



Séparation de sources en imagerie nucléaire

Marc Filippi

► To cite this version:

Marc Filippi. Séparation de sources en imagerie nucléaire. Imagerie médicale. Université Grenoble Alpes, 2018. Français. NNT : 2018GREAT025 . tel-01879639

HAL Id: tel-01879639

<https://theses.hal.science/tel-01879639>

Submitted on 24 Sep 2018

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 25 mai 2016

Présentée par

Marc FILIPPI

Thèse dirigée par **Michel DESVIGNES, Professeur des
Universités, Grenoble INP, GIPSA-lab**, et
coencardée par **Eric MOISAN, Maître de Conférence, Grenoble
INP, GIPSA-lab**,

préparée au sein du **gipsa-lab**
dans l'**École Doctorale Electronique, Electrotechnique,
Automatique, Traitement du Signal (EEATS)**

Séparation de sources en imagerie nucléaire

Thèse soutenue publiquement le **05/04/2018**,
devant le jury composé de :

Monsieur Pascal HAIGRON

Professeur des Universités, Université de Rennes 1, LTSI, Rennes,
Rapporteur

Monsieur Patrick CLARYSSE

Directeur de Recherche, CNRS, CREATIS, Lyon, Président du jury

Madame Pascale PERRET

Chargée de Recherche, INSERM, Radiopharmaceutiques Biocliniques,
Grenoble, Examinatrice

Madame Caroline PETITJEAN

Maître de Conférence, Université de Rouen, LITIS, Examinatrice

Monsieur Michel DESVIGNES

Professeur des Universités, Grenoble INP, GIPSA-lab, Directeur de thèse

Monsieur Eric MOISAN

Maître de Conférence, Grenoble INP, GIPSA-lab, Co-encadrant de thèse



UNIVERSITÉ DE GRENOBLE ALPES
ÉCOLE DOCTORALE EEATS
Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)

THÈSE

pour obtenir le titre de

docteur en sciences

de l'Université de Grenoble-Alpes

Mention : SIGNAL, IMAGE, PAROLE, TÉLÉCOMS

Présentée et soutenue par

Marc FILIPPI

Séparation de sources en imagerie nucléaire

Thèse dirigée par Michel DESVIGNES, et coencadrée par Eric MOISAN

préparée au laboratoire GIPSA-lab

soutenue le 05/04/2018

Jury :

<i>Rapporteurs :</i>	Pascal HAIGRON	-	Université de Rennes 1, LTSI, Rennes
	Grégoire MALANDAIN	-	INRIA Sophia Antipolis, I3S, Nice
<i>Directeur :</i>	Michel DESVIGNES	-	Grenoble INP, GIPSA-lab
<i>Encadrant :</i>	Eric MOISAN	-	Grenoble INP, GIPSA-lab
<i>Examineurs :</i>	Patrick CLARYSSE	-	CNRS, CREATIS, Lyon
	Pascale PERRET	-	INSERM, Radiopharmaceutiques Biocliniques
	Caroline PETITJEAN	-	Université de Rouen, LITIS

Remerciements

Je tiens tout d'abord à remercier tous les membres du jury pour avoir examiné mon travail et fait le déplacement à Grenoble pour assister à la soutenance. Merci pour vos retours et pour les discussions intéressantes qui ont suivi la présentation.

Michel et Eric, un très grand merci à vous deux, pour m'avoir fait confiance depuis le début, et pour avoir bien encadré mes travaux tout au long de ces trois longues années de doctorat. Ce fût un travail difficile, mais vous avez su me guider et réorienter mes travaux lorsque cela était nécessaire. Vous m'avez également permis et proposé de découvrir l'enseignement, ce qui m'ouvre aujourd'hui de nouveaux horizons, que j'ai hâte d'explorer.

Merci également à tous les membres du LRB et du CHU avec qui j'ai eu l'occasion de travailler, toujours dans une ambiance agréable : Pascale, Catherine, Romain, Daniel, Loïc, Gilles. Ça a été l'occasion de confronter deux mondes différents, et j'ai beaucoup appris à vos côtés sur le monde médical et la recherche qui y est associée. Ce fût un challenge de travailler avec des données issues de la réalité clinique et utilisant un traceur encore peu caractérisé en in-vivo chez l'humain. Longtemps la frustration a été présente, et le regret de ne pas aboutir à une application clinique persiste, mais l'importance des enjeux de ces travaux méritaient amplement ces efforts.

Je tiens également à remercier les autres chercheurs qui m'ont aidé durant mon doctorat, notamment Pierre Comon pour sa disponibilité et les échanges éclairants que nous avons pu avoir ensemble au sujet des tenseurs, ainsi que Laurent Condat et Mai pour avoir répondu à mes interrogations sur des problèmes d'optimisation.

Enfin, qu'auraient été ces années sans toutes les personnes qui me les ont rendu très agréables malgré la difficulté et parfois les déconvenues liées au travail :

Alex, Pierre et Antony, c'est avec vous trois que tout a commencé au Gipsa, et on a passé un temps considérable ensemble. Les chemins que l'on mène aujourd'hui sont bien différents et nous voilà désormais aux quatre coins de la France, bien loin les uns des autres. Bon courage dans votre nouvelle vie !

Marion, il me faudrait un chapitre entier pour te remercier dignement, toi et Fabrice. Un grand merci pour ces innombrables week-ends où vous m'avez invité à la Ruchère en montagne, pour m'avoir fait (re-)découvrir les jeux de sociétés, et pour toutes les activités et moments qu'on a pu partagé ensemble. Et aussi (surtout ? !) pour avoir élevé la lapone la plus craquante des environs, qui me donne le sourire à chaque fois !

Pierre Maho, mon co-bureau des dernières heures, qui a su élever son statut de stagiaire à celui de doctorant. Le bureau doit être bien vide maintenant, c'était top de le partager avec toi sur la fin en tout cas. Prend bien soin de ton matériel informatique !

Emmanuelle, on peut dire que tu n'as peur de rien, à part peut-être des gateaux trop cuits

et de la ponctualité des trains. Merci d'avoir partagé ton enthousiasme au labo.

Raph, encore un qui n'a peur de rien, que ça soit de la danse devant des centaines de personnes, des bains glacés, ou des marathons sans aucune préparation. Merci de m'avoir emmener cet hiver dans le Queyras, et de m'y emmener une seconde fois cet été!

Jeanne, impossible de s'ennuyer avec toi, toujours le mot pour rire et les soupirs qui en disent long. Et merci aussi d'avoir organiser la venue de ton Louis ici. Il reste toujours vaincu raquette en main à Grenoble, il va falloir changer ça avant votre départ pour l'Australie!

Geoff, rentré aujourd'hui même pour le far west. Toujours en forme et de bonne humeur, cette année avec toi et tous ces bons moments à l'escalade et à l'ultimate, c'était top! Bravo et merci pour ton apprentissage à vitesse incroyable du français, ça nous a évité de montrer nos talents linguistiques. Si la légende dit vrai, avec ton imitation parfaite de l'accent français tu vas faire fureur sur les plages californiennes.

Florent, merci pour ces petites sorties VTT dans des chemins toujours nouveaux, avec des supers découvertes. À quand le VTT électrique?

Et même si c'est difficile d'être exhaustif, merci à tous les autres doctorants, stagiaires et post-doc du labo, dont de nombreux sont partis parfois depuis bien longtemps, et avec qui j'ai aussi passé des supers moments : Miguel, Pedro, Kevin, Marielle, Théo, Aziliz, Alex H, Thibaud, Louis K, Camille, Francesca, Jeremy, Victor, Quentin, Lucas, Céline, Cindy, Tim&Tim, ...

Merci aussi à mes amis en dehors du labo : Kevin, Fanny, Thomas, Céline, Sherlock et Eric.

À mes futurs collègues profs de Maths : Nelson, Antoine et Tiphaine. On reste en contact pour créer du contenu de qualité pour nos élèves! Merci Tiphaine pour ton partage d'expériences.

Merci à mes collocs, Guillaume, Jade, et Violaine, d'avoir fait de notre appart un lieu plaisant à vivre au quotidien!

Et pour finir, un grand merci à ma famille pour m'avoir soutenu et encouragé de loin (voir de très loin en terme kilométriques) pendant toute la durée de ma thèse. Vivement les prochaines vacances ensemble!

Table des matières

Introduction	1
1 Médecine nucléaire	5
1.1 Introduction à l'imagerie nucléaire	5
1.2 Bruit en imagerie nucléaire	11
1.3 Modèles compartimentaux	13
1.4 Extraction des courbes d'activité	15
1.5 Cas d'étude : Évaluation de l'insulino-résistance à l'aide du 6-DIG	16
1.6 Conclusion	21
2 Séparation de sources	23
2.1 Généralités	24
2.2 Modèle de mélange linéaire en imagerie nucléaire	29
2.3 Spécificités de l'imagerie nucléaire	34
2.4 Contraintes de positivité	37
2.5 Méthodes de séparation de sources en imagerie nucléaire	47
2.6 Approche Tensorielle	57
2.7 Conclusion	67
3 Utilisation des régions d'intérêt pour la séparation de sources en imagerie nucléaire	69
3.1 Les ROI : un a priori acceptable en imagerie nucléaire	70
3.2 Utilisation des ROI comme masques pour les images factorielles	71
3.3 RUDUR : Un algorithme générique de séparation utilisant de façon robuste les ROI.	73
3.4 Conclusion	88

4 Résultats	91
4.1 Jeu de données synthétiques simples	93
4.2 Rénographies synthétiques réalistes	100
4.3 Rénographies cliniques	103
4.4 Évaluation de la sensibilité de la méthode	106
4.5 Utilisation des images d'intérêt pour l'évaluation d'une source de faible intensité	116
4.6 Données scintigraphiques issues de l'étude du 6-DIG	118
4.7 Conclusion sur les résultats	131
Conclusion	133
A Modèle compartimental et système d'équations différentielles du 6-DIG	137
B Calcul des dérivées partielles	139
B.1 Formules de dérivation matricielle	139
B.2 Dérivation selon A	139
B.3 Dérivation selon F	140
C Constantes de Lipschitz des dérivées partielles de f_{RUDUR}	141
C.1 Constante de Lipschitz de la dérivée partielle selon F	141
C.2 Constante de Lipschitz de la dérivée partielle selon A	141
D Séquences d'images	143
D.1 Données synthétiques simples, sans bruit de fond	144
D.2 Données synthétiques simples, avec bruit de fond	145
D.3 Rénographies synthétiques réalistes	146
D.4 Rénographies cliniques	147
D.5 Rénographies cliniques	148
D.6 Données synthétiques simples, avec source de faible intensité	149

D.7 Simulation d'une séquence 6-DIG	150
---	-----

D.8 Exemple de séquence 6-DIG (phase basale)	151
--	-----

Bibliographie	153
----------------------	------------

Introduction

De nombreux jeux de données sont composés d'un mélange de plusieurs sources d'informations, qui sont celles recherchées pour résoudre un problème particulier. Ces informations n'étant pas directement accessibles, il est nécessaire de les séparer afin de les traiter de façon efficace. L'objectif des méthodes de séparation de sources est d'isoler chacune des différentes contributions (les sources) à l'origine du mélange observé, le nombre de sources étant généralement supposé connu. Un exemple simple de problème de séparation de sources est présenté dans la figure 1. Dans cet exemple, la séquence d'image observée est composée de deux sources mélangées. L'objectif est de retrouver les signatures spatiales et temporelles de chacune de ces sources, à partir de la séquence d'images. En pratique, les observations bruitées et le nombre de sources rendent le problème de séparation complexe. Ces problématiques de mélange se retrouvent dans de nombreux domaines où il y a besoin de séparer différentes informations : des voix en traitement du son, des galaxies ou des natures de sol en imagerie hyperspectrale, des molécules en chimiométrie, des utilisateurs en télécommunications, ou encore des organes et des tissus en imagerie médicale et, plus particulièrement, en imagerie nucléaire.

L'imagerie nucléaire est une technique se basant sur le suivi d'éléments radioactifs injectés dans le corps d'un patient. Ce suivi se fait grâce à une gamma-caméra produisant une séquence d'images, chacune représentant la radioactivité détectée pendant un intervalle de temps. Ces images sont une approximation de l'évolution de la distribution du traceur dans le corps du patient. Cette méthode d'imagerie permet de fournir des informations sur l'état de fonctionnement des organes et des tissus. Elle est par exemple utilisée pour étudier la perfusion de différents organes et ainsi diagnostiquer des maladies coronariennes, des embolies pulmonaires ou encore mesurer la capacité de filtration des reins.

Dans la plupart de ces applications cliniques, les diagnostics se font à partir des courbes d'activité d'organes ou de tissus, appelés compartiments physiologiques. Une manière simpliste d'extraire une courbe d'activité est de sélectionner les pixels contenant le compartiment physiologique cible, ces pixels formant ainsi ce qui est appelé une région d'intérêt. Afin de limiter l'influence du bruit, la courbe d'activité est ensuite obtenue en moyennant spatialement l'activité de cette région d'intérêt. Cette méthode manuelle est sujette aux variations intra et inter-observateurs dans la sélection de la région. De plus, les courbes d'activité obtenues de cette manière sont composées d'un mélange de l'activité des différents compartiments présents dans la région dessinée. En effet, ces compartiments peuvent se superposer lorsqu'ils sont projetés sur une image. Dans les séquences 3D, les mélanges existent aussi du fait du volume partiel et de l'imperfection des instruments de mesure. D'autre part, sélectionner une petite région ne contenant a priori que du compartiment physiologique cible n'est pas toujours suffisant à cause du bruit important, ni toujours possible selon les applications visées. Par exemple, le myocarde entourant le ventricule gauche ne peut en général pas se retrouver seul sur un pixel du fait de la résolution spatiale limitée.

Des méthodes de séparation de sources ont donc été proposées afin d'extraire ces courbes

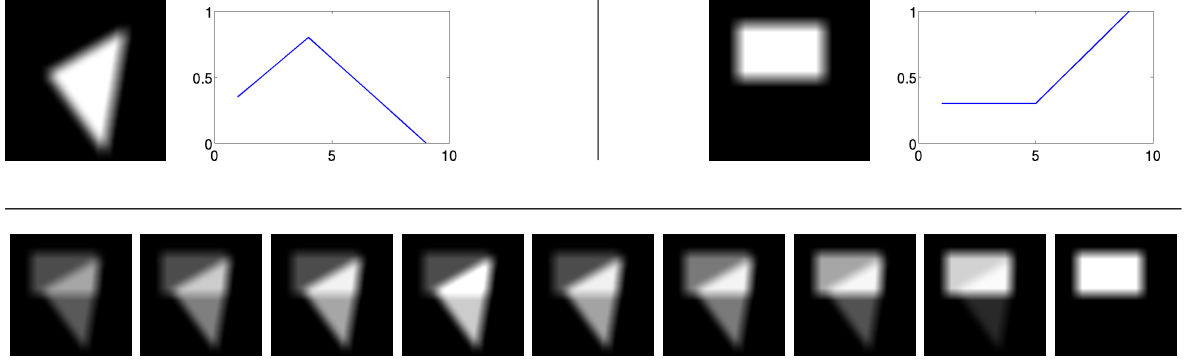


FIGURE 1 – Exemple simple de mélange de deux sources. Les profils spatial et temporel de chacune des sources sont représentés sur la partie haute de la figure, et leur mélange observé aux temps $t=1$ à $t=9$ sur la partie basse. Le mélange est la somme des deux séquences d’images produites par chaque source à partir de leurs profils spatial et temporel. L’objectif des méthodes de séparation de sources est de retrouver les profils de chaque source à partir des données mélangées observées.

d’activité tout en prenant en compte le mélange présent dans les séquences d’images. Les sources à séparer sont ici les compartiments physiologiques. Elles sont caractérisées par leurs signatures temporelle et spatiale, qui sont respectivement leur courbe d’activité et leur carte de présence. Les cartes de présence sont des images invariantes dans le temps représentant la position et l’importance de la source. Les méthodes de séparation se basent sur un modèle de mélange des sources. En imagerie nucléaire on utilise le modèle de mélange linéaire, qui stipule que la courbe d’activité d’un pixel est la somme des courbes d’activité des différentes sources présentes sur ce pixel, pondérées chacune par leur importance sur ce pixel.

Dans sa forme la plus simple, le problème de séparation s’apparente à un problème de factorisation matricielle. Néanmoins, ce problème est mal posé. Il est en effet sous-déterminé, ce qui signifie qu’il n’y a pas unicité de la solution. Parmi les solutions possibles, une seule a le sens physiologique recherché, et rien ne permet aux algorithmes de converger vers cette solution. Pour lever l’indétermination, il est donc nécessaire de rajouter des connaissances a priori sur les sources recherchées. Malgré l’apparente analogie des problèmes de séparation des différents domaines, les connaissances ajoutées sont spécifiques à un domaine ou une application, et il n’est donc généralement pas possible d’utiliser les méthodes développées dans d’autres contextes. Différentes connaissances a priori ont été proposées en imagerie nucléaire. La plus simple est la contrainte de positivité sur les courbes d’activité et les cartes de présence des sources recherchées. Cette contrainte impose de trouver une solution ne contenant aucune valeur négative, ni dans les courbes d’activité, ni dans les cartes de présence. Cette connaissance n’étant pas suffisante pour lever l’indétermination, d’autres connaissances a priori ont été proposées, concernant par exemple la forme des courbes d’activité recherchées ou bien des contraintes sur la parcimonie spatiale des sources.

Malgré cela, les méthodes de séparation en imagerie nucléaire ne sont que très peu utilisées en pratique. La raison est que les algorithmes actuels manquent de fiabilité et souffrent de plu-

sieurs limites dues à la difficulté du problème de séparation. Premièrement, les connaissances a priori utilisées ne sont parfois pas assez informatives ou trop peu réalistes. Deuxièmement, le modèle de mélange linéaire communément utilisé doit être remis en question et sa portée doit être limitée. En effet, il n'est une bonne approximation que dans une zone restreinte des images étudiées. Enfin, l'initialisation de certaines méthodes est aléatoire, alors que la solution obtenue dépend des valeurs de départ fournies aux algorithmes.

En imagerie nucléaire, les régions d'intérêt forment des connaissances accessibles qui apportent des informations particulièrement utiles pour la séparation, notamment sur la position des sources. Elles n'ont pourtant été que très peu utilisées dans ce contexte. Probablement parce que les méthodes de séparation d'imagerie nucléaire ont été développées pour s'affranchir des limites de la méthode simpliste d'extraction de courbes d'activité se basant sur le calcul de l'activité moyenne d'une région d'intérêt.

Objectif de la thèse

L'objectif principal de cette thèse est de proposer des solutions aux limitations des méthodes actuelles afin de rendre la séparation plus fiable et plus efficace. Pour ce faire, l'intégration de diverses connaissances a priori sera étudiée. Une attention particulière sera portée à la manière d'intégrer les régions d'intérêt dans les méthodes de séparation de sources. Ces connaissances a priori doivent permettre de faire face à la sous-détermination du problème initial et faciliter sa résolution. De même, une remise en question du modèle de mélange linéaire est attendue.

Une des motivations de cette thèse est également d'obtenir des résultats sur un jeu de données cliniques utilisant le traceur 6-DIG, pouvant mener à une mesure de l'insulino-résistance, un précurseur du diabète. Afin d'obtenir cette mesure, les courbes d'activité du myocarde et du sang doivent être extraites de séquences d'images très bruitées.

Plan de la thèse

Cette thèse s'articule comme suit :

Chapitre 1 : Le premier chapitre introduit le contexte de l'imagerie nucléaire et ses applications. Il y est montré la nécessité de l'utilisation de méthodes de séparation de sources pour l'extraction des courbes d'activité. Ceci est illustré en fin de chapitre par le détail de l'étude se basant sur des scintigraphies utilisant le 6-DIG.

Chapitre 2 : La problématique de la séparation de sources est présentée. Le modèle de mélange linéaire associé est détaillé dans le cadre de l'imagerie nucléaire ainsi que l'utilisation des contraintes de positivité et les conditions d'unicité de la séparation. L'état de l'art des méthodes de séparation est présenté, en se focalisant sur les méthodes adaptées à l'imagerie

nucléaire et leurs limites. En fin de chapitre est envisagée une approche tensorielle afin de s'affranchir de la sous-détermination du problème de séparation, ce qui constitue une première contribution de cette thèse.

Chapitre 3 : Les principales contributions de cette thèse sont développées dans ce chapitre. Une méthode de séparation de sources générique intégrant différentes connaissances a priori est développée. La connaissance majeure utilisée est la région d'intérêt de chaque source recherchée. Des a priori d'ordre temporels sont également intégrés. La méthode prend la forme d'une fonctionnelle à minimiser, composée d'un terme d'attache aux données, d'une pénalisation relative aux régions d'intérêt et d'un terme de régularisation des courbes d'activité. Une initialisation non-arbitraire utilisant les régions d'intérêt est proposée. Le procédé d'optimisation de cette fonctionnelle est détaillé, et un code source MATLAB commenté est disponible en ligne¹.

Chapitre 4 : Les performances de l'algorithme présenté au chapitre 3 sont évaluées sur différents jeux de données synthétiques et cliniques et comparées aux méthodes principales de l'état de l'art. Une attention particulière est portée à la robustesse de l'algorithme au bruit et vis à vis de la sélection des régions d'intérêt ainsi que du choix des constantes de pondération de la fonctionnelle. Des données synthétiques simulant les scintigraphies 6-DIG sont également créées afin de mieux analyser les jeux de données cliniques utilisant ce traceur. Les tests effectués sur la méthode sont reproductibles grâce aux scripts MATLAB disponibles en ligne¹.

Publications

Les contributions de cette thèse ont donné lieu à un article dans une revue, et deux publications dans des conférences.

Journal international

Filippi, M., Desvignes, M., & Moisan, E. (January 2018). Robust Unmixing of Dynamic Sequences Using Regions of Interest. *IEEE transactions on medical imaging*, 37(1), 306-315.

Conférence internationale

Filippi, M., Desvignes, M., Moisan, E., Ghezzi, C., Perret, P., & Fagret, D. (2016). Factor Analysis of Dynamic Sequence with Spatial Prior for 2D Cardiac Spect Sequences Analysis. In *International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 228-237). Springer International Publishing.

Conférence nationale

Filippi, M., Desvignes, M., Moisan, E., Ghezzi, C., Perret, P., & Fagret, D. (2015). A priori spatiaux et analyse factorielle de séquences scintigraphiques. In *Proc. GRETSI*.

1. <http://www.gipsa-lab.grenoble-inp.fr/~marc.filippi/>

Médecine nucléaire

Sommaire

1.1	Introduction à l'imagerie nucléaire	5
1.1.1	Imagerie fonctionnelle	5
1.1.2	Principe de fonctionnement	6
1.1.3	Type d'imagerie	7
1.1.4	Applications de l'imagerie nucléaire dynamique	10
1.2	Bruit en imagerie nucléaire	11
1.2.1	Bruit de comptage	11
1.2.2	Autres sources de bruit	13
1.3	Modèles compartimentaux	13
1.4	Extraction des courbes d'activité	15
1.4.1	Région d'intérêt (ROI)	15
1.4.2	Mélange des compartiments physiologiques	15
1.4.3	Nécessité de la séparation de sources	16
1.5	Cas d'étude : Évaluation de l'insulino-résistance à l'aide du 6-DIG	16
1.5.1	La résistance à l'insuline	17
1.5.2	Mesure de l'insulino-résistance à l'aide du 6-DIG	19
1.6	Conclusion	21

La médecine nucléaire utilise la radioactivité à des fins de diagnostic et de thérapie. Dans cette thèse, nous nous intéressons uniquement à la médecine nucléaire comme outil de diagnostic. Ce chapitre est une courte introduction à l'imagerie nucléaire et à ses applications. Le besoin de méthodes de séparation de sources y est présenté, ainsi que le détail d'un exemple d'étude avec le 6-DIG, un analogue radioactif du glucose utilisé pour évaluer le degré d'insulino-résistance, un défaut dans la stimulation du transport du glucose par l'insuline.

1.1 Introduction à l'imagerie nucléaire

1.1.1 Imagerie fonctionnelle

L'imagerie nucléaire fait partie de la catégorie des méthodes d'imagerie fonctionnelle (aussi appelée imagerie physiologique). Ces méthodes sont différentes des méthodes d'imagerie anatomiques ou structurales plus classiques. Ces dernières, tels la radiologie X, l'échographie,

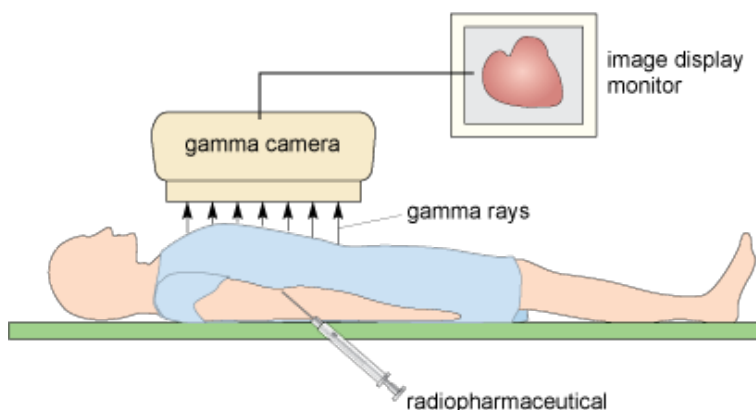


FIGURE 1.1 – En imagerie nucléaire, le traceur radioactif est généralement injecté par voie intraveineuse puis se diffuse dans le corps grâce à la circulation sanguine. L'évolution de sa concentration dans le corps peut être estimée grâce à son activité radioactive, détectée à l'aide d'une gamma-caméra. (Source : <http://prospect.rsc.org/metalsandlife/>)

l'imagerie par résonance magnétique, ou encore le scanner, sont des techniques d'imagerie qui donnent accès aux informations de nature anatomique concernant la structure et la forme des organes et des tissus. Les méthodes d'imagerie fonctionnelle fournissent quant à elles des informations fonctionnelles et métaboliques sur les organes et tissus observés. Elles renseignent sur la physiologie du patient et rendent accessibles des informations quantifiables sur l'état de fonctionnement d'un organe [Gul+10], ce qui n'est pas possible en imagerie anatomique. Grâce à ces informations, des dépistages peuvent être faits bien avant la présence de signes cliniques visibles. Ces méthodes permettent également de confirmer ou d'infirmer des diagnostics demandant vérifications.

1.1.2 Principe de fonctionnement

En imagerie nucléaire, afin d'obtenir des informations fonctionnelles sur les organes et tissus, une petite dose de radiopharmaceutique (traceur) est injectée dans le corps du patient, généralement par voie intraveineuse. Ce traceur est généralement composé d'un vecteur (molécule dont on veut suivre la biodistribution) et d'un marqueur (isotope radioactif). En utilisant un dispositif d'acquisition spécifique dirigé vers le patient, il est possible d'imager l'activité de cet isotope. Ces dispositifs permettent d'obtenir des séquences d'images représentant l'évolution temporelle de l'activité du traceur dans le corps du patient. Comme les isotopes radioactifs sont attachés aux traceurs, ces séquences d'images peuvent être considérées comme une approximation de l'évolution de la biodistribution du vecteur dans le corps du patient. Il est ainsi possible de suivre dynamiquement la concentration du radiopharmaceutique. En choisissant un traceur adapté à l'étude, les images fournissent des informations sur l'état de fonctionnement des organes et tissus étudiés. Le principe général de fonctionnement est illustré en figure 1.1.

Le traceur peut être un simple isotope radioactif, une molécule ou une cellule marquée

par un isotope radioactif. Le choix du traceur dépend du type d'étude. Par exemple, si une tumeur osseuse est recherchée, un traceur se fixant sur les os est utilisé. Une concentration forte du traceur à un endroit inattendu indiquera une présence potentielle de tumeur. Pour visualiser un organe ou un tissu en particulier, on prend généralement un traceur qui se fixe sur l'objet d'intérêt. D'autres traceurs permettent de suivre une molécule spécifique. Par exemple, il existe des analogues radioactifs du glucose, qui se comportent comme le glucose tel le 6-DIG [Hen+97], le 3-OMG [KMPL81] ou le FDG [Gal+78]. Il est donc possible grâce à ces traceurs d'avoir des images approchant la concentration du glucose et de suivre son transport et son métabolisme.

Les faibles doses utilisées en imagerie nucléaire permettent d'observer le fonctionnement des organes et des tissus sans le perturber (doses infra-pharmacologiques). Il est aussi nécessaire que ces doses soient faibles pour que les risques liés aux radiations restent négligeables [Faz+09]. Les techniques de la médecine nucléaire sont considérées comme non-invasives car aucune effraction de la peau n'est commise, à part pour l'injection du traceur.

Deux types d'études existent.

- **Les études statiques :** Dans les études statiques, une seule image est produite pour faire le diagnostic. C'est le cas par exemple dans les scintigraphies osseuses [ES+06], où cette seule image suffit à détecter la présence de tumeurs.
- **Les études dynamiques :** Certains diagnostics ne peuvent se faire qu'en suivant l'évolution de la concentration du traceur dans un organe ou dans un tissu. Dans ces études, il est donc nécessaire d'acquérir une séquence d'images, chacune de ces images représentant la concentration moyenne du traceur pendant un intervalle de temps. C'est le cas de l'imagerie de perfusion qui permet de détecter les maladies coronariennes. Selon la cinétique des phénomènes observés, cet intervalle de temps peut être de l'ordre de la minute ou bien de la seconde.

Des exemples d'images obtenues dans ces deux techniques sont présentés dans les figures 1.2 et 1.3. Dans la suite, nous nous intéressons exclusivement aux études dynamiques.

1.1.3 Type d'imagerie

L'enjeu en imagerie nucléaire n'est pas seulement de mesurer l'activité du traceur, mais aussi d'estimer sa localisation. C'est ce qui constitue toute la difficulté de la construction des images d'activité. Il existe deux méthodes d'imagerie :

- **La scintigraphie :** En scintigraphie, le dispositif d'acquisition est appelé gamma-caméra. Lorsqu'un isotope radioactif se désintègre, il émet un photon gamma. Le rôle des gamma-caméras est de détecter et de localiser ces photons. Les rayons gamma étant très pénétrants, ils peuvent traverser le corps du patient et atteindre la caméra. Un collimateur est intégré à la caméra afin de sélectionner les rayons gamma incidents. Le rôle du collimateur est de laisser passer uniquement les rayons gamma incidents perpendiculaires au plan de la caméra. Ce collimateur est simplement constitué d'un

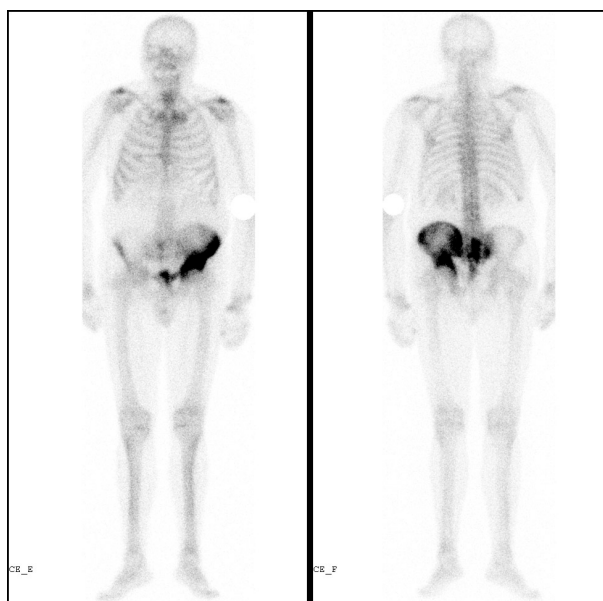


FIGURE 1.2 – Exemple de scintigraphie osseuse planaire statique corps entier. Le traceur utilisé s’accumule dans les os ayant une forte activité ostéoblastique. Les deux images correspondent à la vue antérieure (caméra au dessus du patient) et à la vue postérieure (caméra sous le patient). Elles contiennent l’accumulation de l’activité détectée par les caméras pendant l’étude. Les images montrent une très forte activité ostéoblastique au niveau du bassin gauche et permettent de diagnostiquer une maladie de Paget. (Source : <http://imagerie-moleculaire-poitiers.fr>)

matériau absorbant (le plomb par exemple) comportant des trous parallèles laissant passer les rayons gamma. Ainsi, si le rayon gamma détecté par la caméra n’a pas subi de modification de trajectoire (par diffusion Compton) depuis la désintégration de l’isotope associé, alors il provient nécessairement d’une direction perpendiculaire au collimateur, car sinon il aurait été absorbé par ce dernier. La collimation réalisée permet ainsi de localiser les rayons gamma incidents au prix d’une perte de signal car les rayons gamma sont émis de façon isotrope lors des désintégrations.

Lorsqu’un rayon gamma franchit le collimateur, il est converti en lumière par un cristal scintillateur. Cette lumière est ensuite elle-même convertie en signal électrique par des photomultiplicateurs. Ce signal électrique est ensuite analysé afin de ne sélectionner que les photons gamma ayant l’énergie attendue.

La scintigraphie de base est la scintigraphie planaire, qui permet d’acquérir une séquence temporelle d’images 2D de l’activité. Il est également possible de reconstruire des volumes 3D par tomographie grâce à la scintigraphie en plaçant plusieurs plans de capteurs autour de la caméra ou bien en faisant tourner la caméra autour du patient [Gul+10]. On appelle cette technique la tomographie par émission monophotonique (TEMP, ou SPECT en anglais).

- **La tomographie par émission de positons (TEP) :** Lors de la désintégration de certains isotopes radioactifs, des positons sont émis (désintégration β^+). Après un parcours d’une longueur très faible, en moyenne de l’ordre du millimètre, ce positon

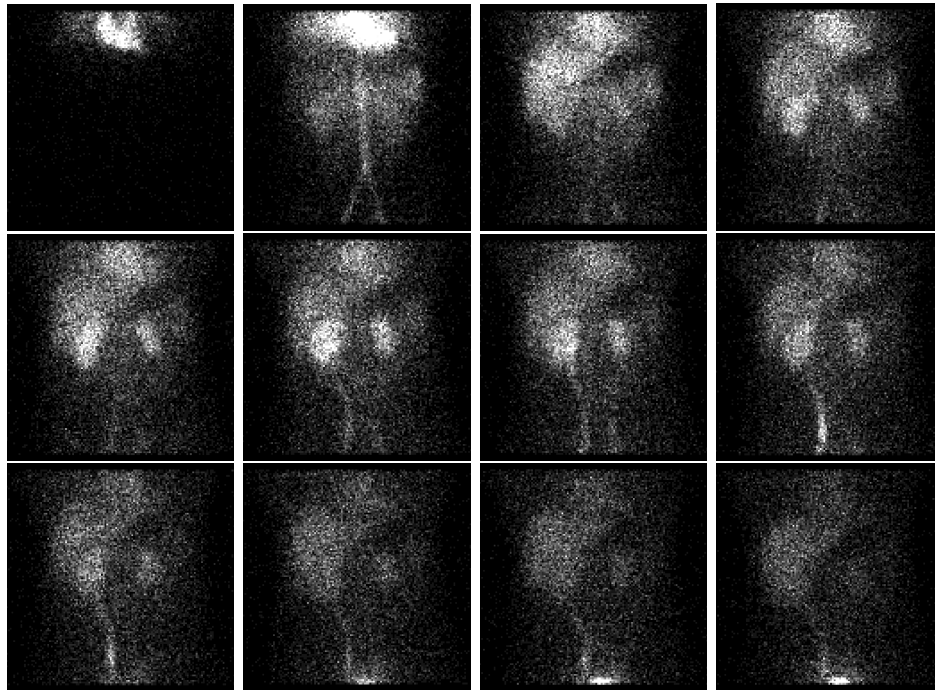


FIGURE 1.3 – Exemple de scintigraphie planaire dynamique. Chaque image est la somme de l'activité reçue pendant 10 secondes (chronologiquement de gauche à droite puis de haut en bas). Le traceur utilisé est du ^{99m}Tc -MAG3, afin d'obtenir des informations sur l'état de fonctionnement des reins. Seulement quelques images de la séquence entière (180 images) sont représentées ici. Le traceur est éjecté du coeur puis est filtré par les reins, et est enfin transporté dans la vessie pour être éliminé. Le rein gauche (à droite dans les images) a une concentration en traceur beaucoup plus faible que le rein droit, il est donc mal perfusé.

s'annihile avec un électron, produisant alors deux rayons gamma de même énergie (511 keV) émis dans des directions opposées. C'est la détection en coïncidence de ces deux photons qui permet de localiser avec une bonne approximation le lieu de la désintégration (voir figure 1.5).

La caméra TEP possède des capteurs situés tout autour du patient [Bai+05]. Lorsque deux capteurs détectent simultanément un événement, c'est à dire un photon, c'est qu'une annihilation, et donc aussi une désintégration, ont eu lieu à un endroit très proche de la ligne reliant ces deux capteurs. L'activité dans le volume observé est ensuite déduite des informations collectées par les capteurs grâce à un algorithme de reconstruction.

La résolution spatiale des caméras en imagerie nucléaire est de l'ordre de quelques millimètres à 1 centimètre.

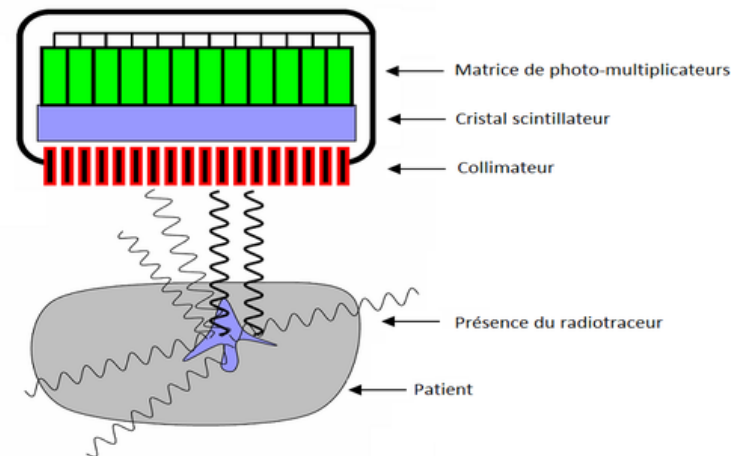


FIGURE 1.4 – Illustration du principe de détection en scintigraphie. Parmi les rayons gamma émis depuis le patient, seuls ceux en gras sont enregistrés par la caméra. Les autres sont atténués dans le collimateur car ils arrivent avec une direction non orthogonale à la caméra, ou bien ne l'atteignent tout simplement pas. Ce système de collimation permet de localiser l'origine des désintégrations détectées.

1.1.4 Applications de l'imagerie nucléaire dynamique

L'imagerie nucléaire trouve de nombreuses applications dans la détection précoce et le suivi des pathologies. Sans être exhaustive, cette section en présente quelques-unes.

Les examens les plus classiques en scintigraphie dynamique concernent la perfusion d'un organe ou d'un tissu donné. La perfusion correspond à la quantification de l'alimentation d'un organe ou d'un tissu en un composé chimique particulier. Ainsi, mesurer la perfusion du myocarde permet de détecter des zones ischémiques où l'afflux sanguin est réduit [Rag+93]. En scintigraphie, on peut également quantifier la capacité de filtration des reins. Ceci permet par exemple de vérifier s'il y a une complication rénale ou si le rein en question peut faire l'objet d'un don, ou bien de faire le suivi d'une greffe rénale [Kra+93]. Des embolies pulmonaires (obstruction de l'artère pulmonaire) peuvent aussi être détectées grâce à l'analyse de la perfusion des poumons [Coc+03]. Les données 3D+t fournies par la TEMP permettent de mieux visualiser la répartition du traceur dans le corps et de faire des diagnostics plus précis que la scintigraphie planaire [MAG93]. La TEMP est aujourd'hui la technologie la plus utilisée pour le diagnostic et le suivi des ischémies myocardiques [Ber+10].

En TEP, on s'intéresse principalement au myocarde, au cerveau et aux cancers. Le traceur le plus utilisé est le fluorodésoxyglucose (FDG) qui est un analogue du glucose marqué au fluor 18 [Gal+78]. Ce dernier se fixe dans les tissus ayant un fort métabolisme et donc ayant un important besoin en glucose. Ceci permet notamment de détecter des régions ischémiques ou nécrosées dans le myocarde et aide ainsi à déterminer l'intervention adéquate [Gho+10]. Le FDG est aussi utilisé en cancérologie [HKK98], car les métastases ont un fort métabolisme. Le FDG se fixe donc en grande quantité sur les tumeurs, ce qui permet de les détecter et aussi

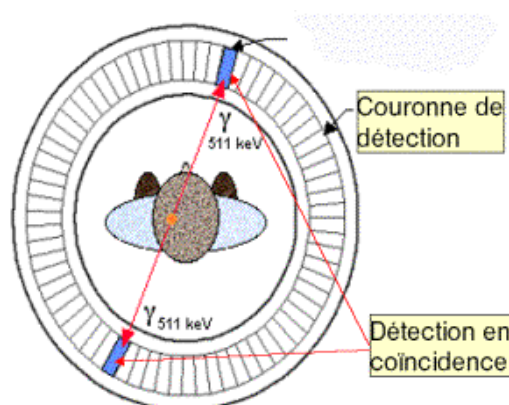


FIGURE 1.5 – Illustration de la détection en coïncidence en TEP. Après collision entre un électron et un positon émis par la désintégration de l'isotope radioactif, deux photons gamma sont émis dans des directions opposées. S'ils sont détectés simultanément par des capteurs sur la couronne de détection, alors la position de la désintégration peut être approchée grâce à la position de ces deux capteurs.

de suivre leur évolution, alors qu'elles sont difficilement détectables avec les autres méthodes d'imagerie. Une autre application majeure de la TEP avec le FDG est l'étude du lien entre différentes pathologies cérébrales, comme la maladie d'Alzheimer ou les déficits cognitifs légers, et le métabolisme des cellules cérébrales [Vid00].

1.2 Bruit en imagerie nucléaire

Les séquences en imagerie nucléaire sont en général perturbées par un bruit important, elles ne représentent donc pas exactement l'évolution de la concentration du traceur dans l'organisme. La grande partie de ce bruit est un bruit de comptage lié à l'approximation de la concentration du traceur par une mesure de la radioactivité. D'autres éléments, comme la diffusion Compton, viennent également ajouter du bruit.

1.2.1 Bruit de comptage

Afin de mieux comprendre l'origine du bruit en imagerie nucléaire, rappelons le principe d'acquisition des images. Le but est d'obtenir l'évolution dynamique de la concentration du traceur injecté dans le corps. Cette concentration ne peut pas être obtenue directement mais peut être estimée grâce à une mesure de la radioactivité du radiopharmaceutique. Ceci constitue une première approximation car pendant un intervalle de temps court, une toute petite partie du traceur se désintègre, on n'a donc accès qu'à une petite partie de l'information. De plus, la radioactivité est mesurée grâce à une gamma caméra. Pour obtenir une carte spatiale de la radioactivité, la gamma caméra ne peut détecter qu'une petite partie des rayons gamma émis par le radiopharmaceutique du fait de la collimation, ce qui constitue une seconde

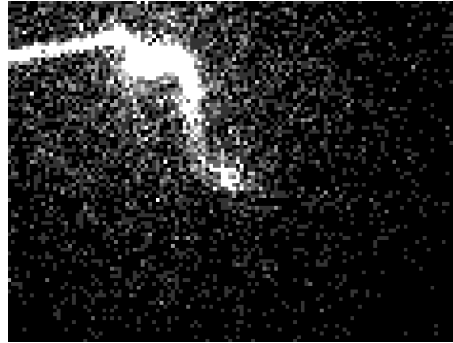


FIGURE 1.6 – Arrivée du traceur (6-DIG) dans le coeur. Le traceur est injecté dans le bras par voie intraveineuse. Dans cette image, le traceur suit le réseau veineux (veine céphalique, veine sous-claviaire et veine cave supérieur) pour rejoindre le ventricule droit. À ce stade le traceur est exclusivement dans le réseau veineux reliant le bras au coeur. Pourtant, de l'activité est détectée en dehors de ces zones. Ceci est dû à la diffusion Compton et aux imperfections de la collimation.

approximation.

Il y a donc deux niveaux consécutifs d'approximation dans l'estimation de la concentration du traceur. Ces deux niveaux d'approximation ne permettent de connaître la position que d'une infime partie des molécules du radiotraceur injecté. Chacun de ces niveaux d'approximation peut être considéré comme un bruit de comptage, c'est à dire un bruit poissonien étant donné le nombre important de molécules de traceur injecté et la faible probabilité de détection. Comme ces deux niveaux d'approximation sont consécutifs il en résulte un bruit poissonien plus fort. Il faut même rajouter un troisième niveau d'approximation, car tous les photons gamma émis dans la direction de la caméra ne vont pas forcément l'atteindre. Ceci est dû à l'atténuation des gamma. Une partie de ces rayons gamma est atténuée par la matière traversée (le corps du patient) en interagissant avec elle.

Une première manière de diminuer le bruit de comptage est d'augmenter les doses injectées. En effet, le bruit poissonien est d'autant plus faible, relativement au signal, que les doses injectées sont importantes. Cependant, plus ces doses sont importantes, plus le patient est exposé aux risques liés à la radioactivité. Les doses sont donc limitées [MS11] ce qui dégrade la qualité des images produites. Une autre manière est d'augmenter la durée d'acquisition de chaque image, ce qui permet d'avoir plus de coups dans l'image et donc de diminuer l'impact du bruit de comptage. Néanmoins ce n'est pas toujours possible, notamment lorsque la dynamique du traceur est très rapide et que sa concentration varie rapidement dans les zones d'intérêt. Il faut donc trouver le bon compromis entre rapport signal à bruit et résolution temporelle du signal. Enfin, il est aussi possible de grouper des pixels afin d'augmenter le nombre de coups dans les super-pixels résultants. Une fois encore cette technique permet d'améliorer le rapport signal à bruit mais dégrade la résolution spatiale. Une bonne résolution spatiale peut être indispensable, notamment lorsque les compartiments physiologiques d'intérêt sont de petite taille, comme le myocarde par exemple.

1.2.2 Autres sources de bruit

Le bruit de comptage est une des principales causes de bruit en imagerie nucléaire mais ce n'est pas le seul. Une autre source de bruit vient de la diffusion des rayons gamma et leurs interactions avec la matière. Les rayons gamma peuvent changer de direction lorsqu'ils entrent en contact avec un électron libre d'un atome. C'est ce qu'on appelle la diffusion Compton. La longueur d'onde du rayon gamma est également modifiée. La conséquence pratique pour l'imagerie nucléaire est que ce rayon gamma sera enregistré (si il l'est) comme étant originaire du lieu de la collision, et non du lieu de la désintégration. Différentes techniques sont utilisées pour corriger ce problème, certaines sont intégrées en routine dans les gamma-caméras [HBB11].

En scintigraphie et en TEMP, la collimation n'étant pas parfaite, des rayons gamma qui ne sont pas parallèles aux tubes du collimateur peuvent quand même être comptabilisés en n'étant pas atténués par le plomb. L'appareil de mesure possède donc une certaine fonction d'étalement du point, affaiblissant la résolution spatiale de l'instrument [Cec+15]. Ainsi une source ponctuelle génèrera de l'activité non pas uniquement sur un pixel, mais dans un voisinage plus important. Les imperfections de la collimation ainsi que la diffusion Compton donnent aux images leur aspect bruité, en élargissant le support spatial des organes et tissus (voir figure 1.6).

Du fait de l'électronique présente dans les caméras, les séquences d'images sont également affectées par un bruit de mesure, souvent considéré comme gaussien. Dans le cas de la TEP et de la TEMP, la reconstruction génère également des erreurs [Niu+11]. Pour la TEP, la détection simultanée de deux photons n'étant pas issue de la même interaction peut aussi générer du bruit. Ce phénomène est appelé la coïncidence fortuite.

1.3 Modèles compartimentaux

En imagerie nucléaire dynamique, les diagnostics peuvent se faire de façon qualitative ou bien quantitative. Dans le premier cas, un spécialiste établit un diagnostic en observant simplement les séquences d'images fournies par la caméra. Cependant, faire des diagnostics de façon qualitative n'est pas toujours possible ni toujours suffisant, à cause du manque de reproductibilité des résultats. Afin de raisonner quantitativement, il faut généralement extraire les courbes d'activité des organes et tissus étudiés. Des indicateurs quantitatifs sont dérivés de ces courbes d'activité et servent ensuite de base aux diagnostics.

Dans un cas d'étude complexe, il peut s'avérer nécessaire de modéliser les interactions entre les différents compartiments physiologiques (organes et tissus) mis en jeu par le traceur afin de calculer ces indicateurs. En imagerie nucléaire, ceci se fait par le biais de modèles compartimentaux [GGC01]. Ces derniers représentent de manière simplifiée la diffusion du traceur entre les organes et tissus d'intérêt. L'hypothèse sous-jacente est que le traceur circule entre ces différents compartiments physiologiques avec une certaine affinité pour chaque compartiment.

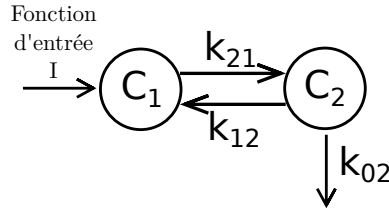


FIGURE 1.7 – Exemple simple de modèle compartimental, à deux compartiments C_1 et C_2 . Le traceur voyage de C_1 à C_2 dans les deux sens. Une fonction d'entrée alimente C_1 , tandis que le compartiment C_2 envoie du traceur vers l'extérieur (compartiment 0).

Un exemple de modèle compartimental simple est disponible en figure 1.7. Dans ce modèle une fonction d'entrée alimente un premier compartiment C_1 , ce qui correspond généralement à l'injection du traceur dans le sang. Les compartiments sont reliés entre eux par une cinétique k_{ij} . Pendant un intervalle de temps infinitésimal, la quantité de traceur se déplaçant de C_i à C_j est proportionnelle à la quantité de traceur dans C_i et à k_{ji} . Le compartiment 0 correspond à l'extérieur.

Les modèles compartimentaux sont des représentations graphiques de systèmes d'équations différentielles. Par exemple, le modèle compartimental de la figure 1.7 peut se résumer au système linéaire suivant :

$$\begin{cases} \frac{dq_1(t)}{dt} = k_{12}q_2(t) - k_{21}q_1(t) + \frac{dI(t)}{dt} \\ \frac{dq_2(t)}{dt} = k_{21}q_1(t) - (k_{12} + k_{02})q_2(t) \end{cases} \quad (1.1)$$

où $q_1(t)$ et $q_2(t)$ sont les quantités de traceur présentes respectivement dans les compartiments C_1 et C_2 , et $I(t)$ la quantité de traceur injecté dans C_1 depuis l'extérieur au temps t . Ce système d'équations différentielles est beaucoup plus intelligible lorsqu'il est représenté sous forme de modèle compartimental.

Les cinétiques k_{12} , k_{21} et k_{02} sont des indicateurs quantitatifs très intéressants pour faire des diagnostics. La connaissance de leur valeur peut mettre en évidence des problèmes liés à la perfusion. Il est aussi possible de suivre l'évolution de la valeur de ces cinétiques suite à l'injection d'un produit, afin d'évaluer l'effet de ce dernier sur le métabolisme [Per+07]. À partir des séquences d'images, il est possible d'extraire les courbes d'activité reliées à q_1 , q_2 ainsi qu'à I . Grâce à ces courbes d'activité et au système d'équations différentielles du modèle compartimental, les cinétiques k_{12} , k_{21} et k_{02} peuvent être calculées. Certains logiciels spécialisés dans la modélisation compartimentale, comme PMOD¹ ou SAAM² permettent de faire ces calculs.

La modélisation des interactions entre les compartiments physiologiques via les modèles compartimentaux est donc très utile en pratique. Il est cependant nécessaire de connaître précisément l'évolution de la radioactivité des compartiments physiologiques pour pouvoir

1. <http://www.pmod.com/web/>

2. <https://tegvirginia.com/software/saam-ii/>

calculer les cinétiques d'échange de façon fiable. La section suivante s'intéresse à l'extraction de ces courbes d'activité.

1.4 Extraction des courbes d'activité

L'extraction de courbes temporelles d'activité des compartiments physiologiques d'intérêt est une tâche courante en imagerie nucléaire. Ces courbes servent de base aux diagnostics futurs, en les exploitant individuellement ou bien en les mettant en lien dans un modèle compartimental.

1.4.1 Région d'intérêt (ROI)

Les séquences d'images étant généralement bruitées, les courbes d'activité ne sont pas calculées à partir d'un unique pixel, mais à partir d'une région d'intérêt (ROI) [Cos85]. Une région d'intérêt est un ensemble de pixels supposés contenir le même compartiment physiologique. La courbe d'activité temporelle de ce compartiment physiologique est calculée en faisant la moyenne spatiale des courbes d'activité des pixels appartenant à cette ROI. Cette région est dessinée manuellement par un spécialiste, ou bien automatiquement dans de plus rares cas.

Il existe plusieurs moyens de tracer manuellement la ROI d'un compartiment physiologique [GPS11]. Si ce compartiment concentre une bonne partie du traceur dans la séquence entière, la ROI peut être tracée sur une image sommée. Cette dernière contient la somme de l'activité reçue par chaque pixel durant la totalité de l'étude. Sur cette image sommée, la région d'intérêt du compartiment physiologique cible peut être tracée en accord avec les connaissances a priori sur la forme de ce compartiment. Si le compartiment concentre une bonne partie du traceur seulement pendant un court moment, la ROI est tracée soit sur une des images où sa concentration est forte, soit sur la somme d'un petit nombre de ces images.

