $K = 3$, soit directement tester les K-moyennes à deux classes.

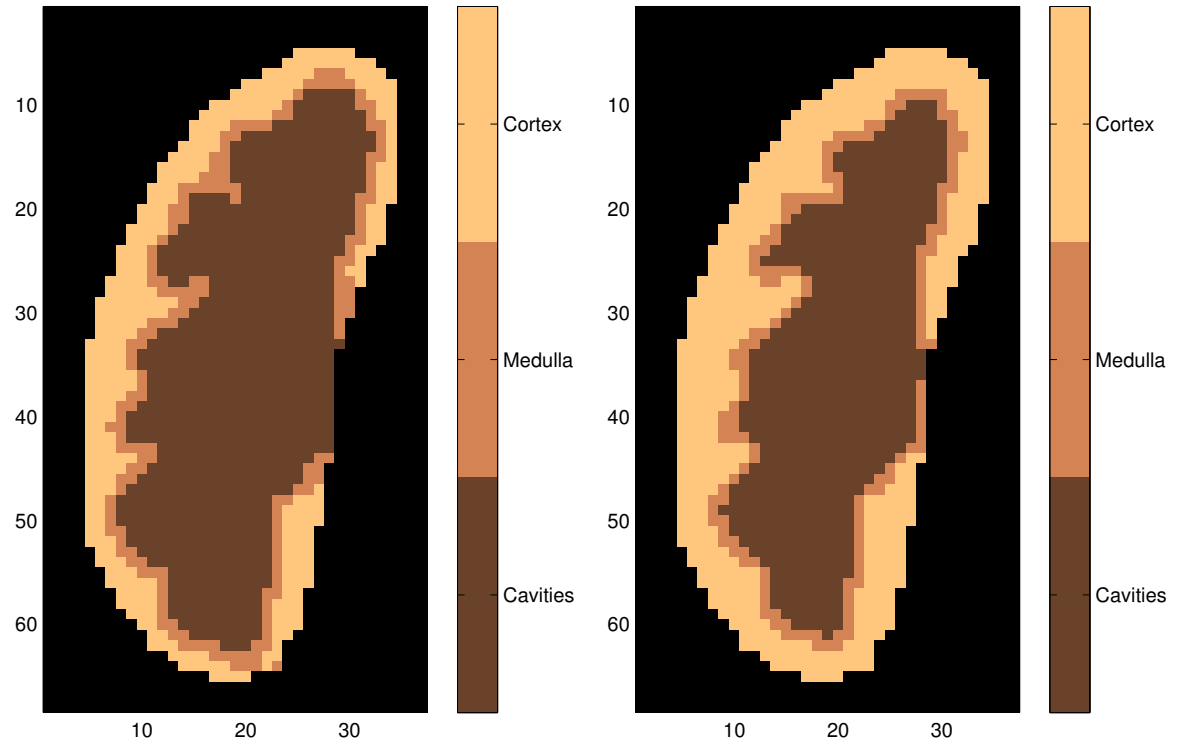


FIG. 5.18 – Segmentation de référence (a) et segmentation obtenue par K-moyennes à 3 classes pour le modèle avec bruit et volume partiel.

K-moyennes à deux classes Pour le modèle avec bruit seul, ou avec bruit et volume partiel, nous obtenons 100% de pixels bien classés (cf. figure 5.19).

Si nous ajoutons le filtre moyenneur, 97% des pixels sont bien classés en cas de recalage parfait ; avec des erreurs de recalage résiduelles, nous atteignons un WCP de 95%, les pixels mal classés étant toujours à la frontière des compartiments de la segmentation idéale (cf. respectivement les figures 5.19(c) et (d)).

Si on souhaite délimiter les trois compartiments, il est possible d'essayer l'algorithme des K-moyennes avec $K > 3$ suivi d'une fusion manuelle.

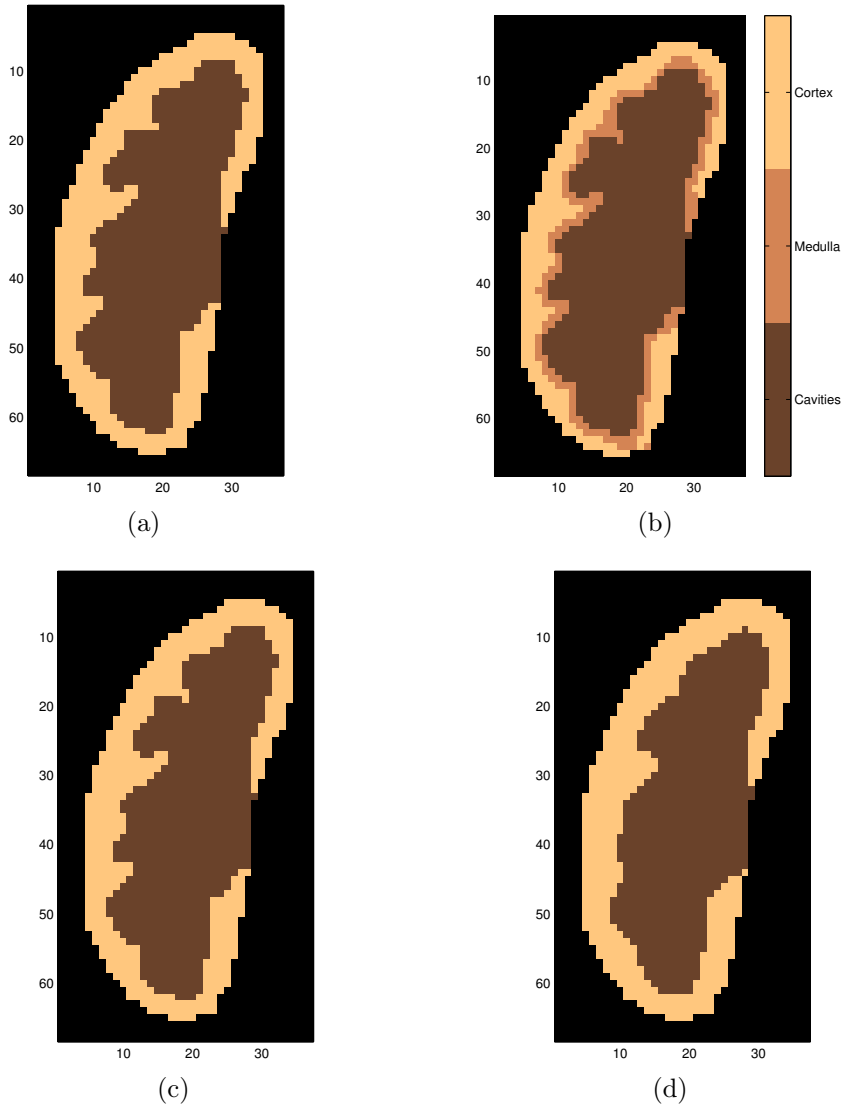


FIG. 5.19 – Segmentation idéale de référence (a) à deux classes (parenchyme et cavités) construite à partir la segmentation idéale (b) à trois classes ; segmentation obtenue par K-moyennes à deux classes pour le modèle avec bruit et volume partiel avec recalage parfait (c) et avec erreurs de recalage (d).

K-moyennes à quatre classes Avec le modèle tenant compte du bruit, du volume partiel et du filtre moyenneur, on obtient WCP = 92% avec un recalage parfait. Sur la figure 5.20, la médullaire est moins bien retrouvée que les deux autres compartiments (recouvrement de 61%, dépassement de 18%), ce qui s'explique par l'effet de volume partiel, particulièrement marqué en raison de la position intermédiaire de ce compartiment entre cortex et cavités, et de son faible volume (14% du volume rénal). L'augmentation

du nombre de classes à $K = 5$ n'apporte pas d'amélioration véritable.

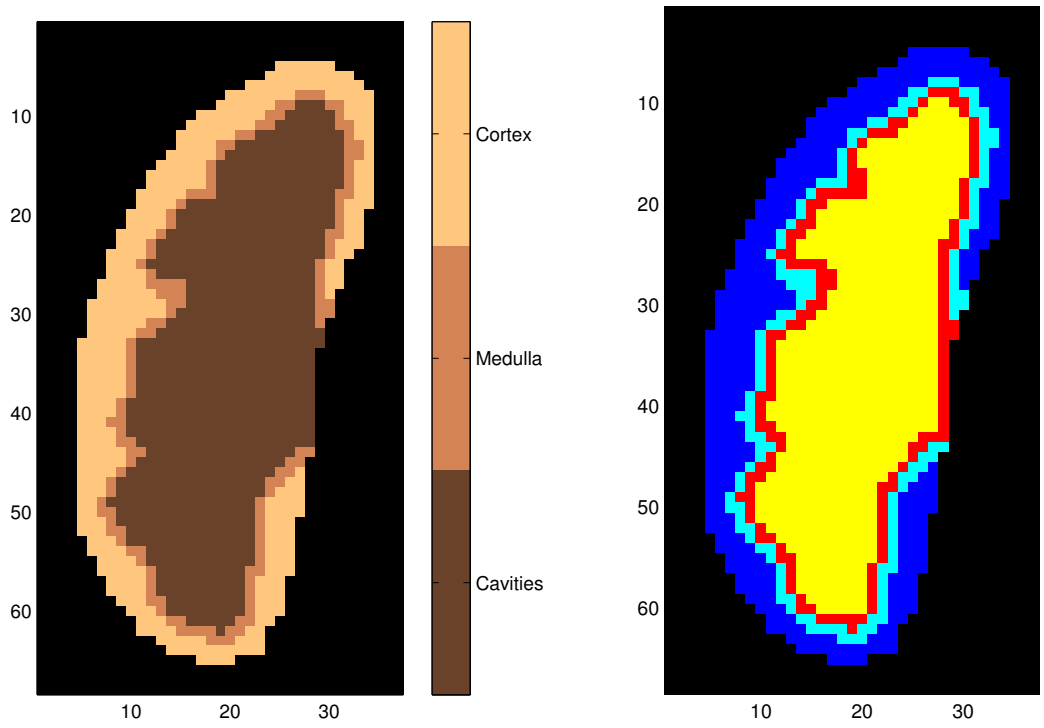


FIG. 5.20 – Segmentation (a) obtenue par regroupement de deux classes de la segmentation initiale (b).

K-moyennes à trois classes, vecteurs normalisés Pour le modèle avec bruit, volume partiel et filtre moyenneur, avec un recalage parfait, le WCP vaut 86% (cf. figure 5.21(a)).

Avec des erreurs de recalage, 83% des pixels sont bien classés (cf. figure 5.21(b)). La médullaire reste difficile à localiser, mais les résultats obtenus sont bien meilleurs qu'avec des vecteurs non normalisés, tout comme dans le cas des reins sains.

A titre d'information, pour mettre en évidence le point soulevé au début de ce paragraphe, à savoir que les segmentations manuelles sont réalisées sur des images de type plutôt fonctionnel et que cela peut les faire différer de la segmentation anatomique sous-jacente, nous avons superposé cette dernière et celle obtenue par les K-moyennes à trois classes sur vecteurs normalisés à des images de la série qui auraient pu être choisies par les experts (cf. figure 5.22). Il nous semble assez évident que l'expert n'aurait pas nécessairement choisi la segmentation anatomique de référence. Cela pourrait expliquer des résultats éventuellement meilleurs sur des données réelles que sur des données de synthèse.

GNG-T semi-automatique Pour le modèle avec bruit, volume partiel et filtre moyenneur, avec un recalage parfait, 96% des pixels sont bien classés (cf. figure 5.23).

Lorsque l'on tient compte des erreurs de recalage, le WCP diminue à 87% : les pixels de la médullaire sont difficilement retrouvés (cf. figure 5.23(b)). Vu la finesse de ce com-

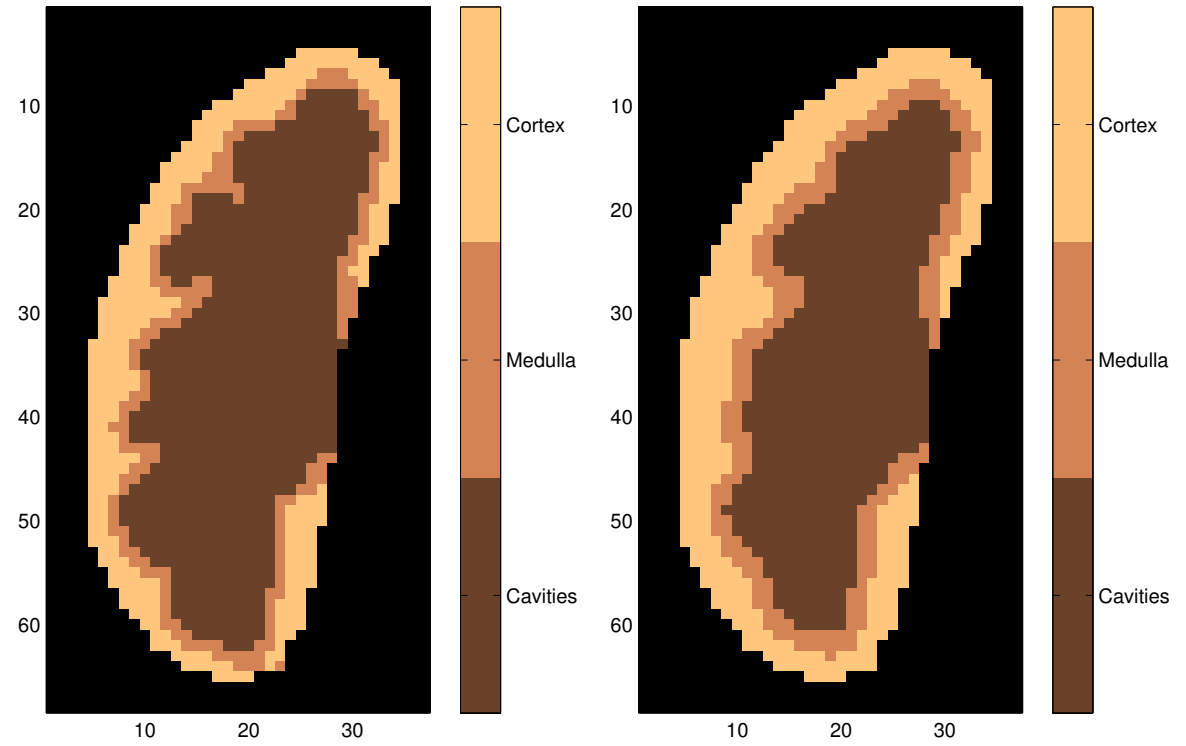


FIG. 5.21 – Segmentations obtenues par K-moyennes à trois classes sur vecteurs normalisés avec le modèle avec bruit et volume partiel et recalage parfait (a) et avec erreurs de recalage (b).

partiment, les évolutions temporelles des voxels de cette zone sont vraiment très instables, puisque la composition de ces voxels change fortement à chaque instant et donne lieu à des comportements hybrides très marqués. Le WCP est néanmoins supérieur à celui obtenu avec les autres méthodes, probablement parce que l'algorithme de GNG-T est moins sensible à la sous-représentation des classes que les K-moyennes (cf. paragraphe 3.1.4).

Cas de cavités ne se remplissant pas

Les résultats sont tout à fait similaires à ceux du paragraphe précédent, mais le critère du GNG-T pour l'extraction des cavités a été modifié : les courbes prototypes après quantification vectorielle sont classées selon leur énergie. Toutes celles dont l'énergie est inférieure à un seuil s_1' sont supposées appartenir à des prototypes de cavités. Le problème lié à la finesse de la médullaire reste le même que précédemment.

Des exemples de segmentation obtenues pour le modèle avec bruit et volume partiel avec recalage parfait ou avec erreurs résiduelles sont présentés respectivement sur les figures 5.24(a) et (b).

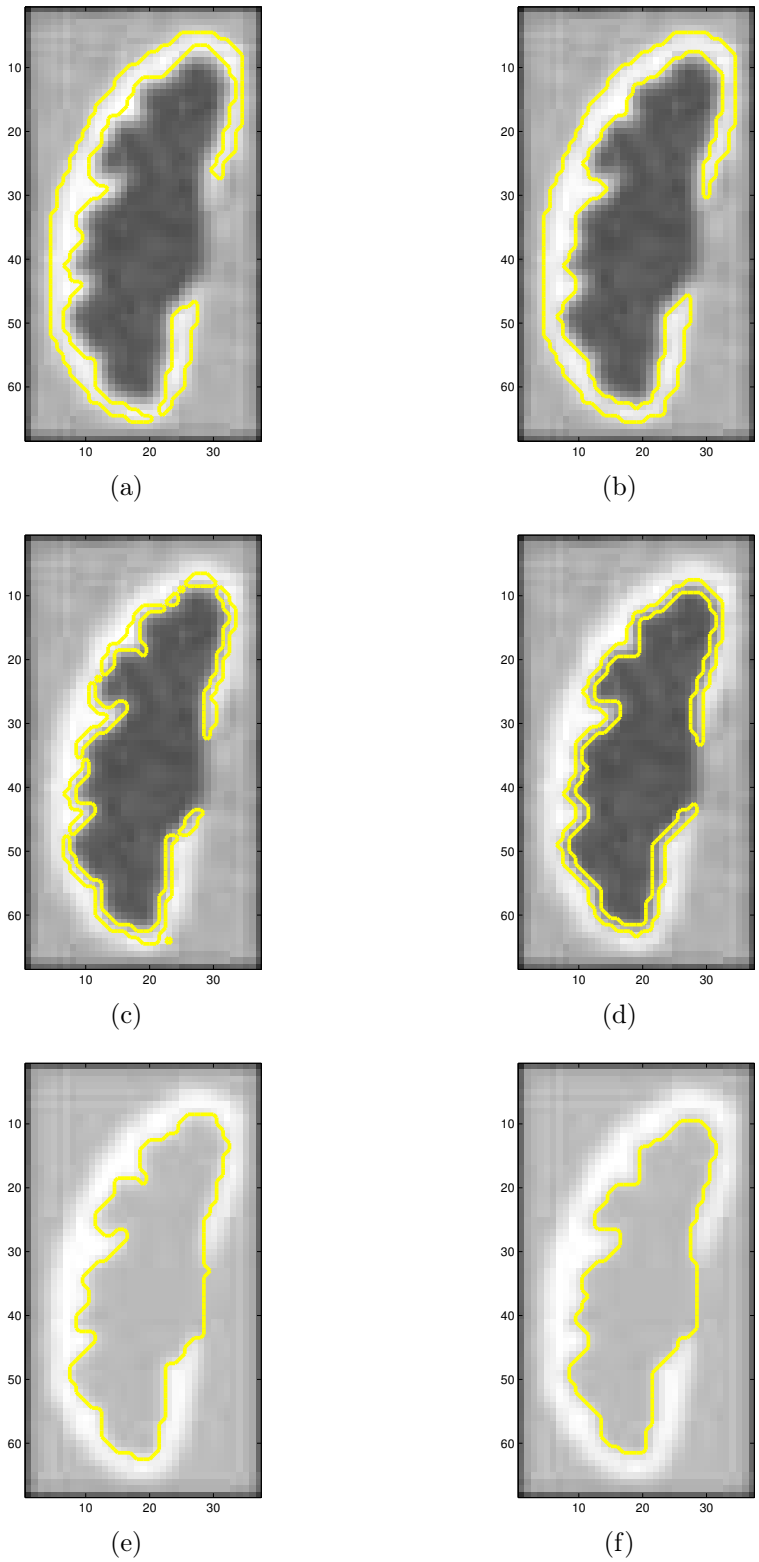


FIG. 5.22 – Segmentations anatomiques de référence pour le cortex (a), la médullaire (c) et les cavités (e), et segmentations obtenues par les K-moyennes à trois classes sur les vecteurs normalisés (respectivement (b), (d) et (f)).