1.4.2 Mélange des compartiments physiologiques

Les courbes d'activité extraites à partir de régions d'intérêt souffrent cependant d'importants problèmes. Premièrement, le choix des régions d'intérêt est difficilement automatisable, et la sélection manuelle est dépendante de l'utilisateur, avec des variations inter-observateur et intra-observateur [Cos85]. Deuxièmement, les courbes d'activité issues d'un moyennage spatial des régions d'intérêt ne contiennent pas uniquement de l'activité issue du compartiment physiologique cible. En effet, l'activité présente sur cette région d'intérêt est composée d'un mélange d'activité de différentes origines. Ceci à cause de plusieurs raisons :

- **Superposition en imagerie planaire.** Les objets étudiés en imagerie nucléaire sont des objets en trois dimensions spatiales. Ce qui est observé en imagerie planaire est une projection 2D de ces objets 3D. Du fait des projections, des objets se situant à des

profondeurs différentes peuvent se chevaucher sur les images produites par la gamma-caméra. Ceci est par exemple particulièrement vrai pour le myocarde et les ventricules cardiaques. Comme le myocarde englobe les ventricules, leurs compartiments physiologiques se retrouvent entremêlés dans les images. Ainsi la région d'intérêt 2D du myocarde contiendra aussi en grande partie des ventricules. La courbe d'activité résultante sera finalement un mélange des courbes d'activité du myocarde et des ventricules.

- **Effet Compton, imperfection de la collimation et coïncidence fortuite.** La superposition des compartiments physiologiques est renforcée par ces trois phénomènes. Ces derniers ont été évoqués dans la section 1.2.2 consacrée au bruit. L'image visible d'un compartiment physiologique se retrouve convoluée par la fonction de transfert associée à ces phénomènes, augmentant ainsi le chevauchement des différents compartiments physiologiques.
- **Volume partiel.** En imagerie 3D, le mélange des compartiments physiologiques est aussi présent du fait des phénomènes cités au point précédent, mais aussi à cause du volume partiel [SBB07]. La résolution spatiale des caméras étant limitée, les voxels situés aux frontières des compartiments physiologiques sont composés d'un mélange de deux compartiments physiologiques ou plus. Sur des compartiments physiologiques de faible épaisseur, comme le myocarde, l'effet de volume partiel ne doit pas être négligé. Les voxels contenant du myocarde contiennent aussi en partie de l'activité issue des ventricules. De même, dans le cerveau, les tissus cérébraux et le sang sont de trop petite épaisseur pour pouvoir être bien séparés dans les voxels.

1.4.3 Nécessité de la séparation de sources

Du fait des mélanges évoqués précédemment, il n'est en général pas possible d'obtenir des courbes d'activité "pures" pour les compartiments physiologiques à l'aide des régions d'intérêt. Les différents compartiments physiologiques doivent donc être séparés afin d'obtenir des courbes d'activité utilisables à des fins de diagnostics [Bar80]. Ceci peut être réalisé grâce à des méthodes de séparation de sources, où ici les compartiments physiologiques jouent le rôle des sources. Le chapitre suivant explore l'état de l'art de la séparation de sources et son application à l'imagerie nucléaire. Le présent chapitre se termine par la présentation d'un cas d'étude d'imagerie nucléaire utilisant le 6-DIG.

1.5 Cas d'étude : Évaluation de l'insulino-résistance à l'aide du 6-DIG

Parmi les applications de l'imagerie nucléaire, il y a le calcul d'un indice d'insulino-résistance, précurseur du diabète [Eri+89]. Les méthodes actuelles d'évaluation de l'insulino-résistance ne sont pas utilisées en routine clinique car elles sont trop lourdes pour le patient [Per+07]. Des études portant sur le 6-DIG ont montré que ce traceur pouvait potentiellement être utilisé pour évaluer l'insulino-résistance avec l'imagerie nucléaire de manière beaucoup

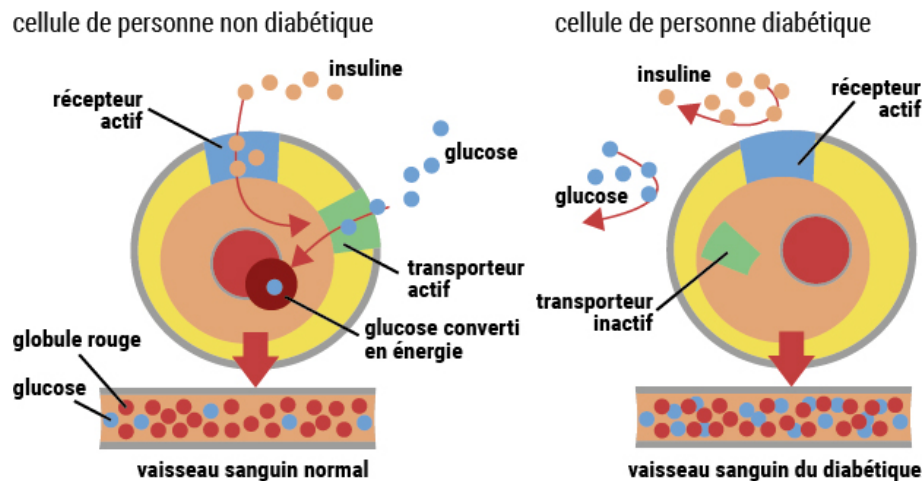


FIGURE 1.8 – Schéma descriptif de l'action de l'insuline sur le transport du glucose. Chez une personne non résistante à l'insuline (à gauche), l'insuline se fixe sur les récepteurs membranaires, ce qui permet de recruter des transporteurs du glucose (GLUT) pour faire entrer le glucose dans la cellule. Chez une personne résistante à l'insuline (à droite), l'insuline se fixe mais la transmission du signal par le récepteur ne se fait pas. Les GLUT4 ne sont donc pas recrutés en nombre suffisant ce qui génère une hyperglycémie et une hyperinsulinémie.

moins contraignante que les méthodes existantes [BR13]. Le rôle de l'insuline dans la régulation de la glycémie est tout d'abord présenté, puis l'étude est détaillée.

1.5.1 La résistance à l'insuline

Le glucose est une molécule utilisée comme combustible par les cellules du corps humain pour produire de l'énergie. Les lieux de haute consommation d'énergie sont les cellules musculaires et les cellules cérébrales. Le glucose est acheminé jusqu'aux cellules grâce au circuit sanguin. La concentration de glucose dans le sang, appelée aussi glycémie, est un élément physiologique très important. Si la glycémie est trop basse, les cellules ne peuvent pas produire assez d'énergie. Ceci peut se traduire par une perte de connaissance et par des troubles plus graves. À l'inverse une glycémie trop haute a des conséquences sur le long terme. Elle favorise la détérioration des vaisseaux sanguins, pouvant entraîner des infarctus du myocarde ou d'autres maladies cardio-vasculaires. Une hyperglycémie peut également entraîner une insuffisance rénale ou même une cécité.

La régulation de la glycémie est donc primordiale pour le bon fonctionnement du corps humain. Lorsque la glycémie est trop faible on parle d'hypoglycémie, et d'hyperglycémie lorsqu'elle est trop élevée. Une glycémie normale à jeun chez un adulte est comprise entre 0,80 et 1,20 gramme de glucose par litre de sang. Lorsque l'hyperglycémie est chronique, ceci est dû à une défaillance des mécanismes de régulation biologique de la glycémie. On parle alors de diabète.

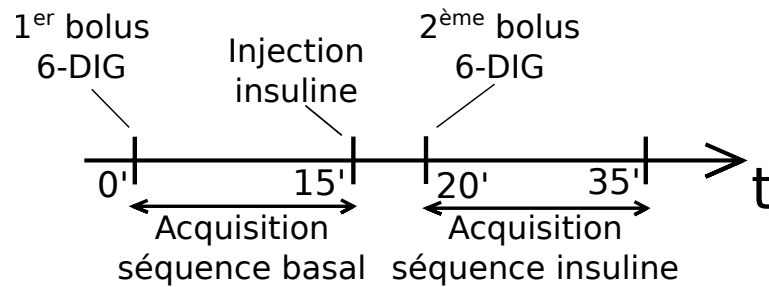


FIGURE 1.9 – Frise descriptive du protocole scintigraphique de l'étude du 6-DIG pour l'évaluation de l'insulino-résistance. Deux séquences d'images sont acquises, une sans injection d'insuline et une avec injection d'insuline. Chaque séquence est composée de 450 images représentant chacune l'activité sur un intervalle de temps de deux secondes. Les acquisitions commencent au moment où le 6-DIG est injecté en bolus (injection rapide).

Une des hormones permettant de réguler la glycémie est l'insuline. C'est une hormone sécrétée par le pancréas. Son rôle dans l'organisme est très important car c'est la seule hormone hypoglycémiante. C'est à dire qu'elle seule permet de réduire le taux de glucose dans le sang. L'insuline favorise le transport du glucose entre le sang et les cellules qui consomment ou qui stockent le glucose. Le glucose entre dans ces cellules grâce aux GLUT (*glucose transporters* en anglais). Les GLUT sont des protéines qui sont rassemblées dans le cytoplasme de la cellule. Certains de ces GLUT (les GLUT4, présents dans les muscles squelettiques et cardiaque ainsi que dans le tissu adipeux) migrent vers la membrane de la cellule lorsque l'insuline envoie un signal, une fois fixé sur les récepteurs membranaires. Une fois fusionnés avec la membrane, les GLUT permettent le passage du glucose du sang à la cellule. Ils facilitent ainsi l'équilibre entre la concentration de glucose à l'intérieur de la cellule et dans le sang. En bref, grâce aux GLUT qu'elle fait migrer, l'insuline favorise l'utilisation du glucose comme source d'énergie ainsi que son stockage dans les cellules. Ceci contribue à diminuer le taux de glucose dans le sang.

On parle d'insulino-résistance lorsque l'insuline ne peut plus favoriser aussi bien le recrutement des transporteurs de glucose [Rea88]. Ceci est dû au fait que les récepteurs membranaires sont moins sensibles à l'insuline, la transmission du signal pour recruter les GLUT4 est moins efficace. La réaction du pancréas est donc de sécréter davantage d'insuline afin de conserver une glycémie convenable. Au bout d'un moment ceci n'est plus suffisant pour réguler correctement la glycémie et apparaît donc un diabète de type 2. L'insulino-résistance est donc un précurseur du diabète [Eri+89]. Une détection précoce de cette insulino-résistance permettrait de prendre des mesures afin d'éviter l'apparition du diabète ou du moins de limiter son ampleur (par exemple adoption d'un mode de vie plus sain avec activité physique et alimentation équilibrée).

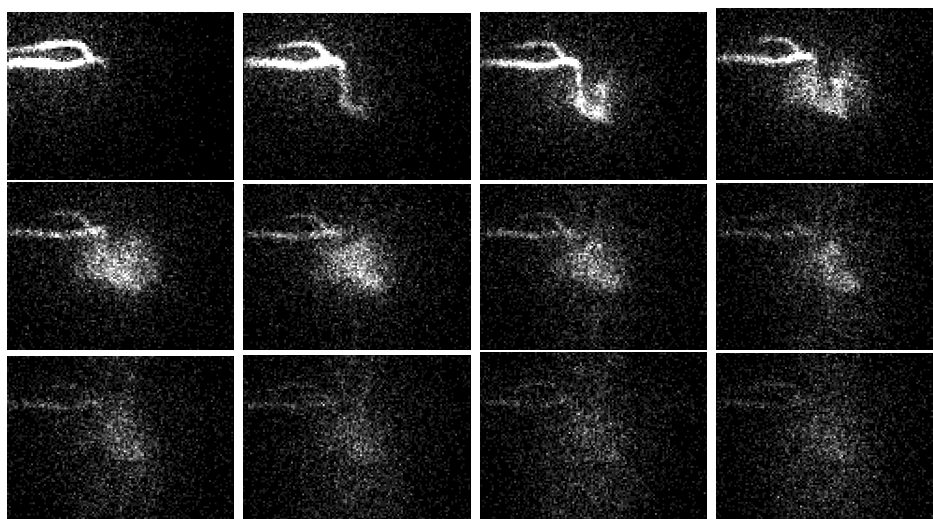


FIGURE 1.10 – Exemple de séquence d'images obtenues en scintigraphie planaire avec le 6-DIG (se lit de gauche à droite puis de haut en bas). Seulement les 12 premières images de la phase basale sont représentées. Chacune de ces images est l'activité enregistrée par la gamma-caméra pendant un intervalle de deux secondes. Le traceur arrive de l'épaule et atteint le ventricule droit. Il est ensuite envoyé dans les poumons avant de revenir vers le ventricule gauche pour se diffuser ensuite dans tout le corps via les artères.

1.5.2 Mesure de l'insulino-résistance à l'aide du 6-DIG

Afin d'évaluer l'insulino-résistance en imagerie nucléaire, il est nécessaire d'utiliser un traceur qui a le même comportement que le glucose. Il existe deux analogues bien connus du glucose : le FDG [Gal+78] et le 3-OMG [KMPL81]. L'inconvénient du FDG est qu'une fois dans une cellule, il y est phosphorylé et ne peut donc plus ressortir, ce qui est problématique pour l'étude de sa cinétique. Le 3-OMG est quant à lui marqué par du carbone 11 qui a une demi-vie de 20 minutes. Un cyclotron à proximité de l'hôpital est donc nécessaire pour le produire afin de l'utiliser en routine clinique. Le 6-DIG est un autre analogue du glucose, marqué lui avec de l'iode 123, à la demi-vie de 13,3 heures. Il a été montré que l'atome d'iode en position 6 ne perturbait pas la reconnaissance du 6-DIG par les transporteurs du glucose [Hen+97]. Contrairement au glucose et au FDG, le 6-DIG ne peut pas être phosphorylé et n'est donc pas dégradé. Il peut donc entrer et sortir librement des cellules grâce aux GLUT. Ainsi le 6-DIG peut être utilisé pour évaluer l'influence de l'insuline sur le transport du glucose dans les cellules, et donc l'insulino-résistance.

Un protocole a donc été développé afin de déterminer un indice d'insulino-résistance grâce à l'analyse de données scintigraphiques planaires en utilisant du 6-DIG [Per+03] ; [Per+07] ; [Bri+07] ; [Per+11] ; [BR13]. Le protocole d'acquisition des images scintigraphiques est résumé dans la figure 1.9. L'idée consiste à séparer les acquisitions d'images en deux phases. Une première séquence est acquise à la suite de l'injection d'une dose de 6-DIG seul. Puis une deuxième séquence est acquise en ayant injecté à la fois de l'insuline et du 6-DIG. L'objectif est de mesurer les différences entre la phase basale (sans injection d'insuline) et la phase

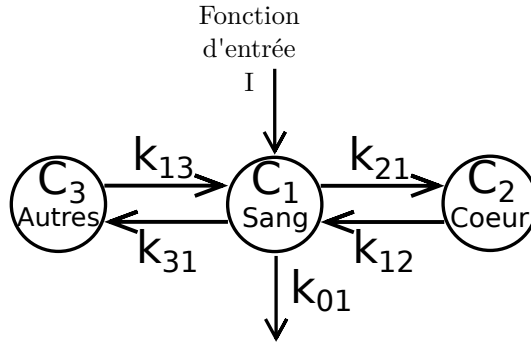


FIGURE 1.11 – Modèle compartimental utilisé pour le calcul de l’insulino-résistance à l’aide du 6-DIG. L’injection du 6-DIG dans le sang est représenté par la fonction d’entrée I . Le compartiment sanguin communique de façon bilatérale avec le myocarde (muscle du coeur), ainsi que les autres tissus (compartiment C_3). Les k_{ij} sont les cinétiques d’échange de 6-DIG du compartiment C_j compartiment C_i . Une fraction du 6-DIG présent dans le sang est aussi sorti du système (modélise les pertes par la vessie et par la décroissance radioactive).

précédée d’une injection d’insuline, afin de déterminer l’impact de l’insuline sur le transport du glucose. Un exemple de séquence d’images obtenue est proposé en figure 1.10.

L’analyse des séquences d’images repose sur un modèle compartimental approchant les interactions entre les différents compartiments physiologiques [BR13]. Les compartiments physiologiques considérés dans cette étude sont le sang, le myocarde et les tissus extérieurs. Le myocarde est choisi comme objet d’étude car il est particulièrement sujet à l’insulino-résistance [Per+07]. Le modèle compartimental choisi est détaillé en figure 1.11. Le système d’équations différentielles associé est présenté en Annexe A.

L’indice d’insulino-résistance IR est calculé à partir de la valeur de la cinétique d’échange k_{21} dans la phase basale (k_{21}^{basal}), et dans la phase insuline (k_{21}^{insu}). Les courbes d’activité du myocarde et du sang s’équilibrent en théorie plus rapidement en présence d’insuline. On s’attend donc à avoir $k_{21}^{basal} < k_{21}^{insu}$. L’indice IR est calculé comme le rapport de ces deux cinétiques :

$$IR = \frac{k_{21}^{insu}}{k_{21}^{basal}} \quad (1.2)$$

Plus cet indice est grand, plus l’insuline influe sur la régulation de la glycémie, et donc moins le myocarde est insulino-résistant. Afin de calculer cet indice à partir des séquences d’imagerie nucléaire et du modèle compartimental, il est nécessaire d’extraire de celles-ci les courbes d’activité du sang et du myocarde. La courbe d’activité du myocarde est particulièrement difficile à obtenir. En effet, si l’on trace la région d’intérêt du myocarde, les pixels de cette région contiendront en majorité de l’activité en provenance des cavités cardiaques, et non du myocarde, du fait de leur proximité. Il est donc nécessaire d’utiliser des méthodes plus sophistiquées pour extraire la courbe d’activité du myocarde. Les méthodes de séparation de sources peuvent répondre à ce problème.

1.6 Conclusion

L'imagerie nucléaire permet d'obtenir de manière non-invasive des informations quantitatives sur la physiologie et l'état de fonctionnement des organes et des tissus. Ces applications sont multiples. Les diagnostics peuvent se baser sur des courbes d'activité temporelles extraites des images produites par la caméra. Toutefois, du fait des mélanges, il n'est parfois pas possible d'extraire les courbes d'activité des compartiments physiologiques proprement avec de simples régions d'intérêt. Des méthodes de séparation ont été développées afin de pallier ce problème, elles sont présentées dans le chapitre suivant.

Séparation de sources

Sommaire

2.1	Généralités	24
2.1.1	Applications présentant des problèmes de séparation	24
2.1.2	Modèle de mélange linéaire	25
2.1.3	Un problème sous-déterminé	27
2.1.4	Principales méthodes de séparation de sources	28
2.2	Modèle de mélange linéaire en imagerie nucléaire	29
2.2.1	Modélisation du problème	30
2.2.2	Signification et hypothèses du modèle de mélange linéaire	30
2.2.3	Impact du bruit en imagerie nucléaire sur le modèle	32
2.2.4	Précisions sur la matrice ϵ du modèle	33
2.3	Spécificités de l'imagerie nucléaire	34
2.3.1	Contraintes de positivité en imagerie nucléaire	34
2.3.2	Régions d'intérêt	35
2.3.3	Images d'intérêt	36
2.3.4	Propriétés statistiques des facteurs recherchés	36
2.3.5	Conclusion sur les spécificités	37
2.4	Contraintes de positivité	37
2.4.1	Interprétation géométrique des contraintes de positivité	37
2.4.2	Un problème encore sous-déterminé	39
2.4.3	Méthodes de séparation de sources avec contraintes de positivité	42
2.5	Méthodes de séparation de sources en imagerie nucléaire	47
2.5.1	Historique de la séparation de sources en imagerie nucléaire	47
2.5.2	Gestion du bruit	52
2.5.3	Principales méthodes	53
2.6	Approche Tensorielle	57
2.6.1	Introduction aux tenseurs	59
2.6.2	Décomposition tensorielle et séparation de sources	59
2.6.3	Application à l'imagerie nucléaire	62
2.6.4	Conclusion sur l'approche tensorielle	66
2.7	Conclusion	67

Ce chapitre introduit progressivement la séparation de sources et son application à l'imagerie nucléaire. Après un tour d'horizon des applications de la séparation de sources, le modèle

de mélange linéaire est présenté. La sous-détermination du problème est abordée ainsi que les différents a priori utilisés classiquement pour contrer cette indétermination. L'insuffisance de ces méthodes et leur inadéquation au cadre de l'imagerie nucléaire nous amènent à présenter les spécificités de ce dernier puis à insister sur l'utilité des contraintes de positivité. Un historique et un état de l'art des méthodes de séparation de sources dédiées à l'imagerie nucléaire sont ensuite détaillés. En fin de chapitre, une partie est consacrée aux modèles tensoriels et à la décomposition tensorielle. Leur utilité pour la séparation de sources est précisée, et des utilisations potentielles pour l'imagerie nucléaire sont également proposées.

2.1 Généralités

Le problème de séparation de sources est un problème inverse classique en traitement du signal. Étant donnée une série d'observations composées chacune de mélanges de sources, l'objectif est de retrouver les signatures des sources à la base de ces mélanges, ainsi que leur poids dans chaque observation. La séparation de sources est nécessaire lorsque les sources d'intérêt sont mélangées et ne sont donc pas directement accessibles. Ces sources doivent avoir des signatures différentes, ainsi que des poids différents dans les observations, afin d'être séparables. La sous-section suivante présente quelques-unes des nombreuses applications pratiques de la séparation de sources.

2.1.1 Applications présentant des problèmes de séparation

Le problème de séparation de sources est un enjeu important dans de nombreux domaines, présentant chacun leurs spécificités, liées à la nature du mélange, des données et à leur méthode d'acquisition. En voici quelques-uns :

- Une des applications bien connues est celle de la soirée cocktail en traitement du son. Dans une ambiance bruyante, où plusieurs conversations ont lieu en même temps, notre cerveau a la faculté de focaliser son attention sur les paroles d'une personne : c'est ce qu'on appelle en psychoacoustique l'effet *cocktail party* [Che53]. Nous réalisons ainsi une séparation de sources. Sous certaines conditions il est donc possible de séparer les paroles d'une personne d'un ensemble de bruits quelconques. En traitement du signal, le problème de la soirée cocktail est donc de reproduire cette capacité, en séparant les paroles de chaque personne de manière automatique grâce à l'enregistrement des conversations à l'aide de plusieurs microphones [Ped+08].
- L'imagerie hyperspectrale est une technique qui permet de photographier une scène suivant un grand nombre de bandes spectrales [SR85]. Cette technique d'imagerie présente de nombreuses applications de séparation de sources [KM02] ; [BD+12], notamment, en télédétection, où des images de surface terrestre à différentes longueurs d'ondes sont prises depuis un satellite. Chaque matériau à la surface de la Terre possède une signature spectrale spécifique. Ainsi les signatures spectrales de l'eau, de la végétation et du goudron sont différentes. Sur un même pixel, différents matériaux peuvent coexister du

fait de la résolution spatiale limitée des instruments de mesure. Le spectre de chaque pixel est donc composé d'un mélange des spectres des différents matériaux présents sur ce pixel. Tout l'enjeu est donc de séparer les spectres des pixels. Cette séparation permet la connaissance de la composition du terrain photographié.

- En astrophysique, toujours à l'aide d'imagerie hyperspectrale, ce n'est pas la Terre mais l'espace qui est imagé (depuis la Terre [Bac+10] ou depuis un satellite [Fre+01]). La séparation de sources permet de séparer les signatures spectrales des étoiles ou des galaxies proches dans les images [Meg+11]. L'analyse de ces spectres donne ensuite des informations sur les objets observés comme leur distance ou leur composition chimique [Bac+15].
- En chimiométrie, la composition de mélanges chimiques peut être estimée grâce à la séparation de sources [SBG05]. Pour cela, ces mélanges sont excités à différentes longueurs d'ondes et les spectres de leurs réponses sont enregistrés. Pour chaque longueur d'onde d'excitation, le spectre réponse des mélanges dépend du spectre réponse de leurs composants. Les méthodes de séparation de sources permettent d'estimer la composition des mélanges ainsi que les réponses spectrales de chaque composant à chaque longueur d'onde [Bro97].
- De nombreuses applications existent également dans le domaine médical. On peut citer notamment les applications pour la séparation de signaux électrophysiologiques de différentes entités [KAS07]. Ainsi la séparation de sources peut s'appliquer à des données issues d'électrocardiogramme (ECG) afin de séparer l'activité du fœtus et de la mère [DLDMV95], ou pour corriger le bruit dû à la respiration [Wis+98]. Elle peut aussi s'appliquer en électromyographie (EMG) [Far+04] afin de diagnostiquer certains syndrômes ou d'évaluer la coordination musculaire. Ou encore à des signaux électroencéphalographiques (EEG ou MEG) pour supprimer les artefacts dus aux clignements des yeux [Vig97]. L'imagerie multispectrale a aussi été utilisée dans le domaine médical pour détecter précocement des cancers de la peau [Dha+09] ; [DD10].
- Comme évoqué dans le premier chapitre, la séparation de sources est aussi nécessaire en imagerie nucléaire afin d'isoler proprement les courbes d'activité des organes et des tissus d'intérêt [DP+82]. Ces organes et tissus sont mélangés à cause du volume partiel, ou bien de la superposition pour l'imagerie planaire. Les travaux de cette thèse sont axés sur les applications de ce domaine.

2.1.2 Modèle de mélange linéaire

Afin de résoudre le problème de séparation de sources, une modélisation mathématique des données et du mélange est faite au préalable. Cette modélisation s'appuie sur la physique des phénomènes étudiés. Dans le cas le plus simple, on utilise le modèle de mélange linéaire [KM02]. Ce modèle stipule que chaque observation est, au bruit près, la somme des sources pondérées par leur poids dans l'observation. Dans le cas où les données possèdent deux diversités (par exemple spatiale et temporelle, le cas où 3 diversités ou plus sont présentes est traité dans la section 2.6 sur les tenseurs), et sont composées de K sources, ceci se traduit par le modèle de

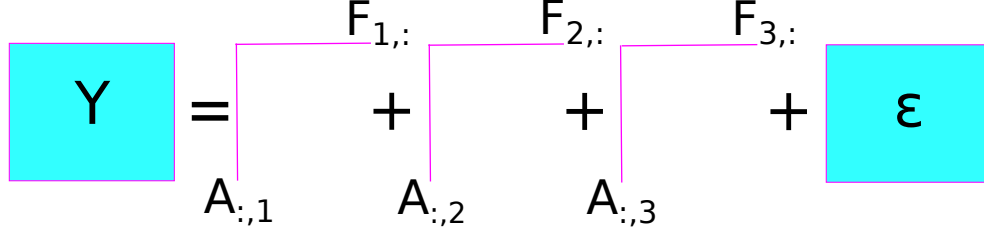


FIGURE 2.1 – Représentation graphique du modèle de mélange linéaire avec $K = 3$ sources. La matrice des données Y est expliquée, au bruit près, par 3 matrices de rang 1 représentant chacune une source. La matrice A est composée des matrices colonnes $A_{:,k}$. La matrice F est composée des matrices lignes $F_{k,:}$. Les matrices de rang 1 sont construites à l'aide du produit matriciel entre une colonne de A et une ligne de F .

mélange linéaire suivant :

$$Y_{i,t} = \sum_{k=1}^K A_{i,k} \times F_{k,t} + \epsilon_{i,t} \quad (2.1)$$

et de manière plus compacte :

$$Y = A \times F + \epsilon \quad (2.2)$$

Dans ces équations, Y est la matrice des données observées, chacune de ses lignes contenant le signal d'une observation. La matrice F contient les signaux purs (i.e non-mélangés) des sources sur ses lignes. La matrice A contient quant à elle dans chacune de ses lignes le poids de chaque source pour une observation donnée. Le bruit, ainsi que les erreurs de modélisation sont contenus dans la matrice ϵ . Les matrices Y et ϵ sont de taille $N \times P$, où N est le nombre d'observations, et P la taille des signaux. Les matrices A et F sont de taille respective $N \times K$ et $K \times P$. Le signal de chaque observation est donc, au bruit près, une combinaison linéaire de signaux purs (i.e non mélangés) contenus dans la matrice F , et dont les poids se retrouvent dans la matrice A . Autrement dit, la matrice des données Y peut s'exprimer, au bruit près, comme une somme de K matrices de rang 1. Chacune de ces matrices de rang 1 représente une source et est construite à l'aide du produit matriciel entre une colonne de A et une ligne de F (cf Figure 2.1).

A titre d'exemple, pour un problème de *cocktail party*, chaque ligne de Y contient le signal sonore reçu par un capteur, chaque ligne de F contient le signal sonore non-mélangé émis par une personne et la colonne de A correspondante représente l'intensité reçue de ce signal par chacun des capteurs. Selon les domaines, les matrices A et F ont des noms différents. Dans la suite, on appellera "facteur" le signal pur d'une source. La matrice F sera alors appelée la "matrice des facteurs". Comme la matrice A contient les cartes d'intensité des facteurs dans les observations, elle sera appelée "matrice des images factorielles". Ces termes sont ceux utilisés en imagerie nucléaire.

Dans le cas d'un mélange linéaire, les matrices A et F caractérisent entièrement chacune des sources. L'objectif des méthodes de séparation de sources est alors de retrouver les matrices A et F à partir des observations Y , le nombre de sources K étant supposé connu dans la

majorité des cas.

Remarque 2.1

Bien qu'elles aient des significations très différentes en pratique, les matrices A et F ont en fait un rôle symétrique dans le modèle. En appliquant simplement l'opérateur transposé à l'équation (2.2), on obtient :

$$Y^T = F^T \times A^T + \epsilon^T \quad (2.3)$$

Il est donc possible d'appliquer une méthode de séparation de sources à Y ou bien à sa transposée.

Généralement, le nombre de sources (K) est faible devant le nombre d'observations (N) et la taille des signaux (P). La séparation de sources permet ainsi de réaliser une compression des données. Cette compression est possible car les observations peuvent être presque entièrement expliquées par un petit nombre de sources. L'information utile contenue dans les données est résumée dans les matrices A et F qui sont de petite taille comparativement à Y . Le taux de compression atteint est donc de $K \times (\frac{1}{N} + \frac{1}{P})$.

Dans certaines applications, le mélange n'est pas linéaire. Cette non-linéarité peut provenir par exemple de l'interaction de la lumière avec les mélanges de matières en imagerie hyperspectrale [NC74]. D'autres modèles de mélange ont donc été proposés en conséquence [BD+12]. Le modèle de mélange linéaire reste très utilisé car il constitue une bonne approximation de la réalité dans de nombreux domaines, et la résolution du problème est facilitée par sa simplicité. Les erreurs de modélisation sont contenues dans la matrice de bruit ϵ . Une modélisation plus fine nécessite une très bonne connaissance des phénomènes physiques mis en jeu, et complexifie généralement le processus de séparation.

Alors que le modèle de mélange linéaire suggère qu'il suffit d'une simple factorisation matricielle de Y pour séparer les sources, la réalité est elle bien plus complexe à cause de la non-unicité de cette factorisation.

2.1.3 Un problème sous-déterminé

Reprenons le modèle de mélange linéaire de l'équation (2.2) et négligeons le terme de bruit. On obtient alors :

$$Y = A \times F \quad (2.4)$$

Le problème de séparation de sources devient un problème de factorisation de la matrice Y en deux matrices A et F . Selon le rang de la matrice des données Y et le nombre supposé de sources K , cette factorisation n'est pas toujours possible. Rappelons que K est le nombre de colonnes de A et le nombre de lignes de F . La condition $\text{rang}(Y) \leq K$ est une condition nécessaire et suffisante d'existence d'une factorisation de la matrice Y en deux matrices A et F . Cependant, si elle existe, cette factorisation n'est pas unique. En effet, considérons une

matrice carrée inversible R , de taille $K \times K$. Si il existe un couple de matrices (A, F) de taille respective $N \times K$ et $K \times P$ vérifiant l'équation (2.4), alors le couple $(A \times R, R^{-1} \times F)$ vérifie lui aussi l'équation (2.4). Ainsi, si une factorisation existe, il en existe une infinité et donc plusieurs mélanges de sources différents peuvent conduire aux mêmes observations. Autrement dit, la matrice Y seule ne caractérise pas le couple (A, F) . Le problème de séparation de sources est donc sous-déterminé, non-identifiable. Trouver une solution au modèle, c'est à dire trouver une factorisation, ne garantit aucunement que cette factorisation soit *la* solution du problème de séparation de sources. D'un point de vue applicatif, cette sous-détermination est problématique car le couple de matrices (A, F) trouvé qui est censé caractériser les sources pures n'a plus le sens recherché, car l'information qu'il contient n'est vraisemblablement toujours pas séparée.

Il existe cependant deux causes de non-unicité qui n'ont pas d'impact sur l'interprétation de la séparation des sources [HS14]. Notons A et F un couple de matrices solution qui contient les sources pures. Lorsque la matrice R est une matrice de permutation, alors $(AR, R^{-1}F)$ contient aussi les sources pures mais dans un ordre différent. De même si R est une matrice diagonale inversible, $(AR, R^{-1}F)$ contient aussi les sources pures mais les matrices A et F sont estimées à un facteur d'échelle près. Ces indéterminations de permutations et d'échelle n'ont que peu de conséquences en pratique, car elles ne mélangent pas les sources. En revanche, les autres formes d'indétermination, que l'on peut assimiler à des indéterminations de rotation grâce à la décomposition en valeurs singulières (SVD en anglais) ne sont pas sans conséquences puisqu'elles mettent en échec la séparation de sources.

Dans le contexte pratique de la séparation de sources, il n'est pas envisageable de chercher une factorisation exacte, car le modèle est bien souvent une approximation de la réalité. De plus, les données sont généralement bruitées et la matrice Y de rang plein, invalidant ainsi la condition $\text{rang}(Y) \leq K$. Le modèle comportant un terme d'erreur (2.2) est donc préféré au modèle exact (2.4). La factorisation est recherchée en minimisant une norme de cette erreur. Les remarques précédentes sur la sous-détermination du problème sont évidemment toujours présentes avec ce modèle si le critère est uniquement de maximiser la fidélité aux données, c'est à dire minimiser une norme de l'erreur résiduelle $(Y - A \times F)$.

La sous-détermination dans le contexte de l'imagerie nucléaire est abordée plus en détail dans la section 2.4.2, en prenant en compte des contraintes de positivité sur les coefficients des matrices A et F . La solution au problème sera considérée unique si elle l'est à une matrice de permutation et/ou d'échelle près.

2.1.4 Principales méthodes de séparation de sources

À cause de la sous-détermination, les méthodes de séparation de sources ne peuvent pas se contenter de maximiser la fidélité aux données. Afin de trouver *la* solution au problème de séparation, il est nécessaire d'injecter des informations supplémentaires sur les sources recherchées, des informations a priori. Il n'existe donc pas de méthode de séparation de sources universelle car les informations ajoutées dépendent de l'application et des données. Les informations a priori ajoutées peuvent cependant être communes à plusieurs applications. Les

méthodes les plus courantes de séparation sont décrites ci-dessous.

- L'analyse en composantes principales [Jol02] (ACP, ou PCA en anglais) minimise la norme euclidienne de l'erreur résiduelle $\|Y - A \times F\|_2$, tout en cherchant des facteurs (lignes de F) orthogonaux entre eux. Elle permet de trouver des facteurs qui expliquent au mieux la variance des données. Bien que n'étant pas à la base une méthode de séparation de sources, elle peut être vue comme telle dans le cas où les sources recherchées sont décorréliées. La recherche d'une solution avec des facteurs orthogonaux n'est pas un problème sous déterminé et la solution se calcule analytiquement. Cependant l'hypothèse d'orthogonalité est une hypothèse forte sur les facteurs, qui n'est pas vérifiée dans tous les domaines d'application. L'ACP est néanmoins à la base de nombreuses autres méthodes de séparation de sources, car elle permet de réduire la dimension de l'espace de recherche en conservant au maximum la variance des données, et est souvent utilisée comme une étape de débruitage [Bar80].
- L'analyse en composantes indépendantes (ACI, ou ICA en anglais) est une généralisation de l'ACP aux ordres statistiques supérieurs [Com94]. Tandis que l'ACP cherche des sources qui sont mutuellement décorréliées (orthogonales), l'ACI cherche des sources qui sont mutuellement indépendantes. L'ACI est très utilisée pour la séparation de sources sonores [Ped+08] et de signaux électrophysiologiques [KAS07].
- L'analyse en composantes parcimonieuses regroupe des méthodes qui traitent le problème de séparation de sources en ajoutant un critère de parcimonie [GL06]. Les facteurs sont recherchés avec comme contrainte d'avoir une représentation parcimonieuse sur un dictionnaire de signaux.
- La factorisation en matrices non-négatives (NMF en anglais) est une famille de méthodes cherchant à minimiser une norme de l'erreur résiduelle $\epsilon = Y - A \times F$, en imposant aux matrices A et F d'avoir des coefficients positifs [PT94]; [LS01]. Ces méthodes sont décrites en détail dans la section 2.4.3.

Ces méthodes générales ne sont toutefois pas adaptées à l'imagerie nucléaire. Les sections suivantes explicitent le modèle de mélange linéaire dans le cadre de l'imagerie nucléaire. Les spécificités de ce cadre sont ensuite détaillées afin de trouver les contraintes adéquates pour résoudre le problème de séparation de sources dans ce contexte.

2.2 Modèle de mélange linéaire en imagerie nucléaire

Dans le cadre de l'imagerie nucléaire, les données prennent la forme d'une séquence temporelle d'images de la radioactivité (cf annexes D). Cette dernière provient d'un traceur injecté au patient. Les différents organes et tissus d'intérêt présents dans la séquence d'images qui ont une courbe d'activité propre constituent les sources. Nous nommerons *compartiments physiologiques* ces sources. En reprenant les notations de la section 2.1.3, K est le nombre de compartiments physiologiques présents dans la séquence d'images, N est le nombre de pixels dans chaque image, et P est le nombre d'images dans la séquence. En pratique, le nombre

de pixels N est de l'ordre de quelques centaines à quelques milliers, et le nombre d'images P ne dépasse que rarement la centaine. Le nombre de compartiments physiologiques K est lui généralement compris entre 2 et 4.

2.2.1 Modélisation du problème

Reprenons le modèle de mélange linéaire de l'équation (2.2), $Y = A \times F + \epsilon$, et explicitons chacun de ces termes dans le cadre de l'imagerie nucléaire. Une illustration de l'équation est fournie dans la Figure 2.2.

- La matrice Y est la matrice des données initiales. Elle contient une séquence d'images décrivant l'évolution de la radioactivité dans le corps du patient au cours du temps. Les images de cette séquence sont vectorisées. Chacune des N lignes de la matrice Y représente la dynamique de la radioactivité mesurée sur un pixel. Chacune des P colonnes de Y est une image vectorisée, représentant la radioactivité perçue par la caméra pendant un intervalle de temps.
- La matrice des images factorielles A contient les informations spatiales des compartiments physiologiques. Chaque colonne de A représente la carte de présence vectorisée d'un des compartiments physiologiques. On retrouve donc sur une ligne de A les poids des différents compartiments physiologiques sur le pixel correspondant. Un coefficient nul dans cette matrice à la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne signifie que le $j^{\text{ème}}$ compartiment physiologique n'est pas présent dans le $i^{\text{ème}}$ pixel. À l'inverse une valeur forte signifie que ce compartiment physiologique est fortement présent dans ce pixel.
- La matrice des facteurs F contient les informations temporelles des compartiments physiologiques. Chacune des lignes de la matrice des facteurs F contient la courbe temporelle d'activité d'un des compartiments physiologiques. Cette matrice possède donc autant de lignes que de compartiments physiologiques dans le modèle, et autant de colonnes que d'images dans la séquence initiale.
- La matrice ϵ est de même taille que la matrice des données Y , et contient les différentes erreurs (erreurs de modélisation, bruit de comptage, bruit de mesure).

2.2.2 Signification et hypothèses du modèle de mélange linéaire

La modélisation des données d'imagerie nucléaire sous la forme d'un modèle de mélange linéaire entraîne de nombreuses hypothèses et approximations. Le modèle de mélange linéaire (2.2) suggère que la radioactivité contenue dans la séquence d'images peut être expliquée presque entièrement (à l'erreur ϵ près) par la radioactivité d'un petit nombre de compartiments physiologiques [NB88]. Comme vu précédemment, les informations temporelles et spatiales des compartiments physiologiques sont séparées dans le modèle, les informations spatiales sont contenues dans la matrice A , et les informations temporelles dans la matrice F . Si bien que l'on pourrait réécrire le modèle sous la forme d'une fonction à variables séparables :

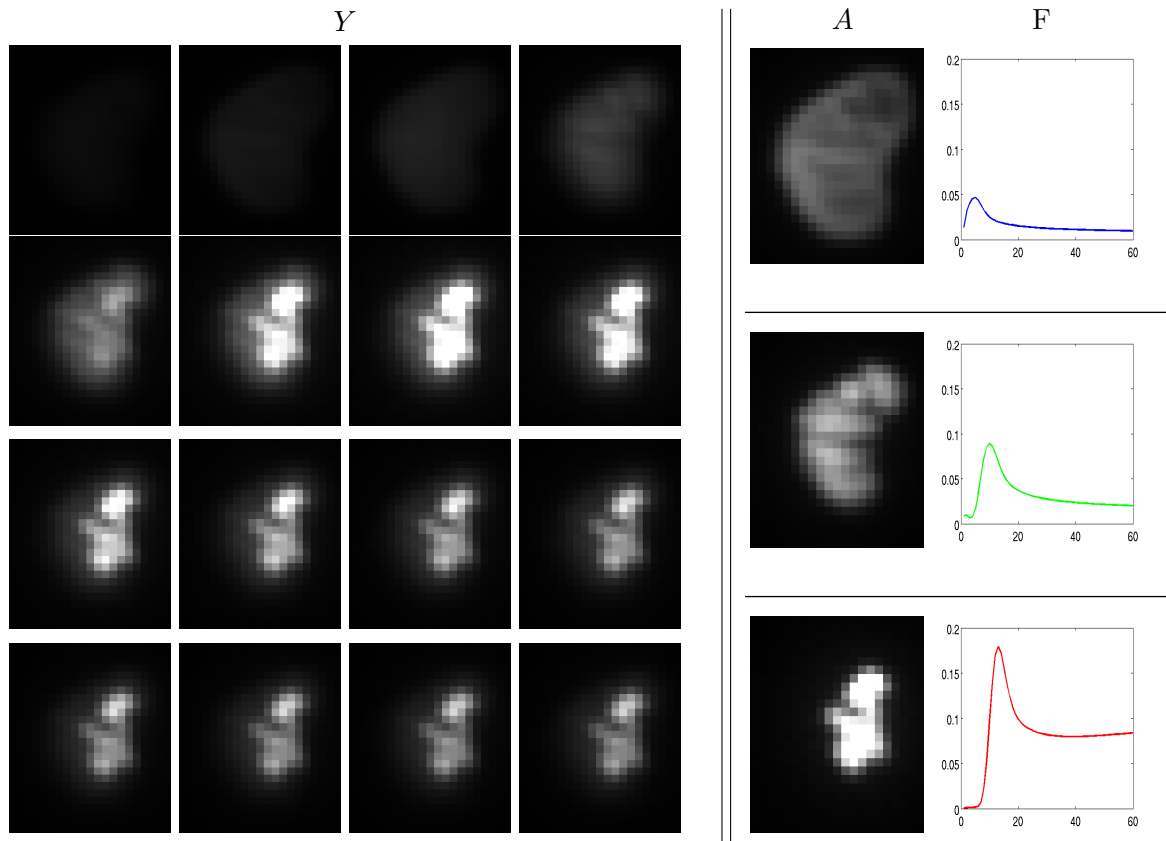


FIGURE 2.2 – Un exemple en imagerie nucléaire de matrices Y , A et F suivant le modèle décrit dans l'équation (2.4), avec $K = 3$ et aucun bruit. Les quatre premières colonnes (à lire de gauche à droite puis de haut en bas) contiennent quelques images du début de la séquence de données Y , aux temps $t=1,3,5,7,\dots,31$. La cinquième colonne contient les images factorielles (colonnes de A). Enfin, la dernière colonne représente les facteurs (lignes de A), avec le temps en abscisse et l'intensité de l'activité en ordonnée. Cette séquence représente la simulation réaliste d'une scintigraphie rénale. Les trois compartiments physiologiques sont (de haut en bas) : le parenchyme, la zone medullaire et le pelvis.

$$y(i, t) = a(i) \cdot f(t) \quad (2.5)$$

où $\cdot$ est le produit scalaire canonique de $\mathbb{R}^K$ dans $\mathbb{R}^K$ et y , a et f sont respectivement des fonctions de $\llbracket 1; N \rrbracket \times \llbracket 1; P \rrbracket$ dans $\mathbb{R}$, de $\llbracket 1; N \rrbracket$ dans $\mathbb{R}^K$ et de $\llbracket 1; P \rrbracket$ dans $\mathbb{R}^K$. Ceci entraîne trois approximations significatives sur les compartiments physiologiques : leur stabilité temporelle, leur homogénéité spatiale et leur faible nombre.

- *Stabilité temporelle des compartiments physiologiques* : La matrice A contient les informations spatiales des compartiments physiologiques, et ne dépend pas du temps. Ceci a pour conséquence directe que les cartes de présence des compartiments physiologiques ne dépendent pas du temps. Autrement dit, les compartiments physiologiques observés

sont supposés non déformables et immobiles. En pratique, cette hypothèse n'est pas vérifiée mais constitue une bonne approximation. Une des causes de mouvement des compartiments physiologiques est le mouvement lié aux battements du coeur ainsi qu'à la respiration du patient. Lorsque la période d'acquisition d'une image contient plusieurs périodes complètes de respiration, on peut cependant considérer que les compartiments physiologiques sont bien non déformables. En effet, ce qui importera sera la moyenne de leur position pendant la période d'acquisition d'une image. Lorsque la période d'acquisition est grande par rapport à la période de la respiration, ce qui est généralement le cas en imagerie nucléaire, on peut considérer que cette moyenne est identique pour chaque image. Ces mouvements n'invalident donc pas l'hypothèse de stabilité temporelle, mais augmentent le mélange des compartiments physiologiques dans la zone thoracique. Si les battements du coeur s'avèrent problématiques pour l'application, il est dorénavant possible de réaliser des acquisitions synchronisées avec le rythme cardiaque du patient via un électrocardiogramme [PN04]. Une autre cause de mouvement des compartiments physiologiques est le mouvement des patients. Ce dernier pouvant rester pendant une heure sous la machine d'acquisition, il est probable qu'il ait des légers mouvements. Ceci déplace généralement toute l'image. Néanmoins, des recalages d'images peuvent limiter l'ampleur de ce problème [Mig99].

- *Homogénéité spatiale des compartiments physiologiques* : La matrice F contient les informations temporelles des compartiments physiologiques, et ne dépend pas de la position spatiale. Le compartiment physiologique est donc ici supposé homogène [NB88], la dynamique de son activité est la même partout, au facteur d'échelle près. Ce facteur d'échelle est contenu dans la matrice A et correspond à l'intensité de présence du compartiment physiologique à la position considérée. L'homogénéité des compartiments physiologiques est une hypothèse forte. En pratique elle peut être mise en cause par plusieurs phénomènes. Notamment, un compartiment physiologique peut lui même être composé de plusieurs sous-compartiments physiologiques ayant des cinétiques d'activité légèrement différentes les unes des autres. Cependant, une meilleure précision impliquerait d'augmenter le nombre K de compartiments physiologiques, ce qui compliquerait grandement la séparation de sources ainsi que son interprétabilité. La variabilité intra-compartiments doit être négligeable face à la variabilité inter-compartiment afin que l'hypothèse d'homogénéité des compartiments physiologiques soit acceptable.
- *Nombre restreint de sources* : La troisième approximation faite par le modèle a déjà été évoquée précédemment : c'est l'explication des données avec uniquement un nombre restreint de compartiments physiologiques. Les nombreuses ramifications du système sanguin permettent au traceur de circuler partout dans le corps. Ainsi, l'hypothèse d'un nombre restreint de compartiments physiologiques est justifiée uniquement si ces derniers concentrent l'essentiel de la radioactivité, c'est à dire l'essentiel du traceur.

2.2.3 Impact du bruit en imagerie nucléaire sur le modèle

L'impact sur le modèle de mélange linéaire du bruit est détaillé point par point ci-dessous.

- Le bruit de comptage, dû au caractère aléatoire de la radioactivité, de la collimation et de l'atténuation des photons est très fort. La gestion de ce bruit de comptage dans le cadre de la séparation de sources en imagerie nucléaire est détaillée dans la section 2.5.2.
- La diffusion Compton, qui modifie la trajectoire des photons gamma, a pour conséquence directe de modifier la matrice des images factorielles A . En effet, les cartes de présence des compartiments physiologiques vont être convoluées avec le filtre associé à la fonction d'étalement engendré par cette diffusion Compton. Ainsi, les compartiments physiologiques se retrouveront d'autant plus mélangés, car chacun va s'étaler spatialement, augmentant ainsi le chevauchement. La matrice F ne subit, elle, aucun changement.
- L'imperfection de la collimation a des conséquences sensiblement identiques à la diffusion Compton : les images factorielles des compartiments physiologiques sont étalées spatialement, et les facteurs dans la matrice F sont épargnés.

La diffusion Compton ainsi que l'imperfection de la collimation modifient donc la matrice A . Les images factorielles obtenues avec une méthode de séparation seront donc modifiées par ces phénomènes. Ceci n'est pas impactant pour les diagnostics, car les images factorielles servent principalement de support visuel pour valider la cohérence de la source estimée. Les courbes d'activité des sources ne sont pas impactées par ce changement.

2.2.4 Précisions sur la matrice ϵ du modèle

Au vu des précédentes remarques sur la signification du modèle en imagerie nucléaire ainsi que l'impact du bruit sur ce dernier, le terme d'erreur ϵ peut être mieux caractérisé.

Pour simplifier, on peut tout d'abord considérer que la matrice A du modèle n'est pas la matrice des vraies cartes de présence, mais la matrice des cartes de présence modifiées par la diffusion Compton et l'imperfection de la collimation. Comme les modifications apportées par la diffusion Compton et l'imperfection de la collimation sur ces cartes de présence n'altèrent pas la reconnaissance des compartiments physiologiques, ceci n'est pas un problème pour les diagnostics. On ne recherche donc plus la matrice des vraies cartes de présence, mais la matrice des cartes de présence modifiées par la diffusion Compton et l'imperfection de la collimation. Ainsi, les erreurs introduites par ces deux phénomènes ne sont plus à prendre en compte, et ne sont donc pas présentes dans la matrice ϵ .

Dans le modèle de mélange linéaire, on approche la concentration du traceur dans la séquence d'images par les concentrations moyennes de quelques compartiments physiologiques (matrice F) accompagnées de leurs cartes de présence (matrice A). Cette approximation génère des erreurs de modélisation comme vu dans la section 2.2.2. Ces erreurs sont notées dans la suite ϵ_{Mod} . Ainsi, la matrice C de la concentration du traceur dans la séquence d'images peut être réécrite $C = \lambda A \times F + \epsilon_{Mod}$, avec λ une constante positive (facteur d'échelle). Les erreurs de modélisation ϵ_{Mod} , dûes aux hypothèses d'homogénéité spatiale et de stabilité temporelle des compartiments physiologiques, et à l'hypothèse d'un nombre très restreint de sources, sont difficilement caractérisables. Elles représentent la concentration du traceur non expliquée par le modèle.

L'information présente dans la matrice Y n'est pas la concentration du traceur, mais une mesure de sa radioactivité. C'est ici qu'interviennent le bruit de comptage, ainsi que le bruit de mesure. Le bruit de comptage opère sur la concentration du traceur, tandis que le bruit de mesure, noté ϵ_{Mes} peut être approché par un bruit additif. La matrice Y , et donc le modèle, peuvent se réécrire :

$$Y \sim \mathcal{P}(A \times F + \epsilon_{Mod}) + \epsilon_{Mes} \quad (2.6)$$

où $\mathcal{P}$ est la loi de Poisson. Cette loi étant stable pour l'addition, ceci se réécrit :

$$Y \sim \mathcal{P}(A \times F) + \mathcal{P}(\epsilon_{Mod}) + \epsilon_{Mes} \quad (2.7)$$

Le terme ϵ du modèle de mélange linéaire initial peut donc être réécrit :

$$\epsilon \sim (\mathcal{P}(A \times F) - (A \times F)) + \mathcal{P}(\epsilon_{Mod}) + \epsilon_{Mes} \quad (2.8)$$

L'erreur ϵ est donc constitué d'un bruit de comptage bien caractérisé, mais également de deux autres bruits très complexes à caractériser (ϵ_{Mod} et ϵ_{Mes}).

2.3 Spécificités de l'imagerie nucléaire

Du fait de la sous-détermination du problème de séparation de sources, il est nécessaire d'ajouter des informations spécifiques aux données et au contexte. L'objet de cette section est de présenter les principales caractéristiques du cadre de l'imagerie nucléaire à considérer lors de la séparation de sources, afin d'extraire les contraintes utiles à la séparation.

2.3.1 Contraintes de positivité en imagerie nucléaire

En imagerie nucléaire, les séquences d'images représentent l'évolution de la radioactivité mesurée. Cette radioactivité correspond à un nombre de désintégrations observées pendant un certain temps. D'un point de vue physique, cette radioactivité ne peut donc pas être négative. Ainsi la matrice des facteurs F , qui contient l'évolution de la radioactivité au sein des différents compartiments physiologiques, ne peut pas contenir de coefficients négatifs [DP+82]. Si la matrice des facteurs F contient des coefficients négatifs, cette matrice n'a plus le sens physique recherché, car le facteur associé ne peut plus être interprété comme l'évolution de la radioactivité d'un compartiment physiologique. Ceci entraîne donc :

$$\forall k \in \llbracket 1; K \rrbracket, \forall t \in \llbracket 1; P \rrbracket, F_{k,t} \geq 0 \quad (2.9)$$

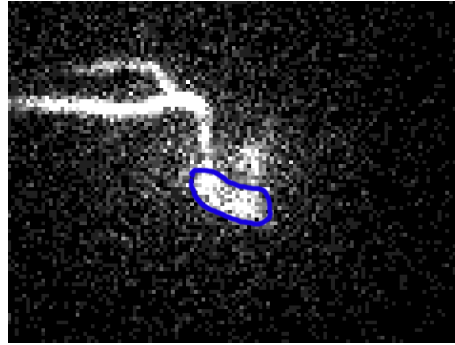


FIGURE 2.3 – Image d’une séquence utilisant du 6-DIG. Sur cette image, le traceur arrive dans le ventricule droit (au centre) par les veines (en haut à gauche). Le ventricule droit est particulièrement visible et sa région d’intérêt aisée à définir, elle est délimitée par le polygone bleu.