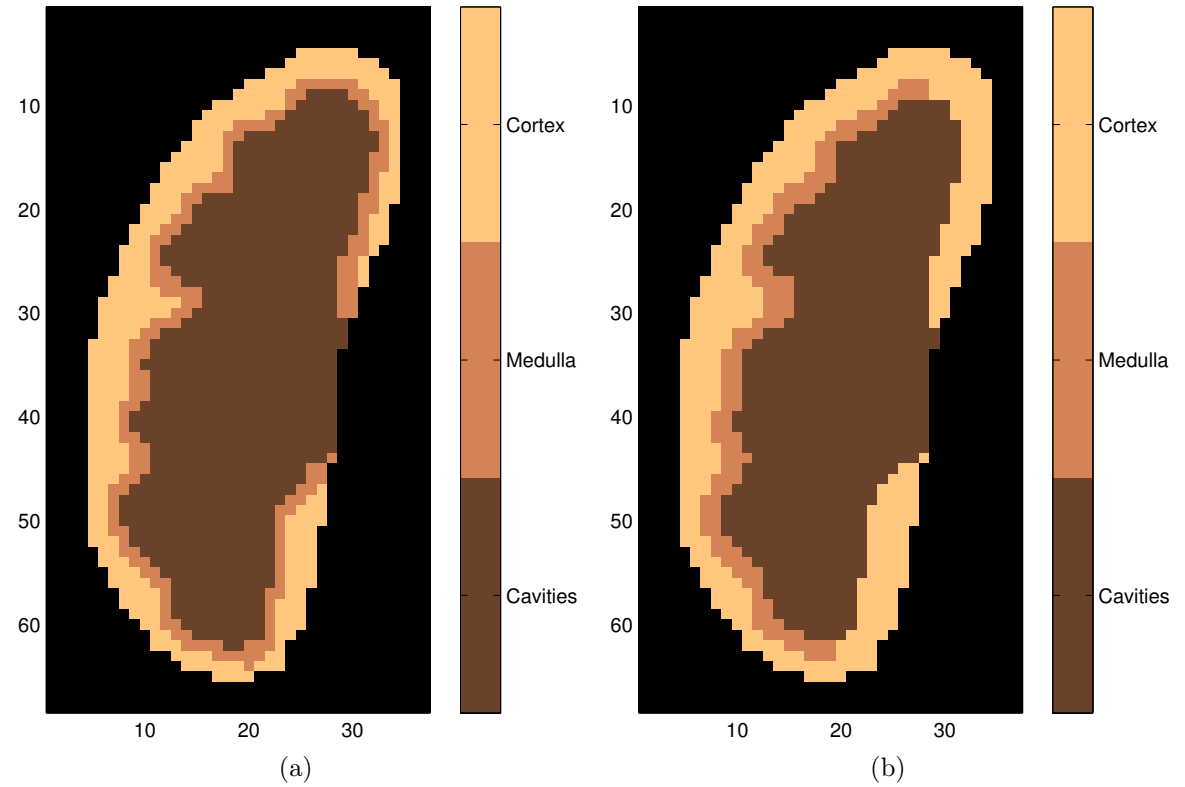


FIG. 5.23 – Segmentations obtenues par GNG-T avec le modèle avec bruit et volume partiel et recalage parfait (a) et avec erreurs de recalage (b).

Cas de cavités se remplissant en partie seulement

Les cas réels rencontrés pouvant être très variés, nous suggérons les pistes suivantes pour adapter notre méthode. Il faudra étudier le nombre de classes optimal pour l'algorithme des K-moyennes : une classe supplémentaire sera certainement nécessaire pour tenir compte du remplissage inhomogène des cavités.

En ce qui concerne le GNG-T, nous proposons d'ajouter un seuil supplémentaire et de procéder en deux étapes au lieu d'une pour extraire les cavités : outre le seuil d'intensité en phase tardive, il est probable qu'un seuil sur l'énergie totale du signal pour chaque voxel permettra de détecter les parties des cavités qui ne se remplissent pas et qui restent donc plus sombres que les autres tout au long de la séquence. Le seuil sur la pente des courbes en phase de filtration pour séparer cortex et médullaire serait conservé. Il y aurait donc trois seuils indépendants à régler au lieu de deux.

Conclusion sur les reins pathologiques

Les performances des classificateurs testés semblent plutôt diminuer en raison de la finesse de la médullaire du modèle, mais pas en raison de l'évolution temporelle par compartiment. Ce point est confirmé par des tests réalisés en associant le modèle anatomique sain à des courbes temps-intensité de reins pathologiques : les résultats sont similaires à

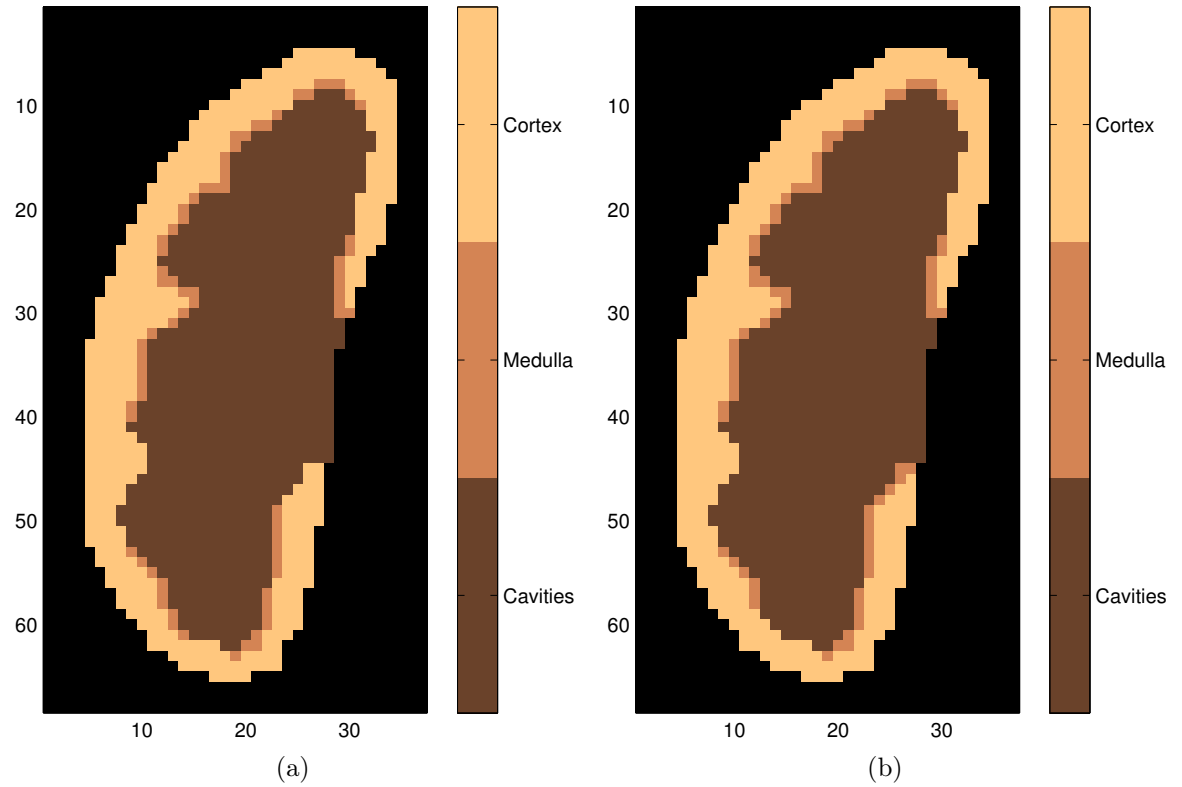


FIG. 5.24 – Segmentations obtenues par GNG-T avec le modèle avec bruit et volume partiel et recalage parfait (a) et avec erreurs de recalage (b).

ceux obtenus pour les reins sains.

En ce qui concerne les cas réels, des tests doivent être réalisés en adaptant la méthode comme nous l'indiquons dans le paragraphe précédent.

5.1.8 Conclusion sur la construction des classificateurs

A cette étape de notre travail, nous avons construit plusieurs classificateurs permettant de segmenter la coupe rénale principale. Nous les avons testés tant sur des données de synthèse que sur des données réelles (uniquement sur reins sains dans le second cas).

Pour les données simulées avec le modèle sans volume partiel, on obtient 100% de voxels bien classés avec l'algorithme des K-moyennes à $K = 3$ classes et par la méthode automatique utilisant GNG-T ; ce modèle est cependant peu réaliste.

Un modèle plus élaboré tenant compte de l'effet de volume partiel et d'un filtrage des données met en évidence la différence entre segmentation anatomique et fonctionnelle : avec les paramètres utilisés, environ 10% des voxels ont des compartiments fonctionnels et anatomiques différents pour des reins sains ; on ne pourra donc pas espérer trouver une coïncidence parfaite entre les deux types de segmentations, le mieux qu'on puisse espérer est un WCP de 90%. Pour le modèle de rein pathologique, la non-correspondance est plus

marquée, surtout pour la médullaire, à cause de la finesse de ce compartiment. Le modèle met en évidence la sensibilité de GNG-T automatique à ces phénomènes : les composantes connexes de la variété sont alors mal séparées. Au vu des simulations réalisées, il n'est pas exclus qu'avec des acquisitions à plus haute résolution spatiale même au détriment du rapport signal à bruit, le classement fonctionnel des voxels soit facilité. Les résultats laissent pourtant supposer qu'un GNG-T semi-automatique pourrait être satisfaisant. L'ajout d'un filtrage d'erreurs de recalage dégrade les performances des K-moyennes à trois classes, le WCP diminuant de 10% ; l'apport des K-moyennes à $K = 4$ classes et du GNG-T semi-automatique est très net puisqu'on atteint la limite de 90% pour le WCP sur des reins sains.

Pour les données réelles, avec l'algorithme des K-moyennes appliqué à des courbes non normalisées, il faut que K soit au moins égal à 5 pour réussir à approcher les segmentations manuelles ; les performances demeurent cependant un peu inférieures.

Parmi les variantes testées, les K-moyennes sur vecteurs normalisés avec $K = 3$ et le GNG-T semi-automatique paraissent les mieux adaptées pour la segmentation des structures rénales internes. En effet, n'importe quelle segmentation anatomique et n'importe quelle segmentation fonctionnelle obtenue par l'une ou l'autre de ces deux méthodes se ressemblent autant que les deux segmentations anatomiques. L'hypothèse de séparation des composantes connexes n'est pas critique pour ces deux méthodes.

Pour les reins sains testés, les résultats donnés par les K-moyennes à trois classes sur des vecteurs normés sont la plupart du temps satisfaisants. La quantification vectorielle des courbes temps-intensité permet ainsi d'aboutir directement à la segmentation attendue, en un processus entièrement automatique. Cependant, en cas de résultat aberrant, il n'y a pas de possibilité de correction par un opérateur. Il est cependant possible de tester un autre nombre de classes (inférieur à 7), dont certaines sont par la suite fusionnées manuellement par un opérateur, ou bien d'utiliser la méthode semi-automatique avec GNG-T. Pour les K-moyennes avec $K \leq 3$, nos résultats sont un peu différents de ceux exposés dans [42] avec une méthode similaire ; sur quatre examens formant la base de données, deux ont été réalisés avec des séquences similaires aux nôtres mais acquises à 3 T au lieu de $1,5$ T, et pour 55 instants au lieu de 256. Les auteurs considèrent que 5 à 7 classes sont nécessaires pour obtenir une segmentation convenable ; seules 2 ou 3 classes sont conservées, les autres étant considérées comme dues au volume partiel ou à des morceaux d'autres organes conservés dans le masque rénal. Pour l'un des cas, la médullaire et les cavités sont regroupées dans une même classe, et pourraient être séparées par un post-traitement parce que les classes sont disjointes sur la segmentation ; ce qui n'est généralement pas le cas sur les séries dont nous disposons. La segmentation est directement réalisée sur les données 3D, sachant que les coupes extrêmes du rein ne sont pas conservées car les images ne sont pas d'assez bonne qualité.

Pour la méthode semi-automatique avec GNG-T, la quantification vectorielle permet, en tenant compte de l'évolution temporelle globale, de réduire la taille des données et de diminuer le bruit, afin que l'étape de fusion, basée sur des propriétés physiologiques, soit facile à faire pour un opérateur : elle consiste uniquement en un réglage assez grossier de deux seuils indépendants ; la méthode peut être une solution alternative offrant davantage de flexibilité que les K-moyennes pour le traitement des cas difficiles, car le nombre de

prototypes représentant les données est nettement plus élevé. On peut même envisager de modifier la valeur de la cible T pour accroître la souplesse de la méthode.

Le gain de temps est considérable : alors qu'il faut une dizaine de minutes pour réaliser une segmentation manuelle, seulement une vingtaine de secondes est nécessaire pour une segmentation fonctionnelle.

Le caractère adaptatif de l'algorithme de GNG-T fait que les prototypes résultants continuent à changer même en fin d'époque. Cependant, en réglant convenablement les taux d'apprentissage α_1 et α_2 de l'équation (3.35), on peut éviter des fluctuations trop importantes. Les résultats obtenus dépendent quelque peu de l'ordre de présentation des données mais, lorsque plusieurs essais sont effectués en modifiant cet ordre, les valeurs des critères de ressemblance entre segmentations ne changent pas de manière significative. Il serait d'ailleurs possible de finaliser les résultats en appliquant un algorithme de K-moyennes dont les conditions initiales seraient les résultats de GNG-T, mais les résultats sont très peu modifiés également.

Des tests supplémentaires devront être menés sur une base de données plus importante contenant des reins sains et des reins pathologiques. Pour ces derniers, comme de nouveaux comportements temporels risquent d'apparaître, une adaptation des critères de fusion pour GNG-T ainsi qu'une recherche d'un nombre de classes K optimal pour les K-moyennes devront être effectuées. Il est probable que la différence de représentativité des voxels due à la dilatation des cavités et à l'amincissement du cortex et de la médullaire, justifiera l'usage de GNG-T plutôt que celui des K-moyennes.

Les courbes fonctionnelles diffèrent très peu en fonction de la technique utilisée ; il faudra cependant également vérifier que c'est également vrai pour les paramètres rénaux fonctionnels qui en sont tirés.

Notons enfin que le WCP traduit bien la qualité des segmentations obtenues, en accord avec les autres critères. Nous pourrons ainsi utiliser l'erreur de classement $1-\text{WCP}$ comme critère de risque pour estimer cette qualité dans le cadre de la généralisation.

5.2 Extensions aux coupes voisines

Après l'étape décrite au paragraphe précédent, nous disposons de M classificateurs avec lesquels nous avons classé les N voxels de la coupe rénale principale (qui, accessoirement, ont servi à les construire, bien que cela n'ait pas d'incidence par la suite). Chacun de ces classificateurs permet d'attribuer une classe aux voxels de la coupe principale, mais aussi à tout autre voxel des autres coupes, puisqu'une partition de $\mathbb{R}^p$ entier a été réalisée. Distinguons deux cas : le premier concerne les données de la base de test, l'autre s'appliquerait davantage au cas réel, où l'on n'aurait pas de segmentation de référence.

5.2.1 Données de la base de test

Pour les reins de la base de test, nous pouvons évaluer les performances de chacun des classificateurs pour les voxels de la coupe principale par rapport à la segmentation d'un radiologue, en se basant sur l'erreur de classification $1 - \text{WCP}$: ce critère traduit convenablement la qualité de la segmentation, il n'atteint pas de valeurs très basses si les

autres critères de similarité ont des mauvaises valeurs. A priori, pour classer les voxels des autres coupes, le radiologue choisit le classificateur « optimal », au sens où il minimise $1 - \mathrm{WCP}$ sur les voxels de la coupe principale. Il s'agit alors d'estimer la probabilité d'erreur de classement lorsqu'on utilise ce même classificateur pour attribuer une classe aux voxels des autres coupes rénales, dont nous avons les masques globaux.

Rappelons que le cadre mathématique dans lequel nous nous plaçons, présenté au chapitre 3.2, est celui de la théorie de la généralisation de Vapnik [79], qui se base sur la minimisation du risque empirique. Nous utilisons ici les mêmes notations que dans ce chapitre. Nous soulignons à nouveau que la manière dont nous l'appliquons est très différente de celle envisagée par Vapnik, mais les résultats peuvent s'appliquer à notre cas. D'autres points de vue sont possibles.

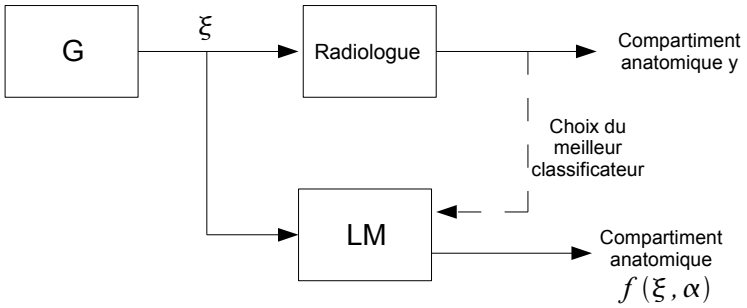


FIG. 5.25 – Généralisation : soit ξ le vecteur temps-intensité d'un voxel x et y le compartiment anatomique que lui associe un radiologue ; l'entité LM permet de choisir, parmi M classificateurs, le meilleur de ceux-ci pour les paires (ξ, y) de la base d'exemples tirés de la coupe rénale principale, c'est-à-dire celui qui minimise l'erreur de classement pour ces paires. Ce choix définitif étant effectué après la présentation de N exemples, la machine doit, pour toute entrée ξ, donner un compartiment anatomique $\widehat{y} = f(\xi, \alpha)$. L'objectif est de retourner le plus souvent possible le même compartiment que le radiologue.

Considérons le système de la figure 5.25. Imaginons que le générateur G fournisse successivement les vecteurs $\xi_1 \ldots, \xi_N$, qui sont les attributs (vecteurs temps- intensité) de voxels de la coupe principale, auxquels on peut associer une classe correspondant à un compartiment anatomique en utilisant :

- soit la segmentation faite par le radiologue, considérée comme la référence ; la classe résultat est représentée par la variable y.
- soit la réponse d'un des M classificateurs construits précédemment, indicés par la variable α, et dont la classe résultat, en sortie de l'entité LM, est représentée par la fonction $f(\xi, \alpha)$.

L'objectif est que LM retourne le plus souvent possible le même compartiment $\widehat{y} = f(\xi, \alpha)$ que le radiologue. Au $\ell^{\text{ième}}$ vecteur ξ présenté, pour chacun des M classificateurs, il est possible de calculer l'erreur de classement commise sur les ℓ paires (ξ, y), qui n'est autre que le risque empirique moyen d'ordre ℓ défini dans l'équation (3.56). A cette étape, le classificateur que l'on peut considérer comme étant le meilleur est celui dont l'erreur de classement est minimale ; l'entité LM rend alors le résultat $f(\xi, \alpha)$ correspondant. A chaque ξ supplémentaire, le classificateur optimum peut changer. Lorsqu'on est arrivé au

N-ième vecteur ξ_N, le classificateur fourni par LM est bien le classificateur « optimal » pour la coupe principale, évoqué ci-dessus.

A priori, on va garder comme classificateur celui qui donne la plus petite erreur de classement par rapport à la segmentation du radiologue pour la coupe de référence, une fois qu'on a présenté les N vecteurs ξ des voxels de cette coupe. On peut alors utiliser ce classificateur pour classer les voxels des autres coupes.

Cependant, deux questions se posent : le classificateur qui donne les meilleurs résultats pour la coupe de référence est-il vraiment le classificateur optimum, c'est-à-dire celui qui minimise la probabilité d'erreur de classement pour des voxels provenant d'autres coupes ? Si le classificateur choisi n'est pas le classificateur optimum, leurs performances sont-elles très différentes ? Par ailleurs, mêmes si elles sont bonnes sur la coupe de référence, restent-elles acceptables si on généralise le classement aux autres coupes ? En effet, on peut par exemple imaginer qu'un classificateur donnant de bons résultats sur un « faible » nombre de voxels, qui ne seraient pas suffisamment représentatifs, voie ses performances se dégrader largement quand on l'essaie sur d'autres échantillons.

Application de la théorie de la généralisation

Dans la suite de l'exposé, ξ désignera un vecteur temps-intensité non normalisé : nous considérerons que l'étape de normalisation est incluse dans le traitement effectué par l'entité LM.

Reprenons le raisonnement ci-dessus en faisant le lien avec les hypothèses du paragraphe 5.1.1 et les notations du paragraphe 3.2.1 page 98. Les N vecteurs ξ_i, représentant les courbes temps-intensité des N voxels de la coupe principale d'un rein donné, peuvent être considérés comme une réalisation d'une séquence de variables aléatoires vectorielles $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ identiquement distribuées de même loi $P_{\mathbf{X}}(\xi)$ inconnue qu'un vecteur aléatoire continu $\mathbf{X} : \Omega \longrightarrow V \subset \mathbb{R}^p$, où $(\Omega, \mathcal{E}, P)$ est un espace probabilisé et V une sous-variété de $\mathbb{R}^p$. Ces vecteurs ξ_i sont issus du générateur G. Grâce à sa segmentation manuelle, le radiologue associe à chaque ξ_i une valeur y valant 1, 2 ou 3, selon la loi de probabilité conditionnelle $F(y|\xi)$. La variable y représente le compartiment anatomique du voxel x représenté par ξ (par exemple, 1 pour les cavités, 2 pour la médullaire, 3 pour le cortex). On pose $\mathbf{z} = (\xi, y)$ et $P_{\mathbf{Z}}(\mathbf{z}) = P_{\mathbf{X},Y}(\xi, y)$. On a pu construire, à l'étape précédente, un ensemble de M classificateurs correspondant aux fonctions $f(\xi, \alpha)$, $1 \leq \alpha \leq M$ à valeurs dans $\{1, 2, 3\}$ et associant à chaque vecteur ξ_i la classe du voxel x_i qu'il représente, tout comme le radiologue.

Il s'agit de choisir, dans l'ensemble $\{f(\xi, \alpha),\ 1 \leq \alpha \leq M\}$, la fonction f qui est la meilleure estimation de la réponse du radiologue. La sélection de cette fonction se fait à partir des performances réalisées par les classificateurs sur les voxels de la coupe rénale principale. On peut se placer dans le cadre d'un problème de reconnaissance de forme, avec un coût d'erreur constant qui ne prend que les valeurs 0 ou 1, la fonction de perte L étant alors définie par :

$$L(y, f) = \begin{cases} 0 & \text{si } y = f, \\ 1 & \text{si } y \neq f. \end{cases} \tag{5.2}$$

Posons $Q(\mathbf{z}, \alpha) = L[y, f(\xi, \alpha)]$, on souhaite minimiser le risque moyen, qui est aussi la

probabilité d'erreur de classement :

$$R(\alpha) = \int Q(\mathbf{z}, \alpha) \mathrm{d}P_{\mathbf{Z}}(\mathbf{z}), \tag{5.3}$$

où la loi $P_{\mathbf{Z}}(\mathbf{z})$ est inconnue mais où un échantillon d'observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$ est donné.

Le classificateur choisi, caractérisé par le paramètre $\widehat{\alpha}_N$, est celui qui minimise le risque empirique moyen d'ordre N (estimateur du minimum du risque empirique ou MRE) :

$$\widehat{R}_N(\alpha) = \frac{1}{N} \sum_{i=1}^{N} Q(\mathbf{z}_i, \alpha). \tag{5.4}$$

Le risque moyen d'ordre N est aussi égal à $1 - \mathrm{WCP}$ évalué sur la coupe rénale principale.

En appliquant les résultats du chapitre 3.2, on obtient que l'estimateur MRE est consistant (cf. paragraphe 3.2.3), c'est-à-dire que le risque d'erreur empirique converge en probabilité vers le risque d'erreur moyen (c'est-à-dire vers la probabilité d'erreur de classement). Soit $\widehat{\alpha}_N$ le paramètre caractérisant la fonction f qui minimise le risque empirique moyen d'ordre N sur les observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$, c'est-à-dire le classificateur minimisant le critère $1 - \mathrm{WCP}$ sur les voxels de la coupe principale.

Faisons l'hypothèse suivante : les ξ des voxels rénaux des autres coupes sont une réalisation d'une suite de variables aléatoires vectorielles $\{\mathbf{X}_{N+1}, \ldots, \mathbf{X}_u, \ldots\}$ indépendantes et identiquement distribuées de même loi $P_{\mathbf{X}}(\xi)$ que pour la coupe principale. Quelle est alors la probabilité d'erreur de classement pour ces voxels ? Est-elle proche du risque empirique moyen d'ordre N ?

Cas général

En se plaçant dans le cas général du paragraphe 3.2.4, on obtient que :

– d'après l'équation (3.95), avec une probabilité au moins égale à $1 - \eta$, la différence entre le risque empirique moyen $R(\widehat{\alpha}_N)$ et le risque empirique $\widehat{R}_N(\widehat{\alpha}_N)$ est bornée par

$$R(\widehat{\alpha}_N) - \widehat{R}_N(\widehat{\alpha}_N) < \frac{\ln M - \ln \eta}{N} \left(1 + \sqrt{1 + 2 \frac{\widehat{R}_N(\widehat{\alpha}_N) N}{\ln M - \ln \eta}} \right), \tag{5.5}$$

– d'après l'équation (3.96), avec une probabilité au moins égale à $1 - 2\eta$, la différence entre le risque moyen obtenu avec le paramètre $\widehat{\alpha}_N$ qui minimise le risque empirique moyen et le risque moyen minimum $R(\alpha_0)$ est bornée par

$$R(\widehat{\alpha}_N) - R(\alpha_0) < \sqrt{\frac{-\ln \eta}{2N}} + \frac{\ln M - \ln \eta}{N} \left(1 + \sqrt{1 + 2 \frac{\widehat{R}_N(\widehat{\alpha}_N) N}{\ln M - \ln \eta}} \right). \tag{5.6}$$

Cas optimiste

Remarquons que, pour des données simulées, on peut atteindre un risque empirique moyen nul avec un ou plusieurs classificateurs, donc se placer dans le cas optimiste du

paragraphe 3.2.4 et resserrer les bornes. On obtient alors que, pour l'ensemble des paramètres $\widehat{\alpha}_N$ qui annulent le risque empirique moyen :

- avec une probabilité au moins égale à $1-\eta$, la différence entre le risque empirique moyen $R(\widehat{\alpha}_N)$ et le risque empirique $\widehat{R}_N(\widehat{\alpha}_N)$ est bornée par

$$R(\widehat{\alpha}_N) \leq \frac{\ln(M-1) - \ln \eta}{N},\ M > 1, \tag{5.7}$$

- avec une probabilité au moins égale à $1-\eta$, la différence entre le risque moyen obtenu avec le paramètre $\widehat{\alpha}_N$ qui minimise le risque empirique moyen et le risque moyen minimum $R(\alpha_0)$ est bornée par

$$R(\widehat{\alpha}_N) - R(\alpha_0) < \frac{\ln(M-1) - \ln \eta}{N}. \tag{5.8}$$

Examinons les bornes de risques que nous sommes susceptibles d'atteindre dans notre cas.

Bornes du risque

D'après respectivement les simulations et les tests sur données réelles, en observant les valeurs obtenues pour le critère $1-$WCP, on peut supposer des risques empiriques d'ordre N variant entre 0% et 15% pour les données simulées, et entre 5% et 25% pour les données réelles. Le nombre de voxels dans la coupe principale varie entre $N = 700$ et $N = 2500$ voxels.

Nous avons choisi de faire varier η entre $0,05$ et $0,2$.

Des exemples de bornes sont tracés sur les figures 5.26 et 5.27 pour $M = 5$, ce nombre étant en accord avec le nombre de classificateurs réellement construits.

Pour $N = 1500$ voxels sur la coupe principale (taille typique), un risque empirique moyen de 15%, avec une probabilité d'au moins 95%, on obtient une probabilité d'erreur de classement maximale de 19%. Baissant de $19,8$ et $17,5$ pour N allant de 800 à 2500 voxels. Comparé aux pourcentages d'erreur obtenus entre deux radiologues, cela peut être considéré comme suffisant.

5.2.2 Cas réel

Dans une démarche plus réaliste, le radiologue ne fournit pas de segmentation manuelle de référence, mais peut choisir le classificateur qui donne le résultat le plus proche de la segmentation qu'il aurait réalisée, c'est-à-dire celui qui minimise le risque empirique d'ordre N, où N est le nombre de voxels du masque rénal de la coupe principale. On peut estimer grossièrement qu'elle est en accord à X% avec la meilleure segmentation fonctionnelle. Notons que les segmentations manuelles de la coupe principale ayant servi aux évaluations des critères de similarité ont été réalisées sans que le radiologue soit influencé par les segmentations fonctionnelles obtenues : lorsqu'on leur montre ces dernières, il est fréquent que certaines leur paraissent au moins aussi satisfaisantes que la segmentation manuelle correspondante. On peut considérer que le $1-$WCP est en réalité inférieur à celui qui est obtenu avec la méthode de validation proposée au paragraphe 4.4.

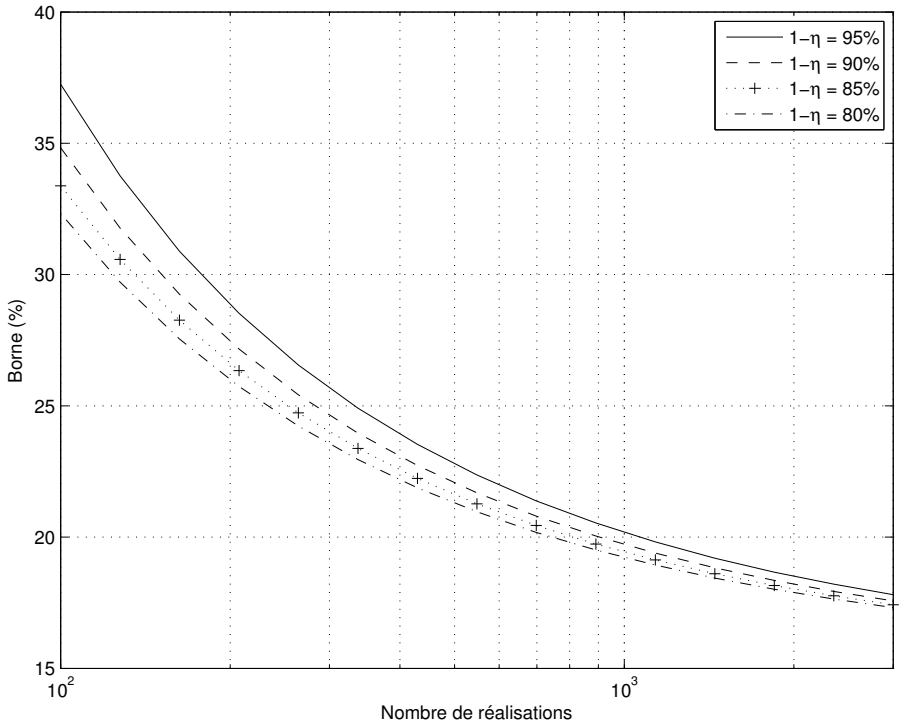


FIG. 5.26 – Borne supérieure du risque moyen $R(\widehat{\alpha}_N)$ en fonction de N dans le cas général pour différentes probabilités $1-\eta$ avec $M=5$ fonctions et $\widehat{R}_N(\widehat{\alpha}_N)=0,15$: estimation pour les données de la base de test (équation (5.5)).

Dans ce cas, on peut même considérer qu'on s'intéresse aux performances d'un unique classificateur, et non de la famille complète des classificateurs, et appliquer les résultats sur les bornes de la différence entre le risque empirique moyen d'ordre N et le risque moyen avec $M=1$. Les résultats correspondants sont donnés sur la figure 5.28. Si le risque empirique est de 15%, on constate que la probabilité d'erreur de classement reste inférieure à 20% avec une probabilité d'au moins 0,85 pour un nombre de voxels N supérieur à 500 (rein de bébé) et à 18% pour $N \leq 1500$ (rein d'adulte). Ceci nous paraît tout-à-fait satisfaisant pour notre application.

5.2.3 Remarques communes

Notons bien que le fait qu'on ait construit les classificateurs à partir des données ξ_i de la coupe principale n'intervient pas ici, il s'agit seulement de mesurer les performances des différents classificateurs sur ces données.

Nous avons supposé que la loi de probabilité est commune aux ξ de toutes les coupes. Si la résolution spatiale est trop faible (c'est-à-dire si les voxels sont trop volumineux), l'effet de volume partiel risque d'être très important sur les voxels en périphérie du rein, et l'hypothèse précédente ne sera peut-être pas très bien vérifiée.

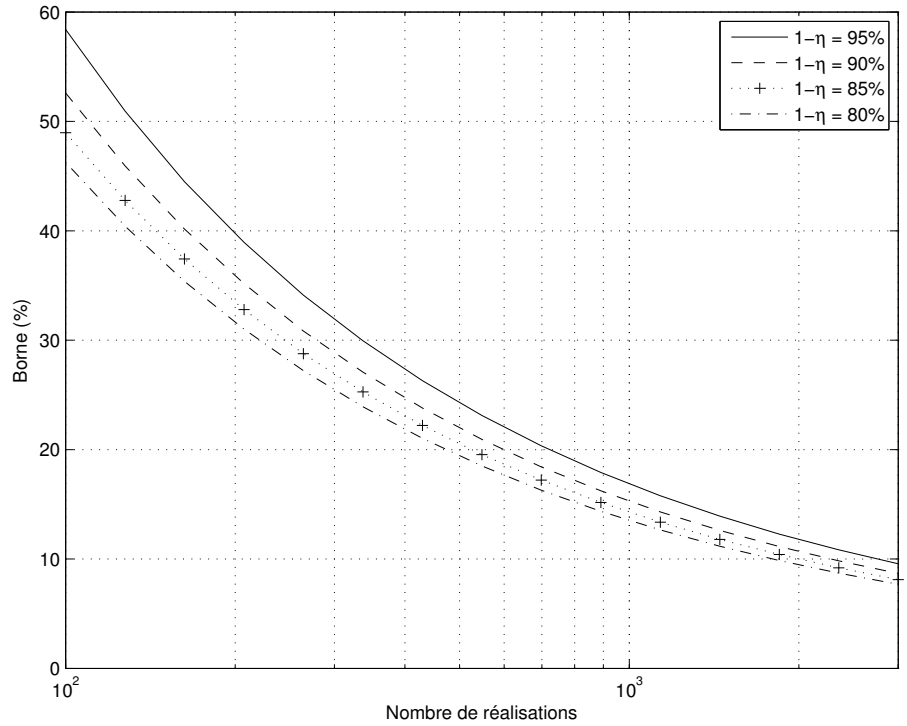


FIG. 5.27 – Borne supérieure de $R(\widehat{\alpha}_N) - R(\alpha_0)$ en fonction de N dans le cas général pour différentes probabilités $1 - \eta$ avec $M = 5$ fonctions et $\widehat{R}_N(\widehat{\alpha}_N) = 0,15$: estimation pour les données de la base de test (équation (5.6)).

5.2.4 Hypothèse d'indépendance des données

Les bornes du risque sont calculées sous l'hypothèse que les ξ_i sont des réalisations de variables aléatoires indépendantes. Cette hypothèse intervient lors de l'étude de la rapidité de convergence, lors de l'application des bornes de Chernoff (équation (3.71)).

Il est difficile de mesurer l'indépendance statistique de variables aléatoires dans un espace de haute dimension. L'évaluation de leur information mutuelle n'est pas envisageable car elle nécessite l'estimation de leur densité de probabilité jointe.

La décorrélation n'implique pas l'indépendance, puisqu'il suffit d'appliquer aux données une transformation comme celle de Karhunen-Loève pour les décorréler sans modifier leur degré d'indépendance. Cependant, des processus physiques décorrélés sont généralement indépendants [89]. Ainsi, nous nous proposons d'évaluer la corrélation entre les données.

Estimation pessimiste

Nous nous plaçons dans le cas le plus défavorable, c'est-à-dire que nous estimons la covariance entre les réalisations issues de pixels voisins, puis distants de m pixels sur une même colonne ou sur une même ligne pour m variant de 1 à m_0. En d'autres termes, nous classons les ξ_i en balayant l'image par colonnes ou par lignes et nous estimons les matrices

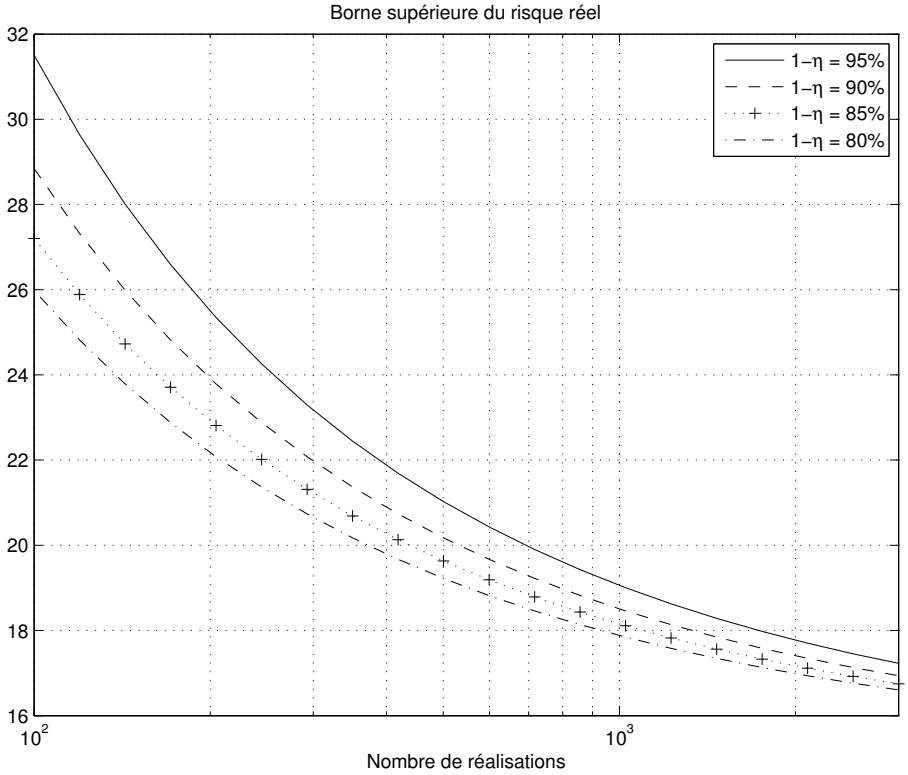


FIG. 5.28 – Borne supérieure du risque moyen $R(\widehat{\alpha}_N)$ en fonction de N dans le cas général pour différentes probabilités $1-\eta$ avec $M=1$ fonction et $\widehat{R}_N(\widehat{\alpha}_N)=0,15$: estimation pour le cas réel (équation (5.5)).

de corrélation $\{\mathbf{M}(m)\}_{1\leq m\leq m_0}$

$$\mathbf{M}(m)=\mathrm{E}[\xi_{i+m}\xi_i^T]-\mathrm{E}[\xi_{i+m}]\mathrm{E}[\xi_i^T], \tag{5.9}$$

en supposant l'ergodicité à l'ordre 2 du signal $(\xi_i)_{1\leq i\leq N}$:

$$\widehat{\mathbf{M}}(m)=\frac{1}{N}\sum_{i=1}^{N-|m|}\xi_{i+|m|}\xi_i^T-\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i\right)^T. \tag{5.10}$$

Nous négligeons ici respectivement l'effet des sauts de colonnes et celui des sauts de lignes. Nous calculons alors à chaque instant j compris entre 1 et p les coefficients $C_j(m)$ définis par

$$C_j(m)=\frac{|\widehat{\mathbf{M}}_{jj}(m)|}{\widehat{M}_{jj}(0)}. \tag{5.11}$$

Si les données sont décorrélées et si l'hypothèse d'ergodicité est satisfaite, les coefficients $C_j(m)$ doivent tendre vers 0. Si elles sont fortement corrélées, ils restent proches de 1. Les résultats obtenus sont présentés sur la figure 5.29. Les données apparaissent comme étant assez fortement corrélées puisqu'il faut un écart de $m=4$ pour que le coefficient $C_j(m)$ atteigne la moitié de sa valeur maximale.

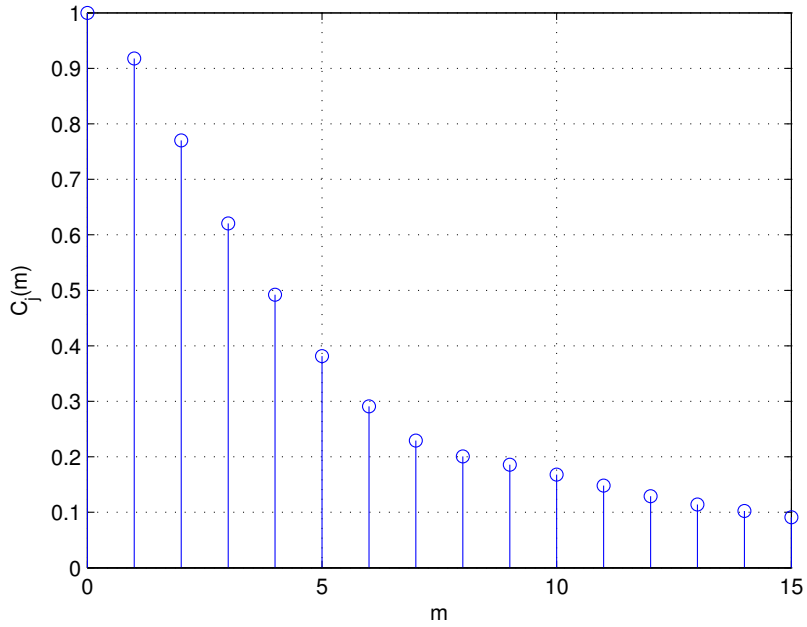


FIG. 5.29 – Moyennes des coefficients $C_j(m)$ sur l'ensemble des p composantes temporelles (estimation pessimiste).

Cette méthode est cependant critiquable parce qu'elle suppose la stationnarité des ξ_i. Nous pouvons en effet considérer, pour simplifier, que la loi de probabilité $P_{\mathbf{X}}(\xi)$ contient trois modes, chacun correspondant à un compartiment donné. Le mode de balayage de l'image colonne par colonne ou ligne par ligne implique que l'hypothèse de stationnarité n'est pas vérifiée puisque, dans une large mesure, des ξ_i successifs sont issus du même mode, correspondant au même compartiment rénal. Les coefficients de covariance sont donc surestimés.

Estimation optimiste

Afin de mieux respecter les hypothèses de stationnarité et d'indépendance, les données ξ_i sont en réalité présentées en balayant le masque rénal de manière aléatoire. En évaluant les coefficients $C_j(m)$ de la même manière que précédemment mais avec un balayage aléatoire, nous obtenons les résultats présentés sur la figure 5.30.

Les données sont alors largement décorrélées puisque les coefficients C_j ne valent que 0,03 même pour $k = 1$.