De la même manière, la matrice A qui représente les cartes de présence des différents compartiments physiologiques ne peut pas avoir de coefficients négatifs, un compartiment physiologique ne pouvant pas être négativement présent sur un pixel [Bar80]. Il est donc nécessaire d’imposer également la contrainte suivante :

$$\forall i \in \llbracket 1; N \rrbracket, \forall k \in \llbracket 1; K \rrbracket, A_{i,k} \geq 0 \quad (2.10)$$

Ces contraintes de positivité ont un intérêt pratique très fort pour la séparation de sources puisqu’elles permettent de réduire le champ des solutions possibles, en se restreignant à la recherche de matrices A et F à coefficients positifs ou nuls. Ceci réduit considérablement la sous-détermination du problème. Cependant, ces contraintes ne sont dans la plupart des cas pas suffisantes pour garantir l’unicité de la solution (voir section 2.4.2).

2.3.2 Régions d’intérêt

En imagerie nucléaire, les diagnostics sont très souvent réalisés grâce au tracé manuel de régions d’intérêt et à l’analyse des courbes d’activité de ces régions [GPS11]. Les spécialistes ont l’habitude de tracer ces régions d’intérêt, que ce soit pour les séquences d’images 2D ou 3D. Des outils et logiciels de visualisation permettent de faciliter cette tâche. L’exemple de la Figure 2.3 montre qu’il est très facile pour un spécialiste de tracer la région d’intérêt du ventricule droit sur une séquence planaire utilisant du 6-DIG, car il ressort particulièrement bien sur certaines images. Les régions d’intérêt des compartiments physiologiques sont donc une information généralement accessible en imagerie nucléaire.

Cependant, comme évoqué dans le premier chapitre, de simples régions d’intérêt ne permettent pas de séparer les sources. Les courbes d’activité obtenues en moyennant les courbes d’activité des pixels de la région d’intérêt peuvent contenir l’activité de plusieurs comparti-

ments physiologiques, du fait des mélanges dans les pixels. Ces courbes d'activité peuvent être corrigées d'un bruit de fond, qui est supposé uniforme, mais cela ne permet pas de supprimer proprement l'influence des autres compartiments physiologiques superposés.

La principale contribution de cette thèse est d'exploiter les informations des régions d'intérêt pour la séparation de sources. Ces régions d'intérêt apportent de précieuses informations sur les cartes de présence des compartiments physiologiques contenues dans la matrice A . Elles se traduisent par des contraintes sur cette matrice A , réduisant ainsi la sous-détermination du problème de séparation de sources. Les différentes propositions d'utilisation des régions d'intérêt pour la séparation de sources sont détaillées dans le chapitre 3 de cette thèse.

2.3.3 Images d'intérêt

En imagerie nucléaire, certaines images contiennent peu d'informations utiles à la séparation de sources, ou ont un rapport signal sur bruit très mauvais. Ces images sont toutefois importantes car elles contiennent l'information de l'activité des compartiments physiologiques pendant certains intervalles de temps. Elles ne doivent donc pas être supprimées de l'analyse, mais jouer un rôle mineur dans la séparation afin de garantir que les autres images, dénomées images d'intérêt, jouent un rôle majeur dans la séparation. Une proposition d'utilisation de cette information est détaillée dans la section 3.3.5.

2.3.4 Propriétés statistiques des facteurs recherchés

De nombreuses méthodes de séparation de sources, telle l'ACI, ont pour hypothèse que les sources, ici les facteurs, sont stationnaires. Dans le cadre de l'imagerie nucléaire, cette hypothèse est fautive. En effet, à cause de la dilution du traceur dans l'organisme, les facteurs ont très souvent des phases transitoires en début de séquence, suivies par des phases d'accumulation ou bien de décroissance d'activité (voir Figure 2.2), ce qui rend l'hypothèse de stationnarité invalide.

Une autre hypothèse commune est celle de l'indépendance des sources. En imagerie nucléaire, les activités des compartiments physiologiques sont généralement intimement liées. Leur liaison peut souvent être modélisée par un modèle compartimental, comme décrit précédemment dans la section 1.3. L'hypothèse d'indépendance des sources n'est donc pas valide.

Les méthodes utilisant comme hypothèses la stationnarité et l'indépendance des sources, pourtant très utiles en traitement du son ou en analyse de signaux électrophysiologiques, ne peuvent donc pas s'appliquer dans le cadre de l'imagerie nucléaire.

2.3.5 Conclusion sur les spécificités

De nombreux enseignements sont à tirer des spécificités de l'imagerie nucléaire afin de réaliser une séparation de sources aussi performante que possible :

1. Des contraintes de positivité doivent être appliquées aux matrices A et F afin de réduire l'espace de recherche.
2. Les régions d'intérêt des compartiments physiologiques constituent une connaissance généralement accessible pour faciliter la séparation de sources.
3. Dans les séquences, certaines images contiennent des informations plus pertinentes que d'autres images pour la séparation de sources et ont un meilleur rapport signal sur bruit.
4. Les méthodes traditionnelles de séparation, de type ACI ou ACP, ne peuvent pas s'appliquer car leurs hypothèses statistiques sur les sources ne sont pas vérifiées en imagerie nucléaire.

Ces différents points seront exploités pour l'élaboration d'une méthode de séparation de sources en imagerie nucléaire dans le chapitre 3. La section suivante se concentre sur l'utilisation des contraintes de positivité, qui se retrouvent dans de nombreux autres domaines.

2.4 Contraintes de positivité

Du fait de la sous-détermination du problème de la séparation de sources, il est nécessaire d'ajouter des connaissances a priori sur les sources recherchées. Par exemple en imagerie nucléaire et en imagerie hyperspectrale, les sources sont connues comme étant positives. Les matrices A et F sont alors estimées avec comme contraintes d'avoir des coefficients positifs. Ceci réduit donc le problème de non-unicité de la solution mais n'est généralement pas suffisant pour rendre le problème identifiable. Ces contraintes de positivité permettent par ailleurs une interprétation géométrique qui est un point d'appui de nombreuses méthodes de séparation. Dans la suite une matrice est appelée positive lorsque tous ses coefficients sont positifs ou nuls.

2.4.1 Interprétation géométrique des contraintes de positivité

Supposons que l'on ait des données non bruitées suivant le modèle de mélange linéaire de l'équation (2.4), avec des matrices A et F positives. Les données contenues dans Y sont donc elles aussi positives, ainsi, toutes les observations (les lignes de Y) sont dans l'orthant positif de $\mathbb{R}^P$. La positivité de F a comme conséquence que les signaux des sources (les lignes de F) sont dans ce même orthant. Le modèle de mélange linéaire stipule lui que chaque observation est une combinaison linéaire de K sources positives (lignes de F). La positivité de la matrice A entraîne donc que cette combinaison linéaire est à poids positifs.

Supposons de plus que les lignes de Y et les lignes de F , c'est à dire les observations et les facteurs, soient normalisées. Comme Y et F sont à coefficients positifs on peut écrire la

normalisation de cette manière :

$$\forall i \in \llbracket 1..N \rrbracket, \|Y_{i,:}\|_1 = \sum_{t=1}^P Y_{i,t} = 1 \quad (2.11a)$$

$$\forall k \in \llbracket 1..K \rrbracket, \|F_{k,:}\|_1 = \sum_{t=1}^P F_{k,t} = 1 \quad (2.11b)$$

Le modèle de mélange linéaire couplé à ces deux normalisations et aux contraintes de positivité a pour conséquence directe que les lignes de A sont elles aussi normalisées :

$$\forall i \in \llbracket 1..N \rrbracket, \|A_{i,:}\|_1 = \sum_{k=1}^K A_{i,k} = 1 \quad (2.12)$$

On peut donc écrire chaque observation de la manière suivante :

$$Y_{i,:} = \sum_{k=1}^K A_{i,k} \times F_{k,:} \text{ avec } \sum_{k=1}^K A_{i,k} = 1 \quad (2.13)$$

La $i^{\text{ème}}$ observation ($Y_{i,:}$) est donc le barycentre des facteurs ($F_{k,:}$) auxquels on a affecté les poids contenus dans la $i^{\text{ème}}$ colonne de A ($A_{i,:}$). La normalisation et la positivité de A entraînent que ces observations sont nécessairement à l'intérieur du simplexe dont les sommets sont les facteurs (les lignes de F) [DP+82]. L'interprétation géométrique du problème est ici très puissante. En effet, la recherche des sources s'apparente à la recherche des sommets d'un simplexe qui contient toutes les observations. La séparation peut donc être vu comme un problème géométrique. Afin de le visualiser dans un espace de dimension convenable, il suffit après normalisation des données, de les centrer et de les observer dans le repère fourni par une analyse en composantes principales. Les $K - 1$ premiers vecteurs propres fournis par l'analyse en composantes principales sont suffisants pour expliquer totalement les données non bruitées centrées et normalisées. Un exemple simple de ce qui serait obtenu pour des données non bruitées contenant $K = 3$ sources est illustré Figure 2.4(a), où ces données sont représentées dans le plan fourni par leur ACP. Les observations sont à l'intérieur du simplexe dont les sommets sont les facteurs. Les coordonnées des observations et des facteurs peuvent être négatives dans le repère fourni par l'ACP dont l'origine est le centre des données, mais les coordonnées sont bien positives si l'on se place dans la base canonique de $\mathbb{R}^P$.

Remarque 2.2

Si l'on ne prend pas en compte le centrage et les normalisations, les observations ne se situent plus dans le simplexe dont les sommets sont les facteurs, mais dans l'enveloppe conique engendrée par ces facteurs [DS04], comme illustré dans la Figure 2.4(b). Cette enveloppe conique contenant l'origine est appelée cône simpliciel.

Remarque 2.3

En considérant l'approche symétrique proposée dans la remarque 2.1, les coordonnées des sommets représentent cette fois les images factorielles et non plus les facteurs.

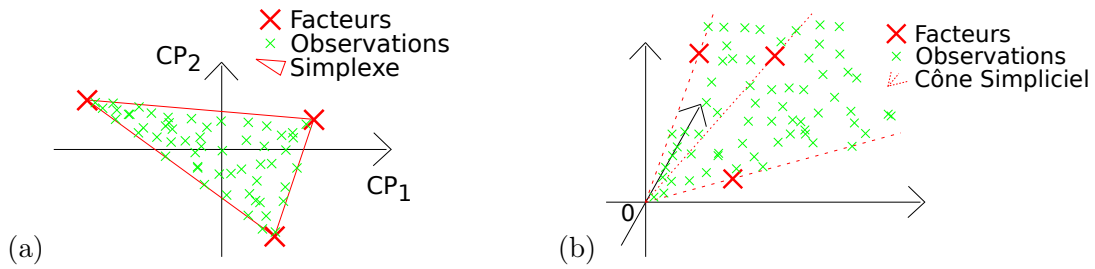


FIGURE 2.4 – (a) Représentation de données centrées réduites, non bruitées et composées de trois sources, dans l'espace fourni par les deux premières composantes principales de leur ACP (CP_1 et CP_2). Les facteurs (croix rouges) forment un simplexe qui englobe toutes les observations (croix vertes). Les observations qui sont sur un des côtés du triangle ne sont composées que de deux sources, celles étant aux extrémités du côté en question. (b) Représentation des données non centrées ni réduites sur trois coordonnées quelconques de la base canonique. Les facteurs (croix rouges) délimitent le cône simpliciel englobant toutes les observations (croix vertes).

2.4.2 Un problème encore sous-déterminé

Malgré l'ajout des contraintes de positivité qui restreignent le champ des solutions possibles, le problème de séparation de sources reste sous-déterminé [Hen87]. En effet, les problèmes d'identifiabilité liés aux rotations sont réduits mais toujours présents (cf figure 2.5).

L'interprétation géométrique des contraintes de positivité permet d'appréhender l'identifiabilité du problème de séparation de sources sous un nouvel angle. Reprenons les normalisations utilisées dans la section précédente. Grâce à ces normalisations le problème de séparation de sources est analogue à celui de la recherche des sommets d'un simplexe. Focalisons nous d'abord sur la contrainte de positivité de la matrice A . Cette contrainte impose que le simplexe formé par les facteurs contienne l'intégralité des observations. En effet, chaque observation peut être vue comme le barycentre des facteurs, dont les poids sont contenus dans A . Si une observation est à l'extérieur du simplexe, alors au moins un des facteurs a un poids négatif pour cette observation. La Figure 2.6(a) illustre cette remarque sur la positivité des coefficients de A . Pour une série d'observations donnée, cette contrainte n'est pas suffisante. Il est en effet possible de construire une infinité de simplexes englobant ces observations, comme suggéré dans la Figure 2.6(b). Cette infinité de simplexes résulte donc en une infinité de solutions pour la matrice F (et donc pour la matrice A aussi), et ces solutions peuvent être très différentes les unes des autres.

La seconde contrainte de positivité, sur la matrice F , permet de réduire le nombre de simplexes admissibles. Cette contrainte implique que les facteurs doivent tous se trouver dans l'orthant positif de $\mathbb{R}^P$. La Figure 2.7 illustre cette contrainte de positivité pour deux jeux de données différents, dans la base fournie par leur ACP. La zone admissible où la contrainte de positivité sur F est vérifiée forme dans cette base un ensemble convexe. Les observations se situant à la frontière de la zone admissible sont celles qui possèdent au moins un zéro dans leur signal [Hou84]. Selon les jeux de données, la contrainte de positivité sur F permet de

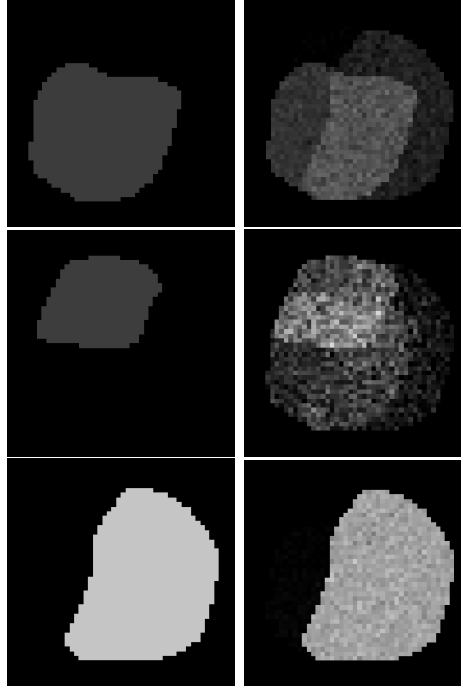


FIGURE 2.5 – Méthode de factorisation non négative appliquée à des données synthétiques. La première colonne contient la vérité terrain des images factorielles. La seconde les images factorielles obtenues par une méthode de séparation. Du fait de l'absence d'a priori, les images factorielles estimées sont mélangées entre elles par rapport à la vérité terrain. On peut voir dans les images factorielles estimées, des fractions plus ou moins importantes des autres images factorielles, montrant ainsi l'imperfection de la séparation.

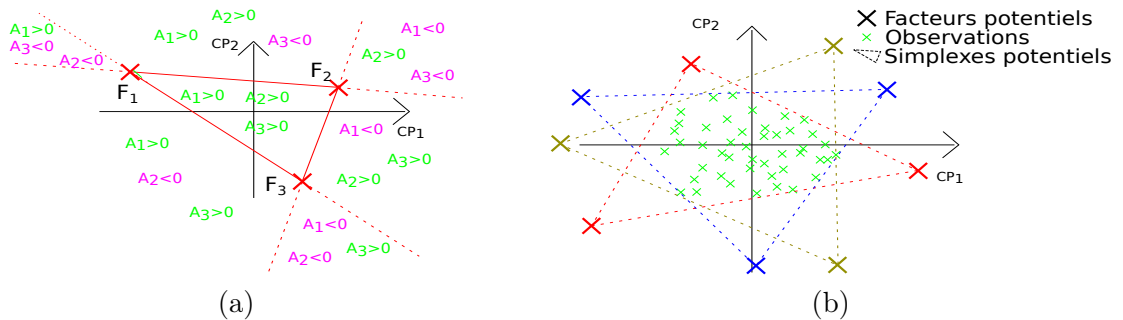


FIGURE 2.6 – (a) Représentation graphique des contraintes de positivité sur A dans l'espace fourni par l'ACP. Les facteurs sont les croix rouges. Le seul espace admissible pour une observation est l'intérieur du simplexe formé par les facteurs. (b) Illustration de l'infinité de simplexes englobant les observations.

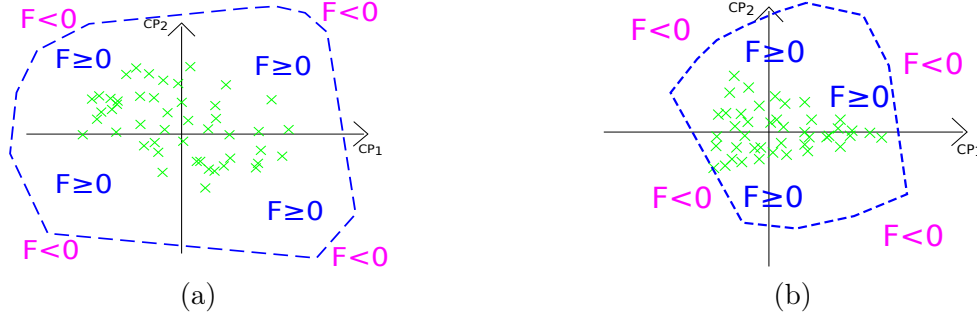


FIGURE 2.7 – (a) et (b) : Représentation graphique de la contrainte de positivité pour F , pour deux jeux de données différents, dans l'espace fourni par leur ACP respective. Les frontières de la zone admissible pour les facteurs sont représentées en pointillés bleus. À l'extérieur de ces frontières, au moins une coordonnée est négative dans la base canonique de $\mathbb{R}^P$.

restreindre plus ou moins les zones admissibles pour les facteurs.

Ainsi, les contraintes de positivité sur A et F agissent de façon complémentaire. Sur A , elles éloignent les facteurs du centre des données, faisant ainsi augmenter le volume du simplexe afin qu'il contienne toutes les observations. Sur F , elles ont un effet opposé : elles restreignent l'espace admissible pour les facteurs à une zone spécifique, limitant ainsi le volume du simplexe. Généralement, le couplage des contraintes de positivité sur A et sur F n'est toutefois pas suffisant pour assurer l'unicité de la solution. L'exemple de la Figure 2.8 permet de s'en convaincre : s'il existe une solution dont au moins un facteur n'est pas à la frontière de la zone admissible pour $F \geq 0$, alors il est possible de construire une infinité de solutions en éloignant ce facteur des autres. En effet, si ce facteur est déplacé dans la zone définie par le prolongement des arêtes incidentes du simplexe et par l'espace admissible qui satisfait $F \geq 0$, alors le nouveau simplexe contiendra bien l'ancien simplexe solution et donc toutes les observations, et sera aussi dans l'espace admissible. À l'inverse, si un simplexe solution existe et qu'aucune observation n'est sur une arête de ce simplexe, alors il est possible de réduire le volume de ce simplexe tout en conservant la condition $A \geq 0$.

Différents travaux se sont attachés à expliciter les conditions nécessaires et/ou suffisantes d'unicité du problème de séparation de sources avec contraintes de positivité. En imagerie nucléaire les premiers travaux ont permis de donner les intuitions de la non-unicité, et de son effet sur l'estimation des images factorielles [Hou84]. Mais c'est dans le cadre de la factorisation en matrices non négatives (NMF) que les travaux théoriques sur la non-unicité se sont principalement développés. Avec tout d'abord un théorème montrant que la solution est unique si et seulement si il existe un unique simplexe contenu dans l'orthant positif de $\mathbb{R}^P$ et englobant toutes les données [DS04] ; [Lau+08]. Ce théorème est démontré en utilisant les outils de l'analyse convexe [Roc15].

Un autre théorème [Lau+08] ; [HSS14] énonce qu'une condition nécessaire pour que la solution soit unique est que chaque colonne de A et chaque ligne de F contiennent au moins un zéro. On peut comprendre ce théorème de façon géométrique. La contrainte sur les zéros des colonnes de A entraîne que si l'on déplace un ou des sommets du simplexe vers l'intérieur,

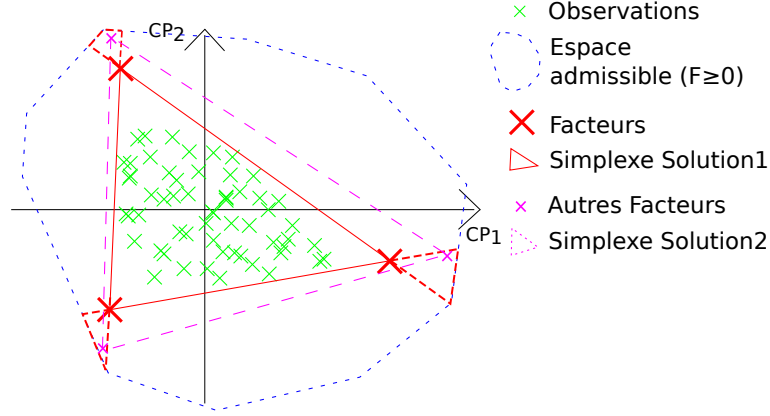


FIGURE 2.8 – Représentation graphique de la non suffisance des contraintes de positivité sur un jeu de données. En connaissant une solution (simplexe et facteurs rouges), il est possible de trouver des simplexes contenant ce premier simplexe et inclus dans l'espace admissible où F est positif. Il suffit pour cela de déplacer un des facteurs dans la zone derrière lui (en tireté rouge) délimitée par le prolongement des arêtes du simplexe et par l'espace admissible. Ces simplexes sont aussi solutions du problème de séparation. Un exemple de ces simplexes est dessiné en magenta.

alors ce dernier n'englobe plus complètement les observations. La contrainte sur les zéros des lignes de F entraîne que les facteurs se trouvent sur la frontière de la zone admissible, et ne peuvent donc pas se déplacer vers l'extérieur. Plus généralement, on peut montrer que si une solution est unique, alors il n'existe pas de couple de colonnes de A tel que les éléments non nuls d'une des colonnes soient aussi des éléments non nuls pour l'autre colonne, et de même pour les lignes de F [HSS14].

Enfin, il a été montré que la connaissance d'un certain nombre de zéros de la matrice des images factorielles A peut permettre de rendre le problème identifiable [Šám+89]. Ceci fournit une motivation supplémentaire pour l'utilisation des régions d'intérêt, qui peuvent aisément se traduire par des contraintes de coefficients nuls dans la matrice A .

2.4.3 Méthodes de séparation de sources avec contraintes de positivité

Il est tout d'abord nécessaire de préciser qu'il n'existe pas de solution analytique au problème de séparation de sources avec contraintes de positivité dans sa forme la plus simple [Lau+08], à savoir chercher A et F en minimisant le carré de la norme de Frobenius de l'erreur résiduelle :

$$(\tilde{A}, \tilde{F}) = \underset{A \geq 0, F \geq 0}{\operatorname{argmin}} \|Y - AF\|_F^2 \quad (2.14)$$

C'est un problème difficile car il fait intervenir des contraintes avec des inégalités (positivité des coefficients de A et F) et des produits d'inconnues. Ce problème étant NP-difficile [Vav09],

des algorithmes ont été développés afin de calculer une approximation de sa solution. On peut distinguer les approches basées sur l'optimisation, et celles basées sur l'interprétation géométrique décrite précédemment.

2.4.3.1 Méthodes d'optimisation

Les méthodes basées sur l'optimisation trouvent une estimation de A et F en minimisant une fonction objectif. Cette fonction est généralement la somme d'un terme de fidélité aux données et de contraintes liées aux a priori sur la solution recherchée. Le terme de fidélité aux données est présent afin que les matrices A et F expliquent au mieux les données, minimisant ainsi l'erreur résiduelle. Cette erreur résiduelle peut être calculée avec une distance euclidienne comme dans l'équation 2.14 lorsque le bruit est gaussien, ou bien avec une divergence de Kullback-Leibler généralisée comme dans l'équation 2.15 lorsque le bruit est poissonien.

$$(\tilde{A}, \tilde{F}) = \underset{A, F}{\operatorname{argmin}} D(Y||AF) = \underset{A, F}{\operatorname{argmin}} \sum_{i,t} \left(Y_{i,t} \times \log \left(\frac{Y_{i,t}}{(AF)_{i,t}} \right) - Y_{i,t} + (AF)_{i,t} \right) \quad (2.15)$$

Cependant, le seul terme de fidélité aux données n'est pas suffisant à cause de la sous-détermination du problème. Il est donc accompagné dans la fonction objectif par un ou plusieurs autres termes contenant des a priori sur la solution recherchée. Les a priori les plus courants sont la parcimonie et la régularité des sources. Selon l'application, d'autres a priori plus spécifiques peuvent être également choisis.

L'a priori de parcimonie suppose que la matrice A est parcimonieuse, c'est à dire qu'elle contient très peu d'éléments non nuls [Hoy04]. Ceci signifie que dans chaque observation, peu de sources sont présentes. Autrement dit, la matrice A est supposée creuse. Cette parcimonie devrait donc se traduire dans la fonction objectif par un terme contenant la pseudo-norme 0 de A ($\lambda \|A\|_0$, avec $\lambda > 0$) qui pénaliserait les coefficients non nuls dans A . Néanmoins, ce terme n'étant pas continu autour des matrices contenant des coefficients nuls, il complexifie grandement l'optimisation de la fonctionnelle. On préfère donc approcher cette parcimonie par un terme contenant une norme 1 ($\lambda \|A\|_1$ avec $\lambda > 0$) qui lui est bien continu partout [OF96]; [Hoy02]; [LZL03]. Ceci nécessite cependant de contrôler la norme de A ou de F [Hoy04], car sinon il est possible de faire tendre ce terme de parcimonie vers zéro tout en conservant la même fidélité aux données (il suffit simplement de multiplier F par une grande valeur, et diviser A par cette valeur). La constante positive λ permet de contrôler l'importance de la parcimonie vis à vis de la fidélité aux données. Plus elle est grande, plus la matrice A sera cherchée parcimonieuse, mais ceci se fait au prix d'une moins bonne fidélité aux données.

Les a priori concernant la régularité peuvent s'appliquer à la fois aux facteurs (F) et aux images factorielles (A). Dans de nombreuses applications, les facteurs et les images factorielles peuvent être supposés lisses. Il est donc pertinent d'introduire dans ce cas un terme de régularisation qui pénalise les facteurs et/ou images factorielles non lisses. En présence

de bruit fort et sans terme de régularisation, les facteurs et images facteurs obtenus ont généralement un aspect bruité. Les deux approches les plus courantes sont la régularisation Tikhonov [Tik63] ; [Ber+07], ou la variation totale [ROF92]. Dans la régularisation Tikhonov, on pénalise la norme 2 du gradient de la matrice supposée lisse. Ainsi, la fonction objectif contiendra le terme $\delta \|\Gamma A\|_F$, avec Γ un opérateur de différence et δ une constante positive, si on souhaite obtenir des images factorielles lisses. De la même manière, avec la variation totale, on pénalise la norme 1 du gradient de la matrice supposée lisse. La constante δ permet de contrôler l'importance de la régularité vis à vis de la fidélité aux données. Plus elle est grande, plus la matrice recherchée sera lisse et la fidélité aux données dégradée. La différence entre la variation totale et la régularisation Tikhonov est que cette dernière pénalise plus fortement les irrégularités fortes grâce à la norme 2, tandis que la variation totale favorise les fonctions constantes par morceaux.

Une fois la fonction objectif choisie, il s'agit maintenant de la minimiser afin de trouver des approximations des matrices A et F . À cause de la bilinéarité du modèle de mélange, le terme de fidélité aux données n'est pas conjointement convexe en A et en F . Cependant, la fonction objectif est généralement convexe en A et F séparément. Ainsi les méthodes utilisent principalement l'optimisation alternée afin de minimiser la fonction objectif. Avec l'optimisation alternée, les matrices A et F sont tour à tour estimées puis gardées constantes pendant que l'autre est estimée, ce processus étant répété en boucle jusqu'à convergence. En gardant une matrice fixée, le problème d'optimisation devient un sous-problème d'optimisation convexe en l'autre matrice. Ces sous-problèmes sont donc plus faciles, mais il faut les résoudre en prenant en compte les contraintes de positivité.

Les contraintes de positivité peuvent être gérées de deux façons. Soit elles sont ignorées dans un premier temps. La solution du sous-problème est alors calculée de manière analytique lorsque cela est possible, ou bien avec une méthode de descente de gradient. Les contraintes de positivité sont appliquées après chaque étape de la descente, en projetant la matrice dans l'orthant positif. Cette méthode est très simple mais la convergence vers un minimum local n'est pas garantie du fait de la projection et des normalisations potentielles. Soit le sous-problème est optimisé en prenant directement en compte les contraintes, en utilisant des outils comme la programmation quadratique qui assure une convergence vers un minimum global, mais le temps de calcul est généralement très long [HS06].

La méthode qui a permis de populariser la NMF est celle utilisant une règle de mise à jour multiplicative [LS01]. Elle utilise l'optimisation alternée, chaque matrice est tour à tour mise à jour en multipliant chacun de ces éléments par un coefficient qui dépend des valeurs des deux matrices A et F . Ceci revient à utiliser une méthode de gradient à faible pas de descente, de sorte que la positivité des coefficients des matrices A et F est assurée. Cette méthode est simple, rapide, et la convergence vers un minimum local est garantie. Elle ne comporte cependant qu'un seul terme dans sa fonction objectif, un terme de fidélité aux données, comme la plupart des autres algorithmes de NMF [Lin07] ; [CP09] ; [KP11].

Enfin, on peut aussi évoquer l'analyse en composantes indépendantes non-négatives [Plu03]. Dans cette méthode, les hypothèses d'indépendance statistique des sources, ainsi que leur positivité sont combinées. Cependant, dans [Plu03], l'hypothèse de positivité s'applique

uniquement à l'une des deux matrices A ou F .

2.4.3.2 Méthodes géométriques

De nombreuses méthodes de séparation de sources avec contraintes de positivité se basent sur la puissante interprétation géométrique de ces dernières. Dans ces méthodes, les facteurs sont généralement recherchés en premier. Les images factorielles sont ensuite calculées en utilisant la matrice des facteurs F obtenue. Il est évidemment possible de choisir l'approche symétrique en utilisant la remarque 2.1. Différentes stratégies sont envisageables pour retrouver les facteurs. Il faut tout d'abord faire le choix de l'interprétation géométrique. Il est possible d'utiliser l'interprétation avec le cône simpliciel ou bien celle avec le simplexe (voir Figure 2.4). Dans la première, aucune normalisation n'est opérée, les facteurs se retrouvent chacun sur une arête du cône simpliciel englobant les données. Une estimation des facteurs est donc équivalente à une estimation des arêtes. Dans la seconde, les matrices Y , A et F sont normalisées, et cette fois-ci les données sont contenues dans un simplexe dont les sommets sont les facteurs. Ce sont donc les sommets du simplexe qui seront recherchés.

Comme vu précédemment, le problème de séparation de sources avec contraintes de positivité reste sous-déterminé. Il est donc nécessaire de faire des hypothèses supplémentaires afin de trouver *le* cône simpliciel ou *le* simplexe recherché. Parmi ces hypothèses, deux sont très souvent utilisées : l'hypothèse de présence d'observations pures, et l'hypothèse de volume minimal.

L'hypothèse de présence d'observations pures (ou pixels purs) suppose que pour chaque source, il existe au moins une observation ne contenant que cette source. Une telle observation est dite pure. Les lignes associées à ces observations pures dans la matrice A sont donc remplies de 0, excepté pour la coordonnée correspondant à la source considérée. L'interprétation géométrique est dans ce cas encore plus forte, puisque le signal de l'observation pure est le même que celui de la source (à un facteur d'échelle près). Ainsi, chaque sommet du simplexe recherché est confondu avec au moins une observation (au bruit près). De même, chaque arête du cône simpliciel recherché contient au moins une observation. Les sommets du cône simpliciel et les arêtes peuvent donc être trouvés parmi les observations elles-mêmes, en cherchant les observations pures. Les observations, et en particulier les observations pures, vérifient ainsi nécessairement la contrainte de positivité sur leur signal (et donc sur F). En revanche, un ensemble de K observations pures prises comme facteurs ne vérifiera pas nécessairement la contrainte de positivité sur la matrice A . Afin de satisfaire cette contrainte, il faut que l'enveloppe convexe engendrée par ces K observations contienne toutes les autres observations. Cette contrainte ne peut généralement pas être satisfaite à cause du bruit. La projection des données dans un sous-espace de petite dimension [BKG95] permet de limiter ce problème.

Étant donné K sources et N observations, il y a $\binom{N}{K}$ combinaisons possibles de K observations, c'est à dire un nombre fini de possibilités. Il n'est cependant pas envisageable de tester la validité de chacune de ces combinaisons rapidement lorsque N est grand. Différents algorithmes ont donc été développés afin de trouver rapidement les observations pures, et donc

les facteurs.

Que ce soit avec le simplexe ou bien le cône simpliciel, les facteurs seront aux extrémités de l'enveloppe convexe engendrée par les observations. Dans l'espace de petite dimension fourni par l'ACP, on peut définir une observation extrême comme étant une observation qui ne peut pas être exprimée comme une combinaison linéaire à poids positifs d'observations différentes. On peut montrer que lorsque l'hypothèse précédente sur les observations pures est vérifiée et que le modèle de mélange linéaire sans bruit est valide, il y a équivalence entre observations extrêmes et facteurs. Cette équivalence est vraie quelque soit l'interprétation géométrique utilisée (simplexe ou bien cône simpliciel). On peut aussi remarquer que le simplexe de volume le plus important ayant comme sommets des observations, est celui ayant comme sommets des observations pures. Autrement dit, le simplexe de volume le plus important formé avec des données est celui que l'on recherche. Ce simplexe est également le simplexe quelconque de volume minimal englobant toutes les données.

De nombreux algorithmes de séparation découlent des remarques précédentes. Afin de trouver les facteurs, les algorithmes N-FINDR [Win99] et SGA [Cha+06] font grossir un simplexe à partir des données afin de trouver le simplexe de volume maximal formé avec des données. Les facteurs sont choisis comme étant les sommets du simplexe final. Avec la même idée, mais avec l'approche du cône simpliciel, l'algorithme SMACC [GRH04] fait grossir un cône grâce aux données, en incrémentant les dimensions progressivement. Le vecteur observation ajouté comme arête au cône étant celui ayant un angle maximum avec le cône de l'itération précédente (cf figure 2.4). Les algorithmes VCA [ND05] et SVMAX [Cha+11] construisent les sommets du simplexe itérativement, en projetant les données dans une direction orthogonale au simplexe en cours. Le vecteur observation ajouté comme sommet est celui ayant la plus grande énergie une fois projeté.

L'hypothèse relative aux observations pures est très pratique pour la séparation. Elle implique que le simplexe ayant comme sommet les facteurs est le simplexe de volume minimal contenant les données, et est aussi le simplexe de volume maximal construit dont les sommets sont des observations. Néanmoins cette hypothèse n'est pas toujours vérifiée, certaines sources peuvent ne pas avoir d'observations pures. Il n'est donc plus possible de rechercher les sommets du simplexe, ou les arêtes du cône simpliciel, parmi les données. On peut cependant faire une autre hypothèse moins contraignante que la première : l'hypothèse de volume minimal. L'hypothèse de volume minimal suppose simplement que le simplexe contenant les facteurs est le simplexe de volume minimal contenant toutes les données. Tandis que l'hypothèse d'observations pures concernait les sommets du simplexe, l'hypothèse de volume minimal est plus souple et concerne plus généralement les frontières du simplexe. Des conditions nécessaires et/ou suffisantes pour que l'hypothèse de volume minimal soit valide ont été démontrées récemment [Lin+15] ; [Ma+15]. Géométriquement, cette hypothèse suggère qu'il y a suffisamment d'observations sur chacun des bords du simplexe, c'est à dire là où au moins une source est absente.

N'étant plus choisis parmi les observations avec cette hypothèse, les sommets du simplexe peuvent se situer dans un nombre infini d'endroits. Le problème de séparation est donc beaucoup plus difficile qu'avec l'hypothèse de pixels purs. Trouver le simplexe de volume minimal

englobant les données est en plus un problème non convexe. Trouver systématiquement un minimum global n'est donc pas envisageable.

Dans la suite, on se place dans l'espace de dimension $K-1$ fourni par l'ACP sur les données centrées réduites. En notant G la matrice des facteurs projetés dans cet espace, le volume du simplexe S dont les sommets sont ces facteurs s'écrit :

$$V(S) = \frac{1}{(K-1)!} \left| \det \begin{bmatrix} 1 & \dots & 1 \\ & G' & \end{bmatrix} \right| = \frac{1}{(K-1)!} \left| \det \begin{bmatrix} 1 & \dots & 1 \\ G'_{:,1} & \dots & G'_{:,K} \end{bmatrix} \right| \quad (2.16)$$

Afin de minimiser ce déterminant tout en gardant les observations à l'intérieur du simplexe, différentes méthodes ont été développées. Dans [Cra94], le point de départ est un grand simplexe qui englobe les données. Puis, itérativement ses sommets sont déplacés, chacun leur tour, afin de minimiser le volume du simplexe tout en conservant les observations à l'intérieur de celui-ci. Des approches similaires pour le cône simpliciel ont aussi été développées [SI89]; [Oue+12]. Afin d'être plus robuste au bruit, les contraintes de positivité ont été assouplies dans [LBD08]; [BD09], permettant ainsi aux observations bruitées d'être légèrement à l'extérieur du simplexe.

Il est également possible de trouver le simplexe de volume minimal en minimisant une fonctionnelle. Un terme pénalisant le volume du simplexe peut accompagner le terme de fidélité aux données afin de régulariser le problème, comme dans la méthode MVC-NMF [MQ07]. Toutefois, le volume du simplexe avec le calcul du déterminant n'étant pas une fonction convexe en la position de ses sommets, la minimisation de la fonctionnelle est délicate. Ce terme peut-être approché par la somme des carrés des distances entre les sommets [Ber+04], qui elle est bien convexe.

Les méthodes géométriques ne sont cependant pas performantes lorsque les sources sont très mélangées et qu'il y a peu de zéros dans la matrice A . Lors de mélanges importants ni l'hypothèse de pixels purs, ni l'hypothèse de volume minimal ne sont valides. Des méthodes utilisant des connaissances a priori fortes sont donc nécessaires pour séparer les sources.

2.5 Méthodes de séparation de sources en imagerie nucléaire

Cette section commence par un historique de la séparation de sources en imagerie nucléaire avec une présentation des différentes méthodes et de leurs évolutions. Cet historique est suivi d'une partie sur la réduction de dimension et sur la gestion du bruit, ainsi que d'une description plus complète des principales méthodes.

2.5.1 Historique de la séparation de sources en imagerie nucléaire

Dans les années 1970, la masse de données fournies par l'imagerie nucléaire a motivé l'utilisation de l'analyse en composantes principales. L'ACP facilitait grandement l'analyse de

ces séquences d'images. D'une part elle permettait de réduire drastiquement la dimension des données en conservant l'information utile [Bar+75], enjeu important étant donné les faibles capacités des ordinateurs de l'époque en terme de stockage et de puissance de calcul. Cette réduction de dimension pouvait aussi servir au débruitage des séquences d'images. D'autre part l'ACP pouvait être utilisée à des fins de diagnostics, afin de classer les études entre cas normaux et cas pathologiques [Bar+75] ; [Opp78] ; [Hou80]. D'autres méthodes d'analyse factorielle, telle l'analyse par correspondance [Ben92] ont aussi été utilisées dans le même but [BPDP70] ; [Sch+77] ; [Sch77]. Cependant, une des faiblesses de ces méthodes est que les facteurs (les composantes principales pour l'ACP) qu'elles fournissent n'ont pas de sens physiologique. Par exemple, les facteurs, qui sont censés représenter des courbes d'activités, pouvaient avoir des coefficients négatifs. L'interprétation physiologique des résultats de ces méthodes était donc limitée.

Parallèlement, l'obtention des courbes d'activité des différents organes et tissus d'intérêt se fait à l'aide de régions d'intérêt dans différentes études [Ron+75] ; [Sho+73]. La sélection de ces régions d'intérêt présente de nombreux problèmes. Premièrement, elle se fait généralement manuellement et il est difficile de l'automatiser. Deuxièmement, la sélection manuelle dépend de l'utilisateur et est sujette à la variabilité intra-observateur et inter-observateur. Troisièmement, la région d'intérêt sélectionnée contient du bruit de fond dû à l'activité dans le sang et dans les organes et tissus voisins ou superposés. Une soustraction de ce bruit de fond peut s'opérer, mais cela se fait via la sélection de régions d'intérêt bien spécifiques et sous l'hypothèse d'uniformité spatiale du bruit [Ken+75] ; [CGK08]. Ces méthodes d'obtention de courbes d'activité n'étant pas satisfaisantes [Bar80] ; [Cos85] ; [NB85] ; [CGK08], il fallait trouver d'autres méthodes pour les extraire.

Puis au début des années 1980, les articles fondateurs de la séparation de sources en imagerie nucléaire sont publiés [Bar80] ; [Baz+80] ; [DP+82]. Les séquences d'images sont supposées être constituées d'un mélange d'un petit nombre de compartiments physiologiques, ayant chacun des signatures propre. Ces signatures spatiales et temporelles, qui sont respectivement les cartes de présence (matrice A) et les courbes d'activité (matrice F), sont supposées à coefficients positifs. Les méthodes développées permettent non seulement d'obtenir des facteurs ayant une signification physiologique, mais aussi les images factorielles qui décrivent la répartition spatiale de ces facteurs. Afin de retrouver ces signatures, les données normalisées et centrées sont projetées dans un espace de petite dimension grâce à une ACP réalisée dans le domaine temporel, puis les sommets du simplexe englobant les données sont recherchés. Ces sommets correspondent aux facteurs, conformément à l'interprétation géométrique des contraintes de positivité évoquée précédemment dans la section 2.4.1. Ils sont estimés conjointement avec les images factorielles grâce à une initialisation puis une mise à jour itérative et alternée jusqu'à convergence : les images factorielles sont fixées pour mettre à jour les facteurs et inversement. Les mises à jour sont faites grâce à des expressions analytiques reliant les données avec les facteurs et les images factorielles. Ces méthodes sont donc constituées de deux grandes étapes : l'analyse orthogonale, puis l'analyse oblique. Dans la première étape, un sous-espace de petite dimension contenant les facteurs est déterminé grâce à une ACP. Dans la seconde étape, les composantes principales de l'ACP subissent une rotation afin qu'elles n'aient plus de coefficients négatifs pour être ainsi interprétées comme des facteurs. Cette rotation

doit également permettre aux images factorielles de ne plus avoir de coefficients négatifs.

Ces premiers travaux ont posé le problème de la séparation de sources en imagerie nucléaire et y ont trouvé une réponse partielle, avant même que le problème de séparation de sources aveugle (qui a par la suite donné naissance à l'ICA) soit formulé [HA84] ; [JT00]. Ces travaux en imagerie nucléaire précédaient également ceux en imagerie multispectrale [ASJ86] et les algorithmes de factorisation en matrices positives [PT94] ou non-négatives [LS01].

Ces méthodes, regroupées sous le terme d'analyse factorielle des séquences dynamiques (FADS en anglais), présentaient l'avantage d'être automatiques et non dépendantes de l'utilisateur. De plus, elles permettaient d'obtenir à la fois les facteurs et les images factorielles. Les images factorielles pouvaient être utilisées pour identifier chacune des sources, et vérifier la cohérence des résultats. Cependant, comme le précisent Nijran et Barber, la solution trouvée est essentiellement créée par l'algorithme utilisé [NB85]. En effet, ces méthodes ne prenaient pas en compte la non-unicité de la solution du problème de séparation, menant à des solutions bien souvent erronées et très dépendantes de l'initialisation à cause de l'existence de minima locaux. Ce problème de non-unicité est pointé par Houston[Hou84]. Les évolutions suivantes se concentrent donc principalement sur l'introduction d'informations a priori dans les méthodes afin de guider l'algorithme vers une solution unique ayant un sens physiologique.

Les informations ajoutées sont de différents types. Certaines concernent la forme et les valeurs des courbes d'activité [NB85] ; [NB86] ; [Buv+93]. Dans [NB85], le problème de séparation est simplifié et un seul facteur est recherché. Les a priori sur ce facteur (paramétrisation de la courbe par exemple) permettent de chercher ce facteur dans un espace restreint. Ce facteur est finalement trouvé comme étant celui le plus proche à la fois de cet espace et de l'espace d'étude défini par l'ACP sur les données d'origine. L'inconvénient est que cette méthode permet de ne trouver qu'un seul facteur, et l'image factorielle associée n'est pas non plus calculée. Dans [NB86], cette méthode est utilisée précédemment à une analyse factorielle classique permettant ainsi de retrouver tous les facteurs et images factorielles. Une idée similaire est présente dans [Buv+93], où les facteurs dont des connaissances a priori sont disponibles, sont tout d'abord déterminés puis utilisés pour trouver les facteurs restants.

Afin de contrer la non-unicité, d'autres auteurs proposent de chercher une solution maximisant l'entropie des facteurs [NSK89] ; [SDBG99] en considérant leur signal comme des distributions de probabilités, mais la motivation de ce critère reste floue. Dans [VD+91], Van Daele suppose lui que le simplexe à trouver est celui de volume minimal, à la manière de certains algorithmes utilisés plus tard en imagerie hyperspectrale [Cra94]. Il suppose également que les images factorielles contiennent essentiellement des 1 et des 0. Cette hypothèse signifie que peu de pixels contiennent des mélanges de différentes sources. Une hypothèse similaire est faite par la suite par Sitek et al., qui introduisent une fonctionnelle comportant un terme pénalisant les images factorielles se ressemblant [SGH02] ; [EF+05]. Cette hypothèse est reprise pour l'appliquer à la fois aux images factorielles et aux facteurs dans [Kle+07]. Il est donc recherché dans cette méthode des sources ayant des signatures spatiales et temporelles n'ayant que peu de support commun.

Dans [Šám+87] ; [Šám+88], la remarque 2.1 sur le rôle symétrique des matrices A et F est

utilisée et l'ACP est réalisée dans le domaine spatial plutôt que dans le domaine temporel. Comme les séquences possèdent généralement moins d'images que de pixels, cette approche permet de diminuer l'impact du bruit ainsi que le temps de calcul [Šám+89]. Il est d'ailleurs montré qu'en utilisant cette approche ainsi que la connaissance de suffisamment de zones qui ne contiennent pas une source, le problème de séparation pouvait être identifiable [Šám+89]. Cette connaissance, qui revient à la connaissance des régions d'intérêt des sources recherchées, permet de forcer à 0 certains coefficients de la matrice A . Une méthode utilisant cette connaissance des ROI a précédemment été développée dans [NB88]. Cependant, elle n'introduit aucune contrainte sur la positivité des coefficients des matrices A et F , et produit donc des résultats contenant des coefficients négatifs et donc ininterprétables.

Tandis que la plupart des méthodes proposent d'ajouter des connaissances a priori uniquement d'ordre spatial ou uniquement d'ordre temporel, Benali et al. proposent une méthode où il est possible d'ajouter des a priori dans ces deux domaines [Ben+94b]. Néanmoins, les connaissances qu'il est possible d'ajouter dans cette méthode sont très limitées car il est uniquement possible de fixer des coefficients à l'avance dans les matrices A et F . Cette méthode utilise également un modèle statistique et physique plus sophistiqué [Ben+93], qui permet de déterminer une métrique optimale pour l'analyse orthogonale et l'analyse oblique. Il est montré qu'avec ce modèle dans le cadre de l'imagerie nucléaire, l'analyse orthogonale optimale revient à faire une analyse factorielle des correspondances [Ben92].

Le pré-traitement des données est lui aussi sujet à quelques évolutions. Afin de réduire le bruit, une agrégation des données peut être effectuée en amont de la réduction de dimension avec l'ACP. Les pixels peuvent être tout simplement regroupés en carrés de 2×2 pixels, 4×4 pixels, ou encore davantage [Buv+93]. Des agrégations plus sophistiquées ont été développées afin d'obtenir des agrégats contenant des pixels ayant des proximités sur le plan spatial mais aussi sur le plan temporel [Aur89], c'est à dire des ressemblances statistiques sur la dynamique de leur activité. L'algorithme d'agrégation CAMIS, utilise également les proximités spatiales et temporelles pour regrouper les pixels en utilisant itérativement un critère de plus proche voisin mutuel [Ben+94a]. Lorsque les données sont agrégées en amont de la séparation de sources, les images factorielles obtenues par l'algorithme de séparation ont une résolution spatiale bien plus basse que les données initiales. Il est donc souhaitable de revenir à la résolution spatiale d'origine. Comme les facteurs eux ne sont pas impactés (car d'après l'hypothèse d'homogénéité spatiale des compartiments physiologiques ils ne dépendent pas de la position), ils sont utilisés avec les données d'origine pour retrouver des images factorielles à la bonne résolution grâce aux moindres carrés. Cependant, le bruit des données d'origine impacte grandement la qualité des images factorielles. Une méthode a donc été développée afin de restaurer les images factorielles à partir des facteurs tout en les régularisant [Fro+99]. La régularisation utilisée permet de lisser les images factorielles tout en préservant les bords des compartiments physiologiques.

Certaines méthodes proposent de fusionner le modèle de mélange avec un modèle compartimental représentant les relations entre les cinétiques des différents facteurs [FBDP89]; [KBA+08]; [Dow+10]. À noter que le modèle compartimental contraint les facteurs, mais pas les images factorielles. Les solutions trouvées doivent donc respecter à la fois le modèle de mélange, et le modèle compartimental. Les paramètres physiologiques liés au modèle compar-

timental peuvent également être estimés conjointement avec les sources dans certaines de ces méthodes [KBA+08] ; [Dow+10].

Plus récemment encore, il a été proposé d'utiliser une approche bayésienne afin d'introduire des a priori qui peuvent être très variés sur les sources [ŠTŠ11] ; [ŠT12] ; [TŠ15]. Le problème est résolu à l'aide de la méthode de Bayes variationnelle [Qui+06] qui permet de simplifier les calculs très complexes engendrés par les modèles bayésiens hiérarchiques utilisés. Ceci permet d'introduire des a priori sophistiqués comme la forme des noyaux de convolution régissant les interactions entre les facteurs et une fonction d'entrée [ŠTŠ11] ; [TŠ15]. Dans [TŠ15], des a priori de parcimonie sur ces noyaux de convolution ainsi que sur les images factorielles sont intégrés à la méthode. Cette méthode tente également de déterminer de manière automatique le nombre de sources présentes en supprimant pendant la séparation les sources de trop faible intensité. La fiabilité de l'estimation du nombre de sources peut toutefois être mise en question [FDM17].

Initialement prévues pour les séquences (2D+t) issues de gamma caméra, les méthodes s'adaptent très bien à l'avènement des caméras (3D+t) (SPECT et PET). En effet, comme dans la plupart des cas la cohérence spatiale n'est pas utilisée, les images sont traitées comme des vecteurs. Les volumes sont donc traités de même en étant vectorisés [Fro+92]. L'approche utilisant une analyse factorielle sur les données complètes (3D+t) est montrée plus performante que l'approche consistant à traiter les volumes en concaténant les différentes couches pour obtenir des séries de données (2D+t) [Fro+93].

Très peu de travaux se sont intéressés à la validité du modèle de mélange linéaire. Ce dernier ne constitue une bonne approximation que dans une petite zone. Il est d'usage de réduire la taille des images manuellement en sélectionnant une fenêtre d'intérêt avant d'appliquer la méthode de séparation sur cette fenêtre. Lorsque la séparation ne donne pas satisfaction, cette étape de pré-traitement sert de paramètre d'ajustement de la méthode. Ainsi, d'autres fenêtres sont choisies jusqu'à ce que la solution de la méthode de séparation soit conforme aux attentes [ŠT12]. Ce processus n'étant pas satisfaisant, une méthode s'affranchissant de la sélection manuelle de fenêtres a été développée dans [ŠT12]. Des régions d'intérêt probabilistes y sont estimées automatiquement. Aucune connaissance a priori n'est utilisée dans cette méthode, ce qui est un inconvénient dans un problème aussi complexe que la séparation de sources.

D'autres méthodes ont été développées afin de séparer des sources de faible intensité mélangées à des sources de plus forte intensité [Šám+86] ; [Šám+98] ; [Ber+99]. Dans ces méthodes, l'algorithme de séparation est appliqué localement, tant sur le plan spatial que temporel. Dans la séquence prise dans sa globalité, la source de signal faible ne contribue que très peu à la variance des données, et est donc difficile à séparer. L'idée est donc d'augmenter l'influence de cette source. Ceci n'est pas simple car elle est mélangée aux autres, on ne peut donc pas appliquer une transformation qui n'a d'influence que sur cette source. Des connaissances a priori sur cette source sont donc utilisées. L'idée est de chercher une zone temporelle et spatiale où cette source est la plus forte, et y appliquer localement un algorithme de séparation. C'est à dire que la séparation se fait sur une quantité réduite d'images de la séquence (localité temporelle), et sur une zone restreinte de ces images (localité spatiale). L'inconvénient de cette méthode est que du fait de ces rognages, une grande quantité d'information de la séquence

d'origine qui aurait pu être utilisée pour la séparation est perdue.