Conclusion

Il est difficile d'évaluer la corrélation des données dans le cas d'un signal non stationnaire à partir d'une seule réalisation $(\xi_i)_{1\leq i\leq N}$. Les données ne peuvent peut-être pas être considérées comme totalement indépendantes, compte tenu du post-traitement plus ou moins équivalent à un moyennage spatial local. Si la première méthode d'évaluation de la corrélation surestime les coefficients, la seconde se rapproche davantage de ce que « voit »

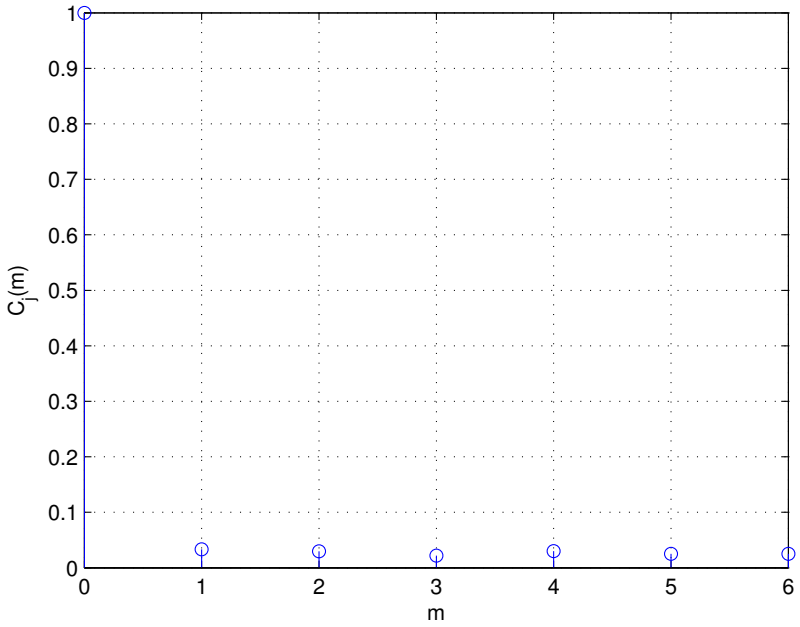


FIG. 5.30 – Moyennes des coefficients $C_j(m)$ sur l'ensemble des p composantes temporelles (estimation optimiste).

l'algorithme d'apprentissage. La réalité est certainement située entre les deux. L'idéal serait de disposer de plusieurs examens du même rein, réalisés dans les mêmes conditions. Même si nous sommes dans l'impossibilité de vérifier l'hypothèse d'indépendance, nous allons évaluer les erreurs de classement dans quelques cas d'extension et les comparer aux valeurs théoriques des bornes obtenues.

5.2.5 Résultats pour l'extension aux autres coupes

Données simulées

Nous avons utilisé le modèle rénal de la coupe principale (1550 voxels) et généré de nouveaux vecteurs ξ en suivant la méthode explicitée au paragraphe 4.1.1. Nous avons calculé le risque empirique moyen sur les voxels de cette coupe pour des classificateurs à base de K-moyennes ($K = 3$ et $K = 4$), puis pour un classificateur à base de GNG-T. Nous avons ensuite utilisé ce même classificateur pour classer 1500 à 10000 voxels supplémentaires, générés de façon similaire. Nous obtenons, pour n'importe quel classificateur, des erreurs de classement (évaluées sur l'ensemble des voxels classés) qui s'éloignent de moins de 1% du risque empirique obtenu sur la coupe d'origine : ce résultat est très satisfaisant et reste inférieur aux bornes théoriques pour $1 - \eta = 95\%$.

Données réelles

Nous disposons de deux reins avec deux coupes différentes. Pour le premier rein, nous disposons d'une segmentation manuelle par coupe. Pour le second, nous avons une segmen-

tation manuelle de la coupe principale et deux segmentations manuelles pour la seconde coupe, réalisées par le même opérateur mais à deux moments différents.

Afin de tester la validité de l'extension, nous évaluons tout d'abord le risque empirique moyen d'un classificateur donné sur la coupe principale en calculant la part de voxels mal classés par rapport à la segmentation manuelle, c'est-à-dire $1 - \text{WCP}$. Nous utilisons alors ce même classificateur pour étiqueter les voxels de la seconde coupe. Nous nous plaçons ainsi dans le cas où $M = 1$. Le risque d'erreur est estimé par le critère $1 - \text{WCP}$ évalué cette fois sur l'ensemble des voxels classés. Les résultats obtenus sont présentés dans le tableau 5.7. Les tests 1 et 2, réalisés sur le premier rein, correspondent à deux classificateurs différents pour les mêmes segmentations manuelles de référence ; la coupe principale contient 1582 voxels rénaux, la coupe secondaire 1518. Les autres tests sont effectués sur le second rein, les résultats des numéros 3 et 4 (respectivement 5 et 6) sont établis à partir d'un classificateur à K-moyennes (respectivement à GNG-T) sur une même coupe principale (1306 voxels) pour chacune des segmentations manuelles de référence disponible pour la coupe secondaire (1300 voxels).

Test	Coupe principale	Ensemble des coupes	Borne sup. du risque réel
1	19,9	21,5	23,5
2	15,0	16,4	18,2
3	23,1	23,0	27,0
4	23,1	23,7	27,0
5	24,4	23,8	28,3
6	24,4	24,7	28,3

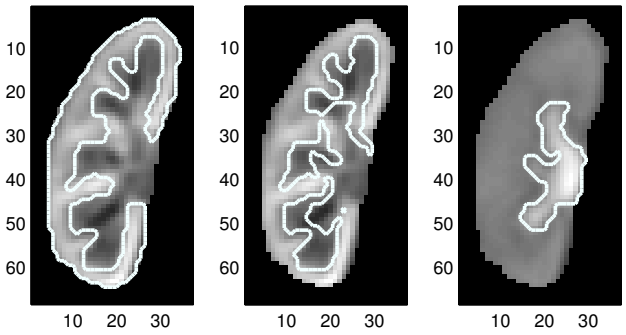
TAB. 5.7 – Comparaisons du pourcentage de voxels mal classés (critère $1 - \text{WCP}$ exprimé en %) sur la coupe principale et sur l'ensemble des coupes pour différents classificateurs, et de la borne supérieure théorique du risque réel pour $1 - \eta = 95\%$.

Nous pouvons constater que, pour tous les cas testés, le pourcentage de voxels mal classés sur l'ensemble des coupes reste inférieur à la borne supérieure théorique du risque réel pour $1 - \eta = 95\%$ et s'éloigne au maximum de 2% du risque empirique obtenu sur la coupe principale. Ainsi, lorsque l'on dispose d'un classificateur permettant d'aboutir à une segmentation semblable à celle réalisée manuellement par un opérateur, on a de bonnes chances que les résultats soient aussi satisfaisants si on utilise ce même classificateur pour segmenter les autres coupes. Des tests supplémentaires doivent bien sûr être menés sur d'autres cas.

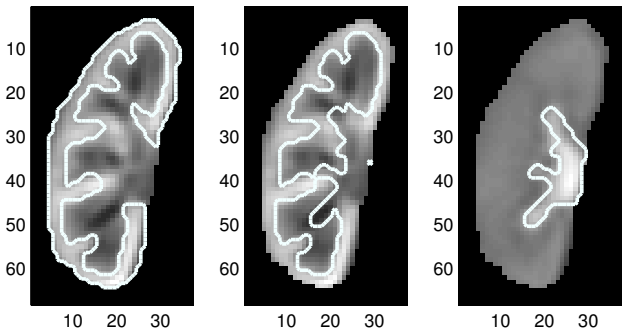
5.3 Conclusion

Les résultats obtenus par généralisation, que ce soit sur des données simulées ou sur les deux acquisitions à deux coupes dont nous disposons sont excellents, puisque les erreurs de classement évaluées sur l'ensemble des voxels classés restent inférieures aux bornes théoriques avec $1 - \eta = 95\%$. Il reste à réaliser des tests supplémentaires sur les données réelles. L'hypothèse d'indépendance, qui est essentielle pour l'utilisation des bornes de Chernoff, est cependant critiquable mais sa validité reste difficile à évaluer. Notons que les

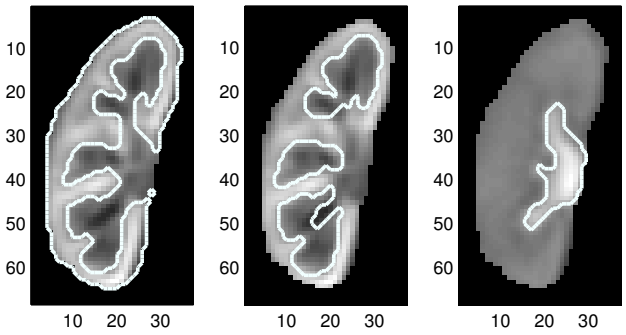
bornes calculées dans le cadre théorique sont larges ; il est donc assez rassurant d'obtenir des valeurs assez proches du risque empirique pour le nombre d'échantillons que nous avons.



(a) Segmentation manuelle anatomique (OP1)



(b) Segmentation manuelle anatomique (OP2)



(c) Segmentation fonctionnelle semi-automatique

FIG. 5.31 – Exemple de segmentation du cortex (à gauche), de la médullaire (au milieu) et des cavités (à droite).

Chapitre 6

Conclusion

Le travail présenté dans cette thèse entre dans le cadre d'un projet national intitulé « Place de l'uro-IRM dans l'évaluation des conséquences fonctionnelles de l'obstruction urinaire de l'enfant et de l'adulte », qui étudie la possibilité de remplacer les examens scintigraphiques, actuellement considérés comme une référence, par des IRM de perfusion rénale à rehaussement de contraste. Cette étude nécessite des tests sur une population nombreuse et variée. Pour rendre les séquences acquises exploitables en vue d'évaluer des paramètres caractéristiques de la fonction rénale, un très long travail de recalage et de segmentation de la part de radiologues est nécessaire, si ces opérations sont effectuées « manuellement ».

Nous proposons un outil destiné à faciliter ces tâches en les automatisant, entièrement pour le recalage, en partie pour la segmentation des structures rénales internes (cortex, médullaire et cavités).

L'une des principales difficultés auxquelles nous nous sommes heurtés est l'absence de vérité-terrain pour valider nos résultats. Nous avons donc développé un modèle rénal suffisamment réaliste permettant de générer des séquences d'images de synthèse d'une coupe rénale pour tester les méthodes proposées. Ce modèle, d'une grande souplesse, permet de mettre en évidence l'influence de différents phénomènes (niveau de bruit réglable, volume partiel plus ou moins important, post-traitement constructeur, erreurs de recalage résiduelles de différentes amplitudes) sur les segmentations fonctionnelles : à notre connaissance, ceci n'a jamais été fait auparavant. Nous avons adapté le modèle pour pouvoir simuler des reins sains ou des reins pathologiques, en définissant une anatomie et des évolutions temporelles par compartiment particulières pour chacun de ces deux cas. Pour tenir compte du volume partiel, chaque voxel est constitué d'un mélange de cortex, médullaire et cavités, et sa courbe temps-intensité est un mélange pondéré des signaux caractéristiques de chaque compartiment.

Le modèle proposé permet de générer des images tests pour le recalage mais aussi d'étudier la robustesse de nos méthodes de segmentation aux erreurs de recalage résiduelles, et de mettre en évidence leurs forces et leurs faiblesses. Il ne prétend pas cependant recouvrir toutes les situations rencontrées dans la réalité. Des essais sur données réelles sont indispensables.

En ce qui concerne le recalage, nous n'avons pas développé de méthode vraiment nouvelle. Nous avons effectué notre choix après une recherche bibliographique critique, en tenant compte de la spécificité de nos données : il s'agit de recaler une séquence complète avec de forts changements de contraste et pouvant contenir quelques images fortement dégradées ; certaines structures apparaissent et disparaissent au cours de la perfusion. Nous nous sommes imposé la contrainte supplémentaire suivante, qui est assez rare dans la littérature concernant l'IRM rénale : nous souhaitons ne pas avoir à extraire un masque rénal avant le recalage. Même si, actuellement, la délimitation des contours du rein est nécessaire pour la segmentation des compartiments rénaux, il se peut qu'à terme, on puisse extraire automatiquement le rein par une technique semblable appliquée sur la séquence recalée.

Remarquons d'abord que les mouvements hors plan de coupe ne sont pas corrigés, c'est-à-dire que l'on effectue le recalage coupe par coupe. Les voxels ont une taille trop importante dans la direction perpendiculaire au plan de coupe pour que nous puissions espérer faire de telles corrections, mais il se peut que des acquisitions avec un autre type de séquence IRM puisse les autoriser. En revanche, avec l'angle de coupe choisi, on peut supposer que les mouvements des coupes voisines seront approximativement les mêmes que ceux de la coupe principale ; les quelques essais que nous avons effectués laissent supposer qu'appliquer les transformations obtenues pour cette dernière aux autres coupes pourrait donner des résultats suffisamment satisfaisants.

Pour les mouvements dans le plan de coupe (essentiellement mouvement respiratoire), nous avons choisi de recaler toutes les images sur une référence choisie au pic cortical, sur laquelle on voit le mieux les structures internes. Nous avons éliminé les méthodes où l'on recale deux images successives et où l'on compose ensuite les transformations en raison des risques importants de dérive rencontrés : il suffit qu'une seule image soit mal recalée pour que le traitement de toutes les suivantes soit compromis.

Nous avons réalisé un recalage subpixel en autorisant les transformations rigides uniquement (rotations et translations), la mesure de ressemblance étant l'information mutuelle, pour tenir compte des changements de contraste entre images issues de phases de perfusion différentes.

La méthode retenue a été testée sur des données de synthèse et sur des données réelles. Les déplacements sont convenablement corrigés, mais on ne tient pas compte d'éventuelles déformations au cours de l'acquisition. Celles-ci sont relativement peu importantes sur les cas traités. Par ailleurs, il nous semble pour l'instant difficilement possible de distinguer les changements de contraste dus à la perfusion de ceux induits par un mouvement non rigide. La question de la validation autre que par « satisfaction visuelle » demeure ouverte. Les résultats sont similaires à ceux obtenus avec la méthode proposée dans [41].

Les erreurs de recalage résiduelles diminuent certes les performances de notre méthode de segmentation, comme le montrent nos tests sur les données de synthèse : le problème le plus délicat est la finesse de la médullaire dans les reins pathologiques. Ces erreurs, difficilement évaluables sur les données réelles (reins sains), sont cependant suffisamment faibles pour que les segmentations obtenues pour celles-ci restent convenables.

Les méthodes de segmentation que nous proposons utilisent toutes les courbes temps-intensité des voxels rénaux pour les classer en différentes catégories. Cette manière de

procéder est présente dans la littérature : le principal algorithme utilisé est celui des K-moyennes, mais on ne trouve que fort peu de détails et de résultats. Nous avons développé de nombreuses nouveautés et approfondi l'analyse des résultats.

Tout d'abord, nous avons défini de manière précise la différence entre segmentation anatomique et segmentation fonctionnelle sur les données de synthèse ; nous avons montré en particulier que l'on ne peut pas espérer que les segmentations fonctionnelles soient exactement identiques à la segmentation anatomique sous-jacente qui permet de générer les images de synthèse. Même s'il est difficile de quantifier ses conséquences sur les méthodes de segmentation semi-automatiques proposées pour des données réelles, nous avons souligné que ce phénomène influence aussi les segmentations manuelles car elles sont réalisées sur deux images différentes contenant à la fois des informations anatomiques et fonctionnelles.

Nous avons testé plusieurs variantes de la méthode de base, qui utilisent toutes la quantification vectorielle des courbes temps-intensité. Nous avons tenté d'en expliquer les faiblesses et d'y remédier. Les tests sur les données de synthèse se sont avérés être d'une aide précieuse dans ce contexte, puisqu'ils permettent d'établir des hypothèses sur les sources d'erreur que constituent le bruit d'acquisition, l'effet de volume partiel, le post-traitement constructeur et les erreurs résiduelles de recalage. Le lecteur se reportera aux paragraphes concernés pour les valeurs numériques du modèle. Pour le modèle de rein sain, si la quantification vectorielle par K-moyennes à trois classes sur des vecteurs non normalisés permet de classer correctement les 90% de voxels ayant les mêmes compartiments anatomique et fonctionnel dominants en l'absence de mouvement, ce pourcentage chute de 10% s'il subsiste des erreurs de recalage. Nous avons alors proposé différentes variantes : K-moyennes à quatre classes et plus, GNG-T semi-automatique ou K-moyennes sur vecteurs normalisés. La méthode qui consiste en une quantification vectorielle par GNG-T suivie d'un réglage de deux seuils indépendants basés sur des critères physiologiques n'a jamais été appliquée dans ce cadre, à notre connaissance. En utilisant ces variantes, nous parvenons à retrouver le compartiment fonctionnel dominant de quasiment tous les voxels rénaux, même en présence d'un faible mouvement résiduel. Les résultats restent qualitativement semblables pour le modèle de rein pathologiques, mais les performances diminuent légèrement, essentiellement en raison de la finesse de la médullaire du modèle anatomique utilisé.

En ce qui concerne les données réelles, nous n'avons réalisé des tests que sur des reins sains. L'absence de vérité-terrain pour la validation des résultats nous a amenés à utiliser une « pseudo vérité-terrain » de référence, constituée par une segmentation manuelle réalisée par un expert, et à évaluer certains critères de ressemblance entre cette segmentation et celles obtenues par les méthodes proposées. L'un de ces critères, à savoir le pourcentage de pixels bien classés, rend compte de la qualité globale de la segmentation. Les autres critères retenus permettent en revanche une analyse par compartiment, de sorte à ne pas défavoriser les cavités par rapport au cortex et à la médullaire, qui ont une plus grande influence dans le critère global car leur volume est généralement bien

plus important sur des reins sains : une segmentation ne peut en effet être considérée comme satisfaisante que si chacun des trois compartiments est convenablement identifié. Comme il est difficile de fixer *a priori* un seuil sur ces critères pour différencier des segmentations convenables ou inacceptables, nous avons évalué ces mêmes critères entre une segmentation manuelle des mêmes données réalisée par un second expert et la même segmentation de référence. Nous avons ainsi fourni, outre une validation qualitative par inspection visuelle des segmentations obtenues, des résultats quantitatifs rarement exposés dans la littérature.

Ces résultats sont très satisfaisants pour certains algorithmes, malgré le bruit et les erreurs de recalage résiduelles mais, quelle que soit la méthode utilisée, par précaution, il est nécessaire qu'un expert suive le déroulement des opérations, afin d'éviter d'éventuels résultats aberrants. La méthode automatique avec l'algorithme des K-moyennes à trois classes appliqué aux vecteurs temps-intensité normalisés est très bien adaptée à la plupart des données dont nous disposions et la segmentation est obtenue instantanément. Cependant, elle ne permet pas à l'observateur de la modifier directement en cas de résultat médiocre, sauf en augmentant la valeur de K et en réalisant une fusion manuelle des classes obtenues. La méthode semi-automatique avec le GNG-T offre une plus grande souplesse d'utilisation et peut être mieux adaptée aux cas délicats : l'intervention de l'opérateur, qui est amené à régler en temps réel deux seuils indépendants, est suffisamment simple pour être réalisée en une vingtaine de secondes et garantir ainsi un gain de temps très appréciable par rapport à la segmentation manuelle qui dure plusieurs minutes.

Dans le cas d'acquisitions volumiques, on possède un ensemble de coupes prises au même instant, qui permettent d'analyser l'intégralité du rein. Afin de faciliter le travail du radiologue, nous suggérons de segmenter tout d'abord la coupe rénale principale, qui est celle contenant le plus grand volume rénal et où, *a priori*, les structures internes apparaissent le mieux. Une fois que cette segmentation est réalisée, on dispose d'un ou plusieurs classificateurs susceptibles d'être utilisés pour classer les voxels rénaux des autres coupes et donc aboutir à une segmentation volumique. Nous avons montré que, sous certaines hypothèses, la probabilité d'erreur de classement est bornée, et nous avons donné des valeurs numériques pour ses bornes, qui laissent supposer que cette approche est valable et peut aboutir à des résultats tout à fait satisfaisants. Des tests effectués sur deux acquisitions corroborent cette conclusion.

La validation des méthodes proposées nécessite évidemment une étude à grande échelle sur une population plus nombreuse constituée à la fois de reins sains et de reins pathologiques. Pour ces derniers, en raison d'une plus grande variété de comportements temporels, il faudra tester les adaptations proposées suite aux essais sur les données simulées : on pourra par exemple ajouter une ou deux classes pour l'algorithme des K-moyennes ; suite à la quantification vectorielle par GNG-T, on pensera à inverser le critère sur le seuil des cavités pour un rein obstrué, voire à ajouter un seuil supplémentaire dans le cas d'un remplissage non homogène des cavités.

Toutes les données réelles sur lesquelles nous avons réalisé nos tests sont issues de la même machine, avec des caractéristiques comparables (instants d'acquisition par rapport à l'injection, résolution spatiale...). La base de données devra contenir des acquisitions

réalisées avec différentes machines, voire avec des séquences variées. Soulignons cependant que les différentes phases de perfusion doivent être « équitablement » représentées, de manière que les vecteurs temps-intensité permettent de distinguer les différents compartiments.

En ce qui concerne la reproductibilité de la méthode, l'algorithme des K-moyennes peut aboutir à un minimum local et non global de la distorsion ; en faisant tourner plusieurs fois l'algorithme avec des conditions initiales différentes et en conservant les prototypes donnant la distorsion la plus faible, on s'affranchit généralement de cette difficulté, de sorte que, dans notre cas, les classificateurs à base de K-moyennes donnent toujours le même résultat pour un même jeu de données. Pour ceux où la quantification vectorielle est faite par GNG-T, l'initialisation et l'ordre des données interviennent, mais nous avons vérifié que les résultats sont modifiés de manière négligeable par rapport à la variabilité inter-opérateur pour des segmentations manuelles ; il pourrait aussi être intéressant de comparer ceci à la variabilité intra-opérateur. En revanche, nous ne disposons pas de données permettant de vérifier la reproductibilité pour un autre examen du même patient.

L'une des faiblesses de notre méthode est qu'elle nécessite pour l'instant la délimitation manuelle du masque rénal. Un traitement similaire à celui appliqué aux voxels rénaux, mais appliqué à l'ensemble de la série recalée, ne donne pas de résultats vraiment satisfaisants, d'une part parce que les vecteurs temps-intensité des organes voisins peuvent être assez semblables à ceux de certains pixels rénaux, d'autre part parce que, le recalage étant fait sur le rein, le mouvement des autres organes n'est pas corrigé, et peut même être amplifié par le recalage : un nombre non négligeable de pixels extérieurs peuvent donc avoir une évolution temporelle de contraste très chaotique, ce qui rend le classement difficile. Si l'extraction du masque nous semble possible sur la coupe principale, elle nous paraît plus difficile sur les coupes périphériques.

Il est aussi envisageable d'inclure d'autres traitements comme l'élimination de l'effet de volume partiel avec les organes voisins du rein [18], voire peut-être entre les compartiments rénaux eux-mêmes, comme les auteurs le suggèrent en tant que prolongement de l'étude dans la même publication.

Rappelons que l'objectif final est d'étudier la possibilité de remplacer les scintigraphies par des IRM en comparant des paramètres fonctionnels comme le GFR et la fonction rénale différentielle évalués à partir des deux types d'examens. Il nous paraît donc indispensable de calculer ces mêmes paramètres à partir des segmentations obtenues par les méthodes semi-automatiques proposées et de les comparer aux valeurs obtenues à partir des segmentations manuelles établies par des radiologues.

Enfin, la validation d'une séparation du cortex et de la médullaire ouvre de nouvelles perspectives par rapport à la scintigraphie, en particulier pour le développement de nouveaux modèles destinés à l'évaluation de la fonction rénale et tenant compte de la distinction entre ces deux compartiments, et dont les tests et l'exploitation seraient ainsi possibles à grande échelle.

Annexe A

Résultats pour le recalage

A.1 Quelques résultats sur les données de synthèse

Si l'on applique une transformation connue à une image et qu'on la recale sur l'image avant transformation, la translation est retrouvée au 20-ième de pixel près.

Dans les tests qui suivent, l'image de référence correspond au pic cortical, elle est la 45$^{\text{ième}}$ d'une série générée grâce au modèle de synthèse (cf. les courbes temps-intensité de la figure 4.4).

A.1.1 Tests en transformant l'image de référence

- Translation seule : figure A.1.
- Rotation seule[1] : figure A.2.
- Translation et rotation : figure A.3.

A.1.2 Tests en transformant une autre image de la série proche du pic cortical

L'image à recaler est située au début de la phase de filtration, il s'agit de la 58$^{\text{ième}}$ d'une série générée grâce au modèle de synthèse.

- Translation seule : figure A.4.
- Rotation seule : figure A.5.
- Translation et rotation : figure A.6.

A.1.3 Tests en transformant une autre image de la série en phase de perfusion tardive

L'image à recaler est située en fin de phase de perfusion tardive, il s'agit de la 256$^{\text{ième}}$ d'une série générée grâce au modèle de synthèse.

[1] Le centre des rotations et de l'image initiale seront toujours confondus, ce qui ne restreint pas les possibilités : toute transformation composée d'une translation et d'une rotation de centre quelconque peut se ramener à une rotation dont le centre est confondu avec celui de l'image combinée à une seconde translation.

– Translation seule : figure A.7.
– Rotation seule : figure A.8.
– Translation et rotation : figure A.9.

A.2 Quelques résultats sur les données réelles

Voir les figures A.10 et A.11.

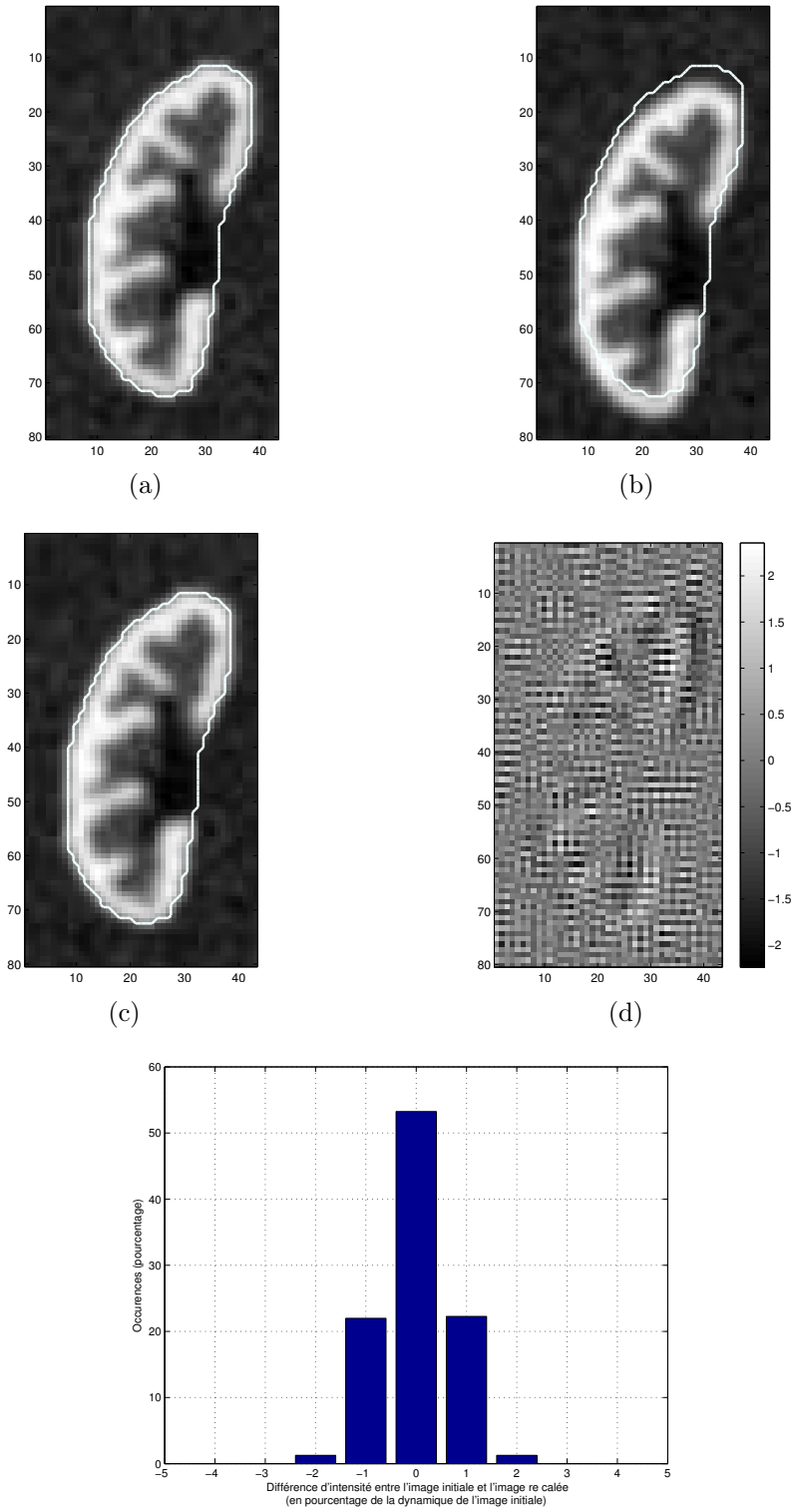


FIG. A.1 – Exemple de recalage sur données de synthèse : image de référence (a) transformée par une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (b) puis recalée (c) ; différence (d) entre (a) et (c) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (a)) et histogramme correspondant (e).

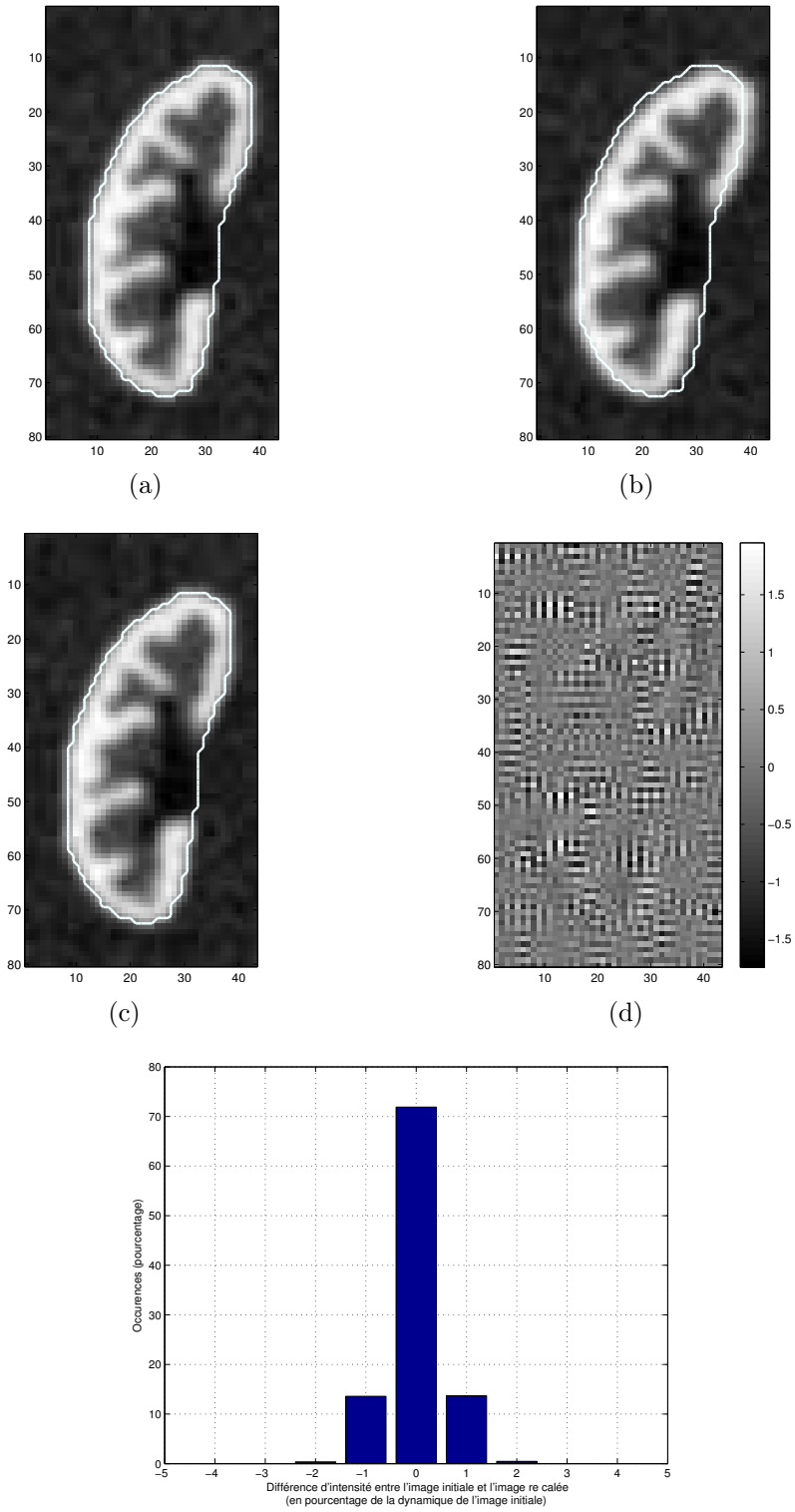


FIG. A.2 – Exemple de recalage sur données de synthèse : image de référence (a) transformée par une rotation de 5° (b) puis recalée (c) ; différence (d) entre (a) et (c) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (a)) et histogramme correspondant (e).

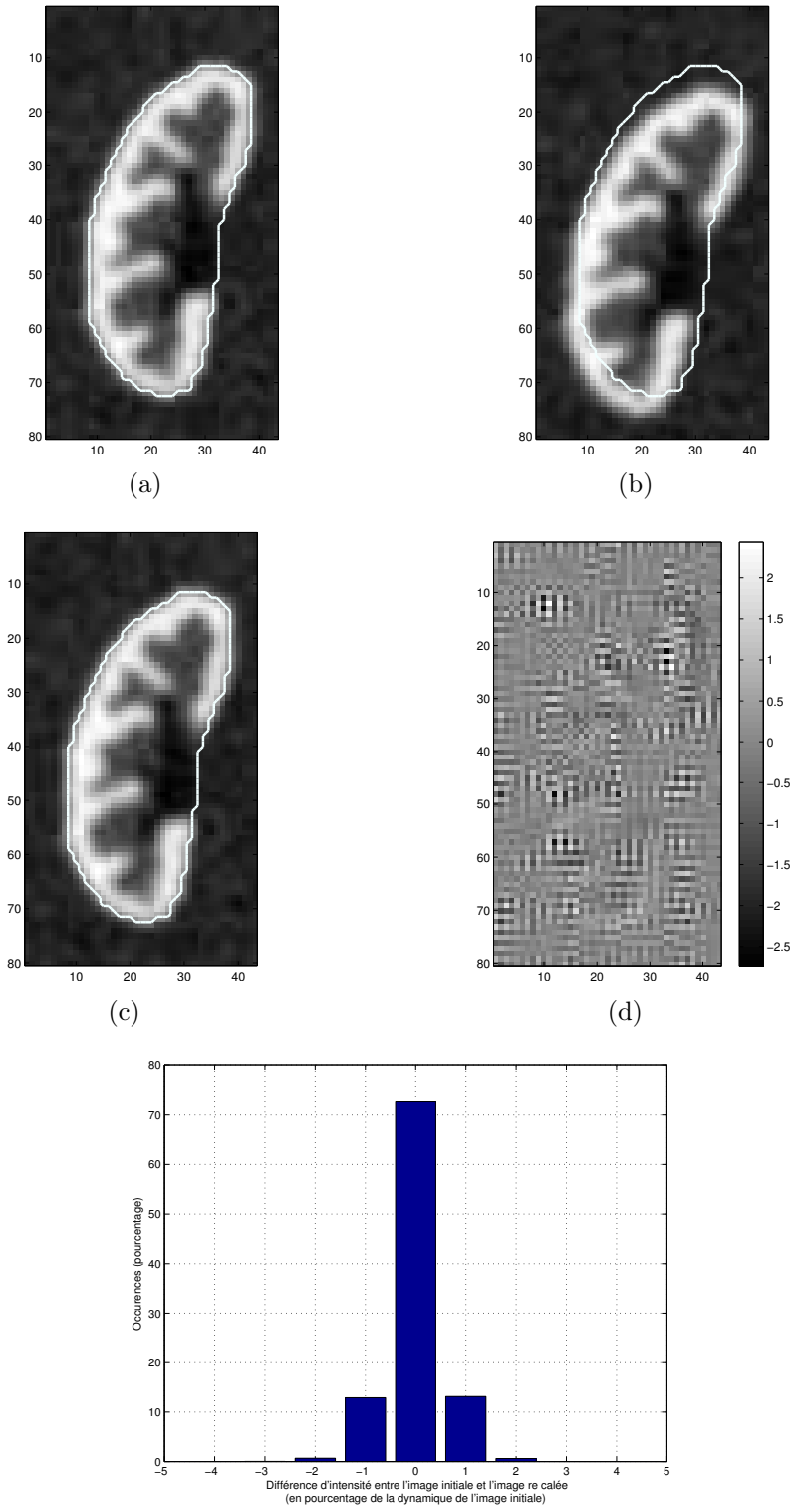


FIG. A.3 – Exemple de recalage sur données de synthèse : image de référence (a) transformée par une rotation de 5° et une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (b) puis recalée (c) ; différence (d) entre (a) et (c) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (a)) et histogramme correspondant (e).

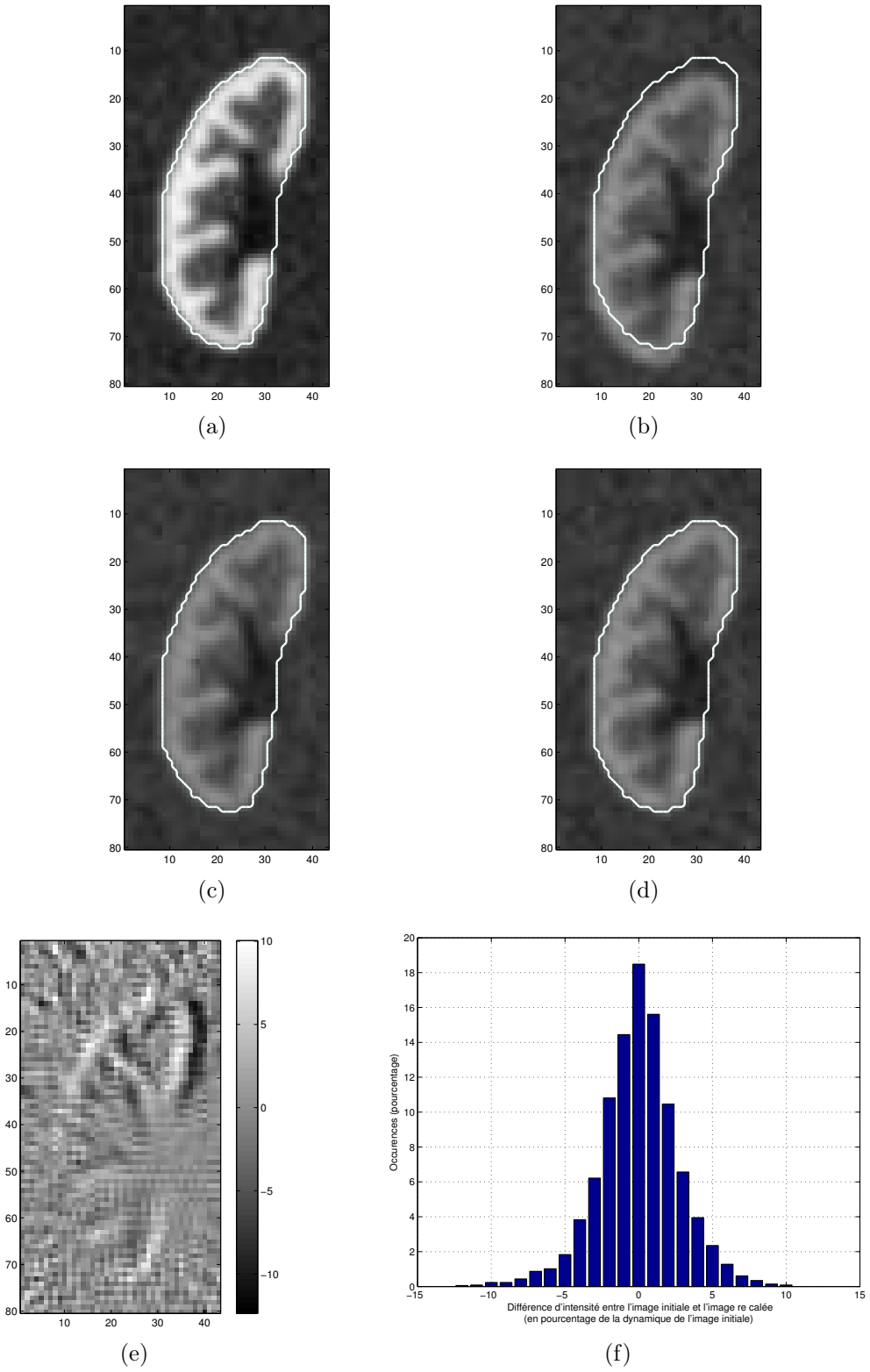


FIG. A.4 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)) et histogramme correspondant (f).