2.5.2 Gestion du bruit

Comme vu dans la section 1.2, le bruit dans les séquences d'images en imagerie nucléaire est principalement du bruit poissonien. L'intensité de ce bruit varie en fonction des doses de traceurs injectées, de leur demi-vie, de la distance à la caméra ou encore de la résolution spatiale et temporelle. Ce bruit a évidemment un impact important sur l'estimation des facteurs et images factorielles. Il est possible de modifier les protocoles d'acquisition afin d'augmenter le rapport signal sur bruit, mais ces modifications ne sont pas toujours faisables ni souhaitables. Différentes stratégies sont envisageables en travaillant directement sur la séquence d'images fournie par la caméra.

Une première stratégie est de débruiter la séquence d'images. Comme le bruit dans les séquences est majoritairement poissonien, un algorithme de débruitage dédié à ce type de bruit est recommandé. La technique la plus courante est d'utiliser en premier lieu une transformation de stabilisation de la variance, qui permet de rendre la variance des pixels non dépendante de leur signal. Le bruit résultant peut alors être considéré comme étant gaussien. Une transformation stabilisatrice de variance classique est la transformée de Anscombe [Ans48] qui transforme chaque coefficient de la matrice des données en un autre qui a un écart type asymptotiquement non dépendant de la moyenne du coefficient initial :

$$Ans : x \mapsto 2\sqrt{x + \frac{3}{8}} \quad (2.17)$$

Il est aussi possible d'utiliser une transformée de Anscombe généralisée lorsque le bruit est de type Poisson-Gauss [SMB98]. Une fois une transformation appliquée, l'étape suivante est de débruiter les images une par une grâce à des algorithmes dédiés au bruit gaussien, tel BM3D [Dab+07] ou d'autres utilisant les ondelettes et leurs dérivées [ZFS08]. Enfin la dernière étape est d'appliquer une transformée inverse pour annuler la transformée stabilisatrice de variance faite au départ, afin que les coefficients retrouvent leur ordre de grandeur initial [MF11].

Les méthodes de débruitage ont toutefois suscitées peu d'intérêt pour la séparation de sources en imagerie médicale. Ces algorithmes ne prenant pas en compte le modèle de mélange linéaire, ce dernier peut se trouver encore moins respecté suite à un débruitage de ce type, rendant ainsi la séparation d'autant plus compliquée. De plus, le débruitage introduit généralement un lissage des images factorielles, et donc une augmentation des mélanges ce qui ne facilite pas la séparation.

Une seconde stratégie est de diminuer la résolution spatiale ou bien temporelle des données. Des techniques d'agrégation de pixels peuvent être utilisées. Ces techniques ont l'avantage de respecter le modèle de mélange linéaire, mais elles ont le défaut de mélanger davantage les images factorielles, ce qui rend le problème de séparation plus complexe. Une autre technique très utilisée est l'analyse orthogonale, à l'aide d'une ACP ou bien d'une analyse factorielle des correspondances, qui permet de réduire très fortement la résolution temporelle ou bien spatiale. Cette diminution de la résolution se fait de telle sorte qu'un maximum de la variance

des données initiales est conservé. Les composantes principales de petites valeurs propres étant supposées contenir principalement du bruit, cette méthode permet de débruiter les données. Cette méthode est également pratique pour accélérer les calculs des méthodes de séparation, car la dimension du problème est réduite. Cependant un des soucis majeurs est que la non-négativité intrinsèque des données n'est pas prise en compte dans la réduction de dimension. Les données projetées dans l'espace réduit peuvent donc se retrouver avec des coefficients négatifs une fois qu'elles sont replongées dans l'espace d'origine. Ce problème peut mettre en difficulté des méthodes de séparation appliquant une non-négativité stricte sur les facteurs et images factorielles, rendant le problème insoluble.

Enfin une dernière stratégie est de régulariser le problème en supposant que les images factorielles et/ou les facteurs sont lisses. L'idée est donc d'optimiser une fonctionnelle comportant un terme de fidélité aux données incluant le modèle de mélange, et un terme régularisant qui pénalise les solutions non lisses. Les deux approches les plus utilisées sont la régularisation Tikhonov [Tik63] ; [Ber+07] et la variation totale [ROF92].

2.5.3 Principales méthodes

Cette section décrit les principales méthodes de séparation de sources utilisées en imagerie nucléaire. Les méthodes développées pendant cette thèse sont qualitativement comparées aux méthodes décrites ci-dessous sur différents jeux de données dans le chapitre 4.

2.5.3.1 FAMIS

Plusieurs méthodes différentes ont pour nom FAMIS, qui est l'acronyme anglais pour "Factor Analysis of Medical Image Sequences", c'est à dire analyse factorielle de séquences d'images médicales. Nous détaillons ici la version décrite dans [Ben+94b]. Cette version, la plus aboutie, autorise l'utilisation conjointe de connaissances d'ordre spatial (sur A) et d'ordre temporel (sur F). Pour ce faire, les auteurs proposent un modèle statistique et un modèle physique des données qui fournissent un cadre théorique solide à la méthode.

Le modèle statistique choisi est le modèle à effet fixe. Dans ce modèle, l'activité de chaque pixel est représentée par un vecteur de taille P (P étant le nombre d'images). Ce vecteur est étudié comme un vecteur aléatoire composé d'un vecteur d'effet fixe, et d'un vecteur d'erreur, qui lui est de moyenne nulle et de variance connue, à un facteur près. Il est montré que dans le cadre de l'imagerie nucléaire, avec un bruit poissonien, la décomposition orthogonale qui approche au mieux l'effet fixe est l'analyse factorielle des correspondances [Ben92] ; [Ben+93]. Cette décomposition orthogonale permet de définir un espace d'étude de petite dimension, à la fois dans le domaine spatial et dans le domaine temporel.

Le modèle physique suggère quant à lui que les facteurs appartiennent tous à l'espace d'étude temporel de petite dimension, et que symétriquement, les images factorielles appartiennent toutes à l'espace d'étude spatial de petite dimension également. Ces deux contraintes

sont compatibles entre elles, et aboutissent à des normalisations, ainsi qu'à des relations permettant d'exprimer les facteurs en fonction des images factorielles et inversement.

Afin de résoudre le problème de séparation, les auteurs proposent une méthode de mise à jour itérative et alternée. L'initialisation des matrices A et F se fait en choisissant des pixels ou des images pertinentes parmi les données dans l'espace d'étude. Les règles de mise à jour se font grâce aux relations qui expriment les facteurs en fonction des images factorielles (et inversement). Ces relations sont obtenues grâce aux modèles statistique et physique. Après chaque mise à jour, les coefficients négatifs dans les matrices des facteurs et des images factorielles sont forcés à zéro, puis ces matrices sont re-normalisées. Au même moment, si des connaissances a priori sont disponibles, il est possible de forcer la valeur de certains coefficients de ces matrices. De telles connaissances peuvent par exemple être qu'un facteur doit être nul à partir d'une certaine image. Ou bien que tel compartiment physiologique n'est pas présent dans certaines parties de l'image, ce qui revient à tracer grossièrement une région d'intérêt. L'algorithme de mise à jour itérative et alternée s'arrête lorsque le nombre de coefficients négatifs des matrices des facteurs et des images factorielles est assez bas.

Le grand avantage de cette méthode est de fournir un cadre théorique solide, menant à une décomposition orthogonale optimale prenant en compte le bruit poissonien. La possibilité d'ajouter des a priori à la fois d'ordre spatial et temporel est également un atout intéressant. Cependant, très peu de types de connaissances peuvent être intégrés. En effet, celles-ci doivent forcément aboutir à une valeur à inscrire directement dans les matrices des facteurs et des images factorielles. Ces connaissances se résument donc le plus souvent à des informations de présence ou d'absence d'un facteur sur un intervalle de temps, ou bien d'une image factorielle sur une zone. Si des facteurs ou des images factorielles peuvent être connus à l'avance il est possible d'intégrer ces informations dans la méthode. Ce sont des connaissances très intéressantes pour la séparation, mais elles sont limitées et ne sont pas intégrées de manière robuste dans FAMIS car elles forcent la valeur de certains coefficients sans adaptation possible.

2.5.3.2 PLS-FADS

PLS-FADS est l'acronyme anglais pour "Penalized Least Squares Factor Analysis of Dynamic Sequences", que l'on pourrait traduire par "Analyse factorielle des séquences dynamiques à l'aide des moindres carrés pénalisés". Cette méthode est détaillée dans [SGH02]. Les auteurs font notamment remarquer que le principal obstacle à la séparation est la non-unicité de la solution. Pour contrer ceci, les méthodes utilisent des connaissances a priori plus ou moins importantes, perdant ainsi en généralité. Les effets visibles de la non-unicité sont que les facteurs et images factorielles obtenus sont encore des mélanges, des combinaisons linéaires des facteurs et images factorielles de la vérité terrain. On retrouve ainsi dans une image factorielle la trace d'une ou plusieurs autres images factorielles. Les auteurs ont donc proposé une fonctionnelle à minimiser qui comporte un terme pénalisant les images factorielles ayant des supports communs. Cette méthode ne demande aucune connaissance spécifique sur les données et est donc très générale. Les compartiments physiologiques recherchés doivent toutefois avoir des supports spatiaux majoritairement disjoints.

La fonction objectif proposée comporte trois termes : un terme de fidélité aux données, un terme pénalisant les solutions à coefficients négatifs, et enfin le terme pénalisant les images factorielles ayant des supports communs. En reprenant les notations utilisées précédemment, avec Y la matrice des données et A et F les matrices des images factorielles et des facteurs, la fonction objectif f_{PLS} s'écrit de cette manière :

$$f_{PLS}(A, F) = f_{LS}(A, F) + a \times f_{neg}(A, F) + b \times f_{uni}(A) \quad (2.18)$$

Avec :

$$f_{LS} = ||Y - AF||_F^2 \quad (2.19)$$

$$f_{neg} = \sum_{i=1}^N \sum_{k=1}^K H(A_{i,k}) + \sum_{t=1}^P \sum_{k=1}^K H(F_{k,t}) \quad (2.20)$$

où

$$H(x) = \begin{cases} x^2, & \text{si } x < 0 \\ 0, & \text{sinon} \end{cases} \quad (2.21)$$

et enfin,

$$f_{uni}(A) = \sum_{k=1}^{K-1} \sum_{l=k+1}^K \sum_{i=1}^N \frac{A_{i,k}}{\sqrt{\sum_{j=1}^N A_{j,k}^2}} \times \frac{A_{i,l}}{\sqrt{\sum_{j=1}^N A_{j,l}^2}} \quad (2.22)$$

Le terme f_{uni} peut également se voir comme une mesure de la corrélation entre les images factorielles normalisées, ou encore une mesure du mélange des images factorielles. La normalisation est présente afin que les images factorielles ne tendent pas vers 0 tout en étant compensées par les facteurs. Les coefficients a et b sont des constantes positives qui gèrent l'importance de la non-négativité et de la non-corrélation entre les images factorielles. Aucune initialisation spéciale des matrices A et F n'est recommandée par les auteurs, excepté qu'elle doit assurer que Y et $A \times F$ doivent avoir une énergie semblable. La fonctionnelle est minimisée grâce à une méthode d'optimisation alternée entre A et F , à la manière des algorithmes de factorisation en matrices non-négatives. Une méthode de gradient conjugué est utilisée afin de trouver la direction de descente, et le pas est calculé avec la méthode de Brent [Bre13]. Selon les auteurs, la méthode n'est que peu sensible au choix de a , qui est fixé à une valeur grande devant 1. La constante b elle n'est en fait pas une constante, puisqu'elle est recalculée après chaque itération afin que $\frac{bf_{uni}}{f_{LS} + af_{neg}} = 0.1$.

On retrouve dans cette fonctionnelle l'idée du simplexe de volume minimal. Dans ces deux méthodes, l'idée est de trouver les facteurs les moins éloignés possibles des pixels, afin dans un cas de minimiser le volume du simplexe engendré, et dans l'autre de minimiser les mélanges. Plus les facteurs sont éloignés des pixels, plus les images factorielles sont mélangées. Aucune

réduction de dimension n'est faite ici. L'avantage majeur de cette méthode est qu'elle n'a besoin d'aucune connaissance a priori. Elle peut donc être appliquée dans beaucoup de situations et est donc très générale. Cependant, il est naturel que les compartiments physiologiques soient mélangés et aient des supports communs. C'est d'ailleurs la raison pour laquelle la séparation de sources est nécessaire afin d'extraire les signatures de ces compartiments. Pénaliser la corrélation entre les images factorielles pénalise donc la *vraie* solution. Si la pondération b est trop importante, les images factorielles trouvées, et donc les facteurs aussi, auront des supports presque disjoints et seront donc mal estimées. De plus cette pondération b doit être assez grande afin de contrer les effets de non-unicité. Il y a donc un compromis à trouver dans la valeur de b . Un autre problème de cette méthode est que le terme de pénalisation f_{uni} est non convexe en A , ce qui rend la méthode très sensible à l'initialisation de A et F .

2.5.3.3 FAROI

L'acronyme anglais FAROI signifie analyse factorielle avec régions d'intérêt probabilistes, cette méthode est détaillée dans [ŠT12]. Comme souligné par les auteurs, les méthodes de séparation procèdent en pratique souvent en deux phases. La première est la sélection de la zone à étudier, et la seconde est l'application de la méthode. La phase de sélection est importante car le modèle de mélange linéaire des méthodes de séparation n'est en général une bonne approximation que dans une zone restreinte de l'image. Dans cette phase de sélection, on définit donc une région d'intérêt (ROI) manuellement, afin de restreindre l'étude à une zone précise. Cette phase a souvent besoin d'être répétée lorsque la méthode de séparation ne donne pas de résultats visuellement convaincants. L'idée de cette méthode est donc d'intégrer cette phase directement à la méthode de séparation, afin d'estimer conjointement la ROI d'étude ainsi que les facteurs et images factorielles.

Pour ce faire, les auteurs proposent une modélisation probabiliste du modèle de mélange et des données. Le bruit est supposé gaussien, et les facteurs et images factorielles à coefficients positifs. Chaque source possède également une région d'intérêt probabiliste. Dans cette dernière chaque pixel contient une valeur dans $[0,1]$. Cette valeur influence le coefficient de l'image factorielle associée. Si cette valeur est proche de 0, alors la densité de probabilité du coefficient de l'image factorielle associée est concentrée vers des valeurs positives proches de 0. Lorsque la valeur est à 1, aucun a priori n'est fait sur le coefficient de l'image factorielle associée, la densité de probabilité de ce coefficient est donc une loi uniforme dans $[0,1]$. Les auteurs souhaitent par cette modélisation encourager la parcimonie dans les images factorielles.

Le calcul des densités a posteriori des matrices A et F étant complexe, les auteurs proposent d'utiliser la méthode de Bayes variationnelle [Qui+06], qui permet de faire de l'inférence en approchant des intégrales en pratique non calculables. Le nombre de sources est estimé grâce au principe de la détermination automatique de pertinence (ARD en anglais), principe utilisé dans de précédents travaux en factorisation de matrices non-négatives [TF09].

Le problème du choix de la région d'étude de départ soulevé par les auteurs est réel et doit en effet être considéré. L'approche des auteurs présente l'avantage d'être automatique.

Cependant comme il est souligné dans [Ond15], cette méthode souffre d'instabilité numérique et le résultat n'est pas fiable. De plus, comme très peu de connaissances a priori sont injectées (juste un encouragement de la parcimonie des images factorielles), le problème de non-unicité de la séparation persiste.

2.5.3.4 S-BSS-vecDC

Cette méthode décrite dans [TŠ15] a été développée par le même groupe de travail que FAROI. S-BSS-vecDC signifie séparation de sources aveugle, parcimonieuse avec déconvolution vectorisée. Elle reprend en partie le modèle bayésien de FAROI, mais sans utiliser les régions d'intérêt probabilistes. L'ajout majeur est un modèle physiologique d'échange du traceur. Les facteurs y sont modélisés comme tous issus de la même fonction d'entrée, mais avec des réponses impulsionnelles différentes. Ceci revient à intégrer à la séparation un modèle compartimental très général qui autorise les rétentions de traceur. Les réponses impulsionnelles prennent la forme de noyaux de convolution. Les facteurs, les images factorielles, la fonction d'entrée, ainsi que les noyaux de convolution et le nombre de sources sont les inconnues du problème. Comme autres connaissances a priori, les images factorielles et les noyaux de convolution sont supposés parcimonieux. La densité a priori du vecteur représentant la fonction d'entrée suit une loi normale tronquée au domaine positif.

Le modèle est résolu en utilisant la méthode de Bayes variationnelle, comme pour FAROI. Le nombre de sources est lui aussi calculé automatiquement conjointement avec les autres paramètres. L'algorithme commence avec un nombre élevé de sources, puis supprime celles dont l'énergie estimée est faible.

L'introduction d'un modèle physiologique au sein de la séparation contraint grandement la séparation. C'est donc un excellent moyen de faire face aux problèmes de non-unicité. La méthode estime également les noyaux de convolution des différents compartiments physiologiques, ce qui peut être utile dans certains diagnostics.

Cependant, comme souligné par les auteurs, cette méthode est très sensible à l'initialisation de la fonction d'entrée. En effet, une modification sur la fonction d'entrée (autre qu'un retard) a des conséquences très fortes sur le choix des noyaux de convolution et donc des facteurs. De plus, le calcul automatique du nombre de sources ne fonctionne pas toujours correctement. Il est donc préférable de fixer le nombre de sources lorsque celui-ci est connu.

2.6 Approche Tensorielle

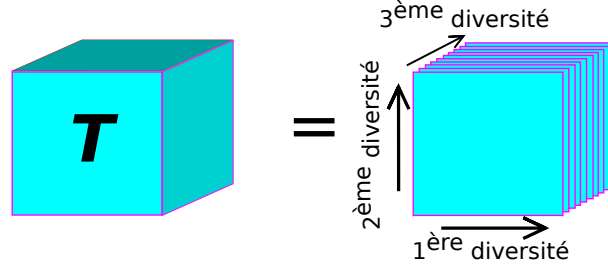
Jusqu'ici, les données considérées étaient des séquences d'images issues d'une caméra quelconque d'imagerie nucléaire. Les données présentaient deux diversités : le temps et l'espace, justifiant ainsi leur représentation sous forme matricielle. D'autres diversités peuvent cependant être considérées en imagerie nucléaire, ce qui permet d'une certaine manière d'ajouter des informations supplémentaires afin de faire face à la non-unicité du problème de sépara-

tion. Dans l'exemple du 6-DIG, il est possible de séparer la séquence d'images en deux : la phase basale (première injection de traceur) et la phase insuline (seconde injection de traceur, précédée de l'injection de l'insuline, cf figure 1.9). Il est aussi possible d'utiliser deux séquences d'images pour chaque patient sur certaines gamma-caméras : celle acquise sur la face antérieure, et celle acquise sur la face postérieure du patient. Dans ces deux exemples, les nouvelles diversités considérées sont de dimension 2. On pourrait aussi considérer la diversité "patient", les données seraient alors un ensemble de matrices, chaque matrice étant la séquence d'images associée à un certain patient. La dimension de cette diversité serait le nombre de patients dans l'étude. Ou encore, en faisant varier la quantité de traceur injecté dans le patient, on obtiendrait également une troisième diversité.

Une représentation de ces données à 3 diversités peut se faire sous la forme d'une seule et unique matrice. Par exemple en concaténant les séquences d'images obtenues dans la même matrice. Cependant, de l'information est perdue lorsque cette concaténation est faite, de la même manière que la cohérence spatiale d'une image est perdue lorsque celle-ci est vectorisée. Il est donc préférable de conserver une représentation faisant bien apparaître les 3 diversités. Dans ce cas, les données ne sont plus représentées par un tableau à double entrée, mais par un tableau tri-dimensionnel. Les tableaux tri-dimensionnels sortent du cadre matriciel et il convient donc d'utiliser des outils liés aux tenseurs pour les analyser. Ces derniers sont des objets mathématiques qui généralisent les notions de scalaires, de vecteurs, et de matrices aux ordres supérieurs.

Comme vu précédemment, dans le cas de deux diversités la séparation de sources en imagerie nucléaire s'apparente à la décomposition de la matrice des données en une somme d'un petit nombre de matrices de rang 1 à coefficients positifs. Chacune de ces matrices de rang 1 représentant une source. La notion de rang matriciel peut être généralisée aux tenseurs [Hit27]. Il existe aussi des méthodes permettant de décomposer ou d'approcher un tenseur en une somme de tenseurs de rang 1 [CLDA09] ; [SVBDL13] (éventuellement à coefficients positifs [CFC15]). Sous certaines conditions les tenseurs de rang 1 obtenus peuvent être interprétés comme des sources. Ces techniques sont donc utilisables pour faire de la séparation de sources et ont trouvé des applications dans un ensemble de domaines variés [Bro97] ; [SH05] ; [Zha+08], dont ne fait pour le moment pas partie l'imagerie nucléaire. Le grand avantage des tenseurs en séparation de sources est que les cas d'unicité de la séparation sont très fréquents contrairement aux matrices. Cette piste mérite donc d'être considérée pour la séparation de sources en imagerie nucléaire, d'autant plus que des résultats d'unicité ont été prouvés très récemment dans le cas où 2 ou 3 sources sont recherchées avec contraintes de positivité [QCL16b] ; [QCL16a]. Comme détaillé plus loin, la troisième diversité ajoutée doit cependant respecter des contraintes afin que les tenseurs de rang 1 obtenus puissent représenter les sources.

Cette section est donc dédiée à la décomposition tensorielle. La dernière partie de cette section est consacrée aux possibles applications de la décomposition tensorielle pour la séparation de sources en imagerie nucléaire et aux différentes tentatives réalisées lors de cette thèse.

FIGURE 2.9 – Un tenseur T d'ordre 3 peut être interprété comme un ensemble de matrices.

2.6.1 Introduction aux tenseurs

Dans cette section, les tenseurs seront traités comme des hypermatrices à valeurs réelles, comme dans [LC09]. C'est à dire qu'un tenseur T d'ordre m (m est le nombre de diversités) peut s'écrire comme un tableau multi-dimensionnel :

$$T = \llbracket t_{i_1, \dots, i_m} \rrbracket_{i_1, \dots, i_m=1}^{d_1, \dots, d_m} \in \mathbb{R}^{d_1 \times \dots \times d_m} \quad (2.23)$$

où d_i est la dimension de la $i^{\text{ème}}$ diversité. À noter que les tenseurs ne sont pas rigoureusement des hypermatrices [Com14]. Par souci de simplicité ils seront considérés comme tels dans la suite. À l'ordre 0, un tenseur sera assimilé à un scalaire, à l'ordre 1 à un vecteur, et à l'ordre 2 à une matrice. Nous nous intéresserons principalement aux tenseurs d'ordre 3 dans la suite. Les tenseurs d'ordre 3 peuvent être vus comme un ensemble de d_3 matrices, où d_3 est la dimension de la $3^{\text{ème}}$ diversité, comme illustré dans la figure 2.9.

Il est possible de définir un produit tensoriel $\otimes$ qui, à deux tenseurs T et T' , de nature potentiellement différente, associe un autre tenseur dont l'ordre est la somme des ordres des deux tenseurs d'entrée :

$$\begin{aligned} \otimes : \mathbb{R}^{d_1 \times \dots \times d_m} \times \mathbb{R}^{d'_1 \times \dots \times d'_{m'}} &\longrightarrow \mathbb{R}^{d_1 \times \dots \times d_m \times d'_1 \times \dots \times d'_{m'}} \\ (T, T') &\longmapsto T \otimes T' = \llbracket t_{i_1, \dots, i_m} \times t'_{j_1, \dots, j_{m'}} \rrbracket_{i_1, \dots, i_m, j_1, \dots, j_{m'}=1}^{d_1, \dots, d_m, d'_1, \dots, d'_{m'}} \end{aligned}$$

Par exemple, dans le cas où $m = m' = 1$, on travaille avec deux vecteurs u et v , et leur produit tensoriel forme la matrice $A = u \otimes v = \llbracket a_{i,j} \rrbracket_{i,j=1}^{d_1, d_2} = \llbracket u_i \times v_j \rrbracket_{i,j=1}^{d_1, d_2}$, de rang 1. Ceci correspond au produit matriciel de u et de la transposée de v .

2.6.2 Décomposition tensorielle et séparation de sources

Avant d'évoquer la notion de décomposition et d'approximation tensorielle, il est nécessaire de définir la notion de rang pour les tenseurs, notion qui s'appuie sur le produit tensoriel précédemment défini.

2.6.2.1 Rang et rang non-négatif d'un tenseur

La notion de rang pour les tenseurs se base sur les tenseurs décomposables. Un tenseur est dit *décomposable* s'il peut être écrit sous la forme d'un produit tensoriel de tenseurs d'ordre 1 (c'est à dire de vecteurs). Plus précisément, $T \in \mathbb{R}^{d_1 \times \dots \times d_m}$ est *décomposable* si et seulement si il existe $(u_1, \dots, u_m) \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m}$ tel que $T = u_1 \otimes \dots \otimes u_m$.

Il est possible de montrer que tout tenseur T peut être décomposé en une somme de tenseurs décomposables :

$$T = \sum_{r=1}^R D_r = \sum_{r=1}^R u_{r,1} \otimes \dots \otimes u_{r,m} \quad (2.24)$$

avec $(D_r)_{r=1..R}$ des tenseurs décomposables, et $(u_{r,j})_{r=1..R, j=1..m}$ des vecteurs. Le *rang* d'un tenseur est défini comme la plus petite valeur de R telle que l'égalité (2.24) est vraie. En particulier, les tenseurs de rang 1 sont les tenseurs décomposables. Cette définition est cohérente avec celle du rang matriciel, et généralise donc cette notion aux ordres supérieurs.

On peut également définir un *rang non-négatif* d'une matrice, ce qui est souhaitable lorsque l'on travaille avec des données positives comme en imagerie nucléaire. Le *rang non-négatif* d'un tenseur est la plus petite valeur de R telle que l'égalité (2.24) est vraie et que les vecteurs $(u_{r,j})_{r=1..R, j=1..m}$ sont à coefficients positifs. Cette définition étant plus contraignante que la définition de rang matriciel, le rang non-négatif d'un tenseur sera toujours supérieur ou égal à son rang.

2.6.2.2 Décomposition et approximation de rang-r d'un tenseur

La définition du rang d'un tenseur n'est pas sans rappeler le modèle de mélange linéaire. Ce modèle décrit dans la section 2.1.2 suppose qu'une matrice de données peut être approchée par la somme d'un petit nombre de matrices de rang 1, où chacune de ces matrices représente une source. De même, le rang non-négatif d'un tenseur est lié au modèle de mélange linéaire avec contrainte de positivité.

Si un tenseur est de rang R , on peut par définition le décomposer en une somme de tenseurs de rang 1. Si certaines conditions sont satisfaites, chaque tenseur de rang 1 peut aussi être interprété comme une source. Si c'est le cas, les vecteurs qui sont à l'origine de ce tenseur de rang 1 contiennent chacun la signature de la source considérée sur une des diversités. La décomposition d'un tenseur en somme de tenseurs de rang 1 peut donc être utilisée pour la séparation de sources. Cette décomposition porte le nom de PARAFAC, CANDECOMP ou bien de décomposition CP [Kie00].

En pratique, les données étant bruitées, une décomposition exacte du tenseur en somme d'un faible nombre de tenseurs de rang 1 n'est pas envisageable. On cherche alors une *meilleure approximation de rang K* du tenseur. Pour un tenseur T d'ordre m , de rang $R > K$, on cherche donc un tenseur X qui minimise la quantité $\|T-X\|$, sous la contrainte que le rang de X est égal à K et où $\|\cdot\|$ est une norme tensorielle. Ce tenseur X a ensuite besoin d'être lui même

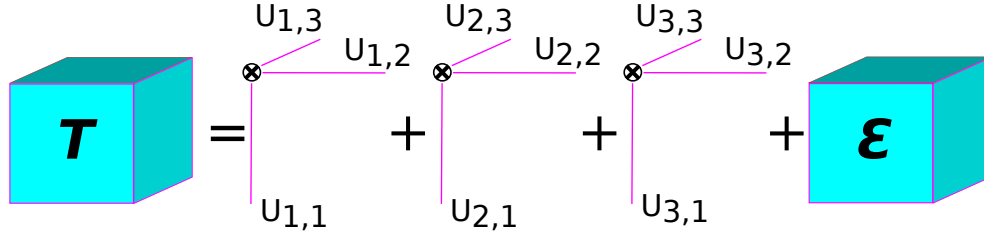


FIGURE 2.10 – Représentation graphique de la décomposition approchée de rang 3 d'un tenseur T d'ordre 3. Ce tenseur est approché comme la somme de 3 tenseurs de rang 1. L'erreur ϵ est ici minimale selon une certaine norme. Ce modèle est une généralisation aux ordres supérieurs du modèle de mélange linéaire vue en début de chapitre (Figure 2.1). Il est également possible d'intégrer des contraintes de positivité aux vecteurs $u_{k,i}$.

décomposé en somme de K tenseurs de rang 1 afin d'être interprétable en séparation de sources. Autrement dit, il faut trouver des vecteurs $(u_{i,k})_{i=1..m, k=1..K}$ tel que :

$$X = \sum_{k=1}^K u_{k,1} \otimes \dots \otimes u_{k,m} \quad (2.25)$$

La recherche de ces vecteurs s'appelle dans la suite la *décomposition approchée de rang K* de T , à bien différencier de X qui est la *meilleure approximation de rang K* de T . Dans le cas où l'on a des données non-négatives, une *meilleure approximation de rang K non négative* du tenseur, avec K strictement inférieur au rang non-négatif du tenseur, suit la même définition, mais avec comme contrainte que les coefficients de X sont positifs. La décomposition approchée associée s'appelle dans la suite *décomposition non-négative approchée de rang K* et généralise aux ordres supérieurs le concept de factorisation en matrices non-négatives [LS01]. On peut aussi parler de factorisation en tenseurs non-négatifs [SH05]. Les vecteurs $(u_{k,i})_{k=1..K, i=1..m}$ Y sont bien évidemment recherchés avec comme contraintes d'avoir tous leurs coefficients positifs.

Différents algorithmes ont été développés afin de trouver une décomposition approchée de rang K d'un tenseur, avec contraintes de positivité ou non [CLDA09]; [SVBDL13]; [CFC15]. Ces algorithmes sont utilisés comme méthode de séparation de sources dans différents domaines, comme en chimiométrie [Bro97] ou en imagerie hyperspectrale [Zha+08].

2.6.2.3 Existence et unicité

Deux questions naturelles se posent sur ces décompositions et approximations : existent-elles ? sont-elles uniques (aux permutations et changements d'échelle près) ? Ces questions ont d'importantes conséquences en pratique, notamment pour la séparation de sources. Les décompositions exactes existent toujours [Com14], mais ce n'est pas le cas pour les meilleures approximations. En effet, dans le cas où le nombre de diversité m est supérieur ou égal à 3, certains tenseurs n'ont pas de meilleure approximation de rang K (où K est inférieur

strictement au rang du tenseur), ce qui peut rendre le problème de séparation mal posé. Il a été cependant montré que dans le cas non-négatif, l'existence d'une meilleure approximation de rang K est garantie [LC09].

Contrairement aux matrices, où la décomposition approchée de rang K n'est généralement pas unique, l'unicité est très courante pour les tenseurs d'ordre supérieur ou égal à 3. Différentes conditions suffisantes d'unicité de la décomposition exacte ont été trouvées [Kru77] ; [DDL13a] ; [DDL13b]. Dans le cas de la non-négativité, deux résultats très encourageants ont été prouvés récemment. Le premier montre que la meilleure approximation de rang K non-négative d'un tenseur à coefficients lui aussi non-négatifs et de rang strictement supérieur à K , est presque toujours unique [QCL16b]. Autrement dit, il existe presque toujours un unique tenseur X à coefficients non-négatifs qui minimise la quantité $\|T - X\|$, et ce quelle que soit la norme utilisée. Cependant ce résultat n'est pas suffisant pour assurer l'unicité de la séparation. En effet, ce tenseur X est peut-être unique, mais qu'en est-il de sa décomposition exacte en tenseurs de rang 1 ? Cette question a trouvé une réponse partielle dans [QCL16a]. Il y est montré (théorème 7.8, [QCL16a]) que presque tous les tenseurs d'ordre 3 à coefficients positifs, dont la dimension de chacune des diversités est au moins de 3, la meilleure approximation non-négative de rang 2 ou 3 est presque tout le temps unique (cf [QCL16b]), et surtout que cette meilleure approximation non-négative possède généralement une unique décomposition non-négative. Dans le cas où l'on travaille avec des données à 3 diversités, et que le nombre de sources est de 2 ou 3, ce théorème assure donc généralement l'unicité de la séparation. Ce théorème montre l'utilité potentielle de la décomposition tensorielle pour la séparation de sources en imagerie médicale, d'autant plus qu'en pratique le nombre de sources présentes est généralement supposé égal à 2 ou 3.

2.6.3 Application à l'imagerie nucléaire

L'objectif de cette partie est d'explorer les possibilités d'utilisation des outils tensoriels pour la séparation de sources en imagerie nucléaire. Ceci est motivé par le théorème d'unicité énoncé dans le paragraphe précédent. Le problème d'unicité est le principal obstacle qui limite les performances et la fiabilité des algorithmes de séparation de sources en imagerie nucléaire. La possibilité de le voir disparaître est donc une opportunité qu'il faut saisir. À notre connaissance, les outils tensoriels n'ont pas encore été utilisés pour la séparation de sources en imagerie nucléaire.

2.6.3.1 Modèle de mélange en imagerie nucléaire

Afin d'utiliser la décomposition tensorielle pour séparer des sources en imagerie nucléaire, il est nécessaire que les tenseurs de rang 1 issus de la décomposition puissent être interprétés comme des sources. En imagerie nucléaire, les sources sont des compartiments physiologiques. Les deux diversités précédemment utilisées étaient le temps et la position. Le modèle de mélange linéaire à 3 sources peut s'illustrer par la figure 2.11 et est compatible avec la décomposition matricielle. Le but est donc de séparer les sources en approchant la matrice des

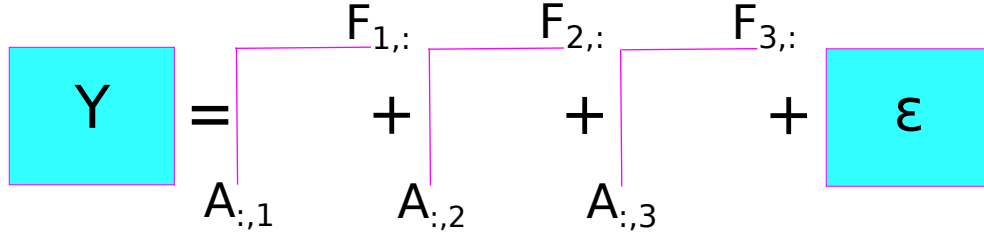


FIGURE 2.11 – Représentation graphique du modèle de mélange linéaire avec $K = 3$ sources. C’est aussi la représentation graphique d’une décomposition matricielle en matrices de rang 1. La matrice des données Y est expliquée, au bruit près, par 3 matrices de rang 1. Chacune des matrices de rang 1 représente une source. Les coefficients des matrices sont positifs. La matrice colonne $A_{:,k}$ ($k^{\text{ème}}$ colonne de A) contient la signature spatiale de la $k^{\text{ème}}$ source. De même, la matrice ligne $F_{k,:}$ ($k^{\text{ème}}$ ligne de F) contient la signature temporelle de la $k^{\text{ème}}$ source.

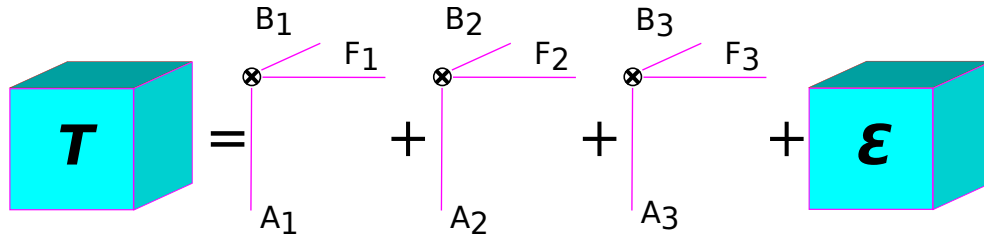


FIGURE 2.12 – Décomposition tensorielle approchée d’un tenseur d’ordre 3. Si le modèle de mélange est compatible, chaque tenseur de rang 1 contient les contributions d’une source. Les signatures spatiales sont contenues dans les vecteurs A_k , les signatures temporelles dans les vecteurs F_k et enfin les signatures sur la troisième diversité dans les vecteurs B_k . Les tenseurs de rang 1 représentant les sources sont construits en faisant le produit tensoriel des signatures de la source associée.

données par une somme de matrices de rang 1. Cette décomposition n’étant pas unique, des connaissances a priori doivent être ajoutées (notamment des contraintes de positivité) afin d’obtenir une décomposition où chaque matrice de rang 1 représentait une et une seule source.

On s’intéresse maintenant au cas où il y a trois diversités. Les données sont alors représentées par un tenseur T d’ordre 3. Afin que le modèle de mélange et la décomposition tensorielle soient compatibles, il faut que chaque tenseur de rang 1 issu de la décomposition représente une source. Cela nécessite par conséquent que chaque source puisse être représentée par un tenseur de rang 1. Les trois vecteurs, dont le produit tensoriel forme ce tenseur de rang 1, doivent contenir les signatures spatiales, temporelles et de la troisième diversité de cette source. Si tel est le cas, les algorithmes de décomposition tensorielle peuvent être utilisés afin de séparer les sources. On peut représenter ceci par la figure 2.12.

Les vecteurs B_k peuvent donc être vus comme des vecteurs qui contiennent les poids des sources ($A_k F_k$) dans chaque sous tenseur $T_{:,j,:}$. Chaque coefficient de T peut être calculé comme une somme de K contributions, chaque contribution étant celle d’une des K sources :

$$T_{i,t,j} = \sum_{k=1}^K A_{i,k} F_{t,k} B_{j,k} + \epsilon_{i,t,j} \quad (2.26)$$

Le tenseur des données est multilinéaire en A , F et B . Le rôle des trois diversités étant de nouveau symétrique, il est possible de les permuter. Les invariances suivant les autres diversités sont également toujours présentes. Dans le cas matriciel, ceci s'exprimait par le fait que les facteurs ne dépendaient pas de la position et que les images factorielles ne dépendaient pas du temps. En généralisant avec les tenseurs, les signatures sur chacune des diversités ne doivent pas dépendre des autres diversités (au facteur d'échelle près). Ces invariances sont à la fois un atout et un inconvénient. En effet, dans le modèle décrit dans l'équation 2.26, les matrices A , F et B sont très contraintes par la multilinéarité et les invariances. Ceci fait que d'un côté, cette multilinéarité associée à des contraintes de positivité permet d'avoir des résultats intéressants d'unicité de la décomposition. Mais d'un autre côté, la troisième diversité ajoutée doit respecter ces contraintes de multilinéarité et d'invariance afin que la décomposition puisse être utilisée pour séparer les sources. La troisième diversité doit donc être choisie méticuleusement afin de vérifier le modèle multi-linéaire de l'équation 2.26. De plus, sur chaque diversité, les signatures des différentes sources doivent être différentes les unes des autres. Dans le cas contraire des problèmes de non-unicité apparaissent comme dans le cas matriciel.

Une autre manière plus simple de considérer ces contraintes est de les voir comme des contraintes de proportionnalité. Si l'on ne prend en compte que les diversités spatiales et temporelles, chaque source produit une séquence d'images bien définie. Cette séquence peut être représentée par une matrice de rang 1 (à l'aide du produit matriciel $A_{:,k} F_{k,:}$). Lorsque l'on considère en plus la troisième diversité, la source est cette fois représentée comme plusieurs séquences d'images, et doit être modélisable par un tenseur de rang 1. Comme ce tenseur est de rang 1, toutes ces séquences d'images doivent être proportionnelles entre elles. C'est à dire que les images factorielles et les courbes d'activité d'une source sont identiques, à un facteur d'échelle près, dans toutes les séquences d'images des données d'entrée. Ces constantes de proportionnalité se retrouvent dans le vecteur B_k . Afin que la troisième diversité apporte des informations qui discriminent les sources, il est nécessaire que les vecteurs B_k soient différents pour chaque source. Ceci est résumé dans la figure 2.13.

2.6.3.2 Ajout d'une troisième diversité

Plusieurs diversités sont proposées ici comme candidates pour la troisième diversité en imagerie nucléaire. Elles sont détaillées une à une et leur intérêt pour la séparation de sources est analysé.

- **Dé-vectorisation des images :** En reprenant les données matricielles de départ, il est possible d'ajouter artificiellement une diversité. En effet, si l'on ne vectorise pas les images, l'information spatiale peut être vue comme deux diversités différentes : lignes et colonnes. On obtient un tenseur d'ordre 3, dont les diversités sont le temps, les lignes

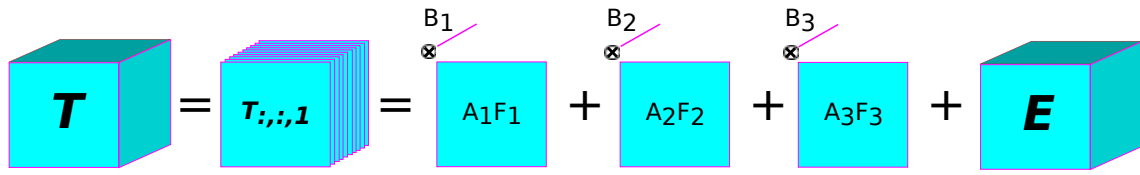


FIGURE 2.13 – Une autre manière de voir la décomposition tensorielle en imagerie nucléaire. Les données sont composées de plusieurs séquences d’images, et chacune de ces séquences est représentée par une matrice $T_{:, :, j}$ dans le tenseur de données. Chaque source ne varie que d’un facteur de proportionnalité d’une séquence d’image à une autre. Ces facteurs de proportionnalité sont contenus dans les vecteurs B_k qui doivent être différents pour chacune des sources afin que la troisième diversité apporte des informations non redondantes par rapport aux deux premières.

et colonnes. C’est la manière la plus naturelle d’introduire une troisième diversité. Cependant, les résultats d’une décomposition tensorielle sur ce tenseur sont inexploitable. En effet, ceci imposerait les images factorielles, vues comme des matrices images 2D, à être de rang 1 ce qui contraint grandement la forme des compartiments physiologiques et n’est pas réaliste. Par exemple, les lignes d’une image factorielle ne sont généralement pas proportionnelles entre elles, alors que cette représentation l’impose. Il est donc nécessaire de conserver l’information spatiale en une seule et unique diversité afin d’obtenir des images factorielles interprétables. La non-vectorisation des images n’est pas une bonne solution pour obtenir une troisième diversité.

- **Quantité de traceur injecté :** Afin d’obtenir plusieurs séquences d’images différentes, il est possible de répéter les acquisitions, en faisant varier la quantité de traceur injecté dans le patient. Ceci introduit bien une troisième diversité, et les sources sont bien proportionnelles d’une séquence à une autre. Cependant, les vecteurs B_k qui contiennent ces constantes de proportionnalité seront identiques entre eux. Ceci est dû au fait que cette troisième diversité n’introduit aucune information permettant de discriminer les sources. De plus, il est difficile d’obtenir cette troisième diversité en pratique car elle nécessite de multiplier les acquisitions sur le même patient.
- **Diversité patient :** Les séquences d’images de différents patients peuvent être agglomérées dans un même tenseur. Ici encore, ceci ne permet pas de faire de séparation de sources. Les formes des compartiments et la physiologie diffèrent généralement d’un patient à l’autre. Ainsi, les patients n’ont aucune raison d’avoir des images factorielles et des courbes d’activités proportionnelles.
- **Diversité basal/insuline :** Dans les études utilisant le 6-DIG, deux séquences d’images sont produites. La première est obtenue en injectant uniquement du 6-DIG. La seconde est obtenue en injectant de l’insuline 5 minutes avant le 6-DIG. L’insuline modifie le comportement du 6-DIG dans les compartiments physiologiques insulino-sensibles. Séparer la séquence basale (sans insuline) et la séquence avec insuline permet d’obtenir une troisième diversité qui apporte des informations non-redondantes grâce à l’effet de l’insuline sur le traceur. De plus, on peut raisonnablement faire l’hypothèse que les images factorielles sont invariantes (au facteur d’échelle près) par rapport à cette nouvelle diversité

en supposant que le patient ne bouge pas entre les deux séquences d'images. Cependant, l'hypothèse que les facteurs sont invariants (au facteur d'échelle près) par rapport à cette nouvelle diversité est contestable. Les facteurs vont effectivement avoir des allures différentes avant et après l'injection de l'insuline. L'insuline modifie la dynamique du traceur dans les muscles et en particulier dans le myocarde. Afin que l'hypothèse soit respectée, il faudrait que cette modification soit uniquement une modification d'échelle sur les facteurs, ce qui n'est pas le cas. L'effet attendu de l'insuline est l'accélération de l'équilibre entre les concentrations du traceur dans le sang et dans les muscles. La relation entre le facteur sans et avec insuline n'est donc pas une relation de proportionnalité mais une relation plus complexe. Autrement dit, avec cette nouvelle diversité, les sources ne peuvent pas être représentées par un tenseur de rang 1. Les méthodes de séparation donneront donc des résultats non interprétables.

- **Diversité antérieur/postérieur :** Les gamma-caméras sont généralement équipées de deux matrices de capteurs : une située au-dessus du patient, et une autre située en-dessous du patient. Les deux séquences d'images produites par ces capteurs peuvent être utilisées afin de construire une troisième diversité. Les deux séquences d'images obtenues apportent bien chacune des informations différentes car les compartiments physiologiques ne sont a priori pas aux mêmes distances des deux matrices de capteurs. Un compartiment physiologique plus proche d'une des matrices de capteurs aura une énergie plus grande sur la séquence d'images associée. Si les sources sont à des distances différentes des capteurs, l'information est bien non-redondante entre les deux séquences d'images. Cependant, lorsque l'on s'intéresse par exemple au myocarde et au sang, ou aux différents compartiments physiologiques du rein, les sources ont des distances aux capteurs très similaires. Dans les exemples cités, certaines sources sont même englobées par d'autres. Les deux séquences d'images auront dans ce cas des informations relativement similaires. De plus, le fait qu'une source puisse être à une distance variable d'une matrice de capteurs du fait de son volume fait que les images factorielles et les courbes d'activité ne respectent pas la contrainte de proportionnalité entre les deux matrices de capteurs.

2.6.4 Conclusion sur l'approche tensorielle

De par ses résultats sur l'unicité de la décomposition, l'approche tensorielle est très séduisante pour la séparation de sources. Néanmoins, cette approche est difficilement applicable à l'imagerie nucléaire. Il est nécessaire que la troisième diversité ajoutée satisfasse certaines contraintes. D'une part, cette troisième diversité doit être non-redondante par rapport aux deux premières : les signatures sur cette diversité doivent être différentes pour chacune des sources. D'autre part, elle doit assurer que les données suivent un modèle de mélange multilinéaire afin que la décomposition tensorielle puisse être utilisée pour séparer les sources. Ces deux contraintes sont difficiles à satisfaire conjointement. Les différentes diversités que nous avons envisagées se sont toutes avérées incompatibles avec la décomposition tensorielle pour faire de la séparation de sources. Deux diversités pourraient cependant être ajoutées dans certaines applications bien spécifiques. La première est la diversité antérieur/postérieur. Celle-ci

est utilisable lorsque les sources sont à des distances bien différentes des deux matrices de capteurs. La seconde est la diversité basal/insuline, qui peut être généralisée à toutes les molécules ayant une action sur le comportement du traceur utilisé. Sous l'action de cette molécule, les courbes d'activité des compartiments physiologiques étudiés ne doivent changer que d'un facteur d'échelle. L'inconvénient de ces deux diversités est qu'elles sont de dimension 2, alors que les hypothèses nécessaires dans le théorème d'unicité de la décomposition imposent une dimension supérieure ou égale à 3. Une des manières d'y remédier est d'acquérir des séquences d'images où la dose de la molécule injectée varie. Cette méthode n'est toutefois pas applicable du fait de la répétition des injections et de la pénibilité engendrée pour le patient.

2.7 Conclusion

La séparation de sources possède des applications dans de nombreux domaines. Cependant, malgré l'apparente analogie des problèmes de séparation, l'imagerie nucléaire a suivi un chemin relativement indépendant et différent des autres domaines. Ceci vient probablement du fait que dans sa forme la plus classique, le problème de séparation de sources est sous-déterminé. Pour résoudre ce problème, il est donc nécessaire d'injecter des informations a priori sur les sources à séparer. Ces informations a priori sont spécifiques à chaque domaine. Il n'existe donc pas de méthodes universelles de séparation de sources. Par exemple, les méthodes se basant sur l'analyse en composantes indépendantes très utilisées pour séparer les signaux électrophysiologiques ne sont pas applicables en imagerie nucléaire.

Le domaine le plus étroitement lié à l'imagerie nucléaire est probablement l'imagerie hyperspectrale. Ces deux domaines partagent les mêmes contraintes de positivité, et des méthodes de séparation similaires notamment grâce à l'interprétation géométrique. La particularité de l'imagerie nucléaire vis à vis de l'imagerie hyperspectrale est le bruit considérable présent dans les séquences d'images. De plus, chaque source en imagerie nucléaire est généralement localisée dans une zone spatiale restreinte et connexe, ce qui permet un usage simple des régions d'intérêt comme détaillé dans le chapitre suivant.

Différentes connaissances a priori ont été utilisées en imagerie nucléaire. Parmi elles on peut noter les a priori sur la forme des facteurs. Plusieurs méthodes cherchent à minimiser un critère supposé défavorable à la solution recherchée. Par exemple la minimisation du volume formé par le simplexe dont les facteurs sont les sommets. La solution peut aussi être recherchée en minimisant le recouvrement entre les images factorielles. Il a également été proposé d'introduire un modèle compartimental conjointement au modèle de mélange, afin de lier les différents facteurs.

Les méthodes actuelles ne sont toutefois que peu utilisées en routine clinique car les courbes d'activité résultantes manquent de fiabilité [Šám+98] ; [ŠT12]. Il nous semble que les résultats sont limités par plusieurs problèmes :

- La non-unicité de la solution est le problème majeur en séparation de sources. Différentes connaissances a priori ont été proposées pour l'imagerie nucléaire, mais cela ne semble pas être suffisant pour résoudre ce problème. Ces a priori proposés sont souvent trop

généraux et donc pas assez contraignants pour assurer l'unicité. Ou alors ils sont ad hoc et leur champ d'application est très limité.

- Les méthodes sont généralement très sensibles à l'initialisation. Ceci vient du fait que le problème de séparation fait généralement intervenir un terme de fidélité aux données non convexe en (A, F) . Les algorithmes convergent ainsi vers un minimum local qui est déterminé par l'initialisation.
- Le modèle de mélange linéaire n'est une bonne hypothèse que dans une zone restreinte de l'image. Ce problème de modèle a été en partie traité en imagerie hyperspectrale où des modèles de mélanges non linéaire ont d'ailleurs été proposés [BD+12]. Cependant en imagerie nucléaire, ce problème est bien soulevé dans [ST12] mais il n'est généralement pas pris en compte.
- Des réductions de dimension sont réalisées dans la plupart des méthodes. Mais dans ce cas, la non-négativité n'est pas prise en compte, ces réductions peuvent donc détériorer la séparation. De plus, elles ne sont efficaces que si le modèle de mélange linéaire est valide.

Afin de remédier au problème d'unicité, nous avons proposé l'utilisation d'une approche tensorielle à la fin de ce chapitre. Dans cette approche, il est nécessaire d'ajouter une troisième diversité aux données et cette diversité doit vérifier des hypothèses assez contraignantes. Il est probable qu'une telle diversité existe dans certaines applications, mais aucune n'est assez pertinente pour les applications visées dans cette thèse.

Nous avons donc développé une méthode de séparation de sources avec contraintes de positivité, utilisant les régions d'intérêt comme connaissances a priori. Les régions d'intérêt des compartiments physiologiques sont accessibles dans la plupart des applications en imagerie nucléaire. Nous montrons dans le chapitre suivant comment ces régions répondent aux différentes limites évoquées précédemment. Elles sont intégrées dans la méthode à plusieurs fins. D'une part, étant liées aux positions des compartiments physiologiques elles permettent d'apporter des informations importantes sur les images factorielles et donc de limiter le problème de non-unicité. D'autre part, le modèle de mélange linéaire peut être relaxé en dehors des régions d'intérêt où sa validité est douteuse. Enfin ces régions d'intérêt peuvent être utilisées pour initialiser l'algorithme de séparation de façon pertinente et non arbitraire. La méthode qui regroupe ces idées est entièrement détaillée dans le chapitre suivant.

Utilisation des régions d'intérêt pour la séparation de sources en imagerie nucléaire

Sommaire

3.1	Les ROI : un a priori acceptable en imagerie nucléaire	70
3.2	Utilisation des ROI comme masques pour les images factorielles . . .	71
3.3	RUDUR : Un algorithme générique de séparation utilisant de façon robuste les ROI.	73
3.3.1	Fidélité aux données	74
3.3.2	Initialisation avec les ROI	75
3.3.3	Pénalisation robuste des images factorielles à l'aide des ROI	76
3.3.4	Relaxation du modèle en dehors des ROI	78
3.3.5	Pondération temporelle de la contrainte de fidélité avec les images d'intérêt	81
3.3.6	Fonction objectif finale	81
3.3.7	Optimisation de la fonctionnelle	84
3.3.8	Flexibilité de la méthode	86
3.4	Conclusion	88

Dans ce chapitre, nous proposons une méthode de séparation de sources générique utilisant les régions d'intérêt (ROI). Celles-ci sont généralement accessibles en imagerie nucléaire. Elles apportent des connaissances très fortes qui peuvent être intégrées de différentes manières dans une méthode de séparation. Les ROI permettent ainsi de répondre aux problèmes soulevés dans la conclusion du chapitre précédent. En effet, les ROI contiennent des informations sur la position des compartiments physiologiques. Elles peuvent donc être utilisées pour contraindre efficacement les coefficients des images factorielles et faire face à la non-unicité. Ces informations sont également utiles pour initialiser les coefficients des images factorielles et des facteurs. Enfin les ROI permettent de cibler les zones intéressantes de l'image. Le modèle de mélange linéaire peut ainsi être relaxé en dehors de ces zones, en diminuant l'impact des pixels qui n'appartiennent à aucune ROI.