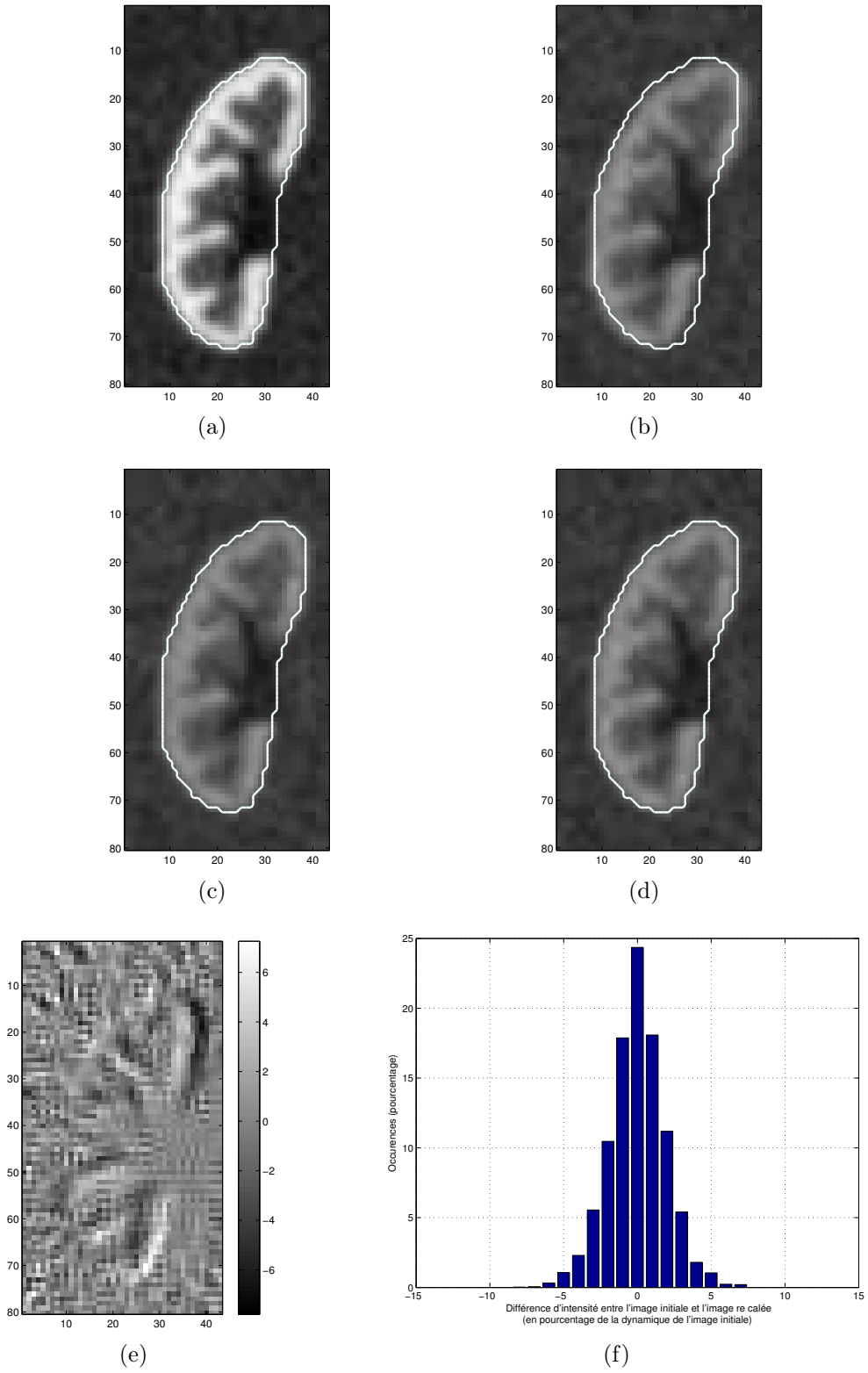


FIG. A.5 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une rotation de 5° (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)).

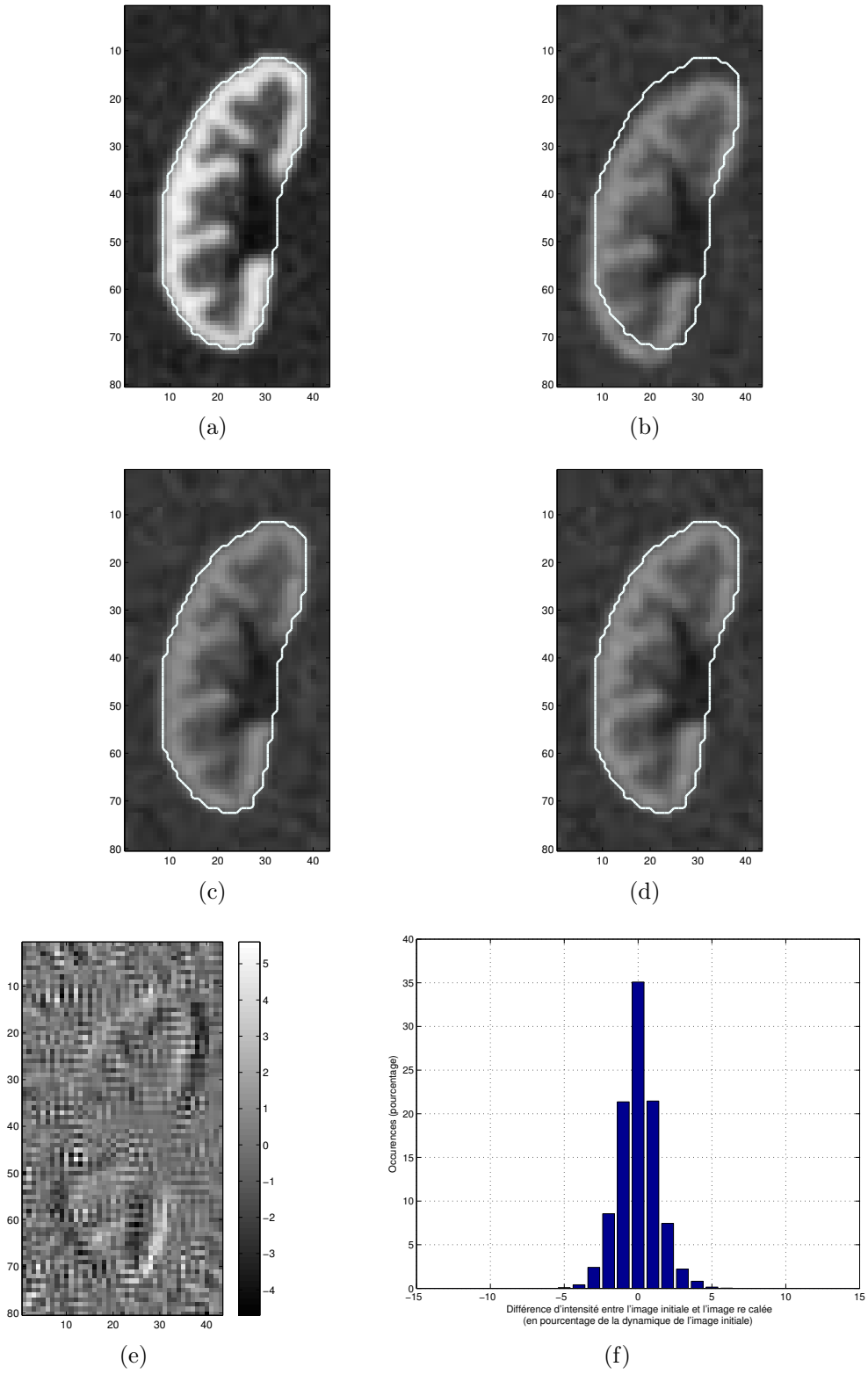


FIG. A.6 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une rotation de 5° et une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)).

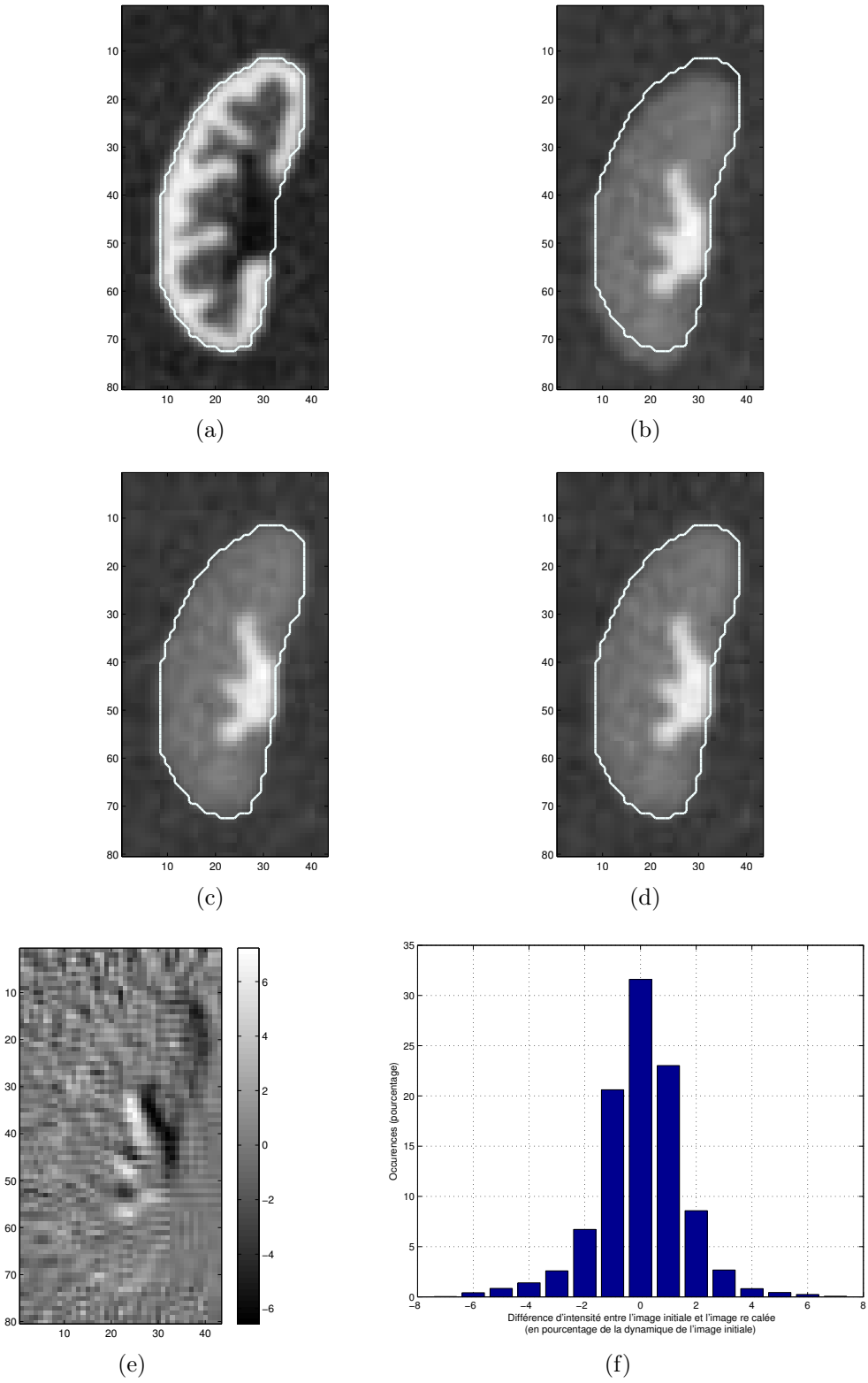


FIG. A.7 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)) et histogramme correspondant (f).

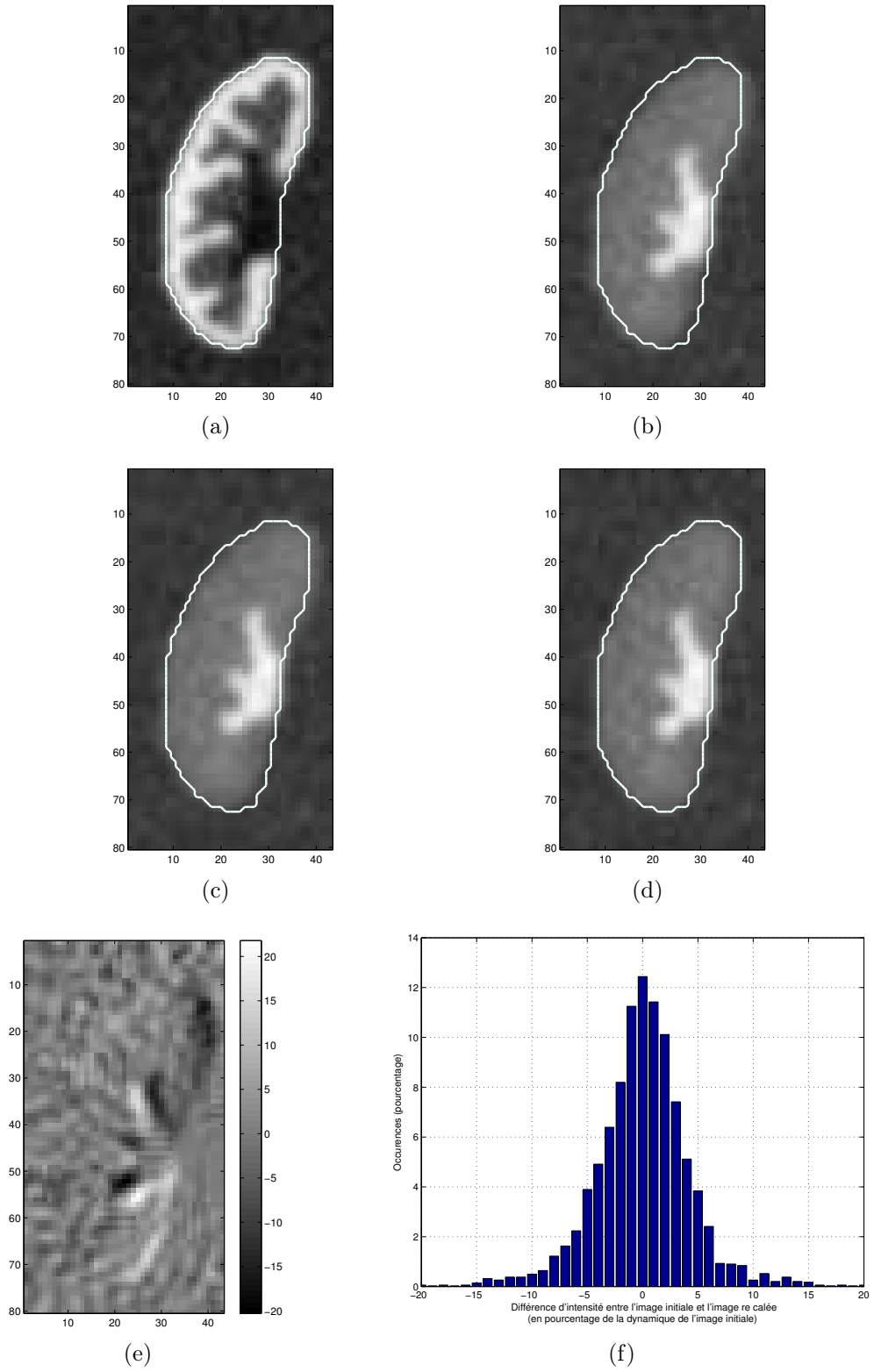


FIG. A.8 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une rotation de 5° (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)).

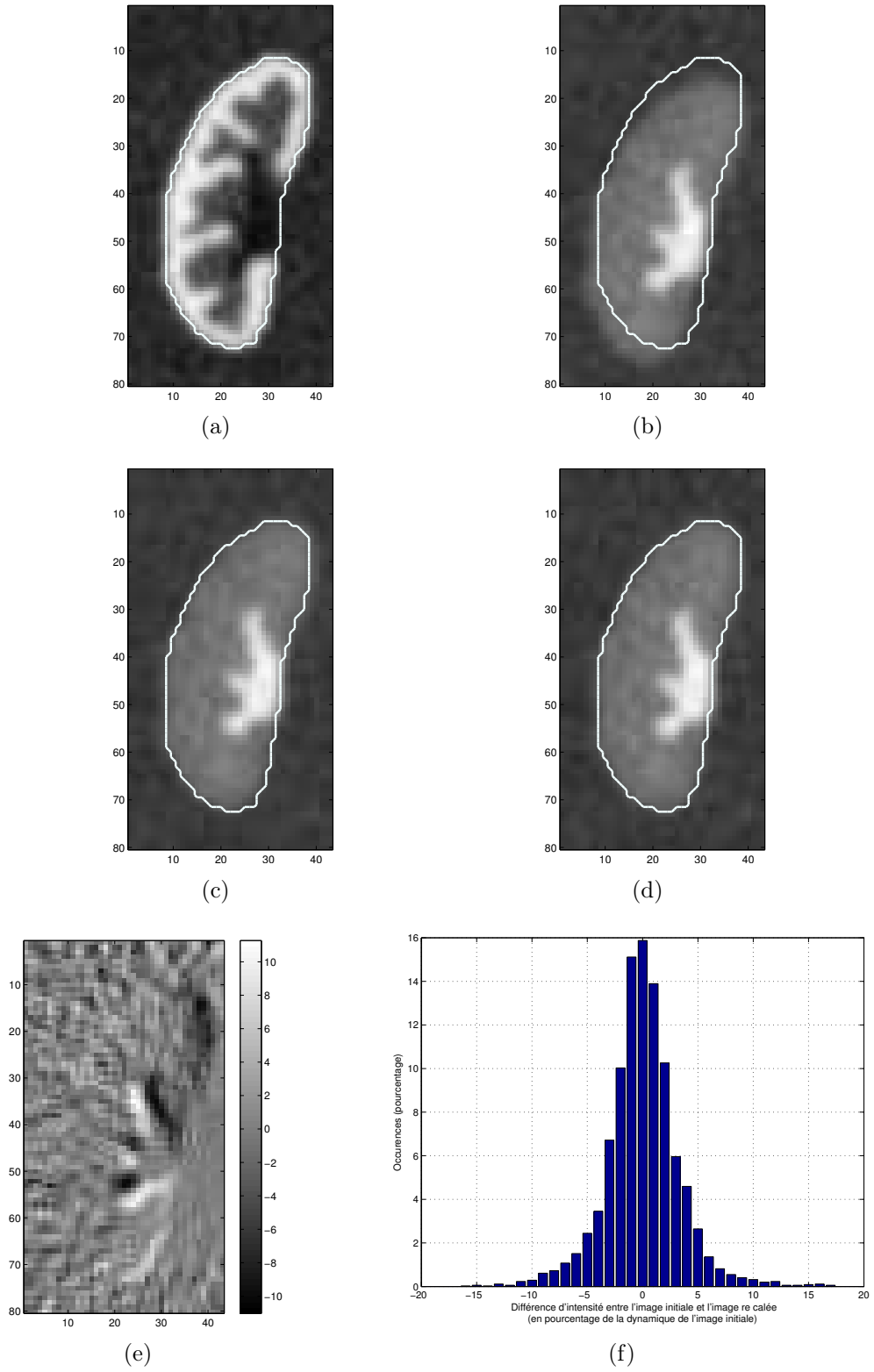


FIG. A.9 – Exemple de recalage sur données de synthèse : image de référence (a) et image test (b), transformée par une rotation de 5° et une translation de 3,5 pixel vers le bas et de 0,75 pixel vers la gauche (c) puis recalée (d) ; différence (e) entre (b) et (d) (exprimée en pourcentage de la différence entre les intensités maximale et minimale de (b)).

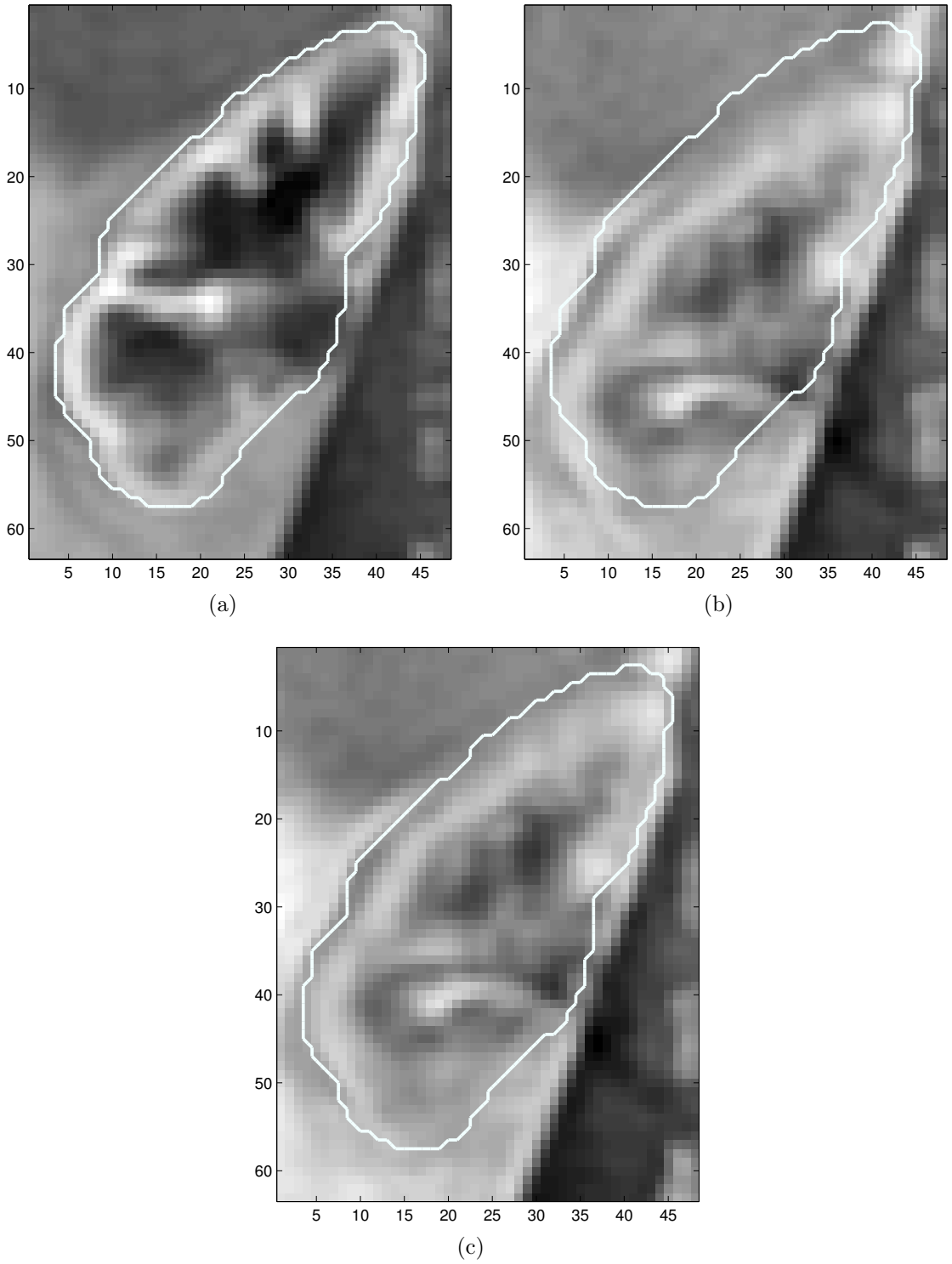


FIG. A.10 – Exemple de recalage sur données réelles : image de référence (a), image en phase de filtration avant recalage (b) et après recalage (c).

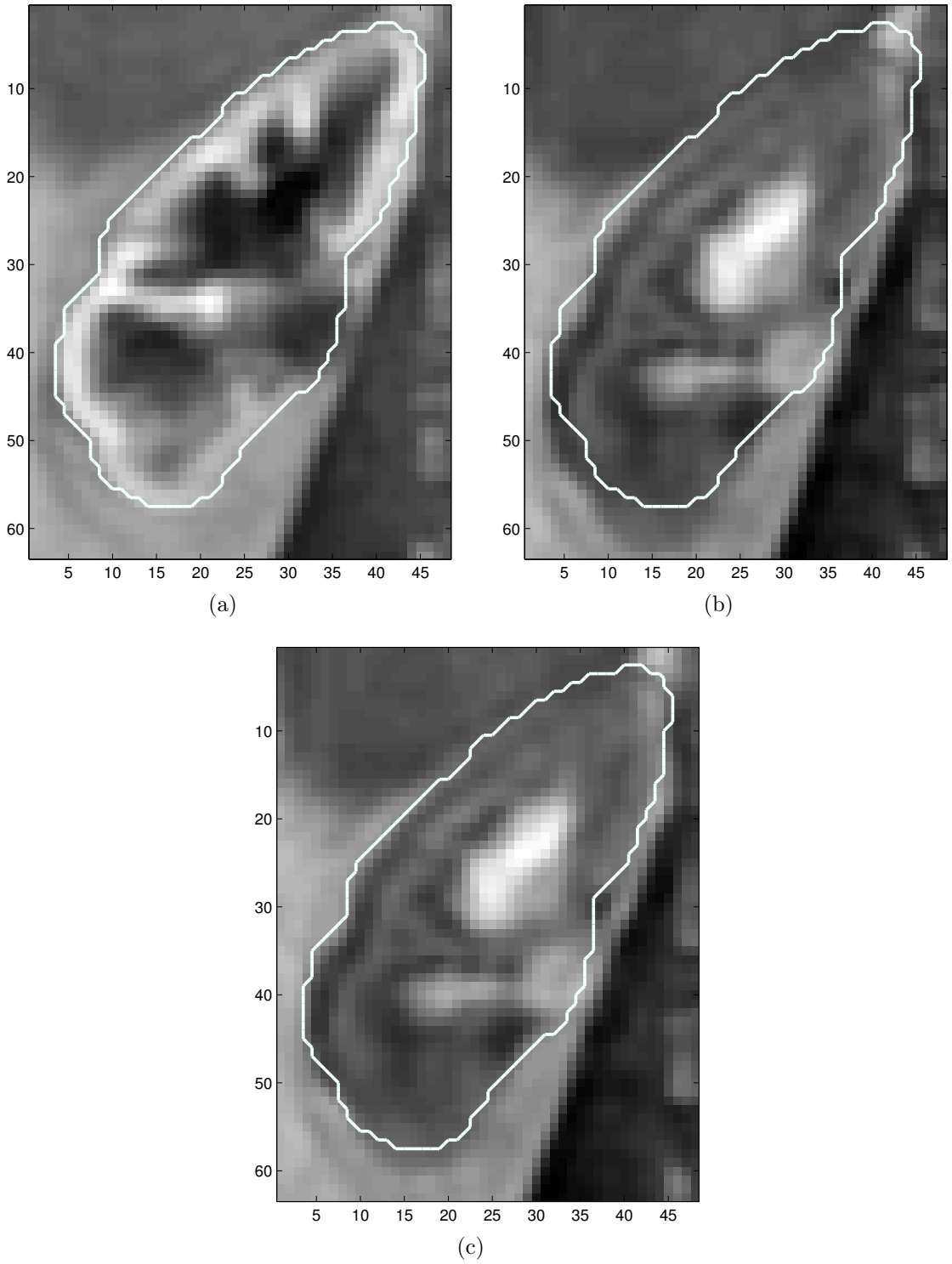


FIG. A.11 – Exemple de recalage sur données réelles : image de référence (a), image en phase de perfusion tardive avant recalage (b) et après recalage (c).

Annexe B

Méthodes d'optimisation sans contraintes

Les méthodes d'optimisation sans contrainte peuvent être classées selon l'information utilisée sur les dérivées successives de la fonction à minimiser ou maximiser. Nous envisagerons uniquement la minimisation, sachant que maximiser une fonction revient à minimiser son opposée. Nous pouvons distinguer les méthodes où seule la fonction à minimiser est évaluée (qui ne seront pas traitées ici) de celles utilisant son gradient (dites du premier ordre) ou son Hessien (dites du second ordre).

B.1 Descente de gradient ordinaire : apprentissage au premier ordre

La méthode de la plus profonde descente, appelée aussi méthode de la plus grande pente ou méthode du gradient à paramètre constant, est l'approche la plus classique : il s'agit de minimiser une fonction

$$\begin{aligned} \Psi : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{w} &\longmapsto \Psi(\mathbf{w}) \end{aligned} \tag{B.1}$$

en utilisant ses dérivées premières. On part d'un point initial $\mathbf{w}(0)$, on calcule le gradient de Ψ en ce point, puis on déplace $\mathbf{w}$ dans la direction opposée à celle du gradient d'une distance convenable. La procédure est répétée en partant du nouveau point $\mathbf{w}$, et ainsi de suite. La règle d'adaptation est la suivante :

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \alpha(t) \left. \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}(t)}, \tag{B.2}$$

le gradient étant évalué au point $\mathbf{w}(t)$. La paramètre $\alpha(t)$ est le pas dans la direction de plus grande pente. L'itération B.2 est répétée jusqu'à la convergence, ce qui arrive en pratique quand la distance euclidienne entre deux solutions consécutives $\|\mathbf{w}(t+1) - \mathbf{w}(t)\|$ devient inférieure à un seuil de tolérance fixé. Cette méthode amène au minimum local le plus proche et ne présente aucun moyen de s'en échapper, donc il est important de choisir des valeurs initiales convenables.

B.2 Descente de gradient stochastique

Soit une fonction

$$\begin{aligned} \mathrm{g} : \mathbb{R}^p \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (\xi, \mathbf{w}) &\longmapsto \mathrm{g}(\xi, \mathbf{w}) \end{aligned} \tag{B.3}$$

et la fonction Ψ définie par

$$\begin{aligned} \Psi : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{w} &\longmapsto \Psi(\mathbf{w}) = \int \mathrm{g}(\xi, \mathbf{w}) \mathrm{d}P_{\mathbf{X}}(\xi) = \mathrm{E}_{\mathbf{X}}[\mathrm{g}(\xi, \mathbf{w})]. \end{aligned} \tag{B.4}$$

La fonction Ψ est donc l'espérance par rapport à $\mathbf{X}$ de la fonction g et son gradient est

$$\begin{aligned} \frac{\partial \Psi}{\partial \mathbf{w}} : \mathbb{R}^p &\longrightarrow \mathbb{R}^p \\ \mathbf{w} &\longmapsto \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}}. \end{aligned} \tag{B.5}$$

Cependant, comme la loi de probabilité $P_{\mathbf{X}}$ est inconnue, on ne peut pas calculer le gradient de la fonction Ψ.

Soit une fonction

$$\begin{aligned} \mathrm{H} : \mathbb{R}^p \times \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (\xi, \mathbf{w}) &\longmapsto \mathrm{H}(\xi, \mathbf{w}), \end{aligned} \tag{B.6}$$

telle que

$$E_X[\mathrm{H}(\xi, \mathbf{w})] = \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}}. \tag{B.7}$$

Supposons qu'on dispose d'une suite $\{\xi_1, \ldots, \xi_N \ldots\}$ de réalisations indépendantes du vecteur aléatoire $\mathbf{X}$. La règle d'apprentissage pour la descente de gradient stochastique sur la fonction Ψ est la suivante [90] :

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \alpha(t)\mathbf{H}[\xi(t), \mathbf{w}(t)], \tag{B.8}$$

où $\alpha(t)$ est soit une suite de nombres positifs ou une matrice définie positive, équation à comparer à l'équation (B.2). A chaque itération , on utilise seulement la dernière réalisation $\xi(t)$ pour évaluer le gradient de Ψ. Si la convergence est généralement moins rapide que dans la version « batch », cet inconvénient est compensé par un coût de calcul bien moindre.

Pour analyser le comportement de la séquence $\mathbf{w}(t)$, on peut utiliser la méthode de l'équation différentielle ordinaire associée [91, 92], qui est basée sur l'analyse d'interpolations à temps continu de la séquence d'approximation stochastique. On associe à l'équation (B.8) une équation différentielle autonome :

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} = -\frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}}. \tag{B.9}$$

Les seuls points de convergence possibles sont les points fixes, ou stationnaires, c'est-à-dire les racines du membre de droite, parce que ce sont les points pour lesquels les variations de $\mathbf{w}$ dans le temps s'annulent. En linéarisant l'équation par rapport à $\mathbf{w}$, on peut analyser la stabilité de ces points fixes. Ceux qui sont dits asymptotiquement stables sont particulièrement importants puisqu'on peut montrer que l'algorithme converge vers l'un d'entre eux, sous réserve de quelques hypothèses sur $\alpha(t)$ et sur H.

B.3 Méthode de Newton : apprentissage au second ordre

Cette méthode est plus efficace que la simple descente de gradient au sens où elle converge en un nombre d'itérations plus faible, mais les calculs sont plus complexes. Elle utilise aussi les dérivées secondes de la fonction coût et nécessite d'évaluer son Hessien à chaque itération. Il s'agit de trouver une direction meilleure que celle du gradient pour converger vers la solution.

En utilisant le développement de Taylor au second ordre de la fonction coût Ψ, on l'approxime dans le voisinage du point $\mathbf{w}$ par une forme quadratique f de la variable $\mathbf{w}'$, telle que

$$\mathbf{f}(\mathbf{w}') = \Psi(\mathbf{w}) + \left(\frac{\partial \Psi}{\partial \mathbf{w}}\right)^T (\mathbf{w}' - \mathbf{w}) + \frac{1}{2}(\mathbf{w}' - \mathbf{w})^T \frac{\partial^2 \Psi}{\partial \mathbf{w}^2}(\mathbf{w}' - \mathbf{w}), \tag{B.10}$$

dont on annule le gradient pour trouver son minimum, ce qui permet de trouver le $\mathbf{w}'$ voisin de $\mathbf{w}$ où l'approximation $\mathbf{f}$ est minimale. On considère donc $\mathbf{w}$ comme fixé, le gradient de f par rapport à $\mathbf{w}'$ est :

$$\frac{\partial f(\mathbf{w}')}{\partial \mathbf{w}'} = \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}} + \frac{\partial^2 \Psi(\mathbf{w})}{\partial \mathbf{w}^2}(\mathbf{w}' - \mathbf{w}), \tag{B.11}$$

où $\frac{\partial^2 \Psi(\mathbf{w})}{\partial \mathbf{w}^2}$ est le Hessien de Ψ, alors $\frac{\partial f(\mathbf{w}')}{\partial \mathbf{w}'}$ s'annule pour

$$\mathbf{w}' = \mathbf{w} - \frac{\partial^2 \Psi(\mathbf{w})}{\partial \mathbf{w}^2}^{-1} \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}}. \tag{B.12}$$

D'éventuels problèmes de conditionnement peuvent nécessiter une régularisation.

Annexe C

Démonstrations

Ce chapitre concerne la convergence de deux algorithmes présentés au paragraphe 3.1.3. Nous montrons que l'algorithme de Lloyd généralisé peut être considéré comme la minimisation de la fonction coût par un algorithme de Newton, alors que le *Neural Gas* est une descente de gradient stochastique sur cette fonction. Les notations utilisées sont celles définies dans le paragraphe 3.1.3.

C.1 L'algorithme de Lloyd généralisé applique un algorithme de minimisation de Newton sur sa fonction coût

Une démonstration de ce résultat se trouve dans [63].

Démonstration. La fonction $\widehat{\Psi}$ définie dans l'équation (3.20) peut également s'écrire

$$\widehat{\Psi}(\mathbf{W}) = \sum_{i=1}^{N} \mathrm{L}(\xi_i, \mathbf{W}), \tag{C.1}$$

avec

$$\mathrm{L}(\xi_i, \mathbf{W}) = \min_j \frac{1}{2}\|\xi_i - \mathbf{w}_j\|^2. \tag{C.2}$$

Le gradient de la fonction $\widehat{\Psi}$ peut s'écrire en K blocs de taille $p \times 1$ de la manière suivante :

$$\frac{\partial \widehat{\Psi}}{\partial \mathbf{W}} = \begin{pmatrix} \frac{\partial \widehat{\Psi}}{\partial \mathbf{w}_1} \\ \frac{\partial \widehat{\Psi}}{\partial \mathbf{w}_2} \\ \vdots \\ \frac{\partial \widehat{\Psi}}{\partial \mathbf{w}_K} \end{pmatrix} = \sum_{i=1}^{N} \frac{\partial \mathrm{L}(\xi_i, \mathbf{W})}{\partial \mathbf{W}}, \tag{C.3}$$

et est égal à la somme des gradients $\frac{\partial \mathrm{L}(\xi_i, \mathbf{W})}{\partial \mathbf{W}}$ de chacun de ses termes. Or, $\mathrm{L}(\xi_i, W)$ ne dépend que du prototype le plus proche de ξ_i (sauf pour les points situés à égale distance

d'au moins deux prototypes, mais qui ont une probabilité nulle) alors pour ξ_i donné, seul le bloc correspondant aux dérivées par rapport à ce prototype $\mathbf{w}_k$ est non nul et vaut

$$\frac{\partial \mathrm{L}(\xi_i, \mathbf{W})}{\partial \mathbf{w}_k} = -2(\xi_i - \mathbf{w}_k). \tag{C.4}$$

De la même manière, le Hessien H de la fonction $\widehat{\Psi}$ peut s'écrire en blocs de taille $p \times p$ de la manière suivante, chaque bloc correspondant à une paire de prototypes :

$$H = \begin{pmatrix} \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_1^2} & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_1 \partial \mathbf{w}_2} & \cdots & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_1 \partial \mathbf{w}_K} \\ \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_2 \partial \mathbf{w}_1} & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_2^2} & \cdots & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_2 \partial \mathbf{w}_K} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_K \partial \mathbf{w}_1} & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_K \partial \mathbf{w}_2} & \cdots & \frac{\partial^2 \widehat{\Psi}}{\partial \mathbf{w}_K^2} \end{pmatrix}. \tag{C.5}$$

Le Hessien H est aussi la somme des Hessiens H_i de chacun de ses termes. Pour ξ_i donnée, seul le bloc correspondant aux dérivées par rapport à ce prototype $\mathbf{w}_k$ est non nul et est égal à la matrice identité $\mathbf{I}$. Ainsi, en sommant les H_i, nous obtenons une matrice diagonale dont les éléments sont égaux aux nombres de prototypes associés à chacun des $\mathbf{w}_k$:

$$H = \begin{pmatrix} N_1\mathbf{I} & 0 & \dots & 0 \\ 0 & N_2\mathbf{I} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & N_K\mathbf{I} \end{pmatrix}. \tag{C.6}$$

D'après l'équation (B.12) , la règle d'adaptation de Newton correspondante est :

$$\mathbf{w}(t+1) = \mathbf{w}(t) - H^{-1}\frac{\partial \widehat{\Psi}(\mathbf{w})}{\partial \mathbf{w}}, \tag{C.7}$$

donc pour un prototype $\mathbf{w}_k$ donné,

$$\mathbf{w}_k(t+1) = \mathbf{w}_k(t) - \frac{1}{N_k} I \frac{\partial \widehat{\Psi}(\mathbf{W})}{\partial \mathbf{w}_k} \tag{C.8}$$

$$= \mathbf{w}_k(t) + \frac{1}{N_k} \sum_{\xi_i \in V_k} (\xi_i - \mathbf{w}_k(t)) \tag{C.9}$$

$$= \frac{1}{N_k} \sum_{\xi_i \in V_k} \xi_i. \tag{C.10}$$

Nous retrouvons l'équation (3.21). □

C.2 L'algorithme de *Neural Gas* applique une descente de gradient stochastique sur sa fonction coût

Soit la fonction Ψ_{NG} définie dans l'équation (3.28). Dans [58], il est démontré que :

$$\frac{\partial \Psi_{\mathrm{NG}}(W)}{\partial \mathbf{w}_j} = 2 \int_X \mathrm{h}_\lambda[k_j(\xi, \mathbf{W})](\xi - \mathbf{w}_j) \mathrm{d}P_{\mathbf{X}}(\xi). \tag{C.11}$$

Soit la fonction

$$\mathbf{H} : \mathbb{R}^p \times \mathbb{R}^{Kp} \longrightarrow \mathbb{R}^{Kp} \tag{C.12}$$
$$(\xi, \mathbf{W}) \longmapsto \mathbf{H}(\xi, \mathbf{W}) = \begin{pmatrix} 2\mathrm{h}_\lambda[k_1(\xi, \mathbf{W})](\xi - \mathbf{w}_1) \\ \vdots \\ 2\mathrm{h}_\lambda[k_K(\xi, \mathbf{W})](\xi - \mathbf{w}_K) \end{pmatrix}.$$

$\mathbf{H}$ vérifie :

$$\mathrm{E}_{\mathbf{X}}[\mathbf{H}(\xi, \mathbf{W})] = \frac{\partial \Psi_{\mathrm{NG}}(\mathbf{W})}{\partial \mathbf{W}}. \tag{C.13}$$

La règle d'apprentissage (3.26) page 67 s'écrit alors :

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \alpha(t)\mathbf{H}[\xi(t), \mathbf{W}(t)], \tag{C.14}$$

qui est la règle d'apprentissage B.8 pour la descente de gradient stochastique sur la fonction Ψ_{NG}.

Bibliographie

[1] C. Carbonnel, V. Seux, V. Pauly, C. Oddoze, C. Roubicek, J.R. Larue, X. Thirion, J. Soubeyrand, and F. Retornaz, "Quelle méthode d'évaluation de la fonction rénale utiliser chez le sujet âgé hospitalisé en unité de court séjour gériatrique ? Comparaison de quatre méthodes," *La Revue de Médecine Interne*, vol. 29, no. 5, pp. 364–369, 2008.

[2] D.W. Cockcroft and M.H. Gault, "Prediction of creatinine clearance from serum creatinine," *Nephron*, vol. 16, pp. 31–41, 1976.

[3] A.S. Levey, J.P. Bosch, J.B. Lewis, T. Greene, N. Rogers, and D. Roth, "A more accurate method to estimate glomerular filtration rate from serum creatinine : a new prediction equation," *Annals of Internal Medicine*, vol. 130, pp. 461–470, 1999.

[4] A. Taylor and J.V. Nally, "Clinical applications of renal scintigraphy," *American Journal of Roentgenology*, vol. 164, pp. 31–41, 1995.

[5] F. P. Esteves, A. Taylor, A. Manatunga, R.D. Folks, M. Krishnan, and E.V. Garcia, "99mTc-MAG3 renography : Normal values for MAG3 clearance and curve parameters, excretory parameters, and residual urine volume," *American Journal of Roentgenology*, vol. 187, no. 6, pp. W610–617, 2006.

[6] C.D. Russell, A.T. Taylor, and E.V. Dubovsky, "Measurement of renal function with Technetium-99m-MAG3 in children and adults," *Journal of Nuclear Medicine*, vol. 37, no. 4, pp. 588–593, 1996.

[7] C.D. Russell and E.V. Dubovsky, "Measurement of renal function with radionuclides," *Journal of Nuclear Medicine*, vol. 30, pp. 2053–2057, 1989.