Cependant, l'inconvénient des régions d'intérêt est que leur sélection est sujette à des variations intra et inter-observateur. De plus leur sélection est bien souvent difficilement auto-

matissable. Les régions d'intérêt doivent donc être intégrées de manière robuste à la méthode de séparation afin de prendre en compte les erreurs potentielles dans leur sélection.

L'usage des ROI nous a tout d'abord permis de développer une première méthode simple, qui utilise les régions d'intérêt comme masque pour les images factorielles et qui a été publiée dans [Fil+15]. Une méthode plus complète et plus robuste au choix des ROI, RUDUR, a ensuite été développée, en intégrant les différentes idées mentionnées précédemment. Une version partielle de la méthode est décrite dans [Fil+16], et la version complète dans [FDM17]. La méthode est applicable même lorsque des régions d'intérêt sont manquantes. L'analogue des régions d'intérêt dans la diversité temporelle, c'est à dire les images d'intérêt, peuvent être utiles dans le cas où une source de très faible intensité est présente. Les images d'intérêt peuvent être intégrées à la méthode générique détaillée dans [FDM17].

Ce chapitre présente ces méthodes. Celles-ci sont ensuite comparées aux méthodes de l'état de l'art dans le dernier chapitre.

3.1 Les ROI : un a priori acceptable en imagerie nucléaire

En imagerie nucléaire, les formes générales des compartiments physiologiques sont globalement connues, même si elles peuvent différer selon les patients. Les spécialistes sont capables de les tracer en parcourant la séquence d'images fournie par la caméra, que ce soit en 2D ou bien en 3D. Les constructeurs de caméra fournissent par ailleurs des logiciels facilitant le tracé de ces régions d'intérêt, notamment pour le cas plus compliqué de la 3D. Les méthodes d'obtention des ROI ont été précédemment détaillées au premier chapitre (section 1.4.1). Une autre manière d'obtenir les régions d'intérêt est d'appliquer une méthode de séparation de sources. Même sans a priori, les images factorielles font généralement apparaître des structures. Bien que ces structures soient très floues et encore mélangées, les spécialistes peuvent y reconnaître la forme des compartiments physiologiques cibles et ainsi dessiner les régions d'intérêt.

Les régions d'intérêt obtenues permettent d'avoir une approximation grossière de la position des compartiments physiologiques. Cette connaissance peut s'avérer très utile pour la séparation de sources mais n'a presque jamais été utilisée dans ce cadre. Cela est probablement dû au fait que les méthodes de séparation de sources en imagerie nucléaire ont précisément fait leur apparition afin de s'affranchir des régions d'intérêt. Les activités des compartiments physiologiques sont classiquement calculées en moyennant l'activité dans une région d'intérêt. Cette région est sélectionnée de manière manuelle sauf dans de rares cas. La sélection de cette région souffre donc de variabilité intra et inter-observateur. De plus, les régions ne contiennent pas uniquement de l'activité due au compartiment physiologique en question. Les méthodes de séparation de sources ont donc été introduites afin d'avoir une méthode automatique pour extraire les courbes d'activité des compartiments physiologiques qui prenne en compte le mélange. Cependant la pertinence des résultats obtenus avec les méthodes de séparation est toujours remise en question, et ces méthodes sont peu utilisées en pratique [Šám+98]; [ŠKZ91]; [Ben+93]. Nous proposons donc de réintroduire ces régions d'intérêt dans la séparation de

sources afin d'améliorer les résultats et la fiabilité des méthodes.

A notre connaissance, seules deux méthodes utilisent explicitement les régions d'intérêt comme connaissance a priori. La première méthode [NB88] utilise les régions d'intérêt pour contraindre les images factorielles. Les coefficients des images factorielles qui sont en dehors de leur ROI associée sont forcés à 0. Dans cette méthode les contraintes de positivité ne sont pas appliquées, et les résultats peuvent donc être physiologiquement ininterprétables. La deuxième méthode est FAMIS [Ben+94b], détaillée à la fin du chapitre précédent. L'utilisation des régions d'intérêt n'y est pas mentionnée explicitement mais cette méthode permet d'intégrer des connaissances brutes sur les images factorielles et les facteurs. Il est possible de fixer à l'avance des coefficients des matrices A et F . En particulier, si des régions d'intérêt sont connues, des coefficients d'images factorielles correspondant à des pixels qui sont en dehors de leur ROI associée peuvent être fixés à 0, à la manière de [NB88].

L'usage des ROI dans ces deux méthodes est très limité car elles utilisent uniquement l'information binaire de présence et d'absence d'un pixel dans les ROI. Comme évoqué dans l'introduction, les régions d'intérêt peuvent être utilisées de manière plus informative et plus robuste (relaxation du modèle de mélange linéaire et initialisation des matrices A et F). De plus, comme le choix des régions d'intérêt est sujet à de fortes variabilités, il est nécessaire que les a priori liés à ces ROI soient intégrés de manière robuste à la méthode de séparation, contrairement aux deux méthodes précédentes.

3.2 Utilisation des ROI comme masques pour les images factorielles

Les régions d'intérêt sont en lien direct avec les images factorielles. Si les régions d'intérêt sont bien tracées, il est possible de savoir en chaque pixel quels sont les compartiments physiologiques présents. Si un pixel fait partie d'une région d'intérêt, il est probable que le coefficient associé dans la matrice A des images factorielles ne soit pas nul, mais sa valeur n'est pas connue directement. En revanche, si un pixel ne fait pas partie d'une région d'intérêt, le coefficient associé dans la matrice A peut être supposé nul, ou bien de très faible valeur. Si on le considère comme nul, on a :

$$A_{i,k} = 0 \quad (\forall (i,k) \text{ tel que } M_{i,k} = 0) \quad (3.1)$$

où M est la matrice binaire contenant les régions d'intérêt :

$$M_{i,k} = \begin{cases} 1 & \text{si le } i^{\text{ème}} \text{ pixel est dans la } k^{\text{ème}} \text{ ROI} \\ 0 & \text{sinon} \end{cases} \quad (3.2)$$

Ces hypothèses peuvent s'avérer très utiles grâce aux contraintes qu'elles véhiculent. Elles introduisent une forme de parcimonie dans les images factorielles si les régions d'intérêt sont

relativement petites par rapport à l'image. Il a même été montré dans [Šám+89] que le problème de séparation peut admettre une solution unique lorsque des coefficients nuls de A sont connus.

La méthode détaillée dans [NB88] utilise les régions d'intérêt de cette manière. Les coefficients de A qui correspondent à des pixels en dehors de la région d'intérêt associée sont forcés à 0. Un défaut majeur de cette méthode est que les contraintes de positivité ne sont pas prises en compte. Les facteurs et les images factorielles produits peuvent avoir des coefficients négatifs, rendant ainsi les résultats inutilisables et incorrects.

Dans [Fil+15], nous avons repris l'idée d'utiliser les régions d'intérêt comme masques pour les images factorielles, mais en y ajoutant les contraintes de positivité. L'algorithme se base sur le schéma classique des mises à jour alternées entre les matrices A et F . Après chaque mise à jour de A , les coefficients correspondant aux pixels en dehors d'une région d'intérêt sont forcés à 0. Les coefficients de A et F négatifs sont également mis à 0. L'algorithme utilisé dans [Fil+15] ainsi que les notations sont détaillés ci-dessous.

On note Y les données d'origine. La matrice X contient les données normalisées :

$$X_{i,:} = \frac{Y_{i,:}}{\sum_{j=1}^P Y_{i,j}} \quad (3.3)$$

où $X_{i,:}$ est la $i^{\text{ème}}$ ligne de X . Notons $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_{i,:}$ le centre des données réduites et $Z = X - \bar{X}$ la matrice des données centrées réduites. Z est ensuite projetée dans le sous-espace affine $\mathcal{S}$ de centre $\bar{X}$ et parallèle à l'espace engendré par les $K-1$ vecteurs propres U_i associés aux $K-1$ plus grandes valeurs propres de la matrice d'autocorrélation $Z^T Z$. On note V_i la projection de $Z_{i,:}$ dans $\mathcal{S}$. Pour des raisons de simplicité calculatoire on note U et V deux matrices de tailles respectives K par P et N par K telles que :

$$U = \begin{pmatrix} \bar{X} \\ U_1 \\ \vdots \\ U_{K-1} \end{pmatrix} \in \mathcal{M}(K, P) \quad (3.4)$$

$$V = \begin{pmatrix} 1 & V_1 \\ \vdots & \vdots \\ 1 & V_N \end{pmatrix} \in \mathcal{M}(N, K) \quad (3.5)$$

où les U_i et V_i sont des vecteurs lignes de tailles respectives P et $K-1$. Les facteurs (contenus dans la matrice F) recherchés sont les sommets du K -simplexe qui englobe les données dans $\mathcal{S}$. On note T_i la projection du $i^{\text{ème}}$ facteur dans $\mathcal{S}$ et T la matrice de taille K

par K des facteurs projetés :

$$T = \begin{pmatrix} 1 & T_1 \\ \vdots & \vdots \\ 1 & T_K \end{pmatrix} \quad (3.6)$$

Dans [Fil+15], nous avons proposé une nouvelle initialisation de cette matrice T . Dans cette initialisation, les facteurs sont initialisés à l'aide des données. Le premier facteur est initialisé comme étant l'observation la plus éloignée du centre des données. Le deuxième facteur est choisi comme étant l'observation la plus éloignée du premier facteur. Puis itérativement, les facteurs suivants sont choisis comme étant les plus éloignés du sous-espace engendré par les facteurs précédemment choisis. Ceci permet d'obtenir une initialisation avec des facteurs placés aux extrémités des observations.

Une fois T initialisée, les matrices A et F des images factorielles et des facteurs s'obtiennent par ces relations :

$$F = T \times U \quad (3.7)$$

$$A = V \times T^{-1} \quad (3.8)$$

La matrice T peut être recalculée à partir de la pseudo-inverse de A :

$$T = (A^T A)^{-1} A^T V \quad (3.9)$$

Les matrices A et F sont normalisées, de telle sorte que la somme de chacune de leur colonne soit égale à 1. L'algorithme de mise à jour alternée est appliqué à la suite de l'initialisation. Les mises à jour continuent tant que les changements relatifs dans la matrice A entre deux itérations sont supérieurs à 10^{-4} , pendant 3 itérations de suite. L'algorithme est détaillé dans la figure 3.1.

Nous avons montré dans [Fil+15] que l'utilisation des ROI comme masque pour les images factorielles permet d'améliorer sensiblement la précision de l'estimation de A et de F . Cette approche est toutefois limitée car les régions d'intérêt y sont traitées comme une information binaire et elle n'utilise pas la cohérence spatiale de ces régions. Comme ces régions sont le plus souvent tracées à la main et sujettes à erreurs, il est préférable de les intégrer de manière plus robuste à l'algorithme de séparation. C'est ce que nous avons proposé par la suite dans [Fil+16] et [FDM17], et fait l'objet des sections suivantes. Une initialisation se basant sur les régions d'intérêt est également proposée.

3.3 RUDUR : Un algorithme générique de séparation utilisant de façon robuste les ROI.

Nous présentons ici RUDUR, un algorithme générique de séparation développé pendant cette thèse. Cet algorithme exploite les connaissances fournies par les régions d'intérêt de multiples manières. Ces connaissances sont introduites dans une fonctionnelle dont les variables sont les images factorielles et les courbes d'activité, de telle sorte que la solution conforme

Algorithme 1. Utilisation des ROI comme masques, avec algorithme de mises à jour alternées (cf [Fil+15]).

- 1: Sélectionner les régions d'intérêt et les enregistrer dans la matrice M .
 - 2: Calculer la matrice X des données centrées normalisées avec (3.3)
 - 3: Calculer les vecteurs propres de $Y^T Y$, et les enregistrer dans U .
 - 4: Projeter les données X dans $\mathcal{S}$, et noter le résultat dans V .
 - 5: Initialiser la matrice T .
 - 6: **Tant Que** Changement dans A **Faire**
 - 7: Mettre à jour A à partir de T avec (3.8).
 - 8: Mettre à 0 les coefficients négatifs de A et ceux en dehors de leur ROI.
 - 9: Normaliser A .
 - 10: Mettre à jour T à partir de A avec (3.9).
 - 11: Calculer F à partir de T avec (3.7).
 - 12: Mettre à 0 les coefficients négatifs de F .
 - 13: Normaliser F .
 - 14: **Fin Tant Que**
 - 15: **Retourner** (A, F)
-

FIGURE 3.1 – Algorithme de séparation utilisé dans [Fil+15]

aux a priori minimise cette fonctionnelle. L'utilisation d'une fonctionnelle est motivée par les a priori de différente nature à introduire dans la séparation et par la flexibilité de cette approche. Les différents termes de la fonctionnelle ainsi que la méthode de minimisation sont détaillés dans la suite.

3.3.1 Fidélité aux données

Dans RUDUR, nous faisons l'hypothèse que les données à séparer suivent le modèle de mélange linéaire $Y = A \times F + \epsilon$, où ϵ est le terme d'erreur, détaillé dans la section 2.2.4. Cette hypothèse se traduit par un terme de fidélité aux données dans la fonctionnelle. Ce terme mesure la proximité entre les données Y et le produit $A \times F$ décrit par le modèle de mélange linéaire. Dans le cas où le terme d'erreur entre Y et $A \times F$ est supposé gaussien, indépendant et identiquement distribué, on utilise une norme de Frobenius pour mesurer la proximité entre Y et $A \times F$:

$$f_{\text{LS}}(A, F) = \|Y - A \times F\|_2^2 = \sum_{i=1}^N \sum_{t=1}^P (Y_{i,t} - \sum_{k=1}^K A_{i,k} \times F_{k,t})^2 \quad (3.10)$$

Dans le cas où l'on observe des réalisations de loi de Poisson on peut utiliser une divergence de Kullback-Leibler généralisée [Dev08]; [LS01] :

$$f_{\text{KL}}(A, F) = \sum_{i,t} \left(Y_{i,t} \times \log \left(\frac{Y_{i,t}}{(AF)_{i,t}} \right) - Y_{i,t} + (AF)_{i,t} \right) \quad (3.11)$$

Nous avons vu dans la section 2.3 que les erreurs dans les images étaient en partie dû à l'aléatoire de réalisations de loi de Poisson qui introduit un bruit de comptage. Mais le terme d'erreur considéré dans le problème de séparation de sources contient également des erreurs de modélisation. En effet, les hypothèses d'homogénéité spatiale, de stabilité temporelle et du nombre restreint de sources sont des approximations de la réalité. Les différences entre la matrice Y et le produit matriciel $A \times F$ ne sont donc pas dûes uniquement au bruit de comptage présent dans les images, mais aussi aux erreurs de modélisation qui sont importantes, et aux erreurs de mesures (cf équation 2.8). Au vu de leur nature, les erreurs de modélisation sont difficiles à caractériser, et dépendent de la séquence d'image étudiée, il n'est donc pas possible de trouver une mesure optimale de l'erreur. Ainsi, dans la suite nous utiliserons faute de mieux le terme f_{LS} qui reste très général, et qui est pratique pour l'optimisation car différentiable et convexe en A et en F .

3.3.2 Initialisation avec les ROI

Dans la plupart des algorithmes de séparation de sources en imagerie nucléaire, le schéma classique est d'estimer les matrices A et F séquentiellement. La matrice A sert à mettre à jour la matrice F , et inversement. Le procédé est itératif jusqu'à convergence et jusqu'à ce que les critères de positivité soient respectés. La mise à jour des matrices peut se faire à l'aide de formules analytiques, ou avec des descentes de gradient. Comme les données sont bilinéaires en A et F , un terme classique de fidélité aux données avec des moindres carrés n'est pas convexe conjointement en A et F et est donc difficile à optimiser. En revanche, les sous-problèmes où l'on considère que l'une des deux matrices est fixée sont bien convexes en l'autre matrice. Ces sous-problèmes ont donc une solution, parfois même analytique.

Les valeurs de départ données aux coefficients des matrices A et/ou F sont très importantes et déterminent en grande partie les estimations finales de A et F . En effet, bien que les sous-problèmes soient convexes, le problème de base ne l'est pas et il comporte des minima locaux. Si l'algorithme est initialisé à proximité d'un minimum local, il est vraisemblable qu'il converge vers ce minimum local. Il est donc primordial d'initialiser l'algorithme proche de la solution désirée. Nous proposons ici d'initialiser grâce aux informations liées aux régions d'intérêt.

En notant A_0 et F_0 les initialisations de A et F , nous proposons l'initialisation suivante :

$$\text{Initialisation : } \begin{cases} A_0 &= M \\ F_{0,k,:} &= \frac{\sum_{i|M_{i,k}=1} Y_{i,:}}{\sum_i M_{i,k}} \quad (\forall k \in \llbracket 1..K \rrbracket) \end{cases} \quad (3.12)$$

où M est la matrice binaire des régions d'intérêt, décrite dans la section précédente. Les

images factorielles sont donc simplement initialisées avec la matrice M . Les facteurs sont initialisés en moyennant spatialement l'activité présente dans la région d'intérêt qui leur est associée. À noter que cette initialisation a du sens car elle correspond à l'ancienne pratique où les activités des compartiments physiologiques étaient calculées en traçant des régions d'intérêt. Si les ROI sont correctement tracées, cette initialisation permet de débiter l'algorithme près de la solution recherchée.

3.3.3 Pénalisation robuste des images factorielles à l'aide des ROI

L'utilisation des régions d'intérêt comme masque pour les images factorielles n'est pas robuste au choix de ces régions. Par exemple deux pixels voisins de part et d'autre d'une frontière de ROI subissent un traitement très différent. L'un a son coefficient dans l'image factorielle associée non contraint, tandis que l'autre est forcé à zéro. Pourtant le choix d'en inclure un dans la ROI et pas l'autre est une décision difficile. D'autres spécialistes auraient peut-être choisi d'inclure les deux dans la région, ou aucun. Cette utilisation des régions d'intérêt n'est donc pas robuste à leur sélection. De plus, à cause de la résolution spatiale, de la diffusion Compton et de l'imperfection de la collimation (voir section 2.2.3), les images factorielles des compartiments physiologiques sont souvent très lisses, avec des valeurs très faibles quand on s'éloigne du centre du compartiment physiologique. Il n'y a donc pas de frontières nettes. Si la région d'intérêt est choisie trop petite, alors certains pixels à l'extérieur contenant un peu de ce compartiment physiologique auront leur coefficient dans l'image factorielle associée forcé à 0. Ceci engendre alors nécessairement des erreurs dans l'estimation de A et de F . Au contraire, si la région d'intérêt est choisie trop grande, cela est beaucoup moins contraignant pour le modèle, et peut ne pas être suffisant pour faire face à la non-unicité.

Ceci montre la limite de l'utilisation des régions d'intérêt comme une information binaire d'appartenance ou non à une image factorielle. Une méthode de séparation utilisant les régions d'intérêt doit être robuste à leur sélection. Nous proposons donc de ne plus interpréter les informations véhiculées par la région d'intérêt comme des informations binaires. Un moyen de corriger ce manque de robustesse est de remplacer cette information par la distance à la région d'intérêt, qui prend en compte la cohérence spatiale de ces dernières. Notons D la matrice de distance aux régions d'intérêt :

$$D_{i,k} = \min_{j|M_{j,k}=1} dist_{euc}(i, j) \quad (\forall i \in \llbracket 1..N \rrbracket, \quad \forall k \in \llbracket 1..K \rrbracket) \quad (3.13)$$

où $dist_{euc}(i, j)$ est la distance euclidienne entre le $i^{\text{ème}}$ et le $j^{\text{ème}}$ pixel. Cette matrice est donc de même taille que la matrice des images factorielles A et la matrice des ROI M . Chaque colonne de D contient la distance euclidienne entre une région d'intérêt et chacun des pixels. En particulier $D_{i,k} = 0$ si et seulement si le $i^{\text{ème}}$ pixel appartient à la $k^{\text{ème}}$ ROI. Contrairement, à M , les informations de cette matrice ne sont pas binaires et sont donc plus complètes.

L'idée de contraindre les pixels en dehors des régions d'intérêt est pertinente, mais ceci doit se faire de manière plus robuste que dans la méthode précédente. Nous proposons ici de

pénaliser les images factorielles qui ont des coefficients non nuls en dehors de la région d'intérêt associée. Cette pénalisation dépend de la distance à la région d'intérêt et de la valeur de ce coefficient. Plus le pixel associé est loin de la région d'intérêt, plus la solution est pénalisée. De même, plus le coefficient est fort, plus la solution est pénalisée. Une telle pénalisation peut s'écrire comme la norme de $h(D) \circ A$, où h est une fonction positive croissante pour chaque coefficient de D et $\circ$ est le produit de Hadamard, c'est à dire le produit matriciel terme à terme. Nous avons donc proposé dans [Fil+16] l'utilisation du carré de la norme de Frobenius :

$$f_{\text{ROI}_2}(A) = \|h(D) \circ A\|_2^2 = \sum_{i=1}^N \sum_{k=1}^K |h(D_{i,k}) \times A_{i,k}|^2 \quad (3.14)$$

La norme 2 est pratique car elle rend le terme f_{ROI} différentiable. Cependant, elle ne pénalise que très peu les petites valeurs de A en dehors de leur ROI et ne favorise donc pas la parcimonie. Or, lorsque les sources sont mal séparées, ceci peut être dû à un mélange des images factorielles entre elles, faisant apparaître de faibles valeurs à la place de zéros dans la matrice A (cf figure 2.5). Cette pénalisation a donc été raffinée dans [FDM17], en utilisant cette fois-ci une norme ℓ_1 lisse, pour corriger ce problème :

$$f_{\text{ROI}}(A) = \|h(D) \circ A\|_{1,\mu} = \sum_{i=1}^N \sum_{k=1}^K \|h(D_{i,k}) \times A_{i,k}\|_{1,\mu} \quad (3.15)$$

où $\|\cdot\|_{1,\mu}$ est une norme ℓ_1 lisse, avec μ une constante positive, définie ainsi (pour une matrice de taille N_1 par N_2 quelconque) :

$$\|X\|_{1,\mu} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \sqrt{X_{n_1,n_2}^2 + \mu^2} - \mu \quad (3.16)$$

Contrairement à la norme ℓ_1 classique, la norme ℓ_1 lisse de l'équation (3.16) est différentiable en 0, ce qui facilite le processus d'optimisation. Lorsque μ est nul, ceci revient à utiliser une norme ℓ_1 classique. Les différences entre la norme ℓ_1 lisse proposée et la norme ℓ_1 classique sont montrées dans la figure 3.2. Ces normes ont des comportements très similaires lorsque μ est proche de 0. Le terme f_{ROI} est très peu sensible à la valeur de μ , du moment que ce dernier est choisi petit par rapport à 1. Dans la suite, on choisira $\mu = 0.1$. À noter que les normes ℓ_1 lisses sont appelées normes abusivement car elles ne respectent ni le critère d'homogénéité ni l'inégalité triangulaire.

La fonction h présente dans la pénalisation f_{ROI} doit être choisie croissante selon chaque coefficient de D , afin de pénaliser davantage les pixels loin des régions d'intérêt. Dans [FDM17] nous proposons d'utiliser une distance élevée au carré, que nous conserverons pour la suite :

$$h(D) = D \circ D \quad (3.17)$$

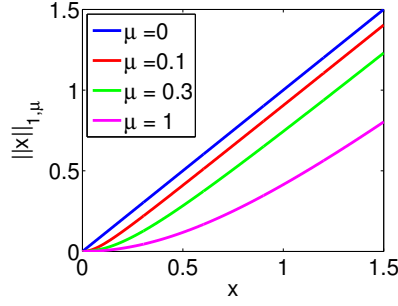


FIGURE 3.2 – Représentation graphique de normes ℓ_1 lisses et de la norme ℓ_1 classique. En abscisse se trouve la valeur d'un réel x , et en ordonnée la norme ℓ_1 (lisse ou non) de ce réel en fonction de la valeur μ . Lorsque μ est différent de 0, $\|\cdot\|_{1,\mu}$ est dérivable en 0. Plus μ se rapproche de 0, plus $\|\cdot\|_{1,\mu}$ se rapproche de la norme ℓ_1 classique.

ce choix est motivé par des résultats expérimentaux qui montrent que les performances sur nos données synthétiques sont meilleures qu'avec une fonction h choisie comme étant l'identité ($h(D) = D$). Finalement, la pénalisation s'écrit :

$$f_{\text{ROI}}(A) = \|D \circ D \circ A\|_{1,\mu} = \sum_{i=1}^N \sum_{k=1}^K \|D_{i,k}^2 \times A_{i,k}\|_{1,\mu} \quad (3.18)$$

À noter que lorsque $D_{i,k} = 0$, il n'y a pas de pénalisation. Lorsqu'un pixel appartient à une région d'intérêt, le coefficient associé dans la matrice A n'est pas pénalisé. En revanche, lorsque $D_{i,k} \neq 0$, le coefficient associé dans la matrice A est pénalisé, et ce d'autant plus que la distance à la région d'intérêt est grande. En effet, plus un pixel est loin d'une région d'intérêt, moins il est vraisemblable qu'il contienne la source associée. Un des autres avantages du terme f_{ROI} est qu'il est convexe en A , et sont ajout ne complique donc pas le processus d'optimisation.

Englobée dans une fonctionnelle comportant un terme de fidélité aux données, cette pénalisation est très contraignante et permet de faire face à la non-unicité. Les connaissances a priori sur les régions d'intérêt sont exploitées de façon robuste grâce à la matrice des distances. Deux pixels voisins de part et d'autre d'une frontière auront un rôle moins différent qu'avec l'utilisation des ROI comme masques. Cette robustesse est étudiée dans le chapitre suivant. La pénalisation présentée ici est un des termes de la fonctionnelle détaillée plus loin dans la section 3.3.6.

3.3.4 Relaxation du modèle en dehors des ROI

Les méthodes de séparation font des hypothèses sur les compartiments physiologiques ainsi que sur leur mélange (voir section 2.2). Les principales hypothèses sont :

- La stabilité temporelle des compartiments physiologiques : ceux-ci sont considérés rigides et non-déformables.
- L'homogénéité spatiale des compartiments physiologiques : l'évolution de l'activité d'un compartiment physiologique est la même en tout pixel, à un facteur d'échelle près.
- Le modèle de mélange des compartiments physiologiques est supposé linéaire et le nombre de sources faible.

Ces hypothèses très fortes permettent d'écrire les données Y sous la forme d'un produit de matrices A et F , à l'erreur ϵ près ($Y = A \times F + \epsilon$). Le modèle de mélange linéaire fait l'hypothèse que la séquence d'images analysée est la somme d'un petit nombre de sources, ayant chacune des signatures spatiale et temporelle propres (une image factorielle et un facteur). La courbe d'activité sur chaque pixel peut donc s'exprimer, au bruit près, à l'aide d'une combinaison linéaire d'un petit nombre de facteurs communs à tous les pixels. Ce modèle est une approximation de la réalité qui n'est valable que dans une zone restreinte des images de la séquence. Les hypothèses précédentes ne sont pas pertinentes dans toute l'image.

Nous proposons ici d'utiliser les régions d'intérêt et la matrice des distances D pour gérer le domaine de validité du modèle de mélange linéaire. Pour les pixels en dehors de ces régions, il y a peu de chances que le modèle de mélange linéaire soit une bonne approximation. En effet, ces pixels ne contiennent vraisemblablement que très peu d'activité des K compartiments physiologiques supposés présents dans l'image. Leur rôle dans la séparation ne doit donc pas être aussi important que ceux appartenant à au moins une des régions d'intérêt. Plutôt que de les supprimer complètement de la séparation, nous proposons de diminuer leur impact dans celle-ci afin de prendre en compte l'information utile qu'ils apportent. Les régions d'intérêt étant tracées manuellement, elles doivent être incluses de façon robuste à la méthode de séparation. Une nouvelle fois, ceci se fera à l'aide de la matrice des distances aux régions d'intérêt.

Nous proposons donc de relaxer le modèle de mélange linéaire pour les pixels en dehors de toutes les régions d'intérêt. Ceci se traduit en pratique par une pondération dans le terme de fidélité aux données. Un terme classique de fidélité aux données avec les moindres carrés s'exprime ainsi :

$$f_{LS}(A, F) = \|Y - A \times F\|_F^2 = \sum_{i=1}^N \sum_{t=1}^P (Y_{i,t} - \sum_{k=1}^K A_{i,k} \times F_{k,t})^2 \quad (3.19)$$

Le modèle de mélange linéaire peut être relaxé pour les pixels en dehors de toutes les régions d'intérêt, en diminuant la pénalisation liée à la mauvaise approximation de ce pixel par le mélange. En notant W la matrice diagonale de pondération des pixels (de taille N par N), un terme pondéré de fidélité aux données avec les moindres carrés peut s'exprimer ainsi :

$$f_{WLS}(A, F) = \|W(Y - A \times F)\|_F^2 = \sum_{i=1}^N W_{i,i}^2 \sum_{t=1}^P (Y_{i,t} - \sum_{k=1}^K A_{i,k} \times F_{k,t})^2 \quad (3.20)$$

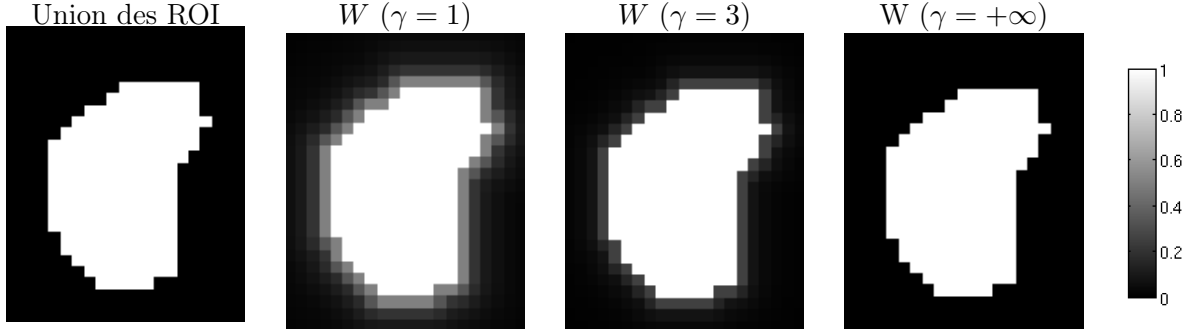


FIGURE 3.3 – Diagonale de la matrice de relaxation W (remaniée en image) en fonction de γ sur des données simulées de scintigraphie rénale (cf figure 2.2). L'union des régions d'intérêt est en première colonne. Les colonnes suivantes contiennent la diagonale de la matrice de pondération résultante (remaniée en image), avec différentes valeurs de γ ($\gamma = 1$, $\gamma = 3$ et $\gamma = +\infty$). La relaxation des pixels à la frontière de l'union des régions d'intérêt est plus ou moins importante selon la valeur de γ .

Grâce à la matrice diagonale W , il est possible de choisir l'impact de chaque pixel dans la séparation. Nous proposons d'utiliser les régions d'intérêt pour construire W , et plus précisément les distances aux régions d'intérêt :

$$W_{i,i} = \frac{1}{1 + \min_{k=1}^K g(D_{i,k})} \quad (3.21)$$

Les coefficients non diagonaux de W sont choisis nuls. Plusieurs fonctions g différentes peuvent être choisies, pourvu qu'elles soient croissantes sur $\mathbb{R}^+$. Nous proposons ici une fonction qui utilise une distance au carré qui pénalise davantage les pixels lointains :

$$g(x) = \gamma \times x^2 \quad (3.22)$$

où γ est une constante positive. Avec cette matrice W , les pixels absents de toute région d'intérêt ont un impact plus faible dans la séparation, dans le sens où la fidélité aux données est relâchée pour ces pixels. Cet impact dépend de leur distance à la plus proche région d'intérêt. Plus cette distance est grande, plus l'impact du pixel dans la séparation est faible. Le paramètre γ gère la vitesse de décroissance de l'impact en fonction de la distance à la plus proche région d'intérêt. Lorsque $\gamma = 0$, W est l'identité et il n'y a donc aucune relaxation du modèle, chaque pixel est traité de la même manière. Lorsque $\gamma = +\infty$, les pixels n'étant dans aucune région d'intérêt ont leur coefficient associé dans W égal à 0. Les pixels en dehors de toutes les régions d'intérêt n'ont donc aucun impact dans la séparation.

3.3.5 Pondération temporelle de la contrainte de fidélité avec les images d'intérêt

Les régions d'intérêt sont une information spatiale. Elles nous renseignent sur la position des compartiments physiologiques. Dans la section précédente, elles nous ont permis d'introduire une pondération spatiale relaxant la contrainte de fidélité aux données. Nous proposons ici de transposer cette idée au domaine temporel grâce à une matrice diagonale Q de taille P par P . Cette matrice pondère l'impact de chacune des images de la séquence dans la séparation. Ainsi, le $t^{\text{ème}}$ coefficient de la diagonale de Q contient le poids souhaité de la $t^{\text{ème}}$ image dans la séparation. Cette matrice est intégrée directement dans le terme de fidélité aux données de f_{RUDUR} :

$$f_{\text{QWLS}} = \|W(Y - AF)Q\|_F^2 = \sum_{i=1}^N W_{i,i}^2 \sum_{t=1}^P Q_{t,t}^2 \times (Y_{i,t} - \sum_{k=1}^K A_{i,k} \times F_{k,t})^2 \quad (3.23)$$

La matrice Q est l'analogue temporel de la matrice W . Tandis que la matrice W est construite à l'aide des régions d'intérêt, la matrice Q peut être construite avec ce que nous appelons les *images d'intérêt*. Ces dernières indiquent pour chaque compartiment physiologique les images où ce dernier présente une activité importante.

Nous montrerons dans la suite (section 4.5), comment l'utilisation de cette matrice de pondération Q permet de faciliter la séparation de sources de très faible énergie.

3.3.6 Fonction objectif finale

Les idées évoquées précédemment sont combinées dans la même fonctionnelle. Nous avons donc proposé dans [FDM17] la fonctionnelle f_{RUDUR} à minimiser sous la contrainte de A et F à coefficients positifs :

$$f_{\text{RUDUR}}(A, F) = f_{\text{QWLS}}(A, F) + \alpha f_{\text{ROI}}(A) + \beta f_{\text{Reg}}(F) \quad (3.24)$$

RUDUR signifie en anglais "Robust Unmixing of Dynamic sequences Using Regions of interest", ce qui peut se traduire par "Séparation robuste de sources dans des séquences dynamiques à l'aide de régions d'intérêt". Le terme f_{QWLS} correspond au terme de fidélité aux données avec les moindres carrés pondérés présenté dans les sections précédentes. Le terme f_{ROI} est la pénalisation des images factorielles à l'aide des régions d'intérêt, présentée dans la section 3.3.3. Le terme f_{Reg} est un terme régularisant qui est détaillé ci-dessous. Enfin les constantes α et β sont les poids qui permettent de contrôler l'importance des différents termes de la fonctionnelle.

Un des effets du bruit dans les données est qu'il dégrade les estimations des facteurs en produisant des facteurs bruités. Le dernier terme, f_{Reg} permet d'y faire face, en pénalisant les

facteurs qui ont trop de variations. La variation totale [ROF92] et la régularisation Tikhonov [Tik63] ; [Ber+07] sont deux manières de pénaliser ces variations. Dans les deux cas, un opérateur de gradient matriciel Γ est nécessaire. Cet opérateur est une matrice bidiagonale de taille P par P :

$$\Gamma = \begin{bmatrix} 0 & 0 & \ddots & \ddots & 0 \\ -1 & 1 & 0 & \ddots & \ddots \\ 0 & \ddots & 1 & 0 & \ddots \\ \ddots & 0 & -1 & \ddots & 0 \\ 0 & \ddots & 0 & -1 & 1 \end{bmatrix} \quad (3.25)$$

Ce qui permet de définir un terme de régularisation Tikhonov :

$$f_{\text{Reg}}(F) = f_{\text{Tik}}(F) = \|S^{0.5} \times \Gamma \times F^T\|_F^2 = \sum_{k=1}^K \sum_{t=2}^P S_{t,t} (F_{k,t} - F_{k,t-1})^2 \quad (3.26)$$

ou bien un terme de régularisation avec la variation totale :

$$f_{\text{Reg}}(F) = f_{\text{VarTot}}(F) = \|S \times \Gamma \times F^T\|_1 = \sum_{k=1}^K \sum_{t=2}^P S_{t,t} |F_{k,t} - F_{k,t-1}| \quad (3.27)$$

où S est une matrice diagonale de taille P par P . Les deux termes présentent l'avantage d'être convexes en F . La régularisation Tikhonov favorise les signaux lisses tandis que la totale variation favorise les signaux constants par morceaux. On préférera donc l'utilisation de la régularisation Tikhonov dans la suite. Un autre avantage de cette régularisation est qu'elle est dérivable, ce qui facilite la minimisation.

Les compartiments physiologiques sont parfois connues comme ayant de fortes variations d'activité, c'est notamment le cas dans le début des séquences à la suite de l'injection du bolus. Il est donc préférable que la contrainte de régularité des courbes d'activités soit relâchée pour ces moments. C'est le rôle de la matrice S , qui permet de pondérer l'importance de la régularisation selon le moment temporelle. Si le $i^{\text{ème}}$ coefficient diagonal de S est nul, alors des variations fortes d'activité pour les facteurs entre les $i^{\text{ème}}$ et la $i-1^{\text{ème}}$ images ne seront pas pénalisées.

La fonction objectif finale contient donc trois termes. La fidélité aux données est assurée par f_{QWLS} , où l'importance de cette fidélité dépend pour chaque pixel de leur distance à la plus proche région d'intérêt. Le terme f_{ROI} assure que les images factorielles n'aient pas de valeurs aberrantes vis à vis des régions d'intérêt. Enfin le terme f_{Reg} promeut les facteurs lisses. Les régions d'intérêt jouent un rôle double dans cette fonctionnelle, à travers les termes

f_{QWLS} et f_{ROI} . Elles sont intégrées de façon robuste à la méthode de séparation grâce à la matrice D des distances aux ROI.

La minimisation de cette fonctionnelle passe par une initialisation des matrices A et F . Cette initialisation se fait grâce aux régions d'intérêt, comme décrit dans la section 3.3.2, ce qui permet d'initialiser l'algorithme de minimisation à proximité de la solution recherchée. Cette fonctionnelle répond aux différentes limites des méthodes évoquées en conclusion du chapitre précédent :

- **Non-unicité** : Des connaissances a priori très fortes sont intégrées dans la méthode de séparation. Les régions d'intérêt permettent de contraindre fortement les images factorielles. Elles sont intégrées de façon robuste grâce à la matrice des distances D et permettent de faire face à la non-unicité.
- **Initialisation** : Une initialisation orientée par les régions d'intérêt est proposée. Cette initialisation coïncide avec la méthode de base pour trouver les activités des compartiments physiologiques en moyennant spatialement le signal de chaque région d'intérêt. Ceci permet d'initialiser l'algorithme à proximité de la solution recherchée, afin de réduire les problèmes liés aux minima locaux.
- **Modèle de mélange** : Le modèle de mélange est relaxé pour les pixels en dehors des régions d'intérêt. L'importance de cette relaxation dépend de la distance à la plus proche région d'intérêt. Plus un pixel est loin des régions d'intérêt, moins il est vraisemblable qu'il respecte le modèle, celui-ci est donc davantage relaxé. Le modèle est également relaxé pour les images éloignées temporellement des images d'intérêt.
- **Réduction de dimension** : Aucune réduction de dimension n'est nécessaire ici, ce qui permet de ne pas perdre d'information avant même d'avoir commencé la séparation, et d'éviter d'introduire des artefacts comme des valeurs négatives dans les données. En revanche, le nombre de sources K est lui bien fixé, comme dans la grande majorité des méthodes de séparation.

De plus les deux termes ajoutés au terme de fidélité aux données (f_{ROI} et f_{Reg}) sont conjointement convexes en A et F . La pondération ajoutée dans le terme de fidélité aux données ne change pas la convexité de ce terme, il reste toujours convexe en chacune des matrices A et F séparément. La fonctionnelle n'est donc pas plus complexe à optimiser qu'une fonctionnelle classique incluant un terme avec des moindres carrés et des contraintes de positivité.

Une des faiblesses de la méthode concerne les paramètres libres à fixer : les constantes α , β et γ . Il n'y a pas de méthode générale pour les choisir. Cependant, les mêmes jeux de paramètres peuvent être utilisés pour différentes séquences d'images issues d'une même étude, celle du 6-DIG par exemple. En appliquant la méthode de séparation sur des données synthétiques simulant l'étude, un jeu de paramètres optimal pour ces données synthétiques peut être trouvé. Si les données synthétiques reproduisent assez fidèlement l'étude, ce jeu de paramètres peut être utilisé par la suite pour des séquences issues de données réelles. Il est aussi possible de chercher un bon jeu de paramètres pour une étude réelle en essayant plusieurs jeux différents et en faisant vérifier qualitativement les résultats par des spécialistes.

Si l'on considère qu'un jeu de paramètres (α, β, γ) pertinent a été trouvé pour un type d'étude, l'application de la méthode est ensuite presque automatique pour ce type d'étude donné. En effet, les seules connaissances nécessaires sont les régions d'intérêt, qu'il faut sélectionner au préalable.

Afin d'estimer les facteurs et les images factorielles, cette fonctionnelle doit être minimisée. C'est l'objet de la section suivante.

3.3.7 Optimisation de la fonctionnelle

Cette section est dédiée à la minimisation de la fonctionnelle f_{RUDUR} . L'algorithme complet est disponible à la fin de la section. Différents éléments déjà évoqués dans la section 2.4.3 consacrée aux méthodes de séparation avec contraintes de positivité sont repris. L'objectif est ici de trouver des estimations $\tilde{A}$ et $\tilde{F}$ de la matrice des images factorielles et de la matrice des facteurs, en trouvant les matrices qui minimisent f_{RUDUR} sous contraintes de positivité :

$$(\tilde{A}, \tilde{F}) = \underset{A \geq 0, F \geq 0}{\operatorname{argmin}} f_{\text{RUDUR}}(A, F) = \underset{A \geq 0, F \geq 0}{\operatorname{argmin}} (f_{\text{QWLS}}(A, F) + \alpha f_{\text{ROI}}(A) + \beta f_{\text{Reg}}(F)) \quad (3.28)$$

Si la constante β est choisie égale à 0, il est vraisemblable que toute l'énergie ira dans les facteurs. En effet, les images factorielles étant pénalisées, elles vont tendre vers 0 tandis que les facteurs eux compenseront en concentrant l'énergie de la source. Ceci est lié à l'indétermination d'échelle déjà évoquée ($A \times F = (\lambda A) \times \frac{1}{\lambda} F$). Pour cette raison, les images factorielles sont normalisées dans la suite de telle sorte que la somme de leurs coefficients soit constante. Pour des soucis d'échelle dans les représentations graphiques des images factorielles, nous avons choisi cette constante égale à $\frac{N}{K}$.

$$\sum_{i=1}^N A_{i,k} = \frac{N}{K} \quad (\forall k \in \llbracket 1..K \rrbracket) \quad (3.29)$$

Lorsque la matrice A est normalisée, la matrice F doit être au préalable contre-normalisée afin que les sources restent inchangées :

$$F_{k,:} = F_{k,:} \times (K \times (N \times \sum_{i=1}^N A_{i,k})^{-1}) \quad (\forall k \in \llbracket 1..K \rrbracket) \quad (3.30)$$

Afin de minimiser f_{RUDUR} , l'approche classique de minimisation alternée a été retenue. Le problème de minimisation revient donc à traiter alternativement et itérativement deux sous-problèmes où l'une des matrices est fixée, et qui ont l'avantage d'être convexes. Ces deux sous-problèmes sont :

$$\hat{A} = \underset{A \geq 0}{\operatorname{argmin}} f_{\text{RUDUR}}(A, F) = \underset{A \geq 0}{\operatorname{argmin}} (f_{\text{QWLS}}(A, F) + \alpha f_{\text{ROI}}(A)) \quad (3.31)$$

et

$$\hat{F} = \underset{F \geq 0}{\operatorname{argmin}} f_{\text{RUDUR}}(A, F) = \underset{F \geq 0}{\operatorname{argmin}} (f_{\text{QWLS}}(A, F) + \beta f_{\text{Reg}}(F)) \quad (3.32)$$

Une descente de gradient est utilisée afin de résoudre ces deux sous-problèmes. Pour chacun des sous-problèmes, plusieurs itérations de descente sont réalisées avant de passer au sous-problème suivant. Afin d'accélérer la convergence, les directions de descente sont choisies comme étant la direction conjuguée du gradient en utilisant la formule de Polak-Ribière [PR69]. Par exemple, pour le premier sous-problème (minimisation selon A), en notant s_n la direction conjuguée du gradient, et A_n la mise à jour de A à la $n^{\text{ème}}$ itération on a :

$$\text{Gradient : } \Delta u_n = -\frac{\partial f_{\text{RUDUR}}}{\partial A}(A_{n-1}, F) \quad (\forall n \geq 1) \quad (3.33)$$

$$\text{Gradient Conjugué : } \begin{cases} s_0 &= 0 \\ s_n &= \Delta u_n + \tau_n \times s_{n-1} \end{cases} \quad (\forall n \geq 1) \quad (3.34)$$

$$\text{Formule de Polak-Ribière : } \tau_n = \frac{\Delta u_n^T (\Delta u_n - \Delta u_{n-1})}{\Delta u_{n-1}^T \Delta u_{n-1}} \quad (3.35)$$

Dans ces formules, les matrices A et F sont vectorisées. Pour le deuxième sous-problème (minimisation selon F), on utilise des formules similaires, où cette fois on dérive non pas par rapport à A mais par rapport à F dans l'équation (3.33). Les formules des dérivées partielles des différents termes de la fonctionnelle en fonction de A et de F sont disponibles en Annexe B.

Afin d'accélérer la convergence, les pas de descente sont eux choisis à l'aide de la constante de Lipschitz de la dérivée de la fonction objectif par rapport à A (ou par rapport à F si l'on minimise selon F). Lorsque l'on minimise selon A , on peut minimiser indépendamment selon chacune des lignes de A . La constante de Lipschitz est différente pour chaque ligne de A . À l'itération n , en notant l_i la constante de Lipschitz de la dérivée partielle de f_{RUDUR} selon la $i^{\text{ème}}$ ligne de A , ou bien l_F la constante de Lipschitz de la dérivée partielle de f_{RUDUR} par rapport à F on a :

$$\text{Constante Lipschitz : } \begin{cases} l_{A,i} &= 2W_{i,i}^2 \|F_n Q^2 F_n^T\|_F \quad (\forall i \in \llbracket 1..N \rrbracket) \\ \text{ou } l_F &= 2(\|A_n^T W^2 A_n\|_F \|Q^2\|_F + \beta \|\Gamma^T S \Gamma\|_F) \end{cases} \quad (3.36)$$

Les calculs des constantes de Lipschitz sont disponibles en Annexe C. En notant $r_{n,i}$ le pas de descente lorsque l'on minimise par rapport à la $i^{\text{ème}}$ ligne de A , et r_n le pas de descente lorsque l'on minimise par rapport à F (à la $n^{\text{ème}}$ itération), on a dans le cas d'une régularisation Tikhonov [Rep+15] :

$$\text{Pas de descente : } \begin{cases} r_{n,i} &= \frac{0.9}{l_i} \\ \text{ou } r_n &= \frac{0.9}{l_F} \end{cases} \quad (\forall i \in \llbracket 1..N \rrbracket) \quad (3.37)$$

Lorsque l'on minimise selon A , le pas de descente est donc un vecteur, car il est différent pour chacune des lignes de A . Les descentes à la $n^{\text{ème}}$ itération se font donc avec les formules suivantes, où A_n est l'estimation mise à jour de A lorsque l'on minimise selon A (et respectivement F_n est l'estimation mise à jour de F lorsque l'on minimise selon F) :

$$\text{Mise à jour : } \begin{cases} A_{n(i,:)} &= A_{n-1(i,:)} + r_{n,i} \times s_{n,i} \\ \text{ou } F_n &= F_{n-1} + r_n \times s_n \end{cases} \quad (\forall i \in \llbracket 1..N \rrbracket) \quad (3.38)$$

où $s_{n,i}$ est la direction du gradient conjugué de la fonction objectif selon la $i^{\text{ème}}$ ligne de A . La résolution des deux sous-problèmes se fait donc en réalisant une descente de gradient jusqu'à convergence. Après chaque étape de la descente, la matrice mise à jour est projetée dans l'orthant positif afin de supprimer les coefficients négatifs. Si la matrice A est mise à jour, elle est également renormalisée afin de satisfaire l'équation 3.29, et la matrice F est contre-normalisée afin que chaque source reste inchangée. En pratique, il n'est pas nécessaire d'aller jusqu'à la convergence dans les sous-problèmes, et seulement 5 (fixé expérimentalement) étapes de descente sont réalisées pour accélérer la convergence globale de l'algorithme.

L'algorithme complet est décrit dans la figure 3.4. Il est initialisé à l'aide des régions d'intérêt. Les calculs et expressions des dérivées partielles sont disponibles en annexe ainsi que le calcul des constantes de Lipschitz. L'algorithme est stoppé lorsque les estimations de A et de F n'évoluent plus beaucoup pendant plusieurs itérations consécutives. En pratique, il est stoppé lorsque pendant 5 alternances consécutives, la fonction objectif a un changement relatif inférieur à 10^{-7} (fixé expérimentalement, correspond à la *Condition* dans l'algorithme en figure 3.4). Du fait des projections dans l'orthant positif et de la renormalisation, cet algorithme n'a pas de garantie de convergence. La convergence a cependant été systématique dans nos différents essais.

Une implémentation MATLAB commentée de cet algorithme est disponible en ligne à l'adresse <http://www.gipsa-lab.grenoble-inp.fr/~marc.filippi/>.

3.3.8 Flexibilité de la méthode

Faire de la séparation en minimisant une fonctionnelle comporte de nombreux avantages pour l'intégration des connaissances a priori. Si une connaissance peut s'exprimer comme un terme de pénalisation convexe en A et F qui favorise les solutions respectant cet a priori,

Algorithme 2. Minimisation de f_{RUDUR}

```

1: Fixer  $K$ 
2: Sélectionner les régions d'intérêt et les enregistrer dans la matrice  $M$ .
3: (Optionnel) Sélectionner les images d'intérêt pour construire  $Q$  (sinon  $Q = Id_P$ ).
4: (Optionnel) Choisir les sous séquences à régulariser (Par défaut  $S = Id_P$ ).
5: Initialiser les matrices  $A$  et  $F$  avec les équations (3.12).
6: Normaliser  $F$  avec l'équation (3.30) puis  $A$  avec (3.29).
7: Tant Que Condition Faire
8:   // Minimisation selon  $F$  avec  $A$  fixée :
9:   Pour  $n=1$  à 5 Faire
10:    Calculer la direction de descente avec (3.33),(3.34) et (3.35).
11:    Calculer le pas de descente avec (3.36) et (3.37).
12:    Opérer une descente grâce à l'équation (3.38).
13:    Projeter  $F$  en remplaçant les coefficients négatifs de  $F$  par des 0.
14:  Fin Pour
15:  // Minimisation selon  $A$  avec  $F$  fixée :
16:  Pour  $n=1$  à 5 Faire
17:    Calculer la direction de descente avec (3.33),(3.34) et (3.35).
18:    Calculer le pas de descente avec (3.36) et (3.37).
19:    Opérer une descente grâce à l'équation (3.38).
20:    Projeter  $A$  en remplaçant les coefficients négatifs de  $A$  par des 0.
21:    Normaliser  $F$  avec l'équation (3.30) puis  $A$  avec (3.29).
22:  Fin Pour
23: Fin Tant Que
24: Retourner  $(A, F)$ 

```

FIGURE 3.4 – Algorithme de minimisation pour f_{RUDUR}

elle est aisément intégrable à la fonctionnelle. Le seul inconvénient est que, comme évoqué dans la section 3.3.6, il faut fixer une constante qui pondère l'importance de ce terme dans la fonctionnelle.

Un autre avantage des méthodes passant par une minimisation de fonctionnelle est qu'elles sont très flexibles. Il est généralement possible de fixer certains paramètres en avance s'ils sont connus. Il est proposé dans FAMIS [Ben+94b] de fixer certains coefficients des facteurs et des images factorielles. Ces coefficients peuvent être connus si par exemple un seul des compartiments physiologiques est présent dans plusieurs images de la séquence, l'image factorielle associée peut être entièrement déterminée. Ces connaissances constituent l'ensemble des a priori intégrables à FAMIS. Les connaissances de ce type sont également intégrables dans RUDUR. En effet, il suffit de fixer ces coefficients dès l'initialisation, et de ne pas les modifier dans la suite de l'algorithme.

L'algorithme RUDUR peut aussi être appliqué dans le cas où la région d'intérêt d'un des compartiments physiologiques n'est pas disponible. Dans ce cas, de très légères modifications doivent être opérées. Nous conseillons de définir la matrice de pondération W en ignorant la

source dont la région est manquante. Ainsi la séparation sera surtout axée sur les compartiments physiologiques dont une région d'intérêt est disponible. Le terme de pénalisation f_{ROI} ignore aussi le compartiment physiologique car aucune connaissance a priori n'est disponible pour ce compartiment. Ainsi, aucune pénalisation n'est induite par l'image factorielle de ce compartiment dans le terme f_{ROI} . Le nombre de sources K doit lui cependant prendre en compte le compartiment sans région d'intérêt. On recherche donc une source de plus que de régions d'intérêt disponibles.