[8] S.R. Cherry, J.A. Sorenson, and M.E. Phelps, *Physics in Nuclear Medicine*, Saunders, 2003.

[9] Y. Inoue, K. Yoshikawa, T. Suzuki, N. Katayama, I. Yokoyama, T. Kohsaka, Y. Tsukune, and K. Ohtomo, "Attenuation correction in evaluating renal function in children and adults by a camera-based method," *Journal of Nuclear Medicine*, vol. 41, no. 5, pp. 823–829, 2000.

[10] N. Grenier, I. Mendichovszky, B. D. de Senneville, S. Roujol, P. Desbarats, M. Pedersen, K. Wells, J. Frokiaer, and I. Gordon, "Measurement of glomerular filtration rate with magnetic resonance imaging : principles, limitations, and expectations.," *Seminars in Nuclear Medicine*, vol. 38, no. 1, pp. 47–55, 2008.

[11] D. Hoa, A. Micheau, G. Gahide, E. Le Bars, and P. Taourel, *L'IRM pas à pas*, Sauramps Campus Medica, 2008.

[12] C. Brochot, B. Bessoud, D. Balvay, C.-A. Cuenod, N. Siauve, and F. Bois, "Evaluation of antiangiogenic treatment effects on tumors microcirculation by Bayesian physiological pharmacokinetic modeling and magnetic resonance imaging," *Magnetic Resonance Imaging*, vol. 24, pp. 1059–1067, 2006.

[13] I.A. Mendichovszky, M. Cutajar, and I. Gordon, "Reproducibility of the aortic input function (AIF) derived from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) of the kidneys in a volunteer study," *European Journal of Radiology*, vol. In Press, Corrected Proof, 2008.

[14] H.S. Teh, E. S. Ang, W.C Wong, S.B. Tan, A. G. S. Tan, S. M. Chng, M.B.K. Lin, and J.S.K. Goh, "MR renography using a dynamic gradient-echo sequence and low-dose gadopentetate dimeglumine as an alternative to radionuclide renography," *American Journal of Roentgenology*, vol. 181, no. 2, pp. 441–450, 2003.

[15] J. Taylor, P.E. Summers, S.F. Keevil, A.M. Saks, J. Diskin, P.J. Hilton, and A.B. Ayers, "Magnetic resonance renography : Optimisation of pulse sequence parameters and Gd-DTPA dose, and comparison with radionuclide renography," *Magnetic Resonance Imaging*, vol. 15, no. 6, pp. 637–649, 1997.

[16] H. Uematsu, T. Matsuda, T. Tsuchida, H. Inoue, K. Hayashi, Y. Yonekura, and H. Itoh, "Semi-quantitative approach to estimating gfr by magnetic resonance imaging," *Magnetic Resonance Materials in Biology, Physics, and Medicine*, vol. 10, no. 3, pp. 171–176, 2000.

[17] V.S. Lee, H. Rusinek, L. Bokacheva, A.J. Huang, N. Oesingmann, Q. Chen, M. Kaur, K. Prince, T. Song, E.L. Kramer, and E.F. Leonard, "Renal function measurements from MR renography and a simplified multicompartmental model," *American Journal of Physiology - Renal Physiology*, vol. 292, pp. 1548–1559, 2007.

[18] D. Rodriguez Gutierrez, K. Wells, O. Diaz Montesdeoca, A. Moran Santana, I.A. Mendichovszky, and I. Gordon, "Partial volume effects in dynamic contrast magnetic resonance renal studies," *European Journal of Radiology*, 2009, In Press, Corrected Proof.

[19] H. Gray, *Anatomy of the Human Body*, Philadelphia : Lea and Febiger, 1918 ; New York : Bartleby.com, http ://www.bartleby.com/107/, 20th edition, 2000.

[20] M. Catala, J.M. Andre, G. Katsanis, and J. Poirier, "Histologie : organes, systèmes et appareils," http ://www.chups.jussieu.fr/polys/histo/histoP2/index.html, 2007, Faculté de médecine Pierre et Marie Curie, Site de la Pitié-Salpêtrière.

[21] R. Martzolff, "Le rein et la formation de l'urine," Vidéo, 2007, http ://video.vulgaris-medical.com/index.php/2007/05/20/27-le-rein-et-la-formation-de-l-urine.

[22] H.J. Michaely, S. Sourbron, O. Dietrich, U. Attenberger, M.F. Reiser, and S.O. Schoenberg, "Functional renal MR imaging : an overview," *Abdominal Imaging*, vol. 32, pp. 758–771, 2007.

[23] N. Grenier, F. Basseau, M. Ries, B. Tyndal, R. Jones, and C. Moonen, "Functional MRI of the kidney," *Abdominal Imaging*, vol. 28, no. 2, pp. 164–175, 2003.

[24] M.F. Bellin, "MR contrast agents, the old and the new," *European Journal of Radiology*, vol. 60, pp. 314–323, 2006.

[25] N. Grenier, M. Pedersen, and O. Hauger, "Contrast agents for functional and cellular MRI of the kidney," *European Journal of Radiology*, vol. 60, no. 3, pp. 341–352, 2006.

[26] N. Michoux, J.P. Vallee, A. Pechere-Bertschi, X. Montet, L. Buehler, and B. Beers, "Analysis of contrast-enhanced MR images to assess renal function," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 19, no. 4, pp. 167–179, 2006.

[27] N. Michoux, X. Montet, A. Pefchere, M.K. Ivancevic, P.Y. Martin, A. Keller, D. Didier, F. Terrier, and J.P. Vallee, "Parametric and quantitative analysis of MR renographic curves for assessing the functional behaviour of the kidney," *European Journal of Radiology*, vol. 54, pp. 124–135, 2005.

[28] M. Dujardin, R. Luypaert, F. Vandenbroucke, P. Van der Niepen, S.Sourbron, D. Verbeelen, T. Stadnik, and J. de Mey, "Combined T1-based perfusion MRI and MR angiography in kidney : First experience in normals and pathology," *European Journal of Radiology*, vol. 69, no. 3, pp. 542–549, 2009.

[29] W.K. Rohrschneider, S. Haufe, M. Wiesel, B. Toenshoff, R. Wunsch, K. Darge, J.H. Clorius, and J. Troeger, "Functional and morphologic evaluation of congenital urinary tract dilatation by using combined static-dynamic MR urography : Findings in kidneys with a single collecting system," *Radiology*, vol. 224, no. 3, pp. 683–694, 2002.

[30] S.P. Sourbron, H.J. Michaely, M.F. Reiser, and S.O. Schoenberg, "MRI-measurement of perfusion and glomerular filtration in the human kidney with a separable compartment model," *Investigative Radiology*, vol. 43, no. 1, pp. 40–48, 2008.

[31] N. Hackstein, C. Wiegand, W.S. Rau, and A.C. Langheinrich, "Glomerular filtration rate measured by using triphasic helical CT with a two-point Patlak plot technique," *Radiology*, vol. 230, pp. 221–226, 2004.

[32] L. Bokacheva, H. Rusinek, J.L. Zhang, Q. Chen, and V.S. Lee, "Estimates of glomerular filtration rate from MR renography and tracer kinetic models," *Journal of Magnetic Resonance Imaging*, vol. 29, no. 2, pp. 371–382, 2009.

[33] B.D. de Senneville, I.A. Mendichovszky, S. Roujol, Gordon I., C. Moonen, and N. Grenier, "Improvement of MRI-functional measurement with automatic movement correction in native and transplanted kidneys," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 4, pp. 970–8, 2008.

[34] J.B.A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[35] G.E. Christensen, R.D. Rabbit, and M.I. Miller, "Deformable templates using large deformation kinematics," *IEEE Transactions on Image Processing*, vol. 5, no. 10, pp. 1435–1447, 1996.

[36] E.L.W. Giele, J.A. De Priester, J.A. Blom, J.A. De Boer, J.M. Van Engelshoven, A. Hasman, and M. Geerlings, "Movement correction of the kidney in dynamic MRI scans using FFT phase difference movement detection," *Journal of Magnetic Resonance Imaging*, vol. 14, pp. 741–749, 2001.

[37] Y. Sun, M.-P. Jolly, and J.M.R. Moura, "Integrated registration of dynamic renal perfusion MR images," in *Proceedings of the International Conference on Image Processing (ICIP 2004)*, Singapore, 2004, vol. Vol. 3, pp. 1923–6.

[38] T. Song, V.S. Lee, H. Rusinek, S. Wong, and A.F. Laine, "Integrated four dimensional registration and segmentation of dynamic renal MR images," in *Proceedings of the 9th Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006)*, Copenhagen, Denmark, 2006, pp. 758–65.

[39] T. Song, V.S. Lee, H. Rusinek, S. Wong, and A.F. Laine, "Four dimensional MR image analysis of dynamic renography," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2006)*, 2006, pp. 3134–3137.

[40] T. Song, V.S. Lee, H. Rusinek, M. Kaur, and A.F. Laine, "Automatic 4-D registration in dynamic MR renography based on over-complete dyadic wavelet and Fourier transforms," in *Proceedings of the 8th Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2005)*, Palm Springs CA, USA, 2005, pp. 205–13.

[41] Y. Sun, J.M.F. Moura, and H. Chien, "Subpixel registration in renal perfusion MR image sequence," in *Proceedings of the IEEE International Symposium on Biomedical Imaging : Macro to Nano (ISBI 2004)*, Arlington, VA, USA, 2004, vol. 1, pp. 700–3.

[42] F. G. Zoellner, R. Sance, P. Rogelj, M.J. Ledesma-Carbayo, J. Roervik, A. Santos, and A. Lundervold, "Assessment of 3D DCE-MRI of the kidneys using non-rigid image registration and segmentation of voxel time courses," *Computerized Medical Imaging and Graphics*, vol. 33, pp. 171–181, 2009.

[43] Y. Sun, M.-P. Jolly, and J.M.F. Moura, "Contrast-invariant registration of cardiac and renal MR perfusion images," in *Proceedings of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004)*, Rennes, France, 2004, vol. 1, pp. 903–10.

[44] P. Rogelj, F.G. Zoellner, S. Kovačič, and A. Lundervold, "Motion correction of contrast-enhanced MRI time series of kidney," in *Proceedings of the 16th International Electrotechnical and Computer Science Conference (ERK 2007)*, Portoroz, Slovenia, 2007, pp. 191–194.

[45] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information-based registration of medical images : a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[46] P. Rogelj, S. Kovacic, and J.C. Gee, "Point similarity measures for non-rigid registration of multi-modal data," *Computer Vision and Image Understanding*, vol. 92, no. 1, pp. 112–140, 2003.

[47] F.J.P. Richard, A.M.M Samson, and C.A. Cuénod, "A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences," *Statistics and Computing*, vol. 19, no. 4, pp. 465–478, 2009.

[48] V.S. Lee, H. Rusinek, M.E. Noz, P. Lee, M. Raghavan, and E.L. Kramer, "Dynamic three-dimensional MR renography for the measurement of single kidney function : Initial experience," *Radiology*, vol. 227, pp. 289–294, 2003.

[49] J.P. Cocquerez, S. Philipp, Ph. Bolon, J.M. Chassery, D. Demigny, C. Graffigne, A. Montanvert, R. Zeboudj, and J. Zerubia, *Analyse d'images : filtrage et segmentation*, Masson, 1995.

[50] C.H. Coulam, D.M. Bouley, and F.G. Sommer, "Measurement of renal volumes with contrast-enhanced MRI," *Journal Of Magnetic Resonance Imaging*, vol. 15, no. 2, pp. 174–179, 2002.

[51] D. Lv, J. Zhuang, H. Chen, J. Wang, Y. Xu, X. Yang, J. Zhang, X. Wang, and J. Fang, "Dynamic contrast-enhanced magnetic resonance images of the kidney," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 5, pp. 36–41, 2008.

[52] J.A. De Priester, A.G. Kessels, E.L. Giele, J.A. De Boer, M.H. Christiaans, A. Hasman, and J.M. Van Engelshoven, "MR renography by semiautomated image analysis : performance in renal transplant recipients," *Journal of Magnetic Resonance Imaging*, vol. 14, no. 2, pp. 134–140, 2001.

[53] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[54] F.G. Zoellner, R. Sance, A. Anderlik, J. Roervik, M. Kocinski, and A. Lundervold, "Towards quantification of kidney function by clustering volumetric MRI perfusion time series," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 19, pp. 103–104, 2006.

[55] T. Song, V.S. Lee, H. Rusinek, J.B. Sajous, and A.F. Laine, "Registration and segmentation of dynamic three-dimensional MR renography based on Fourier representations and k-means clustering," in *Proceedings of the 13th Scientific Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 2005)*, Miami, Florida, USA, 2005, 1 page.

[56] F.G. Zoellner, M. Kocinski, A. Lundervold, and J. Roervik, *Bildverarbeitung für die Medizin 2007*, chapter Assessment of Renal Function from 3D Dynamic Contrast Enhanced MR images Using Independent Component Analysis, pp. 237–241, Informatik aktuell. Springer Berlin Heidelberg, 2007.

[57] H. Rusinek, Y. Boykov, M Kaur, S. Wong, L. Bokacheva, J.B. Sajous, A.J. Huang, S. Heller, and V.S. Lee, "Performance of an automated segmentation algorithm for 3D MR renography," *Magnetic Resonance in Medicine*, vol. 57, pp. 1159–1167, 2007.

[58] T. Martinetz, S. Berkovich, and K. Schulten, "Neural-gas network for vector quantization and its application to time-series prediction," *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 558–569, 1993.

[59] A. Gersho and R.M. Gray, *Vector quantization and signal compression*, Dordrecht, Netherlands, 1992.

[60] T. Martinetz and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7(3), pp. 507–522, 1994.

[61] H. Frezza-Buet, "Following non-stationary distributions by controlling the vector quantization accuracy of a growing neural gas network," *Neurocomputing*, vol. 71, no. 7-9, pp. 1191–1202, March 2008.

[62] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in Neural Information Processing Systems 7*, G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds., pp. 625–632. MIT Press, Cambridge MA, 1995.

[63] L. Bottou and Y. Bengio, "Convergence properties of the k-means algorithms," in *Proceedings of the 9th Conference on Neural Information Processing Systems (NIPS)*, Denver, CO, USA, 1995, pp. 585–92.

[64] M. Cottrell, B. Hammer, A. Hasenfuss, and T. Villmann, "Batch and median Neural Gas," *Neural Networks*, vol. 19, no. 6-7, pp. 762–771, 2006.

[65] T. Kohonen, *Self-organizing maps*, Springer, 2001.

[66] G. Patané and M. Russo, "The enhanced LBG algorithm," *Neural Networks*, vol. 14, no. 9, pp. 1219–1237, 2001.

[67] A. Gersho, "Asymptotically optimal block quantization," *Information Theory, IEEE Transactions on*, vol. 25, no. 4, pp. 373–380, 1979.

[68] C. Chinrungrueng and C. H. Sequin, "Optimal adaptive k-means algorithm with dynamic adjustment of learning rate," *Neural Networks, IEEE Transactions on*, vol. 6, no. 1, pp. 157–169, 1995.

[69] S. Ben-David, D. Pal, and H.U. Simon, "Stability of k-means clustering," in *Proceedings of the 20th Annual Conference on Learning Theory (COLT 2007)*, Berlin, Germany, 2007, pp. 20–34.

[70] M. Ackerman and S. Ben-David, "Measures of clustering quality : A working set of axioms for clustering," in *Proceedings of the 22nd Conference on Neural Information Processing Systems (NIPS 2008)*, Vancouver, B.C., Canada., 2008.

[71] M. Meila, "Comparing clusterings : an information based distance," *Journal of Multivariate Analysis*, vol. 98, no. 5, pp. 873–895, 2007.

[72] R.J.G.B. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 833–841, 2007.

[73] E.B. Fowlkes and C.L. Mallows, "A method for comparing two hierarchical clusterings," *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 553–569, 1983.

[74] J. Kleinberg, "An impossibility theorem for clustering," in *Proceedings of the 16th Conference on Neural Information Processing Systems (NIPS 2002)*, Vancouver, B.C., Canada, 2002, 8 pages.

[75] M. Ackerman and S. Ben-David, "Clusterability : A theoretical study," in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*, 2009, pp. 1–8.

[76] E. Pichon, A. Tannenbaum, and R. Kikinis, "A statistically based flow for image segmentation," *Medical Image Analysis*, vol. 8, no. 3, pp. 267–274, 2004.

[77] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer, "Morphometric analysis of white matter lesions in MR images : method and validation," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 716–24, 1994.

[78] V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, 1998.

[79] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1999.

[80] M. J. Kearns and U.V. Vazirani, *An introduction to computational learning theory*, MIT Press, 1994.

[81] U. Von Luxburg and S. Ben-David, "Towards a statistical theory of clustering," in *Proceedings of the PASCAL Workshop on Statistics and Optimization of Clustering*, London, UK, 2005, 6 pages.

[82] R.D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408–19, 1999.

[83] J. Sijbers, A.J. den Dekker, P. Scheunders, and D. Van Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, 1998.

[84] J. Sijbers, D. Poot, A.J. den Dekker, and W. Pintjens, "Automatic estimation of the noise variance from the histogram of a magnetic resonance image," *Physics in Medicine and Biology*, vol. 52, no. 5, pp. 1335–48, 2007.

[85] M.A.G. Ballester, A. Zisserman, and M. Brady, "Estimation of the partial volume effect in MRI," *Medical Image Analysis*, vol. 6, pp. 389–405, 2002.

[86] D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 1, pp. 120–128, 2003.

[87] B. Chevaillier, Y. Ponvianne, J.L. Collette, D. Mandry, M. Claudon, and O. Pietquin, "Functional semi-automated segmentation of renal DCE-MRI sequences," in *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 2008, pp. 525–528.

[88] B. Chevaillier, Y. Ponvianne, J.L. Collette, D. Mandry, M. Claudon, and O. Pietquin, "Functional semi-automated segmentation of renal DCE-MRI sequences using a growing neural gas algorithm," in *Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne (Switzerland), 2008, 4 pages.

[89] J. Harthong, *Probabilités et statistiques, de l'intuition aux applications*, Diderot éditeur, Arts et Sciences, Paris, 1996.

[90] L. Bottou, "Online algorithms and stochastic approximations," in *Online Learning and Neural Networks*, David Saad, Ed. Cambridge University Press, Cambridge, UK, 1998.

[91] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[92] H.J. Kushner and Yin G.G., *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, 2003.

Résumé

Pour l'évaluation de la fonction rénale, l'Imagerie par Résonance Magnétique (IRM) dynamique à rehaussement de contraste est une alternative intéressante à la scintigraphie. Les résultats obtenus doivent cependant être évalués à grande échelle avant son utilisation en pratique clinique. L'exploitation des séquences acquises demande un recalage de la série d'images et la segmentation des structures internes du rein. Notre objectif est de fournir un outil fiable et simple à utiliser pour automatiser en partie ces deux opérations. Des méthodes statistiques de recalage utilisant l'information mutuelle sont testées sur des données réelles. La segmentation du cortex, de la médullaire et des cavités est réalisée en classant les voxels rénaux selon leurs courbes temps-intensité. Une stratégie de classification en deux étapes est proposée. Des classificateurs sont d'abord construits grâce à la coupe rénale principale en utilisant deux algorithmes de quantification vectorielle (K-moyennes et Growing Neural Gas). Ils sont validés sur données simulées, puis réelles, en évaluant des critères de similarité entre une segmentation manuelle de référence et des segmentations fonctionnelles ou une seconde segmentation manuelle. Les voxels des autres coupes sont ensuite triés avec le classificateur optimum pour la coupe principale. La théorie de la généralisation permet de borner l'erreur de classification faite lors de cette extension. La méthode proposée procure les avantages suivants par rapport à une segmentation manuelle : gain de temps important, intervention de l'opérateur limitée et aisée, bonne robustesse due à l'utilisation de toute la séquence et bonne reproductibilité.

Abstract

Dynamic Contrast-Enhanced Magnetic Resonance Imaging has a great potential for renal function assessment but has to be evaluated on a large scale before its clinical application. Registration of image sequences and segmentation of internal renal structures is mandatory in order to exploit acquisitions. We propose a reliable and user-friendly tool to partially automate these two operations. Statistical registration methods based on mutual information are tested on real data. Segmentation of cortex, medulla and cavities is performed using time-intensity curves of renal voxels in a two step process. Classifiers are first built with pixels of the slice that contains the largest proportion of renal tissue : two vector quantization algorithms, namely the K-means and the Growing Neural Gas with Targeting, are used here. These classifiers are first tested on synthetic data. For real data, as no ground truth is available for result evaluation, a manual anatomical segmentation is considered as a reference. Some discrepancy criteria like overlap, extra pixels and similarity index are computed between this segmentation and a functional one. The same criteria are also evaluated between the reference and another manual segmentation. Results are similar for the two types of comparisons. Voxels of other slices are then sorted with the optimal classifier. Generalization theory allows to bound classification error for this extension. The main advantages of functional methods are the following : considerable time-saving, easy manual intervention, good robustness and reproducibility.