3.4 Conclusion

Étant donné la complexité que représente la tâche de séparer des sources en imagerie nucléaire, il est nécessaire que la méthode de séparation intègre des connaissances a priori fortes sur les sources à séparer. Les méthodes de séparation en imagerie nucléaire se sont développées en partie pour pallier les inconvénients des régions d'intérêt [Bar80]. Très peu de méthodes de séparation ont donc utilisé les régions d'intérêt, et celles les utilisant ne sont pas robustes à leur sélection et n'ont pas rencontré un grand succès en imagerie nucléaire. Les régions d'intérêt apportent pourtant des informations essentielles pour la séparation, grâce à leur lien direct avec les images factorielles.

Le chapitre précédent se concluait par les principales limites des méthodes de séparation actuelles pour l'imagerie nucléaire. Dans ce chapitre, nous avons proposé une méthode, RUDUR, répondant à ces principales limites, en utilisant de plusieurs façons la connaissance des régions d'intérêt. Les régions d'intérêt sont intégrées de façon robuste à la méthode grâce à l'utilisation d'une matrice contenant les distances de chaque pixel aux régions d'intérêt. La méthode minimise une fonctionnelle comportant trois termes. Le premier est un terme de fidélité aux données, dont l'importance de chaque pixel et chaque image est pondérée respectivement par les régions d'intérêt et par les images d'intérêt. Le second terme pénalise les images factorielles ayant des valeurs aberrantes vis-à-vis des régions d'intérêt grâce à la matrice des distances aux ROI. Enfin le troisième terme est une régularisation sur les facteurs permettant de faire face au bruit en pénalisant les facteurs peu lisses.

Les régions d'intérêt apportent ainsi des informations pertinentes et permettent de faire face à la non-unicité du problème de séparation, au bruit, et aux limites du modèle de mélange linéaire. Quelques variantes de la méthode ont été proposées. Notamment une adaptation au cas où il n'est pas possible de tracer la région d'intérêt d'un des compartiments physiologiques. Il est également possible d'ajouter des connaissances brutes sur la valeur des coefficients des images factorielles et des facteurs, comme dans FAMIS [Ben+94b]. Ces variantes sont applicables avec très peu de modifications sur la fonction objectif et montrent ainsi que la méthode proposée est flexible et s'adapte très bien aux différents problèmes que l'on peut rencontrer en imagerie nucléaire. Enfin une méthode est proposée pour chercher une source de très faible signal, grâce aux images d'intérêt, analogue des régions d'intérêt dans le domaine temporel.

RUDUR est donc une méthode générique, pouvant exploiter différentes connaissances a priori pour la séparation. Dans le chapitre suivant les performances de RUDUR et de ses

variantes sont évaluées sur des données synthétiques et des données cliniques. Les performances sont comparées aux méthodes de l'état de l'art présentées dans la section 2.5.3. La robustesse des méthodes à la sélection des régions d'intérêt est évaluée, ainsi que la robustesse à la sélection des méta-paramètres α , β et γ . Une section est dédiée à l'application de cette méthode sur les données obtenues avec le traceur 6-DIG.

Résultats

Sommaire

4.1	Jeu de données synthétiques simples	93
4.1.1	Présentation des données	93
4.1.2	Résultats et analyse	93
4.1.3	Ajout d'une source de bruit fond	96
4.2	Rénographies synthétiques réalistes	100
4.2.1	Présentation des données	100
4.2.2	Résultats et analyse	100
4.3	Rénographies cliniques	103
4.3.1	Présentation des données	103
4.3.2	Résultats et analyse	103
4.4	Évaluation de la sensibilité de la méthode	106
4.4.1	Sensibilité au choix des paramètres	106
4.4.2	Impact de la pénalisation des images factorielles	108
4.4.3	Impact de la régularisation	108
4.4.4	Robustesse au choix des régions d'intérêt	112
4.5	Utilisation des images d'intérêt pour l'évaluation d'une source de faible intensité	116
4.6	Données scintigraphiques issues de l'étude du 6-DIG	118
4.6.1	Résultats de séparation sur les données cliniques	119
4.6.2	Simulation par modélisation des séquences d'images 6-DIG	123
4.6.3	Résultats de séparation sur les données simulées	127
4.6.4	Conclusion sur les données 6-DIG	129
4.7	Conclusion sur les résultats	131

Les performances de RUDUR sont évaluées dans ce chapitre. Ces performances sont tout d'abord comparées à celles des autres méthodes de l'état de l'art sur un jeu de données synthétiques simples, sur un jeu de données synthétiques réalistes simulant des rénographies, ainsi que sur des scintigraphies cliniques. La robustesse de RUDUR dans le choix des régions d'intérêt et des paramètres est ensuite étudiée. L'impact des différents termes de la fonctionnelle est également analysé. Le cas où une source de signal très faible est présente est utilisé comme application des images d'intérêt. Enfin, les performances des différents algorithmes sur les données du 6-DIG sont analysées.

Dans ce chapitre, les différentes méthodes sont comparées quantitativement avec les erreurs d'estimation des courbes d'activité (lignes de F) et des images factorielles (colonnes de A) de chaque source. Comme il n'est pas possible de s'affranchir de l'indétermination d'échelle, les matrices A obtenues ainsi que celle de la vérité terrain, sont au préalable normalisées afin que chacune de leurs colonnes soit de somme égale à $\frac{N}{K}$. Les matrices F sont contre-normalisées de telle sorte que chaque source (produit matriciel $A_{:,k}F_{k,:}$) reste inchangée. De même, afin de s'affranchir des indéterminations de permutation, les colonnes de A et de F sont au préalable réordonnées afin de faire correspondre à chaque facteur et image factorielle de la vérité terrain, le facteur et l'image factorielle estimés les plus proches de la solution trouvée. Les erreurs d'estimation sur les matrices A et F sont ensuite calculées.

Deux mesures sont utilisées : l'écart absolu moyen et l'erreur quadratique moyenne. Ces erreurs sont normalisées et s'expriment ainsi :

$$NMSE(F_k^{GT}, \hat{F}_k) = \frac{\sum_{t=1}^P (F_{k,t}^{GT} - \hat{F}_{k,t})^2}{\sum_{t=1}^P (F_{k,t}^{GT})^2} \quad (4.1)$$

$$NMAE(F_k^{GT}, \hat{F}_k) = \frac{\sum_{t=1}^P |F_{k,t}^{GT} - \hat{F}_{k,t}|}{\sum_{t=1}^P |F_{k,t}^{GT}|} \quad (4.2)$$

où F_k^{GT} est le $k^{\text{ème}}$ facteur de la vérité terrain, et $\hat{F}_k$ le $k^{\text{ème}}$ facteur estimé. Des formules similaires sont utilisées pour mesurer les erreurs d'estimation sur les images factorielles.

Les méthodes de séparation utilisées à des fins de comparaison dans ce chapitre ont précédemment été détaillées dans la section 2.5.3. Il s'agit de FAMIS [Ben+94b], PLS-FADS (renommé PFADS par simplicité dans la suite) [SGH02], FAROI [ŠT12] et S-BSS-vecDC [TŠ15] (renommé S-BSS dans la suite). FAMIS a été retenue pour les comparaisons car tout comme RUDUR elle utilise les régions d'intérêt comme connaissance a priori. PLS-FADS est une méthode qui a souvent été utilisé, et qui procède à la séparation, comme RUDUR, en minimisant une fonctionnelle. FAROI et S-BSS-vecDC ont également été retenues car elles utilisent un cadre bayésien pour la séparation et qu'elles font partie des méthodes les plus récentes. De plus, S-BSS-vecDC introduit dans le processus de séparation un modèle compartimental adapté aux données traitées ici. Sauf mention contraire, les paramètres utilisés dans ce chapitre pour RUDUR sont $\alpha = 1$, $\beta = 10$, $\gamma = 3$, $\mu = 0.1$. Les régions d'intérêt sont choisies manuellement et sont communes à RUDUR et FAMIS. Les connaissances sur l'aspect temporel des sources ne sont pas utilisées sauf dans les sections 4.5 et 4.4.3 où les matrices Q et S sont choisies différentes de l'identité.

4.1 Jeu de données synthétiques simples

4.1.1 Présentation des données

Ce premier jeu de données contient des séquences de 50 images de 2500 pixels (50 lignes et 50 colonnes). Les séquences sont construites à partir du modèle de mélange linéaire (cf équation 2.2) et comportent $K = 3$ sources. Les matrices A et F qui caractérisent ces sources sont les mêmes pour toutes les séquences. Les séquences sont donc au bruit près identiques. Les courbes d'activité des 3 sources sont le résultat de la convolution d'une fonction d'entrée (commune aux 3 sources) et d'un noyau (différent pour chaque source). Les données vérifient ainsi les hypothèses faites par la méthode S-BSS-vecDC [TŠ15]. Les données s'écrivent sous la forme $Y = A \times F + \epsilon$. Les images factorielles et les facteurs utilisés sont disponibles dans la deuxième ligne de la figure 4.1. La variance du bruit est proportionnelle au signal, approchant ainsi un bruit poissonien, avec : $\epsilon_{i,t} = \sigma \times e_{i,t} \sqrt{Y_{i,t}^0}$ où $Y^0 = A \times F$, et $e_{i,t}$ est une réalisation aléatoire d'une loi normale de moyenne nulle et d'écart type 1. Le paramètre σ permet de contrôler le niveau de bruit. Deux niveaux de bruit sont testés dans la suite, $\sigma = 0.5$ et $\sigma = 1$, ce qui correspond à l'ordre de grandeur que l'on peut observer sur des données cliniques. Un exemple de séquence d'images de ce jeu de données, avec $\sigma = 0.5$, est disponible en annexe D.1.

4.1.2 Résultats et analyse

Un exemple de courbes d'activité et de cartes de présence obtenues par les différentes méthodes de séparation sur une telle séquence d'images (avec $\sigma = 0.5$) est présenté dans la figure 4.1. Pour toutes les méthodes, hormis RUDUR et FAMIS qui utilisent les régions d'intérêt, les images factorielles estimées sont encore mélangées : il est possible de voir dans ces images factorielles les contributions des autres sources, en particulier sur la deuxième source. De plus, on observe une moins bonne estimation des courbes d'activité pour ces méthodes, même si ces courbes sont dans l'ensemble bien estimées. L'utilisation des régions d'intérêt dans RUDUR et FAMIS permet ici d'améliorer nettement l'estimation à la fois des images factorielles et des courbes d'activité.

Afin d'obtenir des résultats quantitatifs, 100 séquences d'images pour chaque niveau de bruit ont été créées. Les différences entre les séquences d'images sont dues à l'aléatoire du bruit. Les résultats présentés dans les tables 4.1 et 4.2 résument les erreurs d'estimation moyennes obtenues sur ces 100 séquences, pour chaque niveau de bruit. Les erreurs obtenues sont fidèles aux résultats de la figure 4.1. RUDUR et FAMIS présentent les meilleures estimations pour les images factorielles et pour les courbes d'activité. Quand le niveau de bruit (σ) augmente, les erreurs d'estimation augmentent naturellement pour toutes les méthodes. Cependant elles augmentent significativement plus pour FAMIS que pour RUDUR. Ne possédant aucun terme de régularisation, FAMIS est plus sensible au bruit que RUDUR. Si bien que pour $\sigma = 1$, RUDUR présente l'écart moyen et l'erreur quadratique moyenne les plus bas pour toutes les courbes d'activité et toutes les images factorielles. Les autres méthodes ont quant à elles de

TABLE 4.1 – Erreurs d'estimation des facteurs (F). Jeu de données simples.

Mesure	Algorithme	sigma=0.5			sigma=1		
		F1	F2	F3	F1	F2	F3
NMSE	<i>RUDUR</i>	0.003 $\pm$ 0.001	0.014 $\pm$ 0.004	0.000 $\pm$ 0.000	0.015 $\pm$ 0.003	0.092 $\pm$ 0.025	0.002 $\pm$ 0.000
	<i>FAMIS</i>	0.002 $\pm$ 0.001	0.014 $\pm$ 0.006	0.000 $\pm$ 0.000	0.067 $\pm$ 0.101	0.368 $\pm$ 0.308	0.014 $\pm$ 0.063
	<i>PFADS</i>	0.062 $\pm$ 0.012	1.046 $\pm$ 0.167	0.015 $\pm$ 0.002	0.344 $\pm$ 0.053	16.604 $\pm$ 3.854	0.076 $\pm$ 0.011
	<i>S-BSS</i>	0.117 $\pm$ 0.143	2.299 $\pm$ 2.587	0.042 $\pm$ 0.047	0.608 $\pm$ 0.380	8.791 $\pm$ 5.751	0.145 $\pm$ 0.083
	<i>FAROI</i>	0.040 $\pm$ 0.007	0.700 $\pm$ 0.062	0.009 $\pm$ 0.001	0.095 $\pm$ 0.141	2.363 $\pm$ 3.408	0.012 $\pm$ 0.010
NMAE	<i>RUDUR</i>	0.056 $\pm$ 0.007	0.228 $\pm$ 0.031	0.016 $\pm$ 0.002	0.127 $\pm$ 0.012	0.553 $\pm$ 0.076	0.037 $\pm$ 0.004
	<i>FAMIS</i>	0.043 $\pm$ 0.004	0.127 $\pm$ 0.022	0.017 $\pm$ 0.002	0.177 $\pm$ 0.110	0.578 $\pm$ 0.209	0.063 $\pm$ 0.076
	<i>PFADS</i>	0.242 $\pm$ 0.023	1.285 $\pm$ 0.131	0.084 $\pm$ 0.007	0.523 $\pm$ 0.051	6.600 $\pm$ 1.101	0.198 $\pm$ 0.019
	<i>S-BSS</i>	0.318 $\pm$ 0.175	2.448 $\pm$ 1.968	0.147 $\pm$ 0.115	0.868 $\pm$ 0.324	6.871 $\pm$ 2.718	0.332 $\pm$ 0.121
	<i>FAROI</i>	0.195 $\pm$ 0.016	0.749 $\pm$ 0.044	0.066 $\pm$ 0.004	0.253 $\pm$ 0.089	1.264 $\pm$ 0.602	0.081 $\pm$ 0.017

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples. *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

TABLE 4.2 – Erreurs d'estimation des images factorielles (A). Jeu de données simples.

Mesure	Algorithme	sigma=0.5			sigma=1		
		A1	A2	A3	A1	A2	A3
NMSE	<i>RUDUR</i>	0.035 $\pm$ 0.002	0.161 $\pm$ 0.013	0.005 $\pm$ 0.000	0.129 $\pm$ 0.007	0.461 $\pm$ 0.032	0.020 $\pm$ 0.001
	<i>FAMIS</i>	0.320 $\pm$ 0.005	0.832 $\pm$ 0.040	0.050 $\pm$ 0.001	0.509 $\pm$ 0.489	1.397 $\pm$ 0.308	0.098 $\pm$ 0.056
	<i>PFADS</i>	0.076 $\pm$ 0.005	0.307 $\pm$ 0.014	0.006 $\pm$ 0.000	0.274 $\pm$ 0.034	0.721 $\pm$ 0.119	0.023 $\pm$ 0.002
	<i>S-BSS</i>	0.298 $\pm$ 0.339	0.341 $\pm$ 0.168	0.073 $\pm$ 0.087	0.937 $\pm$ 0.534	0.980 $\pm$ 0.303	0.188 $\pm$ 0.107
	<i>FAROI</i>	0.068 $\pm$ 0.002	0.348 $\pm$ 0.008	0.006 $\pm$ 0.000	0.167 $\pm$ 0.020	0.542 $\pm$ 0.024	0.021 $\pm$ 0.002
NMAE	<i>RUDUR</i>	0.158 $\pm$ 0.004	0.338 $\pm$ 0.014	0.057 $\pm$ 0.001	0.309 $\pm$ 0.009	0.596 $\pm$ 0.021	0.114 $\pm$ 0.002
	<i>FAMIS</i>	0.516 $\pm$ 0.004	0.788 $\pm$ 0.016	0.166 $\pm$ 0.003	0.619 $\pm$ 0.104	0.972 $\pm$ 0.086	0.241 $\pm$ 0.053
	<i>F_{PLS}</i>	0.289 $\pm$ 0.011	0.758 $\pm$ 0.025	0.063 $\pm$ 0.001	0.550 $\pm$ 0.034	1.251 $\pm$ 0.110	0.127 $\pm$ 0.004
	<i>S-BSS</i>	0.464 $\pm$ 0.308	0.721 $\pm$ 0.171	0.195 $\pm$ 0.171	0.962 $\pm$ 0.347	1.354 $\pm$ 0.259	0.391 $\pm$ 0.156
	<i>FAROI</i>	0.289 $\pm$ 0.004	0.884 $\pm$ 0.011	0.072 $\pm$ 0.001	0.452 $\pm$ 0.028	1.111 $\pm$ 0.031	0.137 $\pm$ 0.006

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples. *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

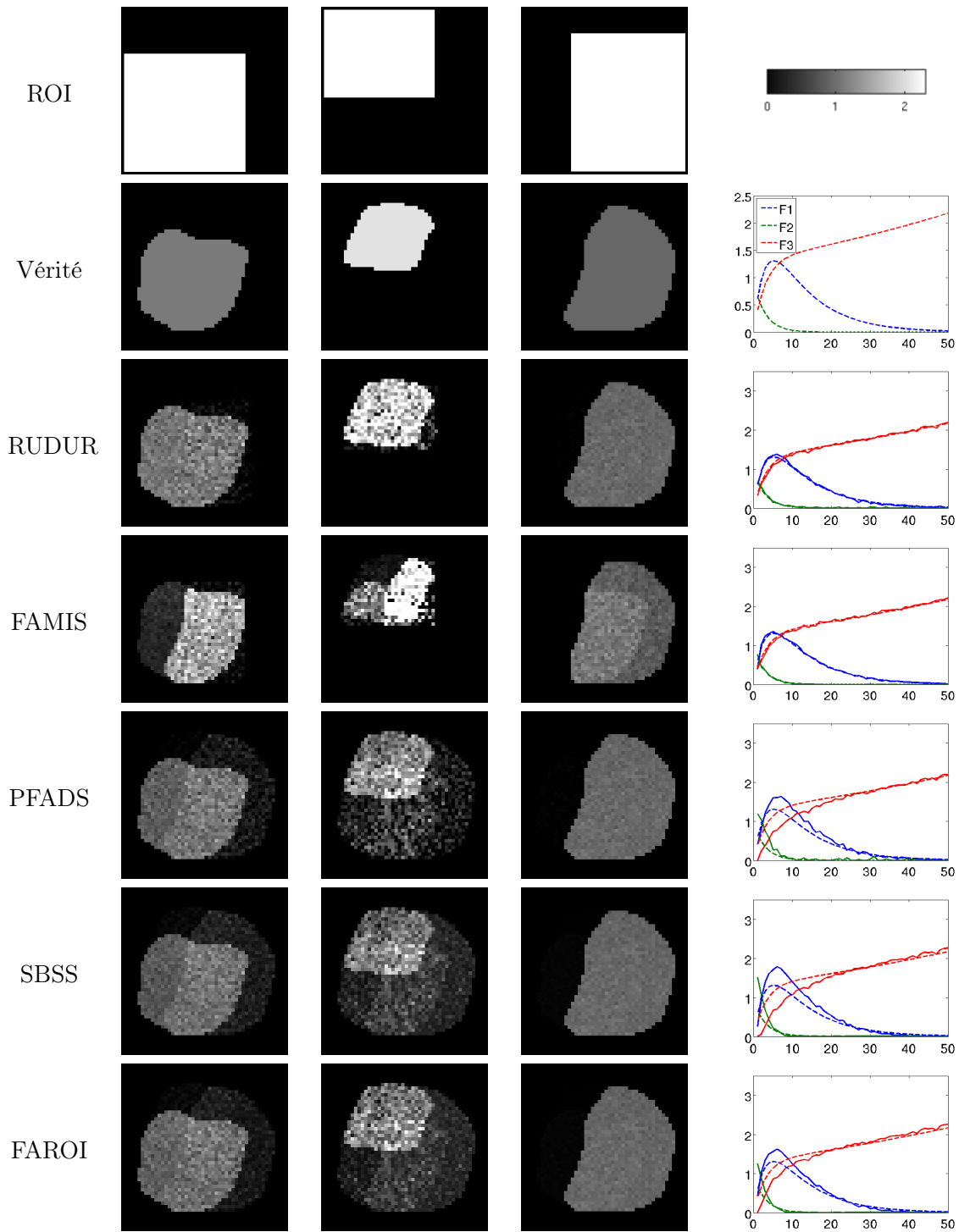


FIGURE 4.1 – Résultat des différentes méthodes sur une séquence du jeu de données simple, avec $\sigma = 0.5$. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS, et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tiret, avec le temps en abscisse, et l'activité en ordonnée.

moins bons résultats. Les connaissances a priori qu'elles utilisent ne sont pas assez informatives pour pouvoir faire face à la non-unicité du problème de séparation. Par ailleurs, la contrainte de non recouvrement des supports spatiaux de PLS-FADS n'est pas pertinente ici car les sources partagent de nombreux pixels en communs.

4.1.3 Ajout d'une source de bruit fond

Afin de complexifier les données précédentes, nous y avons ajouté une nouvelle source. Cette source s'assimile à un bruit de fond présent dans toute l'image, ce qui est commun en imagerie nucléaire. Ce bruit est assimilable à une source et son intensité dépend à la fois de la position et du temps. L'image factorielle et le facteur de ce bruit de fond sont disponibles dans la deuxième colonne de la figure 4.2. Un exemple de séquence d'images de ce jeu de données, avec $\sigma = 0.5$, est disponible en annexe D.2. Le protocole de test est identique au protocole utilisé dans le jeu de données précédent, à la différence près que les méthodes recherchent cette fois-ci $K = 4$ sources, la 4^{ème} source étant le bruit de fond. La région d'intérêt choisie pour le bruit de fond est l'image entière. Cette nouvelle source complexifie grandement la séparation, car une source supplémentaire à estimer est présente dans toute l'image et introduit ainsi du mélange dans tous les pixels où les autres sources sont présentes.

Les résultats obtenus sur une séquence sont présentés en figure 4.2, pour $\sigma = 0.5$. Pour toutes les méthodes, l'estimation des courbes d'activité est nettement moins bonne en présence de ce bruit de fond. Ces courbes sont davantage bruitées, en particulier pour les méthodes FAMIS et PLS-FADS qui ne comportent pas de terme régularisant. La quatrième source est mal estimée par toutes les méthodes. On retrouve dans l'image factorielle associée obtenue la structure de l'image factorielle de la troisième source. Avec RUDUR, tandis que les deux premières sources sont bien estimées, les troisième et quatrième sources obtenues sont mal séparées. Ceci est dû au fait que les régions d'intérêt choisies pour ces deux sources se ressemblent beaucoup, la connaissance de ces régions n'est donc pas d'une grande aide pour la séparation dans ce cas-ci.

Les tables de résultats contenant les moyennes des erreurs d'estimation sur 100 séquences d'images pour chaque niveau de bruit sont disponibles en figures 4.3, 4.4, 4.5 et 4.6. Du fait de l'addition d'une quatrième source qui recouvre toutes les autres, les erreurs d'estimation sont plus élevées que dans le cas plus simple de $K = 3$ sources. Les observations qualitatives faites précédemment sont cohérentes avec les indices quantitatifs présents dans ces tables. On peut noter que RUDUR obtient de meilleurs résultats pour l'estimation des courbes d'activité des troisième et quatrième source lorsque le bruit est fort ($\sigma = 1$), comparativement aux autres méthodes. De nouveau, FAMIS est plus sensible au bruit que RUDUR à cause de l'absence de terme régularisant.

FAROI et PLS-FADS obtiennent de mauvais résultats car ils n'utilisent pas de connaissances a priori assez informatives pour résoudre ce problème de séparation. Les connaissances a priori qu'ils utilisent s'avèrent de plus fausses, comme par exemple l'hypothèse de peu de recouvrement entre le support spatial des images factorielles faite par PLS-FADS. Aidée par

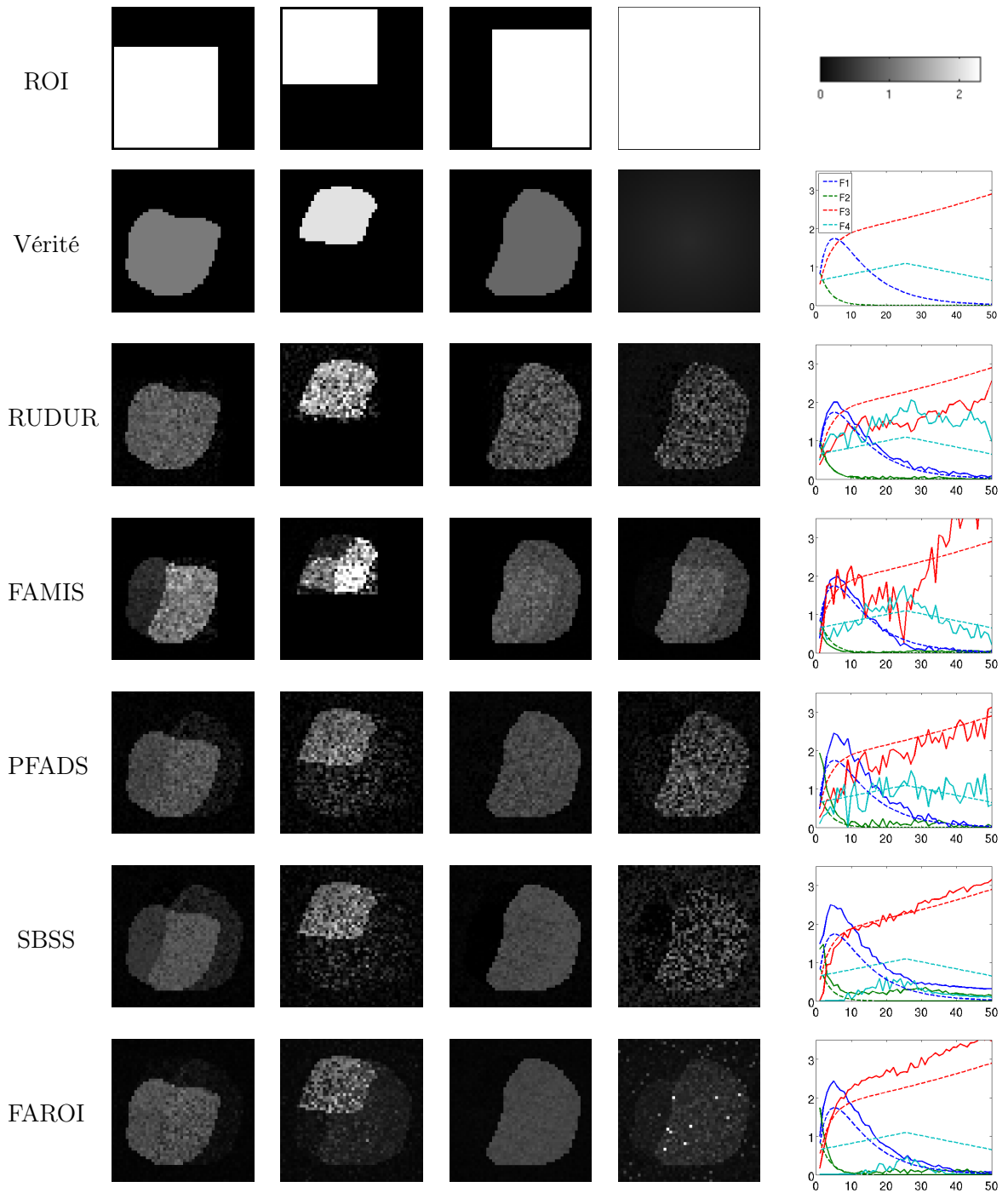


FIGURE 4.2 – Résultat des différentes méthodes sur une séquence du jeu de données simple avec bruit de fond, avec $\sigma = 0.5$. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS, et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 4) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tirets, avec le temps en abscisse, et l'activité en ordonnée.

TABLE 4.3 – Erreurs d’estimation des facteurs (F). Jeu de données simples, avec bruit de fond ($\sigma = 0.5$).

		sigma=0.5			
Mesure	Algorithme	F1	F2	F3	F4
NMSE	<i>RUDUR</i>	0.012 ± 0.004	0.113 ± 0.033	0.088 ± 0.024	0.485 ± 0.128
	<i>FAMIS</i>	0.033 ± 0.013	0.143 ± 0.031	0.117 ± 0.036	0.082 ± 0.010
	<i>PFADS</i>	0.167 ± 0.027	2.217 ± 0.296	0.036 ± 0.011	0.202 ± 0.063
	<i>S-BSS</i>	0.402 ± 0.119	2.308 ± 0.600	0.015 ± 0.012	0.708 ± 0.154
	<i>FAROI</i>	0.129 ± 0.019	0.932 ± 0.077	0.051 ± 0.003	0.843 ± 0.016
NMAE	<i>RUDUR</i>	0.121 ± 0.026	0.784 ± 0.147	0.258 ± 0.038	0.553 ± 0.084
	<i>FAMIS</i>	0.191 ± 0.042	0.727 ± 0.065	0.281 ± 0.043	0.236 ± 0.016
	<i>PFADS</i>	0.416 ± 0.039	2.647 ± 0.277	0.150 ± 0.022	0.370 ± 0.066
	<i>S-BSS</i>	0.658 ± 0.037	3.694 ± 0.465	0.095 ± 0.026	0.709 ± 0.021
	<i>FAROI</i>	0.375 ± 0.036	1.679 ± 0.068	0.216 ± 0.008	0.824 ± 0.006

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l’écart-type de la mesure d’erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples avec bruit de fond ($\sigma = 0.5$). *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

TABLE 4.4 – Erreurs d’estimation des images factorielles (A). Jeu de données simples, avec bruit de fond ($\sigma = 0.5$).

		sigma=0.5			
Mesure	Algorithme	A1	A2	A3	A4
NMSE	<i>RUDUR</i>	0.049 ± 0.005	0.173 ± 0.014	0.234 ± 0.036	1.780 ± 0.280
	<i>FAMIS</i>	0.252 ± 0.006	0.597 ± 0.027	0.061 ± 0.003	1.396 ± 0.014
	<i>PFADS</i>	0.103 ± 0.048	0.344 ± 0.013	0.044 ± 0.007	1.190 ± 0.089
	<i>S-BSS</i>	0.603 ± 0.401	0.632 ± 0.322	0.025 ± 0.003	1.242 ± 0.200
	<i>FAROI</i>	0.079 ± 0.002	0.389 ± 0.009	0.032 ± 0.001	1.094 ± 0.008
NMAE	<i>RUDUR</i>	0.208 ± 0.009	0.407 ± 0.016	0.596 ± 0.076	1.122 ± 0.111
	<i>FAMIS</i>	0.456 ± 0.006	0.676 ± 0.013	0.188 ± 0.005	1.081 ± 0.007
	<i>PFADS</i>	0.393 ± 0.057	0.891 ± 0.021	0.269 ± 0.012	0.851 ± 0.027
	<i>S-BSS</i>	0.869 ± 0.413	1.216 ± 0.371	0.197 ± 0.012	0.933 ± 0.082
	<i>FAROI</i>	0.375 ± 0.004	1.023 ± 0.012	0.264 ± 0.002	0.933 ± 0.004

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l’écart-type de la mesure d’erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples avec bruit de fond ($\sigma = 0.5$). *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

TABLE 4.5 – Erreurs d'estimation des facteurs (F). Jeu de données simples, avec bruit de fond ($\sigma = 1$).

		sigma=1			
Mesure	Algorithme	F1	F2	F3	F4
NMSE	<i>RUDUR</i>	0.035 ± 0.007	0.360 ± 0.073	0.088 ± 0.016	0.471 ± 0.089
	<i>FAMIS</i>	0.220 ± 0.295	0.476 ± 0.278	0.421 ± 0.195	0.645 ± 0.158
	<i>PFADS</i>	0.540 ± 0.069	26.72 ± 4.569	0.185 ± 0.050	0.711 ± 0.123
	<i>S-BSS</i>	0.540 ± 0.249	8.650 ± 3.641	0.125 ± 0.054	0.360 ± 0.125
	<i>FAROI</i>	0.030 ± 0.056	2.136 ± 0.343	0.091 ± 0.115	0.637 ± 0.041
NMAE	<i>RUDUR</i>	0.187 ± 0.022	1.246 ± 0.165	0.244 ± 0.025	0.536 ± 0.055
	<i>FAMIS</i>	0.404 ± 0.239	1.181 ± 0.233	0.541 ± 0.135	0.722 ± 0.111
	<i>PFADS</i>	0.664 ± 0.054	8.952 ± 1.072	0.358 ± 0.065	0.693 ± 0.055
	<i>S-BSS</i>	0.775 ± 0.209	6.418 ± 1.635	0.317 ± 0.080	0.529 ± 0.089
	<i>FAROI</i>	0.148 ± 0.085	2.676 ± 0.260	0.261 ± 0.087	0.758 ± 0.018

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples avec bruit de fond ($\sigma = 1$). *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

TABLE 4.6 – Erreurs d'estimation des images factorielles (A). Jeu de données simples, avec bruit de fond ($\sigma = 1$).

		sigma=1			
Mesure	Algorithme	A1	A2	A3	A4
NMSE	<i>RUDUR</i>	0.156 ± 0.015	0.445 ± 0.029	0.234 ± 0.022	1.953 ± 0.096
	<i>FAMIS</i>	0.811 ± 1.117	1.133 ± 0.365	0.222 ± 0.429	3.132 ± 2.408
	<i>PFADS</i>	0.285 ± 0.046	0.727 ± 0.112	0.090 ± 0.030	1.157 ± 0.067
	<i>S-BSS</i>	0.519 ± 0.228	0.922 ± 0.197	0.116 ± 0.065	1.232 ± 0.184
	<i>FAROI</i>	0.272 ± 0.016	0.593 ± 0.055	0.034 ± 0.010	0.854 ± 0.378
NMAE	<i>RUDUR</i>	0.376 ± 0.016	0.670 ± 0.023	0.574 ± 0.027	1.135 ± 0.014
	<i>FAMIS</i>	0.676 ± 0.297	0.891 ± 0.091	0.329 ± 0.242	1.267 ± 0.107
	<i>PFADS</i>	0.644 ± 0.050	1.342 ± 0.107	0.347 ± 0.044	0.879 ± 0.027
	<i>S-BSS</i>	0.870 ± 0.233	1.517 ± 0.201	0.355 ± 0.094	0.853 ± 0.052
	<i>FAROI</i>	0.691 ± 0.032	1.286 ± 0.079	0.237 ± 0.034	0.619 ± 0.180

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples avec bruit de fond ($\sigma = 1$). *S-BSS* correspond à la méthode *S-BSS-vecDC* et *PFADS* à la méthode *PLS-FADS*.

l'exploitation de l'hypothèse que les données sont générées à partir d'un modèle compartimental, S-BSS-vecDC réussit à obtenir ici de bons résultats.

4.2 Rénographies synthétiques réalistes

4.2.1 Présentation des données

Le jeu de données de cette section est entièrement détaillé dans [BGL13]. Il est composé de simulations réalistes de rénographies utilisant du radiopharmaceutique ^{99m}Tc -MAG3. Ces simulations se basent sur un modèle compartimental complexe représentant la cinétique du ^{99m}Tc -MAG3 dans le corps, et plus spécifiquement dans le rein. En imposant une fonction d'entrée, le modèle compartimental permet d'obtenir la courbe d'activité de chaque compartiment physiologique considéré (matrice F). Les hypothèses de la méthode S-BSS-vecDC sont donc bien compatibles avec ce jeu de données.

Le modèle compartimental est couplé au fantôme 3D XCAT [Seg+10]. Ce fantôme donne les localisations spatiales des différents compartiments physiologiques, ce qui permet d'obtenir la matrice A des images factorielles. Quelques changements sont opérés sur le fantôme XCAT initial, afin d'avoir une zone rénale la plus réaliste possible [BGL13].

Le modèle compartimental et le fantôme permettent de connaître la concentration en tout point et en tout temps du radiopharmaceutique. Ces informations sont ensuite transmises à SIMIND, un programme permettant de simuler l'acquisition d'une gamma-caméra [Lju12]. Ce programme prend notamment en compte l'aléatoire de la radioactivité, l'atténuation des photons, l'effet Compton, et la collimation. Au final, les séquences créées sont des rénographies réalistes, pour lesquelles il est possible de faire varier des paramètres physiologiques comme l'absorption rénale relative (entre rein gauche et rein droit) et le temps de transit rénal en modifiant les cinétiques du modèle compartimental. Un exemple de séquence d'images de ce jeu de données est disponible en annexe D.3.

4.2.2 Résultats et analyse

Six séquences d'images ont été produites à l'aide du protocole décrit ci-dessus. Pour chacune de ces séquences, l'absorption rénale relative et/ou le temps de transit rénal sont différents, ceci afin d'explorer plusieurs scénarios. Dans chaque séquence, deux sous-séquences sont extraites, une pour chaque rein. Ces sous-séquences sont des zooms autour des zones rénales gauche et droite. Elles sont de taille 26 par 21 pixels et contiennent 60 images. L'essentiel de l'activité présente dans chacune des sous-séquences est dû à 3 compartiments physiologiques différents qui sont des subdivisions du rein : le cortex rénal, la médullaire rénale, et le pelvis. Dans le modèle utilisé pour générer ces données, chacun de ces compartiments physiologiques possède une carte de présence et une courbe d'activité. Le but est donc de retrouver pour chacune de ces 12 sous-séquences les matrices A et F à l'origine de ces données. Les erreurs

TABLE 4.7 – Erreurs d’estimation des facteurs (F) - Rénographie réaliste

Measure	Algorithm	Rein droit			Rein gauche		
		F1	F2	F3	F1	F2	F3
NMSE	<i>RUDUR</i>	0.033 ± 0.007	0.037 ± 0.005	0.022 ± 0.013	0.063 ± 0.020	0.033 ± 0.008	0.046 ± 0.030
	<i>FAMIS</i>	0.056 ± 0.006	0.014 ± 0.005	0.400 ± 0.161	0.078 ± 0.008	0.051 ± 0.093	0.199 ± 0.037
	<i>PFADS</i>	0.120 ± 0.042	0.148 ± 0.104	0.043 ± 0.028	0.110 ± 0.048	0.028 ± 0.009	0.061 ± 0.026
	<i>S-BSS</i>	0.037 ± 0.020	0.112 ± 0.117	0.176 ± 0.068	0.099 ± 0.086	0.205 ± 0.333	0.155 ± 0.068
	<i>FAROI</i>	0.295 ± 0.235	0.055 ± 0.013	0.208 ± 0.064	0.231 ± 0.161	0.059 ± 0.022	0.246 ± 0.073
NMAE	<i>RUDUR</i>	0.130 ± 0.017	0.130 ± 0.008	0.120 ± 0.029	0.191 ± 0.034	0.128 ± 0.016	0.176 ± 0.065
	<i>FAMIS</i>	0.170 ± 0.011	0.096 ± 0.221	0.442 ± 0.096	0.210 ± 0.016	0.159 ± 0.144	0.326 ± 0.036
	<i>PFADS</i>	0.271 ± 0.048	0.317 ± 0.133	0.146 ± 0.039	0.265 ± 0.050	0.110 ± 0.018	0.177 ± 0.043
	<i>S-BSS</i>	0.167 ± 0.066	0.211 ± 0.133	0.301 ± 0.080	0.274 ± 0.133	0.281 ± 0.291	0.288 ± 0.086
	<i>FAROI</i>	0.378 ± 0.138	0.149 ± 0.024	0.331 ± 0.054	0.386 ± 0.128	0.155 ± 0.040	0.376 ± 0.058

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l’écart-type de la mesure d’erreurs (NMSE ou NMAE) sur 6 séquences simulées, pour chaque rein (gauche et droit). Chaque séquence possède une absorption rénale relative et/ou un temps de transit rénal différent. F1, F2 et F3 correspondent respectivement aux courbes d’activité du cortex, de la médullaire rénale et du pelvis.

d’estimation de A et de F des différentes méthodes de séparation sont détaillées dans les tables 4.7 et 4.8. Un exemple de courbes d’activité et de cartes de présence obtenues pour une sous-séquence de rein droit est disponible en figure 4.3. Les mêmes ROIs sont utilisées pour RUDUR et FAMIS, elles sont disponibles dans la première ligne de cette figure.

TABLE 4.8 – Erreurs d’estimation des images factorielles (A) - Rénographie réaliste

Measure	Algorithm	Rein droit			Rein gauche		
		A1	A2	A3	A1	A2	A3
NMSE	<i>RUDUR</i>	0.197 ± 0.051	0.106 ± 0.043	0.099 ± 0.041	0.275 ± 0.069	0.109 ± 0.027	0.261 ± 0.191
	<i>FAMIS</i>	0.737 ± 0.293	1.019 ± 0.106	1.487 ± 0.091	0.331 ± 0.045	0.842 ± 0.106	2.514 ± 0.479
	<i>PFADS</i>	0.195 ± 0.082	0.101 ± 0.034	0.427 ± 0.060	0.221 ± 0.087	0.091 ± 0.031	0.628 ± 0.139
	<i>S-BSS</i>	0.450 ± 0.180	0.255 ± 0.382	0.511 ± 0.094	0.518 ± 0.256	0.319 ± 0.499	0.681 ± 0.132
	<i>FAROI</i>	0.322 ± 0.095	0.142 ± 0.067	0.550 ± 0.199	0.337 ± 0.155	0.179 ± 0.092	0.797 ± 0.201
NMAE	<i>RUDUR</i>	0.473 ± 0.070	0.341 ± 0.055	0.427 ± 0.063	0.597 ± 0.094	0.358 ± 0.045	0.586 ± 0.124
	<i>FAMIS</i>	0.755 ± 0.106	0.869 ± 0.020	1.091 ± 0.008	0.519 ± 0.031	0.800 ± 0.012	1.197 ± 0.018
	<i>PFADS</i>	0.472 ± 0.117	0.371 ± 0.069	0.890 ± 0.084	0.522 ± 0.113	0.357 ± 0.075	1.129 ± 0.137
	<i>S-BSS</i>	0.707 ± 0.166	0.540 ± 0.419	1.109 ± 0.148	0.772 ± 0.229	0.604 ± 0.506	1.240 ± 0.163
	<i>FAROI</i>	0.599 ± 0.091	0.459 ± 0.120	1.092 ± 0.185	0.613 ± 0.143	0.517 ± 0.142	1.276 ± 0.157

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l’écart-type de la mesure d’erreurs (NMSE ou NMAE) sur 6 séquences simulées, pour chaque rein (gauche et droit). Chaque séquence possède une absorption rénale relative et/ou un temps de transit rénal différent. A1, A2 et A3 correspondent respectivement aux images factorielles du cortex, de la médullaire rénale et du pelvis.

Sur ces données réalistes, les résultats montrent une nette amélioration de l’estimation des courbes d’activité et des cartes de présence avec RUDUR. C’est la seule méthode qui obtient des erreurs d’estimation faibles pour toutes les courbes d’activité et toutes les cartes de présence. Toutes les autres méthodes estiment mal au moins une des sources présentes.

D’un point de vue qualitatif, contrairement à FAMIS qui n’utilise pas de régularisation, les courbes d’activité obtenues avec RUDUR sont relativement lisses et ne comportent pas de

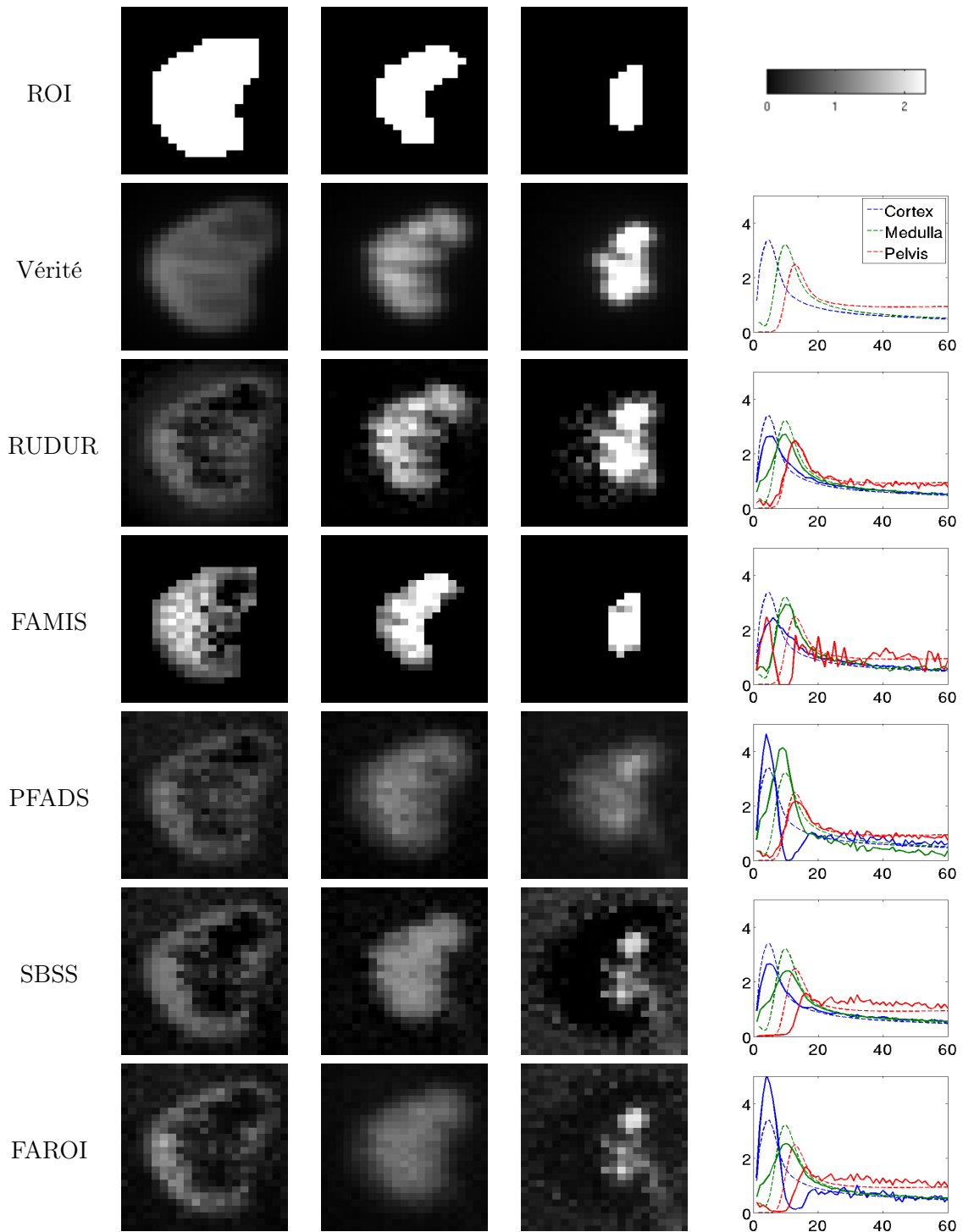


FIGURE 4.3 – Résultat de la séparation sur une séquence du jeu de données de simulations de rénographies réalistes (rein droit). Les trois premières colonnes correspondent dans l'ordre au cortex rénal, à la médullaire rénale et au pelvis. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS, et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tirets, avec le temps en abscisse, et l'activité en ordonnée.

sauts brusques d'activité. Les méthodes n'utilisant pas les régions d'intérêt trouvent des images factorielles qui sont spatialement trop étendues et bruitées sur les bords. RUDUR et FAMIS n'ont pas ce problème. Cependant, l'utilisation des ROIs dans FAMIS est trop restrictive et n'est pas robuste à leur choix. Ceci se traduit par une mauvaise estimation des mélanges dans les pixels périphériques à ces ROI, où les images factorielles sont censées avoir des valeurs faibles mais non nulles. Comme FAMIS force à 0 ces coefficients, les images factorielles ainsi que les courbes d'activité sont mal estimées. Dans RUDUR, les ROIs sont introduites de manière plus souple, et les images factorielles et courbes d'activité obtenues sont fidèles à celles de la vérité-terrain.

4.3 Rénographies cliniques

4.3.1 Présentation des données

Le jeu de données utilisé ici est constitué de données cliniques. Ce sont des rénographies composées chacune de 180 images, où chaque image représente l'activité mesurée pendant un intervalle de temps de 10 secondes. Ces données sont disponibles en ligne¹. Un exemple de séquence d'images de ce jeu de données est disponible en annexe D.4. Comme il s'agit de données réelles, il n'y a pas de vérité terrain pour les matrices A et F . Cependant, les courbes d'activité du cortex rénal peuvent être calculées à partir des séquences. En effet, certains pixels ne contiennent presque que du cortex rénal. Ils sont situés sur la périphérie extérieure du rein. Afin de comparer les méthodes de manière quantitative, nous avons donc extrait la courbe d'activité du cortex rénal, à l'aide de la procédure recommandée en utilisant une image sommée [GPS11] (cf deuxième ligne de la figure 4.5). Même si cette courbe d'activité de référence n'est pas parfaite, elle permet de faire des comparaisons. Les erreurs d'estimation rapportées dans la table 4.9 sont celles obtenues en comparant cette courbe d'activité de référence et celles obtenues par les méthodes de séparation. Comme pour les rénographies synthétiques réalistes, les méthodes de séparation sont utilisées sur des sous-séquences d'images ne comportant que la zone rénale. Le nombre de sources en entrée des algorithmes est fixé à $K = 3$.

4.3.2 Résultats et analyse

Afin de s'assurer d'obtenir des courbes d'activité de référence fiables, nous avons présélectionné uniquement les séquences dans lesquelles le cortex rénal était suffisamment visible pour que la courbe d'activité puisse être extraite correctement. 19 séquences d'images ont donc été sélectionnées (les séquences 1, 2, 8, 16, 21, 28, 29, 32, 37, 38, 39, 45, 58, 59, 74, 76, 89, 93 et 94). Des tests ont également été réalisés sur les autres séquences, mais la sélection de la région d'intérêt était trop complexe et trop peu fiable pour être utilisée à des fins de comparaisons. Les erreurs d'estimation moyennes réalisées sur les séquences sélectionnées sont rapportées dans la table 4.9. Elles sont calculées en moyennant les erreurs d'estimation de 38 courbes

1. Database of dynamic renal scintigraphy - www.dynamicrenalstudy.org

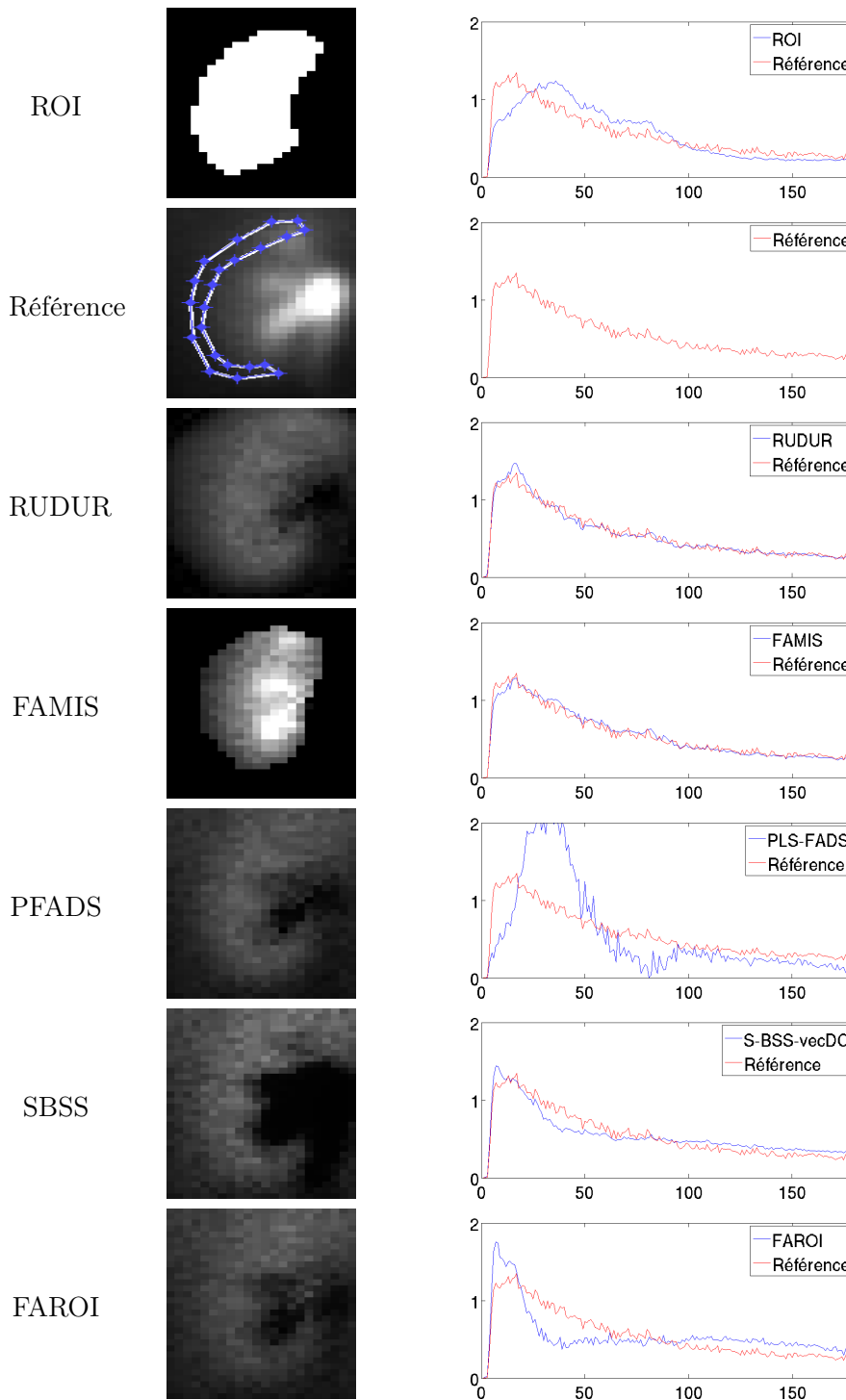


FIGURE 4.4 – Résultat de la séparation sur une séquence du jeu de données de rénographies cliniques (rein droit). Sur la seconde ligne est présentée l'image sommée ainsi que la région d'intérêt (polygone blanc avec sommets en bleu) utilisée pour obtenir la courbe d'activité de référence du cortex rénal située sur la deuxième colonne. La région d'intérêt du cortex rénal utilisé pour RUDUR et FAMIS est affichée dans la première ligne. La courbe d'activité moyenne de cette région est disponible dans la seconde colonne. Les lignes 3 à 7 contiennent les images factorielles du cortex rénal obtenues par chacune des méthodes (colonne 1), ainsi que les courbes d'activité associées (colonne 2) avec le temps en abscisse et l'activité en ordonnée.

TABLE 4.9 – Erreurs d’estimation sur les courbes d’activité de 38 cortex rénaux - Données cliniques

Méthode	NMSE	NMAE
<i>RUDUR</i>	0.0119 ± 0.0219	0.0762 ± 0.0430
<i>FAMIS</i>	0.0207 ± 0.0369	0.1004 ± 0.0745
<i>F_{PLS}</i>	0.0522 ± 0.0679	0.1523 ± 0.1103
<i>S-BSS-vecDC</i>	0.0128 ± 0.0085	0.0875 ± 0.0264
<i>FAROI</i>	0.0710 ± 0.0818	0.1804 ± 0.1088

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l’écart-type de la mesure d’erreurs (NMSE ou NMAE) sur 38 courbes d’activités de cortex rénaux. Les erreurs sont mesurées pas rapport à la courbe d’activité du cortex rénal obtenue en moyennant l’activité dans une région d’intérêt tracée selon les procédures recommandées [GPS11].

d’activité de cortex rénal (2 par séquence, rein gauche et droit). Ces résultats doivent toutefois être analysés avec précaution puisque les courbes d’activité utilisées comme références ne sont pas parfaites. Un exemple de courbes et d’images factorielles obtenues est disponible en figure 4.5. Une fois encore, les méthodes utilisant les régions d’intérêt fournissent de bonnes performances. Les méthodes donnant les estimations les plus proches des courbes d’activité de référence sont RUDUR et S-BSS-vecDC, suivies par FAMIS. Les autres méthodes présentent des erreurs d’estimation plus importantes sur la courbe d’activité du cortex rénal. On constate que les images factorielles estimées à l’aide des méthodes n’utilisant pas les régions d’intérêt ont des valeurs importantes sur les pixels du bord de l’image alors que ces pixels contiennent vraisemblablement très peu d’activité liée au cortex rénal. RUDUR n’a pas ce problème grâce à la relaxation du terme de fidélité aux données, les images factorielles estimées ne sont que très peu impactées par les pixels du bord de l’image. FAMIS n’est pas non plus sensible à ce problème grâce à son utilisation non permissive des régions d’intérêt. Sur la première ligne de la figure 4.5 est tracée la ROI utilisée par RUDUR et FAMIS, ainsi que la courbe d’activité moyenne de cette ROI. Cette courbe d’activité est significativement différente de la courbe d’activité de référence, ce qui montre la présence de mélange dans cette ROI. L’utilisation d’une méthode de séparation comme RUDUR, FAMIS ou S-BSS-vecDC, permet d’améliorer significativement l’estimation de la courbe d’activité du cortex rénal comparativement à celle obtenue avec la ROI.

Les courbes de parenchyme obtenues sur les séquences non-sélectionnées pour la comparaison ont une cinétique assez similaire à celles présentées en figure 4.5. La principale différence est que la décroissance de la courbe est beaucoup plus lente, ce qui montre un dysfonctionnement dans la filtration du traceur. Un exemple de courbe obtenue est disponible en figure ?? (la séquence d’image associée est disponible en annexe D.5).

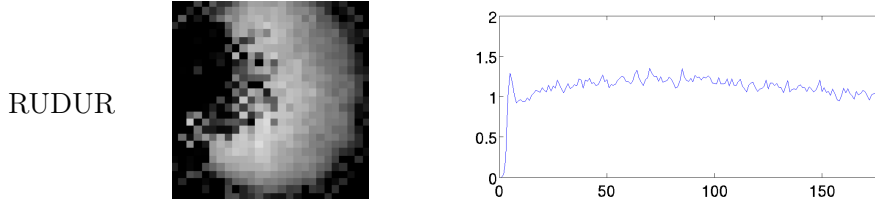


FIGURE 4.5 – Résultat de la séparation sur une des séquences non retenues du jeu de données de rénographies cliniques (rein gauche) avec RUDUR. La courbe d’activité estimée du parenchyme est sur la droite, et l’image factorielle associée sur la gauche.

4.4 Évaluation de la sensibilité de la méthode

Afin d’utiliser RUDUR, il est nécessaire de fixer au préalable les paramètres α , β et γ de la fonctionnelle. Des régions d’intérêt doivent également être choisies. Dans cette section, la robustesse de RUDUR au choix des paramètres et des régions d’intérêt est évaluée. Des tests sont également effectués pour observer l’effet de la présence ou de l’absence d’un des termes de la fonctionnelle utilisée par RUDUR (f_{ROI} et f_{Reg}).

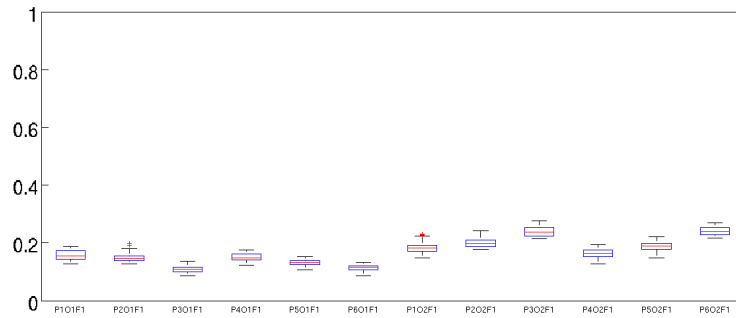
4.4.1 Sensibilité au choix des paramètres

La fonctionnelle à optimiser dans RUDUR comporte trois paramètres à fixer. Les constantes α et β qui pondèrent les différents termes de la fonctionnelle, et la constante γ qui pondère l’importance des pixels en dehors des régions d’intérêt dans la séparation. Il est très complexe de déterminer de manière automatique des valeurs pour ces paramètres. Cela nécessite d’avoir connaissance de nombreuses caractéristiques des sources (intensité, mélange ...) et une très bonne caractérisation des erreurs de modélisation, ce que nous ne sommes pas en mesure de faire. Afin d’évaluer la sensibilité de RUDUR au choix de ces paramètres, nous avons testé différents jeux de paramètres sur des séquences d’images identiques. Celles de rénographies réalistes ont été choisies afin d’avoir à la fois une vérité terrain et des données proches de la réalité.

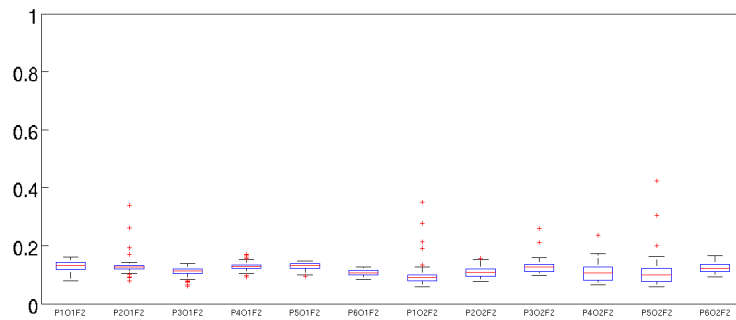
Pour ce test, le paramètre α peut prendre 6 valeurs possibles (1, 2, 5, 10, 20, 50). Le paramètre β peut prendre quatre valeurs différentes (1, 2, 5, 10). Le paramètre γ peut prendre 4 valeurs différentes (0.5, 1, 3, 5). En prenant en compte toutes les possibilités, il y a donc 96 jeux de paramètres différents. Chacune des 12 sous-séquences de rénographies réalistes a été traitée par RUDUR avec ces 96 jeux de paramètres possibles. Les résultats obtenus sont présentés dans la figure 4.6, où chacun des trois graphiques correspond à la distribution des erreurs d’estimation (NMAE) d’un des trois facteurs. Sont représentés dans ces graphiques les premier, deuxième et troisième quartiles de ces distributions, ainsi que les valeurs minimales et maximales et les données aberrantes, pour chacune des sous-séquences.

Les seules estimations vraiment sensibles au choix des paramètres sont les estimations du facteur du pelvis du rein gauche de quelques simulations (PxO2F3 dans le troisième gra-

(a) Distribution de l'écart moyen (NMAE) sur la courbe d'activité du cortex rénal (F1)



(b) Distribution de l'écart moyen (NMAE) sur la courbe d'activité de la médullaire (F2)



(c) Distribution de l'écart moyen (NMAE) sur la courbe d'activité du pelvis (F3)

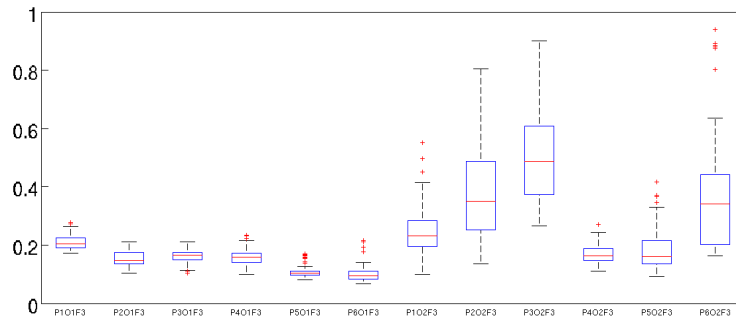


FIGURE 4.6 – Robustesse aux paramètres - Distribution des erreurs d'estimation (NMAE) des courbes d'activité du cortex rénal (a), de la médullaire rénale (b) et du pelvis (c) pour chaque sous-séquence des données synthétiques réalistes. Pour construire ces distributions, 96 jeux de paramètres différents ont été utilisés ($\alpha \in \{1, 2, 5, 10, 20, 50\}$, $\beta \in \{1, 2, 5, 10\}$, $\gamma \in \{0.5, 1, 3, 5\}$). L'axe des ordonnées correspond à l'erreur d'estimation (NMAE) de la courbe d'activité considérée. Chaque boîte correspond à la distribution de l'erreur d'estimation de la courbe d'activité d'une des sous-séquences. Chaque boîte résume ainsi l'erreur d'estimation pour l'ensemble des 96 jeux de paramètres possibles. Plus la boîte et les moustaches associées sont grandes, plus la méthode est sensible aux choix des paramètres. Sur l'axe des abscisses, les noms du type $PxOyFz$ correspondent à la $z^{\text{ème}}$ courbe d'activité issue du $y^{\text{ème}}$ rein de la $x^{\text{ème}}$ simulation. Les boîtes à moustaches sont générées à partir de la fonction *boxplot* de MATLAB. La ligne rouge correspond à la médiane, et les extrémités de la boîte bleue aux premier et troisième quartiles. Les extrémités de la ligne en tirets noire correspondent aux valeurs minimales et maximales, et les croix rouges aux données aberrantes.

phique). On constate que les autres distributions des erreurs d'estimation sont pour la plupart concentrées autour de la médiane et contiennent très peu de données aberrantes. Étant donnée la grande diversité dans les jeux de paramètres possibles (le minimum et le maximum possibles de chaque paramètre ont un rapport d'au moins 10), ceci indique une faible sensibilité de RUDUR au choix des paramètres sur ces séquences. Il n'est donc pas nécessaire de trouver ici des paramètres optimaux pour utiliser RUDUR. Pour un type d'étude donnée, il suffit d'avoir une idée de l'ordre de grandeur de chaque paramètre. Ces ordres de grandeurs peuvent être estimés par exemple sur des données synthétiques proches de l'étude en question.

4.4.2 Impact de la pénalisation des images factorielles

Nous montrons dans cette section l'intérêt du terme f_{ROI} dans la fonctionnelle utilisée par RUDUR. Pour ce faire, des tests ont été réalisés en faisant varier le paramètre α qui pondère l'importance de ce terme dans la séparation, les autres paramètres étant restés à leurs valeurs de base ($\beta = 10$, $\gamma = 3$). Les résultats obtenus sur le jeu de données synthétiques simples sont présentés en figure 4.7. Lorsque α est égal à 0, les erreurs moyennes (NMAE) obtenues sont beaucoup plus importantes que lorsque α n'est pas nul. Lorsque $\alpha = 0$, le terme f_{ROI} ne joue plus aucun rôle dans la séparation, ainsi, les images factorielles ne sont plus du tout pénalisées et les régions d'intérêt ne sont utilisées que pour l'initialisation. Ceci n'est pas suffisant pour bien séparer les sources présentes dans ces données du fait de la sous-détermination du problème. On remarque également dans la figure 4.7, la robustesse de RUDUR au choix du paramètre α . Des variations de la valeur de ce paramètre entre 0.2 et 50 ne changent que peu les erreurs d'estimation moyennes obtenues.

Le terme f_{ROI} est donc fondamental dans la fonctionnelle de RUDUR car il permet d'introduire des a priori fort sur les images factorielles. L'absence de ce terme entraîne des séparations très mauvaises car RUDUR n'est plus en mesure de faire face à la sous-détermination du problème. Ceci se constate également dans la figure 4.8 qui montre des exemples d'images factorielles et de courbes d'activité obtenues avec et sans le terme f_{ROI} . Quand ce dernier est absent, les images factorielles estimées sont encore mélangées, entraînant de mauvaises estimations des courbes d'activité.

4.4.3 Impact de la régularisation

Cette section est dédiée à l'étude de l'impact de la régularisation. Afin d'évaluer l'effet de cette régularisation sur la séparation, RUDUR a été utilisé sur les données synthétiques rénales réalistes en faisant varier le paramètre β qui pondère l'importance de la régularisation dans la séparation. Un exemple de résultats obtenus sur une séquence avec $\beta = 0$, $\beta = 10$ et $\beta = 50$ est disponible en figure 4.9.

Conformément aux attentes, plus la régularisation est importante, plus les courbes d'activité obtenues sont lisses. Pour $\beta = 0$, la courbe d'activité estimé du pelvis est assez bruitée, tandis que pour $\beta = 50$, l'estimation est beaucoup plus lisse. Ce lissage a cependant aussi

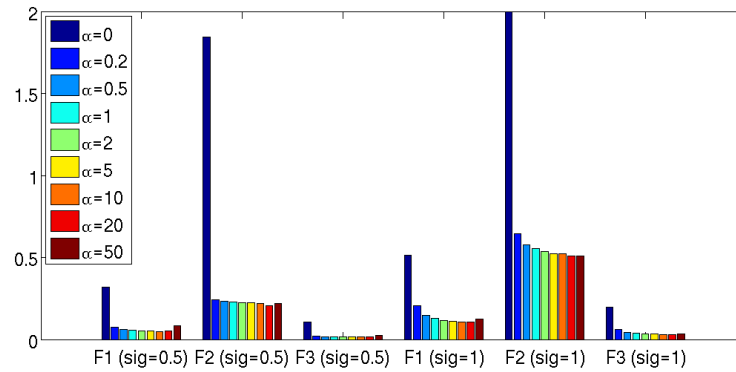


FIGURE 4.7 – Importance de la pénalisation f_{ROI} - Erreurs d'estimation moyennes (NMAE) des courbes d'activité du jeu de données synthétiques simples, pour différentes valeurs de α . Chaque bâton est la moyenne de l'erreur d'estimation obtenue par RUDUR pour une courbe d'activité particulière et un α particulier sur 100 simulations du jeu de données synthétiques simples (avec $\sigma = 0.5$ ou $\sigma = 1$). Les valeurs des autres paramètres sont les mêmes pour tous les tests : $\beta = 100$ et $\gamma = 3$. Lorsque $\alpha = 0$, le terme f_{ROI} n'a plus aucun poids dans la séparation. Dans ce cas-ci, la séparation est très mauvaise et les courbes d'activité mal estimées du fait du manque de connaissances a priori utilisées.

des conséquences sur les pics des courbes d'activité. Ceux-ci sont en général sous-évalués et donc mal estimés lorsque la régularisation est trop forte ($\beta = 50$), car ils introduisent de trop grandes variations dans le signal. Les moyennes des erreurs d'estimation des courbes d'activité (NMAE) pour différentes valeurs de β et pour différents compartiments physiologiques sont relevées dans la figure 4.10. Ces résultats montrent que la régularisation raisonnable est bénéfique pour l'estimation des courbes d'activité du pelvis et de la médullaire rénale. La dégradation des résultats pour le cortex rénal est due à la mauvaise estimation des pics d'activité comme montré dans la figure 4.9.

Une solution pour corriger le problème du lissage du transitoire est de ne pénaliser les variations des courbes d'activité que sur certains intervalles temporels où ces courbes sont connues comme variant peu. Ceci peut se faire grâce à la matrice S qui introduit des connaissances sur les aspects temporels des courbes. Dans la figure 4.11 sont représentés les résultats obtenus de la même manière que dans la figure 4.9, mais cette fois-ci en ne pénalisant les variations dans les courbes d'activité qu'à partir de la vingtième image ($S_{t,t} = 0$ pour $t \leq 20$). Cette fois-ci le lissage a toujours un effet débruitant, mais ne perturbe plus l'estimation des pics d'activité qui sont conservés, même pour $\beta = 50$. Les moyennes des erreurs d'estimation des courbes d'activités (NMAE) pour différentes valeurs de β reportées dans la figure 4.12 montrent une nette amélioration par rapport au cas sans pondération. Cette fois-ci, une valeur forte de régularisation ne diminue les performance de séparation sur aucun des compartiments physiologiques. La courbe d'activité du cortex n'est que très peu affectée par la régularisation car elle est naturellement peu bruitée, du fait de la présence importante de ce compartiment dans la séquence d'images. La régularisation bénéficie en revanche à la médullaire rénale et au pelvis, car ces compartiments sont de petite taille dans la séquence, et donc sans régularisation leur

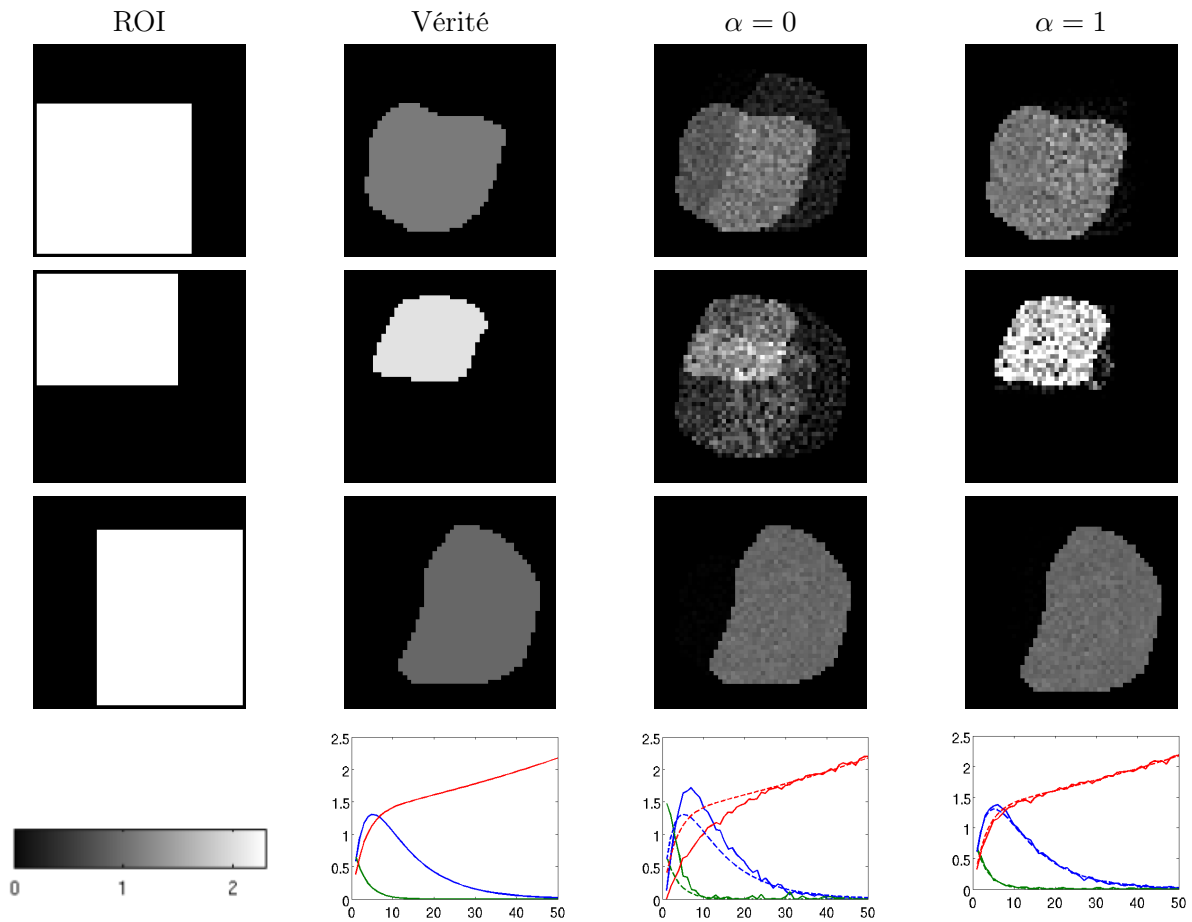


FIGURE 4.8 – Importance de la pénalisation f_{ROI} - Exemple d'images factorielles et de courbes d'activité obtenues par RUDUR avec et sans le terme f_{ROI} sur le jeu de données synthétiques simple ($\sigma = 0.5$). La première colonne contient les régions d'intérêt utilisées par RUDUR pour chaque source. La seconde colonne contient la vérité terrain des images factorielles et des courbes d'activité. Les troisième et quatrième colonnes contiennent les images factorielles et les courbes d'activité obtenues par RUDUR respectivement avec et sans le terme f_{ROI} .

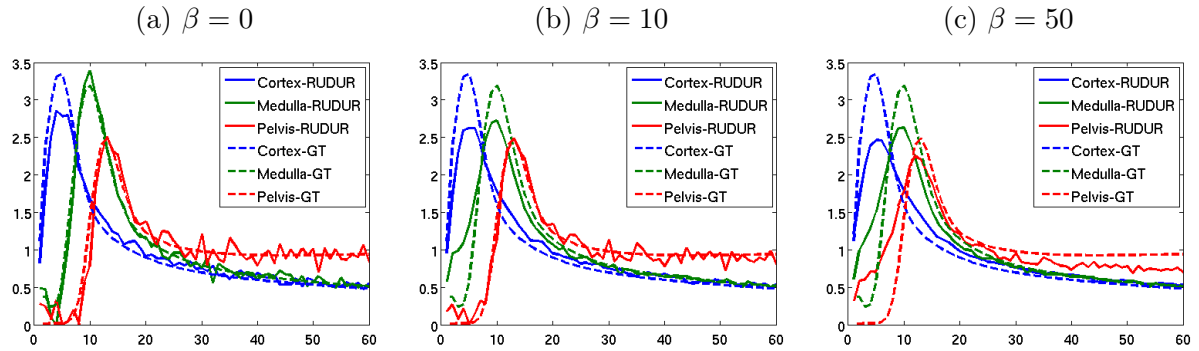


FIGURE 4.9 – Importance de la régularisation - Exemple de courbes d'activité obtenues avec RUDUR sur une séquence simulée de rénographie synthétique réaliste. Trois valeurs sont testées pour la constante pondérant la régularisation : $\beta = 0$ (pas de régularisation), $\beta = 10$ et $\beta = 50$. Les courbes d'activité obtenues par RUDUR sont en trait plein, tandis que celles de la vérité terrain sont en tiretés. En l'absence de régularisation, les estimations des courbes d'activité sont très bruitées. La régularisation permet de limiter l'impact de ce bruit. En revanche, une trop forte régularisation détériore la qualité de l'estimation en lissant les pics présents dans les courbes de la vérité terrain.

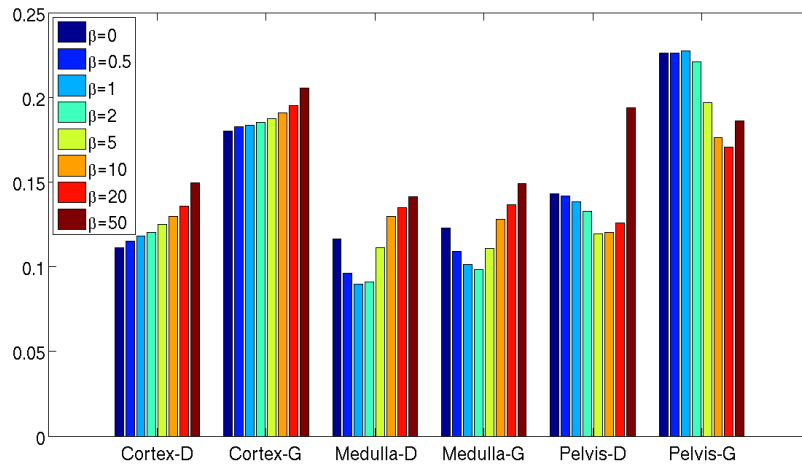


FIGURE 4.10 – Importance de la régularisation - Erreurs d'estimation moyennes obtenues avec RUDUR sur les données simulées de rénographies synthétiques réalistes pour différentes valeurs de β . Chaque bâton du diagramme correspond à la moyenne des erreurs d'estimation (NMAE) des courbes d'activité dans toutes les séquences pour un compartiment physiologique. Sur l'axe des abscisses, le "D" et le "G" signifient respectivement droite et gauche.

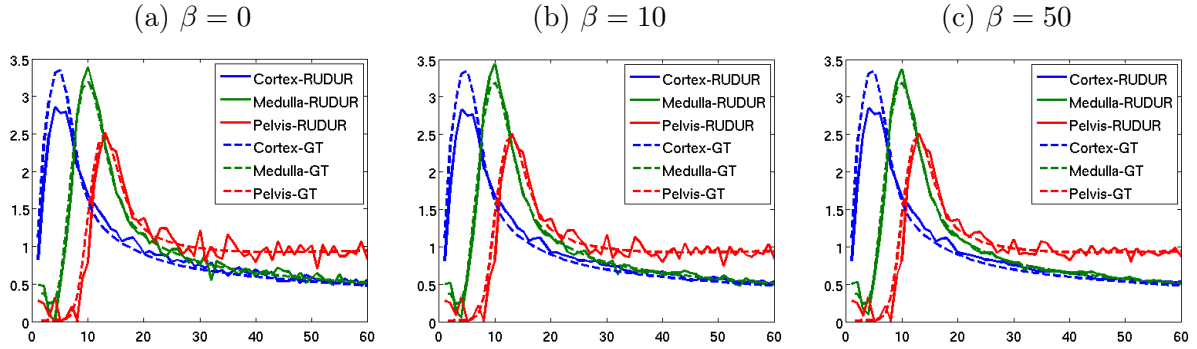


FIGURE 4.11 – Exemple de courbes d’activité obtenues avec RUDUR sur une séquence simulée de rénographie synthétique réaliste, en appliquant la régularisation Tikhonov uniquement sur les images 20 à 60 ($S_{t,t} = 0$ pour $t \leq 20$). Trois valeurs sont testées pour la constante pondérant la régularisation : $\beta = 0$ (pas de régularisation), $\beta = 10$ et $\beta = 50$. Les courbes d’activité obtenues par RUDUR sont en trait plein, tandis que celles de la vérité terrain sont en tirets.

estimation est très peu lisse à cause du bruit.

4.4.4 Robustesse au choix des régions d’intérêt

Les régions d’intérêt étant en pratique souvent choisies manuellement, elles peuvent être l’objet de variations inter et intra-observateurs. Il est donc nécessaire que RUDUR soit robuste à leur sélection. Ainsi, de légères variations dans ces choix ne doivent pas changer les estimations des sources. Nous avons donc appliqué RUDUR et FAMIS sur une même séquence de rénographie synthétique réaliste en faisant varier les régions d’intérêt utilisées. Les mêmes régions que dans la section 4.2 ont été utilisées, et de légères perturbations y ont été appliquées afin de simuler les variations inter et intra-utilisateurs dans leur sélection. Les perturbations retenues sont des translations de la région d’intérêt de un pixel, horizontalement (translation vers la gauche, vers la droite, ou aucune) et/ou verticalement (translation vers le haut, vers le bas, ou aucune). La région d’intérêt peut également être dilatée. Ceci a permis de créer $3 * 3 * 2 = 18$ régions d’intérêt différentes pour chaque organe, menant ainsi à $18^3 = 5832$ triplets de régions d’intérêt différents en combinant toutes les possibilités. Quelques exemples de ces régions d’intérêt sont disponibles en figure 4.13. La séquence étudiée étant composée de 60 images de 26 par 21 pixels, des translations verticale et/ou horizontale de un pixel produisent des régions d’intérêt significativement différentes.

RUDUR et FAMIS ont donc été appliquées sur la séquence d’images avec chacun des triplets de ROI construits, menant ainsi à 5832 séparations différentes pour chaque méthode. La dispersion des erreurs d’estimation (NMAE) de chacune de ces méthodes est présentée en figure 4.14, et les valeurs numériques des moyennes et écarts types sont dans la table 4.10. RUDUR présente des erreurs d’estimation des courbes d’activité et des images factorielles beaucoup plus faibles que FAMIS pour tous les compartiments physiologiques. De plus, la

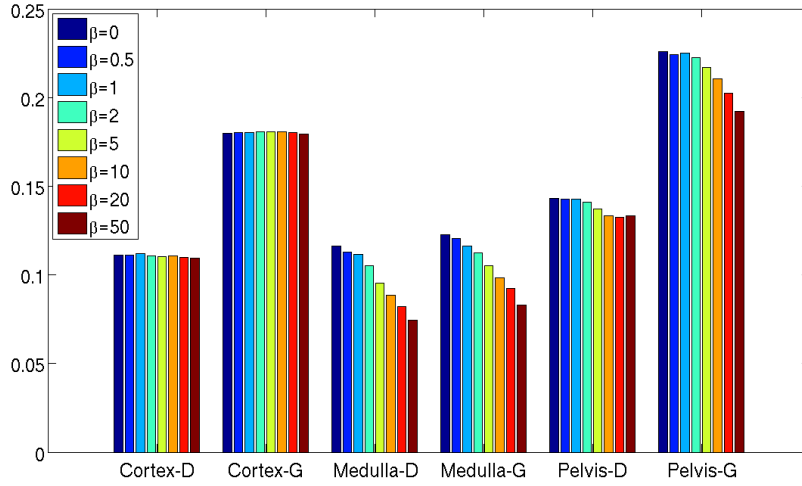


FIGURE 4.12 – Erreurs d’estimation moyennes obtenues avec RUDUR sur les données simulées de rénographies synthétiques réalistes pour différentes valeurs de β . Aucune régularisation n’est opérée pour les 20 premières images ($S_{t,t} = 0$ pour $t \leq 20$). Chaque bâton du diagramme correspond à la moyenne des erreurs d’estimation (NMAE) des courbes d’activité dans toutes les séquences pour un compartiment physiologique. Sur l’axe des abscisse, le "D" et le "G" signifient respectivement droite et gauche.

TABLE 4.10 – Robustesse au choix des ROI : Erreurs d’estimation (NMAE) moyenne des courbes d’activité et des images factorielles avec 5832 triplets de ROI différents - Rénographie synthétique réaliste

Méthode	F1	F2	F3	A1	A2	A3
<i>RUDUR</i>	0.154 ± 0.006	0.158 ± 0.012	0.103 ± 0.009	0.500 ± 0.024	0.332 ± 0.022	0.504 ± 0.039
<i>FAMIS</i>	0.199 ± 0.024	0.313 ± 0.204	0.396 ± 0.169	0.871 ± 0.228	0.807 ± 0.321	0.802 ± 0.278

Les résultats sont sous la forme *moy* $\pm$ *std*, où *moy* et *std* sont respectivement la moyenne et l’écart-type de l’erreur d’estimation (NMAE) en utilisant 5832 triplets de ROI différents. Ces résultats sont calculés sur une unique séquence de rénographie synthétique réaliste.

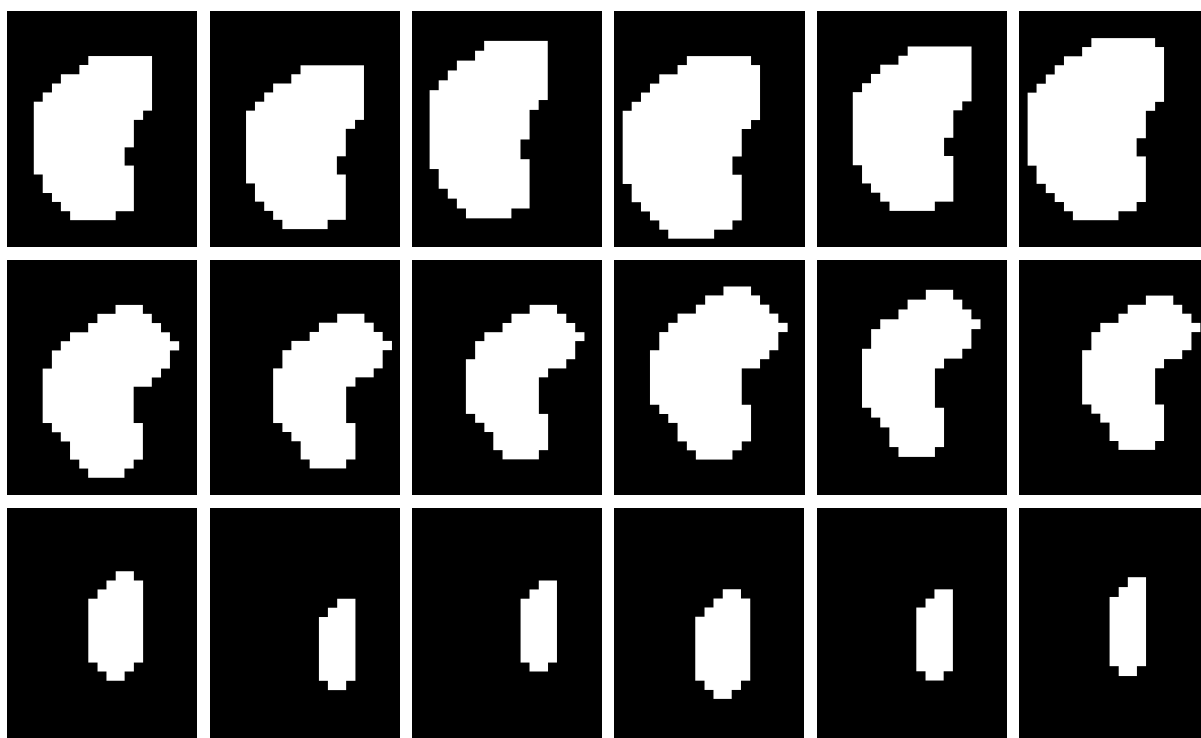


FIGURE 4.13 – Robustesse au choix des ROI - Exemples de régions d'intérêt obtenues en perturbant les régions d'intérêt originales par des translations horizontales, verticales ainsi que par une dilatation. La première ligne correspond aux régions d'intérêt du cortex rénal, la seconde à celles de la médulla rénale, et la troisième à celles du pelvis.

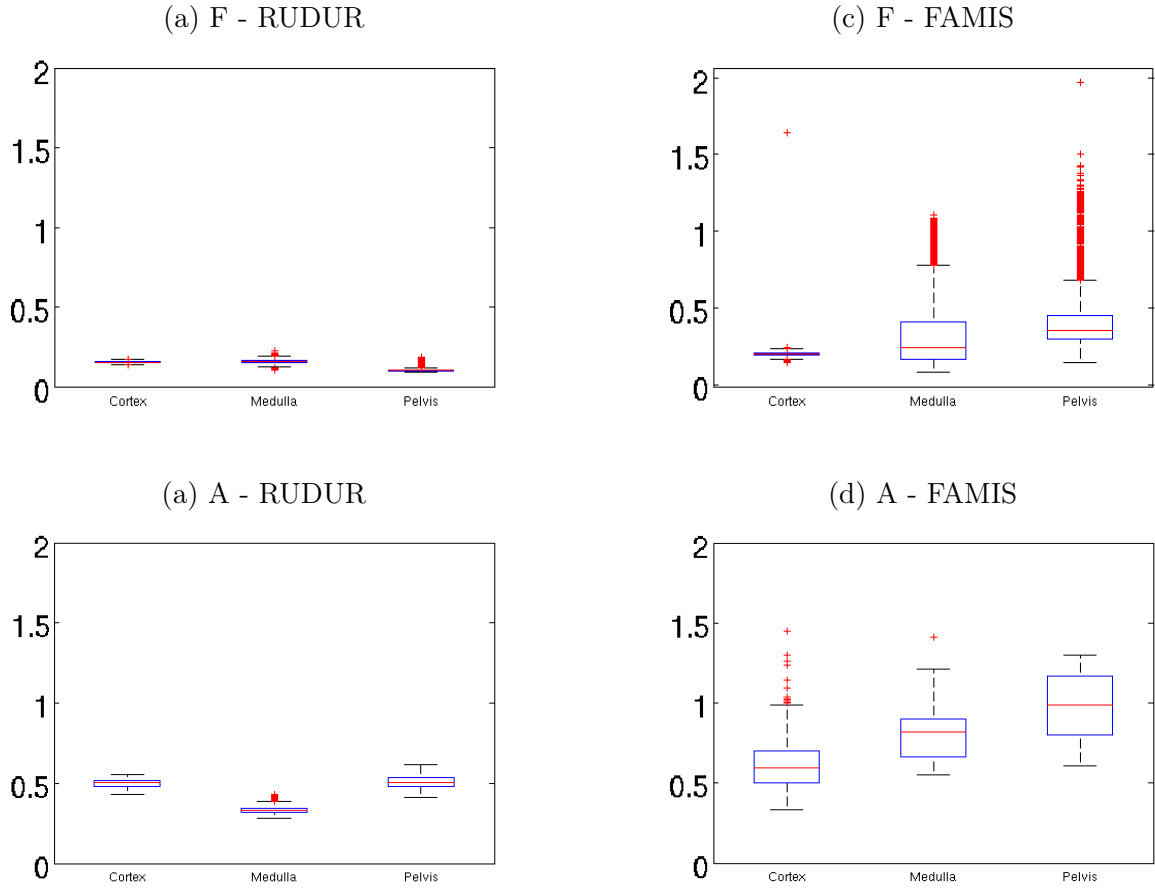


FIGURE 4.14 – Robustesse aux choix des ROI - Distribution des erreurs d'estimation (NMAE) des courbes d'activité (première ligne) et des images factorielles (seconde ligne) en utilisant 5832 triplets de ROI différents pour une séquence de rénographie synthétique réaliste. L'axe des ordonnées correspond à l'erreur d'estimation (NMAE) de la courbe d'activité ou de l'image factorielle considérée. Chaque boîte correspond à la distribution de l'erreur d'estimation de la courbe d'activité ou de l'image factorielle d'un des compartiments physiologiques. Elle résume ainsi l'erreur d'estimation pour l'ensemble des 5832 triplets de ROI possibles. Plus la boîte et les moustaches associées sont grandes, plus la méthode est sensible aux choix des ROI. La ligne rouge correspond à la médiane, et les extrémité de la boîte bleue au premier et troisième quartile. Les extrémités de la ligne en tiretée noire correspondent aux valeurs minimales et maximales, et les croix rouges aux données aberrantes.

variance de l'erreur d'estimation est très faible avec RUDUR alors qu'elle est forte avec FAMIS. Les courbes d'activité de la médullaire rénale et du pelvis sont particulièrement sensibles au choix des régions d'intérêt avec FAMIS. Ceci peut s'expliquer par le fait que les régions d'intérêt de ces deux compartiments sont petites et que FAMIS utilise ces régions de manière binaire. De légères modifications de ces régions sont donc relativement importantes compte tenu de la taille d'origine de ces régions. Les valeurs numériques de la table 4.10 confirment ces différentes observations. Ces tests montrent dans l'ensemble une sensibilité importante de FAMIS à de légères perturbations des régions d'intérêt, tandis que RUDUR y est plutôt robuste.

Ces résultats s'expliquent par le fait que contrairement à FAMIS, RUDUR utilise les régions d'intérêt de manière souple en se concentrant non pas sur la présence ou l'absence de régions d'intérêt sur un pixel, mais plutôt sur la distance aux régions d'intérêt. Cette distance a l'avantage d'être peu sensible aux légères variations dans le choix des ROI, et permet ainsi à RUDUR d'être robuste au choix de ces dernières.

4.5 Utilisation des images d'intérêt pour l'évaluation d'une source de faible intensité

Lorsqu'une source de faible intensité moyenne mais de forte intensité ponctuelle est présente dans la séquence d'images, les méthodes de séparation échouent à extraire cette source car son signal est en moyenne trop faible vis à vis du bruit et des autres compartiments physiologiques. C'est le cas par exemple du myocarde dans les séquences de scintigraphie utilisant du 6-DIG. Ce type de source est difficile à estimer, surtout lorsque le bruit est important. Un jeu de données synthétiques possédant une telle source a été générée afin de tester les performances de RUDUR avec et sans l'utilisation des images d'intérêt. Ce jeu de données est identique au jeu de données synthétiques simples présenté dans la section 4.1.1, à ceci près que la seconde source a été remplacée par une nouvelle source. Cette dernière est d'intensité faible, sauf pendant un intervalle de temps de 5 images. La courbe d'activité ainsi que l'image factorielle de cette source sont disponibles dans la seconde ligne de la figure 4.15 (seconde source). Cette source est d'autant plus difficile à estimer qu'elle est mélangée à la fois spatialement et temporellement aux deux autres sources. Il n'y a pas de pixels ou d'images où cette source est seule ou très dominante. Un exemple de séquence d'images de ce jeu de données est fourni en annexe D.6.

L'introduction de connaissances a priori sur le dérouls temporel de la source de faible intensité à travers la matrice Q peut permettre de résoudre ce problème. Dans tous les tests précédents, la matrice Q était choisie comme étant l'identité car aucun a priori d'ordre temporel n'a été utilisé. Nous testons ici RUDUR sur ce jeu de données en faisant varier les coefficients diagonaux de Q . Nous avons choisi de faire varier uniquement les coefficients diagonaux de Q associés aux images 13 à 19, c'est à dire là où la seconde source est de plus forte intensité (les images d'intérêt de cette source), afin d'en augmenter l'importance dans la séparation. Les autres coefficients diagonaux de Q sont laissés à 1. Afin de compenser l'augmentation de l'importance du terme de fidélité aux données par la matrice Q , les paramètres α et β sont

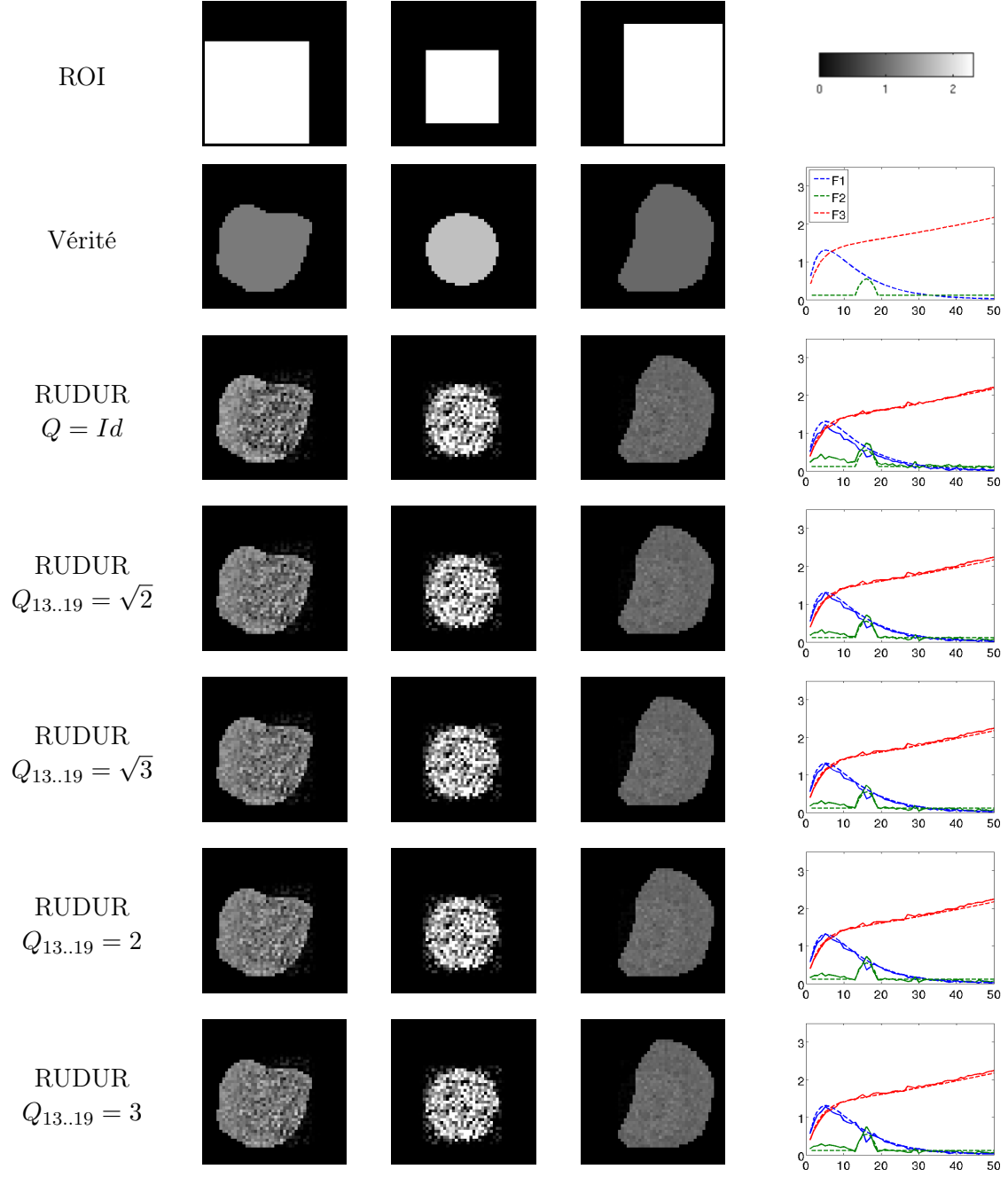


FIGURE 4.15 – Résultat de RUDUR sur une séquence du jeu de données simples contenant une source de faible intensité, avec $\sigma = 0.5$, en utilisant différentes valeurs pour la matrice Q . La première ligne contient les régions d'intérêt utilisées par RUDUR et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tiretés, avec le temps en abscisse, et l'activité en ordonnée. Les coefficients diagonaux 13 à 19 de Q valent 1, $\sqrt{2}$, $\sqrt{3}$, 2 ou 3. La source de faible intensité est la seconde source.

multipliés par la moyenne du carré des coefficients diagonaux de Q . Les résultats obtenus sur un exemple sont disponibles en figure 4.15 (avec $\sigma = 0.5$). Les moyennes des erreurs obtenues pour différentes matrices Q sur 100 séquences du jeu de données synthétiques avec source de faible intensité sont présentées dans les tables 4.11 et 4.12.

Les résultats de la figure 4.15 montrent que lorsque Q est choisie égale à l'identité, la première source ainsi que la source de faible intensité sont mal séparées. Le début des courbes d'activité de ces deux sources est mal estimé. Les images factorielles, mais aussi et surtout les courbes d'activité de ces deux sources, sont mieux estimées lorsque la matrice Q est choisie différente de l'identité. Ceci se confirme quantitativement dans les tables 4.11 et 4.12. Lorsque $\sigma = 0.5$, les erreurs d'estimation de la source de faible intensité sont nettement améliorées, tandis que les deux autres sources ont des erreurs d'estimation similaires pour les différentes valeurs de Q choisies. Ces tableaux montrent également que comme attendu, la source la moins bien estimée est la source de faible intensité.

Au delà d'une certaine valeur pour les coefficients diagonaux 13 à 19 de Q , comprise entre 2 et 3, les erreurs d'estimation globales semblent s'accroître. Le même phénomène est présent lorsque $\sigma = 1$, sauf que cette valeur est cette fois-ci plus proche de 1. Des valeurs trop fortes sur les coefficients diagonaux 13 à 19 de Q dégradent la séparation. Ceci peut s'expliquer par le fait que lorsque Q contient des valeurs importantes sur sa diagonale, il devient primordial pour minimiser la fonctionnelle d'expliquer fidèlement avec les matrices A et F les images d'intérêts associées à ces coefficients diagonaux importants, quitte à expliquer les erreurs qu'elles contiennent. Par conséquent, les autres images seront mal expliquées, et les informations qu'elles contiennent seront moins bien exploitées pour la séparation. Plus le bruit est fort, plus ce phénomène s'accroît. C'est pourquoi la valeur *optimale* des coefficients diagonaux 13 à 19 de Q est plus faible pour $\sigma = 1$ que pour $\sigma = 0.5$. L'utilisation des images d'intérêt pour construire Q est donc efficace pour la séparation des sources de faible intensité, à condition que les images d'intérêt ne prennent pas trop d'importance dans la séparation, au regard du bruit et des erreurs qu'elles contiennent.

Les méthodes autres que RUDUR échouent totalement à estimer correctement cette source, hormis FAMIS qui obtient des résultats légèrement moins bons que RUDUR (avec $Q = Id$) grâce à son utilisation des régions d'intérêt. La source étant de très faible intensité, il est nécessaire d'apporter des a priori considérables afin de pouvoir l'extraire des séquences d'images, ce qui explique l'échec des autres méthodes.

4.6 Données scintigraphiques issues de l'étude du 6-DIG

Les différentes méthodes de séparation ont également été évaluées sur les scintigraphies cliniques utilisant le 6-DIG. La difficulté principale dans ces données est d'extraire la courbe d'activité du myocarde, nécessaire pour le calcul d'un indice d'insulino-résistance.

TABLE 4.11 – Erreurs d'estimation des facteurs (F). Jeu de données simples, avec source de faible intensité.

Mesure	Algorithme	sigma=0.5			sigma=1		
		F1	F2	F3	F1	F2	F3
NMSE	$RUDUR-Q = Id$	0.009 ± 0.006	0.278 ± 0.430	0.002 ± 0.004	0.034 ± 0.007	0.709 ± 0.249	0.005 ± 0.003
	$RUDUR-Q = \sqrt{2}$	0.011 ± 0.008	0.185 ± 0.127	0.001 ± 0.000	0.038 ± 0.009	0.598 ± 0.113	0.003 ± 0.001
	$RUDUR-Q = \sqrt{3}$	0.010 ± 0.008	0.157 ± 0.116	0.001 ± 0.000	0.047 ± 0.011	0.689 ± 0.134	0.003 ± 0.001
	$RUDUR-Q = 2$	0.008 ± 0.006	0.124 ± 0.091	0.001 ± 0.000	0.056 ± 0.015	0.784 ± 0.182	0.004 ± 0.001
	$RUDUR-Q = 3$	0.009 ± 0.004	0.125 ± 0.053	0.001 ± 0.000	0.098 ± 0.021	1.383 ± 0.785	0.008 ± 0.006
NMAE	$RUDUR-Q = Id$	0.097 ± 0.029	0.447 ± 0.344	0.030 ± 0.029	0.188 ± 0.019	0.751 ± 0.167	0.055 ± 0.016
	$RUDUR-Q = \sqrt{2}$	0.105 ± 0.036	0.366 ± 0.117	0.019 ± 0.002	0.183 ± 0.019	0.597 ± 0.063	0.045 ± 0.004
	$RUDUR-Q = \sqrt{3}$	0.098 ± 0.035	0.334 ± 0.116	0.019 ± 0.002	0.191 ± 0.021	0.627 ± 0.083	0.045 ± 0.004
	$RUDUR-Q = 2$	0.088 ± 0.030	0.300 ± 0.088	0.019 ± 0.002	0.202 ± 0.026	0.675 ± 0.131	0.046 ± 0.008
	$RUDUR-Q = 3$	0.091 ± 0.020	0.305 ± 0.060	0.019 ± 0.002	0.289 ± 0.040	1.016 ± 0.342	0.071 ± 0.032

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples, avec source de faible intensité. Selon la ligne, les coefficients diagonaux 13 à 19 de Q valent 1, $\sqrt{2}$, $\sqrt{3}$, 2 ou 3. Les autres coefficients diagonaux de Q sont égaux à 1.

La source de faible intensité est la seconde source.

TABLE 4.12 – Erreurs d'estimation des images factorielles (A). Jeu de données simples, avec source de faible intensité.

Mesure	Algorithme	sigma=0.5			sigma=1		
		A1	A2	A3	A1	A2	A3
NMSE	$RUDUR-Q = Id$	0.050 ± 0.023	0.317 ± 0.037	0.012 ± 0.012	0.144 ± 0.022	0.551 ± 0.058	0.030 ± 0.010
	$RUDUR-Q = \sqrt{2}$	0.062 ± 0.025	0.297 ± 0.022	0.007 ± 0.000	0.140 ± 0.015	0.519 ± 0.043	0.024 ± 0.001
	$RUDUR-Q = \sqrt{3}$	0.057 ± 0.024	0.299 ± 0.023	0.007 ± 0.000	0.153 ± 0.023	0.528 ± 0.045	0.024 ± 0.002
	$RUDUR-Q = 2$	0.051 ± 0.020	0.303 ± 0.024	0.007 ± 0.000	0.173 ± 0.030	0.535 ± 0.048	0.025 ± 0.003
	$RUDUR-Q = 3$	0.059 ± 0.013	0.328 ± 0.029	0.008 ± 0.000	0.264 ± 0.044	0.529 ± 0.051	0.038 ± 0.013
NMAE	$RUDUR-Q = Id$	0.182 ± 0.036	0.498 ± 0.031	0.080 ± 0.030	0.322 ± 0.022	0.703 ± 0.035	0.138 ± 0.017
	$RUDUR-Q = \sqrt{2}$	0.202 ± 0.040	0.484 ± 0.019	0.068 ± 0.002	0.316 ± 0.017	0.693 ± 0.030	0.127 ± 0.004
	$RUDUR-Q = \sqrt{3}$	0.193 ± 0.039	0.487 ± 0.020	0.068 ± 0.002	0.329 ± 0.023	0.704 ± 0.032	0.128 ± 0.004
	$RUDUR-Q = 2$	0.183 ± 0.032	0.492 ± 0.021	0.069 ± 0.002	0.351 ± 0.029	0.712 ± 0.034	0.129 ± 0.006
	$RUDUR-Q = 3$	0.197 ± 0.020	0.513 ± 0.024	0.071 ± 0.002	0.434 ± 0.035	0.713 ± 0.048	0.154 ± 0.024

Les résultats sont sous la forme $moy \pm std$, où moy et std sont respectivement la moyenne et l'écart-type de la mesure d'erreurs (NMSE ou NMAE) sur 100 simulations du jeu de données synthétiques simples, avec source de faible intensité. Selon la ligne, les coefficients diagonaux 13 à 19 de Q valent 1, $\sqrt{2}$, $\sqrt{3}$, 2 ou 3. Les autres coefficients diagonaux de Q sont égaux à 1.

La source de faible intensité est la seconde source.

4.6.1 Résultats de séparation sur les données cliniques

La zone d'intérêt dans les scintigraphies 6-DIG est la zone cardiaque. Les méthodes de séparation ont donc été utilisées sur des sous-parties de la séquence contenant environ 500 pixels. Les 30 premières images de la séquence ont été conservées pour la séparation, car les images suivantes contiennent essentiellement du bruit et très peu d'informations utiles. Le nombre de sources recherchées a été fixé à $K = 3$: le ventricule droit, le ventricule gauche et le myocarde. Il est nécessaire de dissocier ventricule gauche et droit car pendant la phase transitoire de dilution du traceur dans l'organisme, ces deux compartiments physiologiques

ont des courbes d'activité bien différentes. En effet, le pic d'activité du ventricule droit à lieu quelques secondes en amont de celui du ventricule gauche. De plus, le pic du ventricule gauche est moins intense et plus étalé. Le pic d'activité du myocarde est lui attendu à la suite de celui du ventricule gauche. Un exemple de séquence scintigraphique utilisant du 6-DIG est disponible en annexe D.8.

Les résultats de séparation de chaque méthode sur trois séquences différentes sont visibles dans les figures 4.16, 4.17 et 4.18. L'interprétation de ces résultats est délicate car ces données ne possèdent pas de vérité terrain. On peut toutefois juger qualitativement ces résultats en fonction de ce qui est attendu. Les positions des ventricules droits et gauche et du myocarde sont à peu près connues. Ce dernier doit se situer dans le coin inférieur droit des images, en enveloppant le ventricule gauche. De même, l'ordre d'apparition des pics d'activité doit lui aussi être conforme à ce qui a été évoqué dans le paragraphe précédent.

Dans ces trois exemples, les images factorielles estimées du myocarde sont très peu réalistes pour les méthodes n'utilisant pas les ROI (PFADS, S-BSS-vecDC et FAROI). En effet, la forme caractéristique en U n'est pas visible, et les images factorielles du myocarde ne sont pas plus intense dans le coin inférieur droit qu'ailleurs. Pour les deux autres méthodes, RUDUR et FAMIS, qui utilisent elles les ROI, les images factorielles du myocarde sont plus réalistes dans la forme et la position. Ceci ne constitue cependant pas une validation de la séparation car dans ces méthodes les images factorielles sont influencées par les ROI et sont donc contraintes de leur ressembler.

Au niveau des courbes d'activité, dans de nombreux cas l'estimation de la courbe d'activité du myocarde possède un pic avant le pic d'activité du ventricule gauche, ce qui n'est pas réaliste. Dans d'autres estimations, la courbe d'activité semble croissante tout au long de la séquence, (S-BSS-vecDC premier exemple, et FAROI troisième exemple). Ces courbes d'activité sont probablement influencées par le foie, qui se situe à l'extérieur du coin inférieur gauche des images, et non par le myocarde.

Il n'y a au final que peu d'estimations réalistes où la séparation a été potentiellement efficace. Dans le deuxième exemple, RUDUR et S-BSS-vecDC fournissent des estimations du myocarde réalistes et qui sont cohérentes entre elles. Faute de vérité terrain, on ne peut toutefois pas affirmer que les résultats de la séparation sont bons dans ces deux cas. Dans les deux autres exemples, aucune des méthodes ne fournit d'estimation réaliste du myocarde. Les résultats ne sont pas meilleurs sur les autres séquences disponibles dans le jeu de données.

Le problème de séparation sous-jacent est particulièrement complexe, car les séquences d'images sont très bruitées, le myocarde a un signal très faible et il est également fortement mélangé aux ventricules. Il est donc très difficile d'extraire les informations du myocarde dans ce contexte. Ceci peut expliquer les échecs des méthodes, même dans le cas où la connaissance des régions d'intérêt est utilisée. La présence d'un pic d'activité dans le myocarde précédant celui du ventricule gauche indique que le myocarde estimé est un mélange de ventricules gauche et droit et de bruit. En effet, le myocarde étant de signal très faible, il est probable que la non homogénéité spatiale des ventricules ainsi que les autres compartiments physiologiques non pris en compte soient des facteurs plus importants que le myocarde. Les méthodes vont

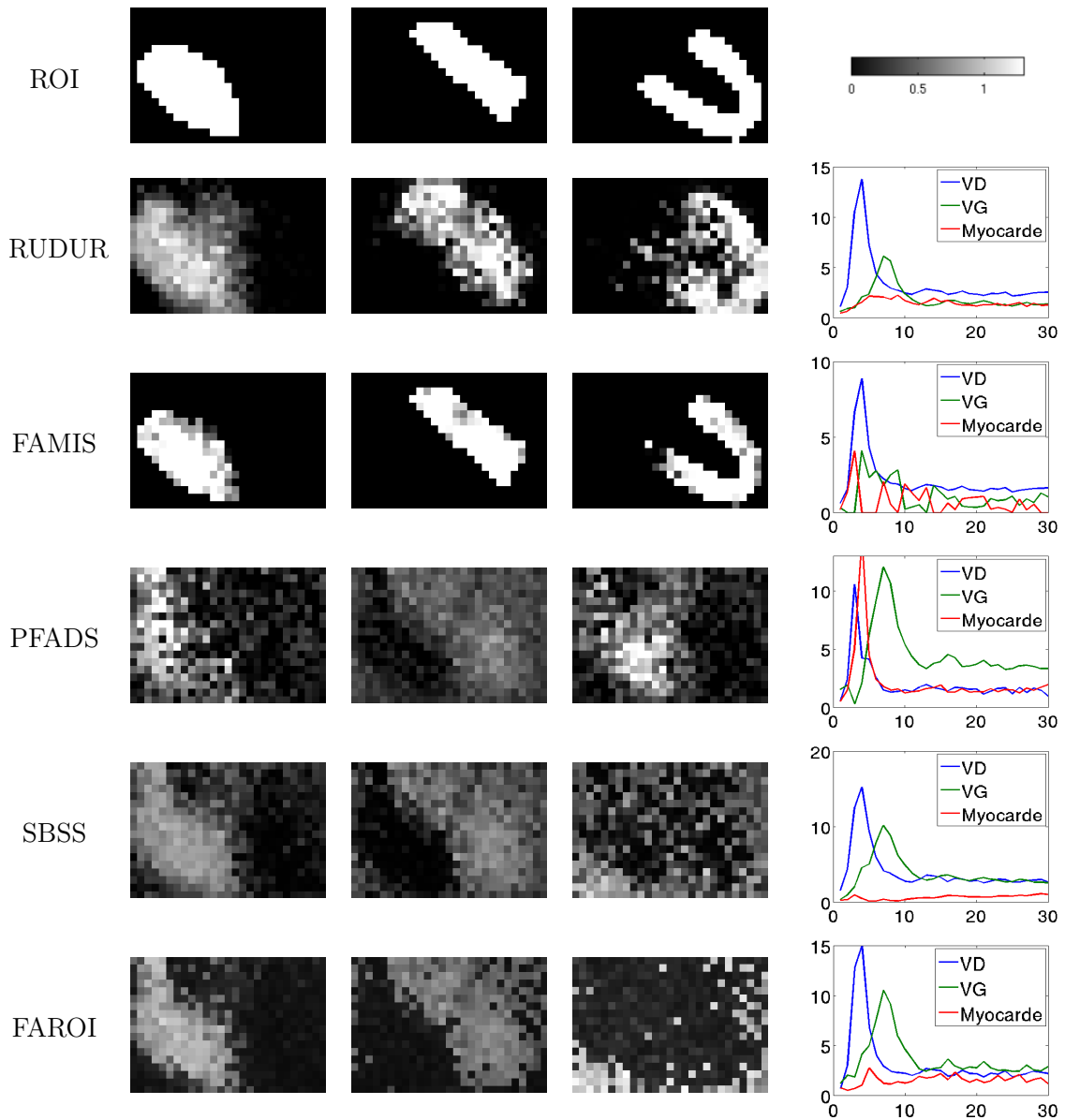


FIGURE 4.16 – Résultat de la séparation sur un exemple de séquence 6-DIG clinique (premier exemple). Les trois premières colonnes correspondent dans l'ordre au ventricule droit, au ventricule gauche et au myocarde. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne, avec le temps en abscisse et l'activité en ordonné.

donc chercher à expliquer ces facteurs plutôt qu'à expliquer le signal du myocarde avec la troisième source. Les ventricules gauches et droits ayant un signal plus fort, leur estimation est généralement cohérente avec ce qui est attendu.

Nous proposons dans la suite un protocole pour simuler simplement des séquences d'images

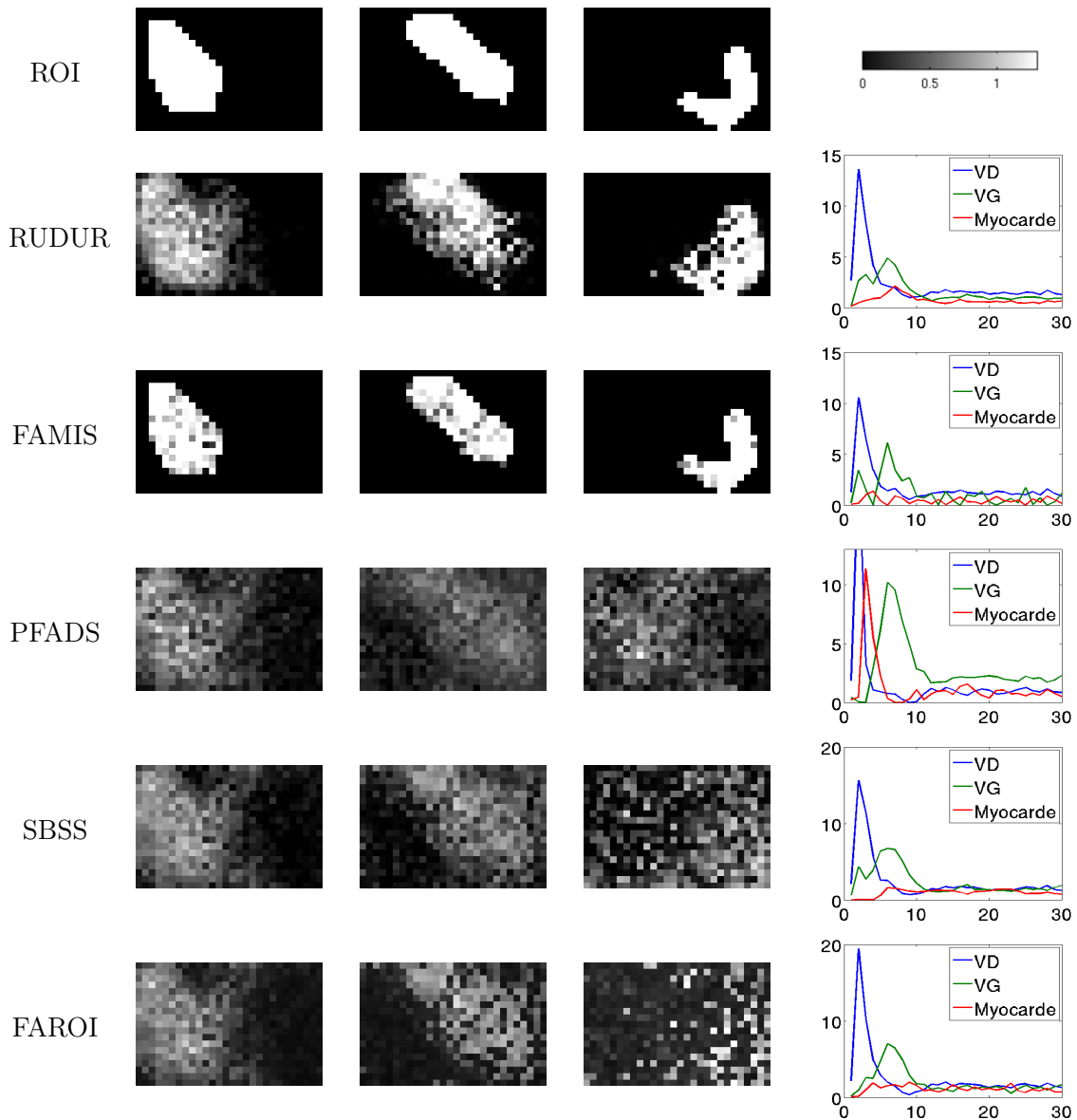


FIGURE 4.17 – Résultat de la séparation sur un exemple de séquence 6-DIG clinique (deuxième exemple). Les trois premières colonnes correspondent dans l'ordre au ventricule droit, au ventricule gauche et au myocarde. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne, avec le temps en abscisse et l'activité en ordonné.

ressemblantes à ces données cliniques. La vérité terrain de ces séquences étant connue, les causes d'échecs de la séparation sur les données cliniques peuvent ainsi être identifiées et analysées.

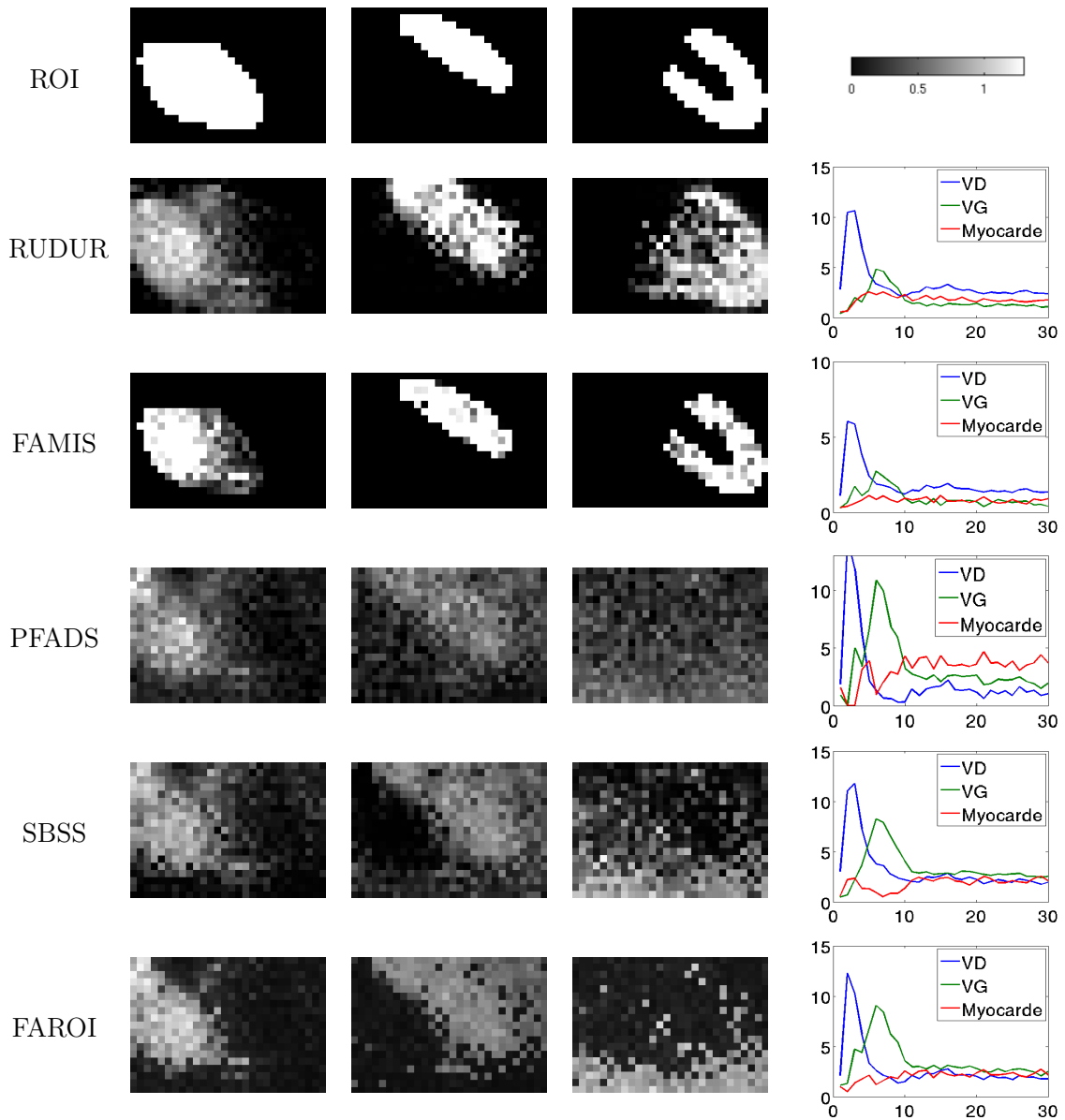


FIGURE 4.18 – Résultat de la séparation sur un exemple de séquence 6-DIG clinique (troisième exemple). Les trois premières colonnes correspondent dans l'ordre au ventricule droit, au ventricule gauche et au myocarde. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne, avec le temps en abscisse et l'activité en ordonné.

4.6.2 Simulation par modélisation des séquences d'images 6-DIG

Afin de construire les données simulées, nous avons développé un protocole similaire à celui utilisé pour générer les données de rénographies réalistes. Un modèle compartimental ainsi que

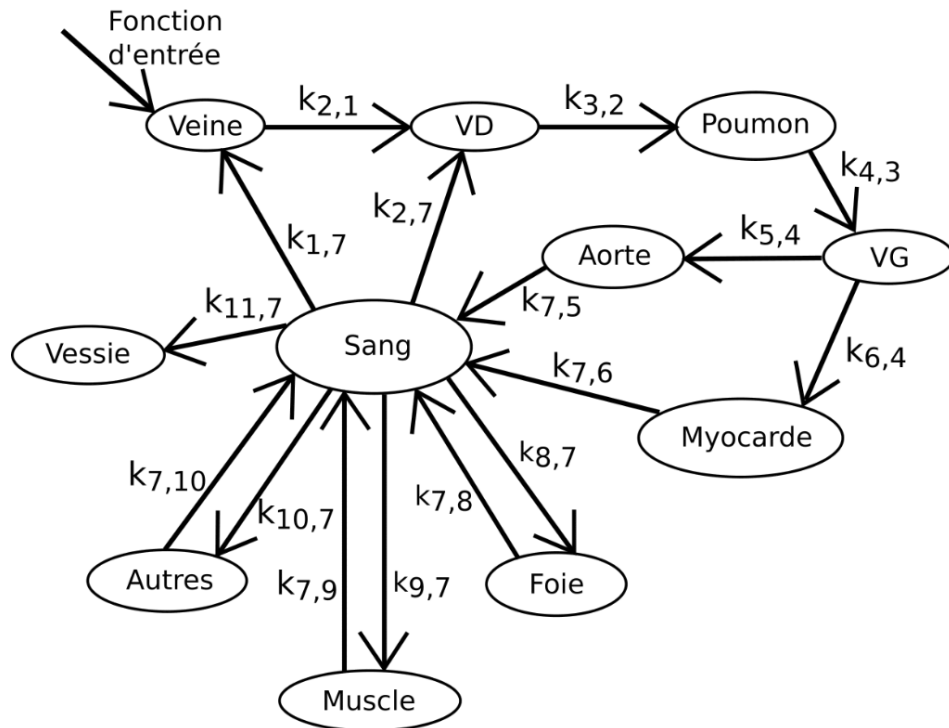


FIGURE 4.19 – Modèle compartimental utilisé pour la simulation des données 6-DIG. Le traceur entre dans le corps via la veine sous-clavière, puis se dirige vers le ventricule droit, les poumons puis le ventricule gauche. Le traceur se dirige partiellement vers le myocarde (5%) tandis que le reste se diffuse dans le corps via le système sanguin. Une fois dans la vessie le traceur y est stocké sans retour possible dans le reste de l'organisme.

les cartes de présence des compartiments ont été choisis. Le comportement de la caméra est simplifié pour prendre en compte la collimation, et les autres phénomènes sont modélisés par un lissage passe bas.

Le modèle compartimental utilisé pour définir les courbes d'activité (matrice F) est détaillé dans la figure 4.19. Les compartiments physiologiques y sont au nombre de 11. La veine sous-clavière est alimentée par la fonction d'entrée, c'est à dire l'injection du traceur radioactif. Cette dernière est choisie comme une exponentielle à décroissance très rapide, dont la somme est égale à la quantité de traceur radioactif injecté dans les études du 6-DIG, c'est à dire environ 7×10^{12} molécules de 6-DIG radioactif. La fonction d'entrée contient une deuxième exponentielle, de même allure mais retardée de 15 minutes, qui correspond à la seconde injection de 6-DIG après celle de l'insuline. Les valeurs choisies par défaut des cinétiques $k_{i,j}$ du modèle compartimental sont détaillées dans la table 4.13. Le myocarde ne reçoit en moyenne que 5% du débit cardiaque, ce qui explique que la cinétique entre le ventricule gauche et le myocarde est faible. Cette cinétique peut-être choisie plus grande pour la phase suivant l'injection d'insuline, selon le degré d'insulino-résistance souhaité. Le modèle compartimental ainsi que les cinétiques ont été définis à partir des connaissances sur la circulation du sang, et des observations faites sur les données cliniques afin de les reproduire fidèlement.

TABLE 4.13 – Cinétiques par défaut du modèle compartimental pour la simulation de données 6-DIG (en s^{-1}).

$k_{2,1}$	$k_{3,2}$	$k_{4,3}$	$k_{5,4}$	$k_{6,4}$	$k_{7,5}$	$k_{7,6}$	$k_{8,7}$
0.3	0.4	0.4	0.45	0.025	1	0.5	0.003
$k_{7,8}$	$k_{9,7}$	$k_{7,9}$	$k_{10,7}$	$k_{7,10}$	$k_{11,7}$	$k_{1,7}$	$k_{2,7}$
0.01	0.003	0.01	0.5	0.3	0.001	0.005	0.03

$k_{i,j}$ correspond à la cinétique de déplacement du traceur du compartiment j vers le compartiment i (exprimée en s^{-1}). Les compartiments physiologiques sont numérotés dans l'ordre : 1-Veine sous-clavière ; 2-Ventricule Droit (VD) ; 3-Poumons ; 4-Ventricule Gauche (VG) ; 5-Aorte ; 6-Myocarde ; 7-Sang ; 8-Foie ; 9-Muscle ; 10-Autres ; 11-Vessie.

L'évolution de la quantité de traceur dans chacun des compartiments est déterminée grâce au modèle compartimental, aux cinétiques et à la fonction d'entrée. Les pertes dues à la décroissance radioactive sont prises en compte à la suite de cette étape : la quantité présente dans chaque compartiment à un temps t est multipliée par le rapport entre la quantité totale de traceur non encore désintégré et la quantité initiale de traceur injecté. Au temps t , ce rapport est égal à $e^{-\frac{\log(2)}{t_{1/2}} t}$, où $t_{1/2}$ est la demi-vie du traceur, c'est à dire 13.3 heures.

Au facteur d'échelle près, les quantités obtenues sont les courbes d'activité des compartiments physiologiques du modèle (matrice F). Ce facteur d'échelle correspond au rapport entre la quantité de rayons gamma détectée pendant un intervalle de temps dt par la caméra, et la quantité de molécules de traceur radioactif présentes. C'est donc le produit de la probabilité qu'une molécule de traceur se désintègre pendant un intervalle de temps dt , et de la probabilité qu'un rayon gamma émis depuis le corps du patient atteigne la gamma caméra en passant la collimation. En première approximation, la probabilité de désintégration est égale à $\frac{\log(2)}{t_{1/2}} \times dt$. Afin d'estimer la probabilité de détection d'un rayon gamma, on suppose que la désintégration se passe à 30 cm de la collimation de la caméra, et que le rayon est détecté si il arrive dans un cercle de 5 mm de rayon placé à la verticale du lieu de désintégration sur la caméra. En supposant de plus que toutes les directions d'émission de rayons gamma sont équiprobables, le calcul de l'angle solide associé permet d'obtenir que la probabilité de détection est de $\frac{1-\cos(\frac{r}{h})}{2}$. À titre d'exemple, pour une durée $dt = 0.1s$, le facteur d'échelle est environ égal à 10^{-10} .

La matrice A contient les cartes de présence des compartiments physiologiques et est définie approximativement grâce aux connaissances sur l'anatomie humaine. Afin de prendre en compte l'imperfection de la collimation, l'effet Compton et le volume partiel, ces cartes de présence sont convoluées avec un noyau gaussien bidimensionnel.

Les courbes d'activité et les cartes de présence des principaux compartiments physiologiques obtenues avec ce protocole sont disponibles en figure 4.20

La séquence d'images simulée Y est obtenue grâce à la réalisation de simulations de lois de

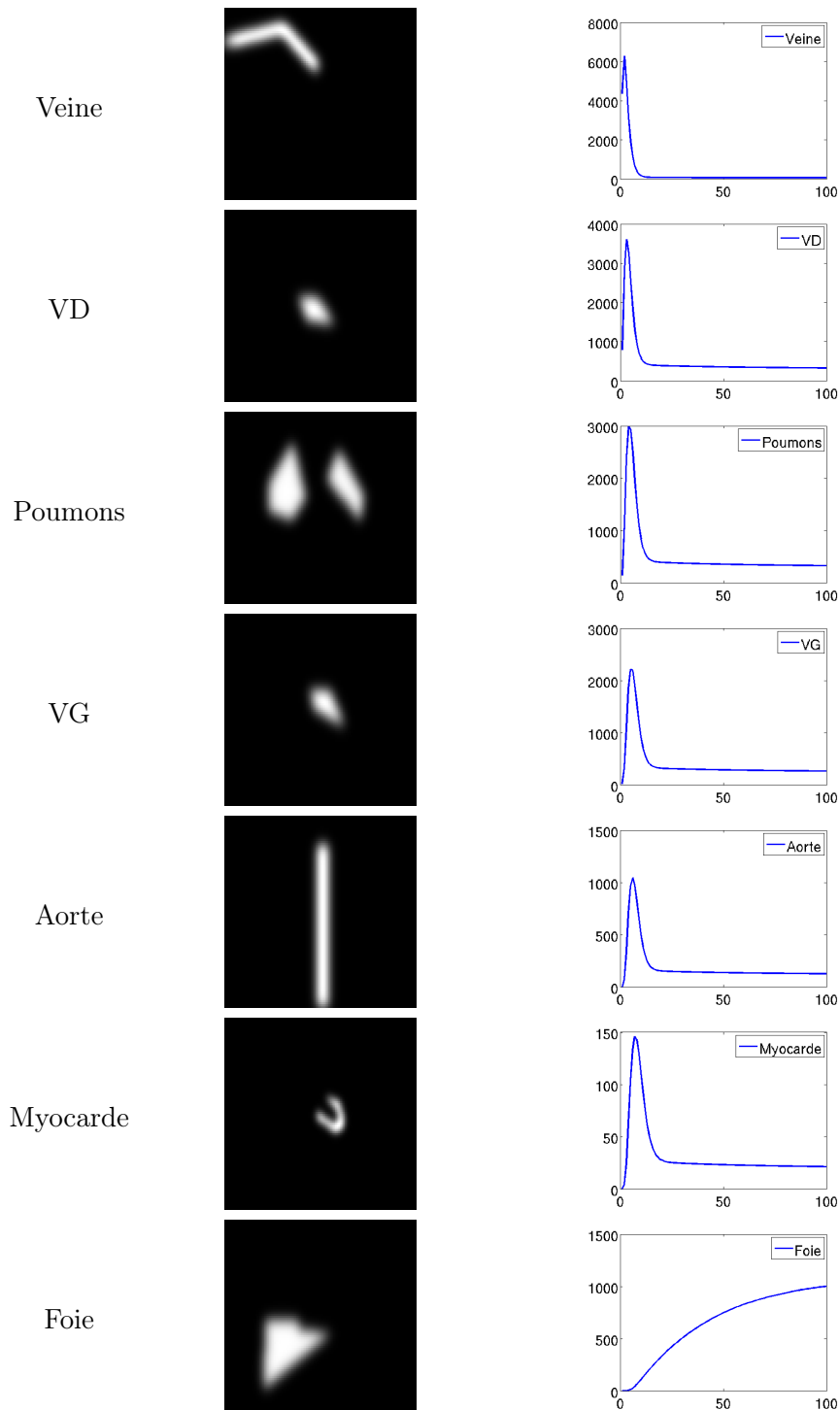


FIGURE 4.20 – Cartes de présence (matrice A), et courbes d'activité (matrice F) des principaux compartiments physiologiques de la simulation du 6-DIG. Le modèle compartimental est corps entier, mais les cartes de présence ne sont montrées que dans la zone imagée par la caméra. La vessie n'est pas affichée car elle est en dehors du champ de la caméra. Les autres compartiments physiologiques non affichés (Muscle, Sang, Autres) ont des cartes de présence homogène dans la zone visible par la caméra.

Poisson : $Y = \mathcal{P}(A \times F)$. Les compartiments physiologiques sont supposés homogènes spatialement et stables temporellement. En pratique, dt est choisi petit devant les constantes de temps du modèle compartimental ($dt = 0.1s$). Un rééchantillonnage permet ensuite d'obtenir une séquence où chaque image représente une durée de deux secondes, comme celles des cas cliniques. Un exemple de séquence d'images obtenue est disponible en annexe D.7. Le protocole et les paramètres utilisés ne sont évidemment pas parfaits, mais les séquences simulées ressemblent aux séquences obtenues cliniquement. La différence notable entre les données cliniques et les données simulées est que ces dernières font l'hypothèse que les compartiments physiologiques sont homogènes spatialement. Cette hypothèse est en pratique fausse pour toute la phase transitoire en début de séquence lorsque le traceur se dilue dans l'organisme.

4.6.3 Résultats de séparation sur les données simulées

Ces données simulées ont été séparées à l'aide du même protocole et des mêmes algorithmes que ceux utilisés précédemment pour les données de 6-DIG. Un zoom a donc été fait dans la zone cardiaque, et $K = 3$ sources ont été recherchées. Les 30 premières images de la séquence ont été utilisées. Le résultat de chacun des algorithmes sur une des séquences simulées est visible en figure 4.21. Aucune des méthodes ne fournit de bonnes estimations pour le myocarde. La méthode ayant la meilleure estimation de la courbe d'activité du myocarde est FAMIS, grâce à son utilisation des régions d'intérêt. Cependant cette estimation reste très peu satisfaisante car elle comporte d'importantes erreurs, notamment sur le pic d'activité du myocarde et sur la fin de la séquence étudiée. Les méthodes n'utilisant pas la connaissance des régions d'intérêt en tant qu'a priori fournissent de très mauvaises estimations de la courbe d'activité et de l'image factorielle du myocarde. S-BSS-vecDC et FAROI estiment notamment une courbe d'activité du myocarde croissante tout le long de la séquence, ce qui est contradictoire avec la vérité terrain. RUDUR fournit ici des estimations plus mauvaises que FAMIS pour le myocarde. Ceci peut venir du fait que pour être robuste au choix des ROI, RUDUR est plus permissif que FAMIS. En effet, FAMIS force à 0 une grande partie des coefficients des images factorielles au regard des ROI ce qui explique pourquoi l'image factorielle estimée par FAMIS est plus proche de la vérité terrain que celle estimée par RUDUR. L'allure de l'estimation de FAMIS est toutefois assez plate et bruitée, tandis que la véritable courbe d'activité du myocarde est lisse et possède une décroissance qui n'est pas présente dans l'estimation de FAMIS.

La très faible intensité du myocarde le rend très difficile à séparer du ventricule droit et du ventricule gauche qui sont eux dans l'ensemble bien estimés. Cette très faible intensité vient du fait que seulement 5% du débit cardiaque est dirigé vers le myocarde, engendrant ainsi une cinétique $k_{6,4}$ faible. Dans ces conditions, il est très compliqué d'estimer fiablement ce compartiment physiologique, ce qui explique les mauvais résultats précédents obtenus sur les données cliniques. On retrouve dans les simulations les mêmes résultats que ceux observés dans les données cliniques. En particulier, la présence du foie dans le coin inférieur gauche des images perturbe l'estimation du myocarde par les méthodes n'utilisant pas les régions d'intérêt. La présence de pic d'activité précoce du myocarde se retrouve également dans les données simulées en utilisant RUDUR. La courbe d'activité du myocarde obtenue par RUDUR avec les deux pics d'activité suggère que cette estimation n'est pas la véritable courbe d'activité

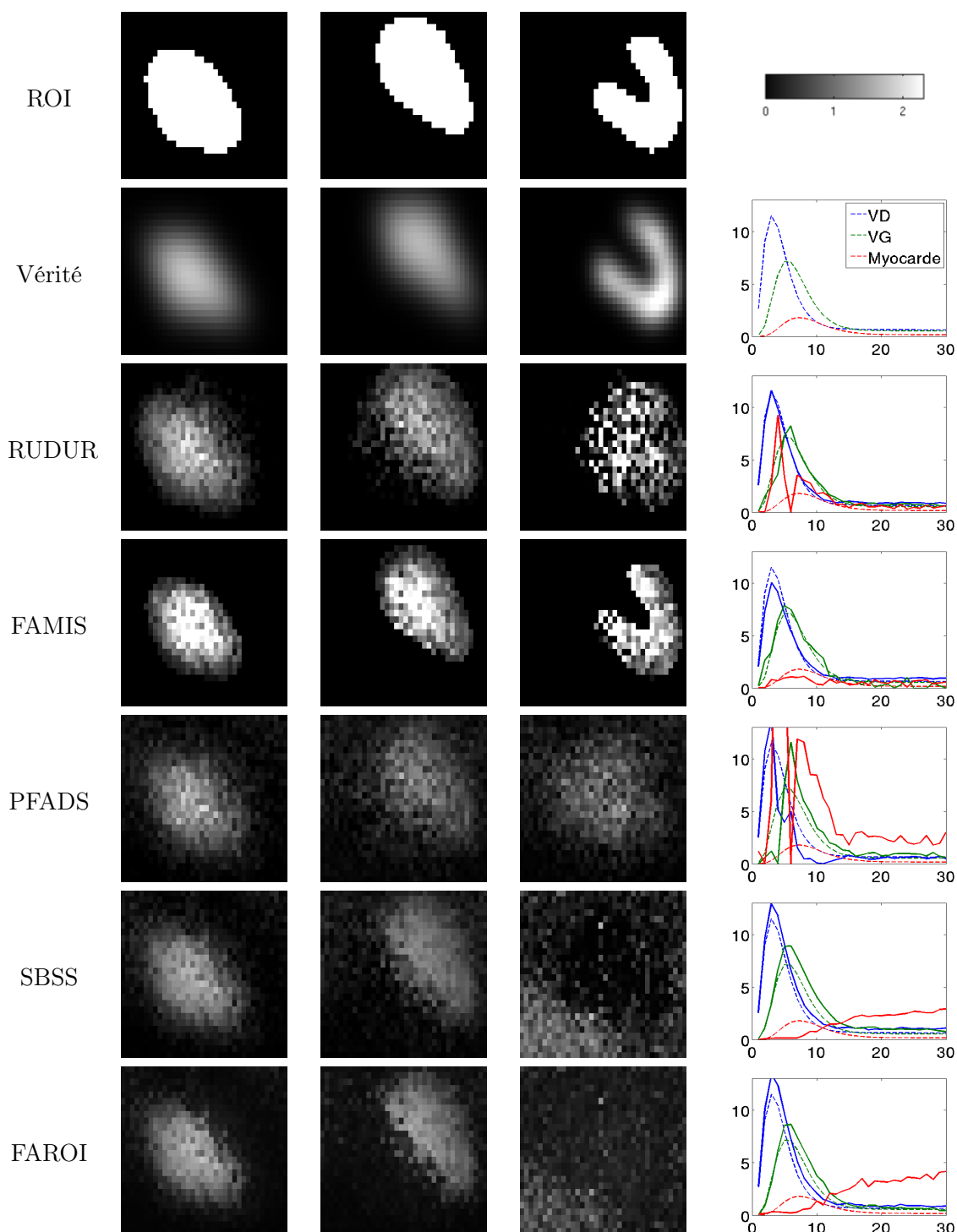


FIGURE 4.21 – Résultat de la séparation sur une séquence du jeu de données de simulations simulations d'études 6-DIG. Pour des soucis de visibilité, les valeurs des courbes d'activité du myocarde (vérité terrain et estimation) ont été multipliées par 3. Les trois premières colonnes correspondent dans l'ordre au ventricule droit, au ventricule gauche et au myocarde. La première ligne contient les régions d'intérêt utilisées par RUDUR et FAMIS, et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par chacune des méthodes. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par chaque méthode sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tiret, avec le temps en abscisse, et l'activité en ordonnée.

du myocarde mais un mélange des courbes d'activité des deux ventricules.

Une activité plus forte dans le myocarde permettrait de mieux séparer ce dernier. Pour le vérifier, la cinétique $k_{6,4}$ a été artificiellement multipliée par 3, 5 et 7 sur une séquence simulée de 6-DIG. Les résultats de RUDUR sur les séquences obtenues sont disponibles en figure 4.22. On peut y constater que plus la cinétique est importante, plus l'estimation de l'image factorielle du myocarde fait ressortir la forme en U caractéristique. De plus, les estimations des courbes d'activité du myocarde sont elles aussi meilleures lorsque la cinétique augmente. La qualité de l'estimation des courbes d'activité des ventricules en revanche est moins bonne. En effet, quand le myocarde est de signal très faible, le problème peut être considéré comme un problème simple de séparation à 2 sources et les ventricules sont donc bien estimés. Mais lorsque le myocarde est de signal plus fort, il ne peut plus être négligé et vient donc perturber l'estimation des ventricules car le problème de séparation est désormais à trois sources.

4.6.4 Conclusion sur les données 6-DIG

Les données scintigraphiques utilisant le 6-DIG proposent des problèmes de séparation particulièrement complexes. Ceci est principalement dû à la difficulté d'extraire le myocarde dans ces séquences. À cause de la faible proportion du débit cardiaque qui s'y dirige, la quantité de traceur dans le myocarde est très faible comparativement à celle présente dans les ventricules. Le myocarde est également fortement mélangé spatialement à ces ventricules ce qui ne facilite pas son extraction, tout comme le contexte très bruité de ces séquences. Enfin, les cinétiques du myocarde et des ventricules se ressemblent beaucoup. Ces différents points empêchent l'introduction de connaissances temporelles à l'aide de la matrice Q . En effet, comme vue dans la section 4.5, l'utilisation des images d'intérêt n'améliore pas les résultats en présence de bruit important. L'utilisation de la matrice S a également été testée sur ces données afin de ne pas régulariser les pics d'activité, mais n'a apporté aucune amélioration sur la séparation. Des séquences d'images 3D utilisant la TEMP ont aussi été réalisées avec le 6-DIG, afin de minimiser les mélanges dans les voxels et ainsi faciliter la séparation du myocarde. À cause du mouvement du myocarde lors des battements cardiaques, celui-ci reste néanmoins mélangé de façon importante dans les données 3D. De plus, ces dernières sont très bruitées, ce qui rend complexe la sélection de la région d'intérêt du myocarde. Les résultats obtenus sur ces séquences d'images 3D n'ont pas été meilleurs que ceux obtenus avec les données d'origine.

Les problèmes rencontrés sur les données cliniques se retrouvent avec les données simulées. D'après ces dernières, pour obtenir une séparation de meilleure qualité il faudrait que la cinétique d'échange entre le myocarde et le ventricule gauche soit significativement plus rapide et que le débit cardiaque se dirigeant vers le myocarde soit plus important.

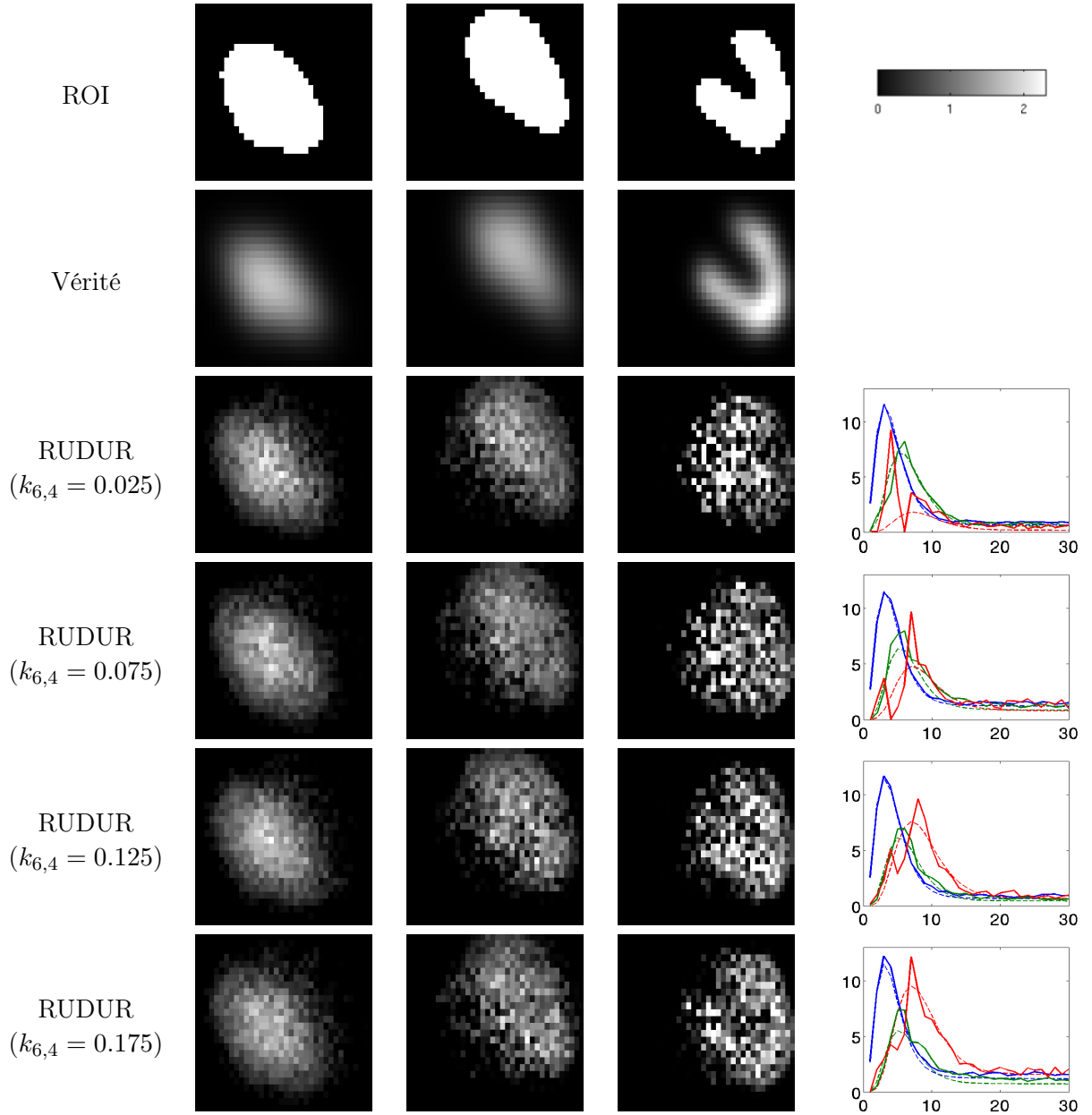


FIGURE 4.22 – Résultat de la séparation sur des simulations d'études utilisant le 6-DIG, construites avec différentes valeurs pour la cinétique ventricule gauche-myocarde ($k_{6,4}$). Pour des soucis de visibilité, les valeurs des courbes d'activité du myocarde (vérité terrain et estimation) ont été multipliées par 3. Les trois premières colonnes correspondent dans l'ordre au ventricule droit, au ventricule gauche et au myocarde. La première ligne contient les régions d'intérêt utilisées par RUDUR, et la seconde contient la vérité terrain des images factorielles et des courbes d'activité. Les lignes suivantes (colonnes 1 à 3) représentent les images factorielles obtenues par RUDUR sur chacune des séquences. L'échelle de couleur utilisée pour les images factorielles est dans le coin supérieur droit de la figure. Les courbes d'activité obtenues par RUDUR sont affichées dans la dernière colonne. Les estimations sont tracées en trait plein, tandis que la vérité terrain est tracée en tirets, avec le temps en abscisse, et l'activité en ordonnée. Le ventricule droit est en bleu, le ventricule gauche en vert et le myocarde en rouge.

4.7 Conclusion sur les résultats

Les différents tests réalisés sur RUDUR dans ce chapitre sont reproductibles en exécutant les scripts MATLAB disponibles en ligne². Plusieurs enseignements sont à tirer de ces résultats :

- L'intégration des différentes connaissances a priori dans RUDUR lui permet d'obtenir de très bons résultats de séparation comparativement aux autres méthodes de l'état de l'art, tant sur des données synthétiques que cliniques.
- La sensibilité de RUDUR au choix des constantes de pondération de la fonctionnelle est raisonnable, faire évoluer ces constantes d'un ordre de grandeur ne change que peu les résultats obtenus. Ceci permet de conserver des jeux de paramètres identiques pour des séquences d'images similaires.
- Pénaliser les images factorielles non cohérentes avec les régions d'intérêt permet de faire face efficacement à la sous-détermination du problème de séparation. Les régions d'intérêt sont en effet très informatives et permettent de contraindre efficacement les images factorielles.
- Contrairement à FAMIS, RUDUR est robuste à la sélection des régions d'intérêt grâce à l'utilisation de l'information de distance à ces régions et non à l'unique critère de présence ou d'absence. De petites modifications dans la sélection des régions ne change que peu les distances entre les pixels et ces régions, et par conséquent ne change que peu les résultats de séparation de RUDUR.
- Utiliser un critère de régularisation sur les courbes d'activité modifie les variations fortes et rapides qui sont parfois des indicateurs importants. Dans RUDUR, la matrice S exprime les périodes temporelles pendant lesquelles il faut conserver les discontinuités et les périodes où il faut lisser les courbes d'activité. En pratique, la matrice S est construite à l'aide de connaissances a priori sur la localisation temporelle des pics d'activité, et permet de les conserver tout en lissant le reste des courbes.
- Lorsque des sources de forte intensité et de faible intensité sont présentes dans une séquence à des instants différents, elles peuvent être séparées avec RUDUR grâce à l'utilisation de la matrice Q . Cette dernière exprime la localisation temporelle des images d'intérêt de la source faible. Elle permet d'augmenter l'importance de la fidélité aux données sur ces images. Lorsque les données sont très bruitées, cela amplifie l'influence du bruit présent sur les images d'intérêt de la source faible, et limite donc l'utilisation de cette connaissance.
- Malgré l'ajout de connaissances a priori, RUDUR ne parvient pas à obtenir des résultats de séparation satisfaisants sur les données de scintigraphies utilisant du 6-DIG. Les simulations de ces scintigraphies montrent que la séparation échoue notamment à cause de la très faible activité du myocarde, dont le comportement temporel est proche des ventricules. Le fort bruit des images de la séquence ne peut être compensé par les informations a priori injectées dans le modèle.

2. <http://www.gipsa-lab.grenoble-inp.fr/~marc.filippi/>

Conclusion

Les travaux de cette thèse ont porté sur la séparation de sources en imagerie nucléaire. Ils ont été motivés par des données scintigraphiques utilisant du 6-DIG pouvant permettre l'accès à un indice d'insulino-résistance. Afin de calculer cet indice, il est nécessaire d'extraire les courbes d'activité du myocarde et du ventricule gauche. À cause des recouvrements spatiaux importants entre ces deux compartiments, cette extraction passe par l'utilisation de méthodes de séparation de sources. Cependant, différents éléments font que la séparation s'avère très complexe sur ces données et que les méthodes échouent à obtenir des séparations cohérentes. Premièrement, le myocarde est fortement mélangé spatialement avec les ventricules. Deuxièmement, l'allure de sa courbe d'activité est proche de celle du ventricule gauche. Troisièmement, les images de la séquence sont très bruitées. Enfin, du fait de la faible proportion du débit cardiaque dirigé vers le myocarde, ce dernier est de signal très faible.

Plus généralement, le principal obstacle à la séparation en imagerie nucléaire est que le problème sous-jacent est sous déterminé, même en prenant en compte les contraintes de positivité sur les courbes d'activité et les cartes de présence des sources. Les méthodes de séparation doivent intégrer des connaissances a priori suffisamment informatives sur les sources afin de faire face à cette sous-détermination. Malgré l'intégration de certains a priori, les méthodes actuelles sont bien souvent trop peu informatives pour les problèmes complexes. En outre, la séparation se base sur le modèle de mélange linéaire, dont la validité dépend fortement de la zone de la séquence considérée.

Les différents défis proposés par les données 6-DIG nous ont amené à considérer certaines connaissances a priori spatiales et temporelles sur les sources afin de faciliter leur séparation. Ces connaissances se basent notamment sur les très informatives régions d'intérêt, et ont été intégrées à une fonctionnelle à minimiser, sous la forme de pénalisations ou bien de relâchements de contraintes. En se basant sur les régions d'intérêt, nous avons également proposé une initialisation non-arbitraire et proche de la solution recherchée. Par ailleurs, la validité du modèle de mélange linéaire a été remise en question et nous avons limité sa portée en relâchant la contrainte de fidélité aux données dans certaines parties de la séquence d'images.

Grâce à l'intégration de ces connaissances, la méthode développée permet d'obtenir de bons résultats de séparation sur des données synthétiques et cliniques, comparativement aux autres méthodes de l'état de l'art. La méthode est également robuste à la sélection des régions d'intérêt, de légères modifications de ces dernières ne changent que peu les estimations des sources. Malgré cela, les résultats de séparation sur les données 6-DIG montrent des incohérences et ne sont toujours pas satisfaisants. Même en ajoutant de nombreuses connaissances a priori, l'estimation de la courbe d'activité du myocarde est mauvaise. Nos simulations sur des données synthétiques montrent que la séparation est mise en échec ici principalement à cause de la faiblesse du signal du myocarde. Il est très difficile d'obtenir les informations de ce dernier car il ne reçoit qu'une toute petite partie du débit cardiaque.

Contributions

Plusieurs contributions découlent de ces travaux de thèse, dont voici les principales :

- **Décomposition tensorielle** : Nous avons étudié la décomposition tensorielle pour la séparation de sources en imagerie nucléaire (section 2.6). Lorsqu'une troisième diversité ayant de bonnes propriétés est présente dans les données, le formalisme tensoriel permet d'avoir un problème de séparation bien posé avec une unique solution. Nous avons en revanche montré que cette condition n'est pas valide pour l'étude du 6-DIG.
- **Connaissances a priori spatiales** : L'utilisation des **régions d'intérêt** améliore la séparation. Ces régions sont intégrées de différentes manières dans le processus de séparation.
 - Pénalisation des images factorielles non cohérentes avec les régions d'intérêt. Cette pénalisation dépend de la distance entre les pixels et les régions d'intérêt afin d'être robuste à la sélection de ces dernières (section 3.3.3). L'intégration de cette pénalisation permet de faire face à la sous-détermination du problème d'origine en favorisant les solutions en accord avec les régions d'intérêt.
 - Relâchement de la contrainte de fidélité aux données sur les pixels éloignés des régions d'intérêt, où le modèle de mélange linéaire n'est probablement pas valide. Le relâchement dépend pour chaque pixel de sa distance à la plus proche région d'intérêt, focalisant ainsi la séparation sur les régions d'intérêt, tout en étant robuste à leur sélection (section 3.3.4).
 - Initialisation des images factorielles et des courbes d'activité à l'aide des régions d'intérêt. Cette initialisation permet de débiter l'algorithme de façon non arbitraire et au plus près de la solution (section 3.3.2).
- **Connaissances a priori temporelles** : Intégration d'informations temporelles pour faciliter la séparation.
 - **Images d'intérêt** : Introduction d'une pondération temporelle de la fidélité aux données (section 3.3.5). En définissant cette pondération à l'aide des images d'intérêt, il est possible d'extraire des sources de faible signal en focalisant le processus de séparation sur les images où cette source est la plus intense (section 4.5).
 - **Régularisation adaptée** : Proposition d'un critère de régularisation Tychonoff adaptée afin de favoriser l'obtention de courbes d'activité lisses sans sous-estimer leurs pics d'activité. La connaissance a priori utilisée ici est la position temporelle approximative des pics d'activité, qui sont généralement en début de séquence car provoqués par la dilution du bolus de traceur (section 3.3.6).
- **Étude des séquences 6-DIG** : Simulation de séquences scintigraphiques utilisant du 6-DIG en adoptant une modélisation simplifiée (section 4.6). Ces simulations ont permis de mettre en évidence la difficulté de l'extraction de la courbe d'activité du myocarde du fait de la faible fraction du débit cardiaque dirigé vers ce muscle.

Ces différentes contributions sont rassemblées dans une seule et même méthode, RUDUR, dont le code source MATLAB commenté est intégralement disponible en ligne³. Les tests de

3. <http://www.gipsa-lab.grenoble-inp.fr/~marc.filippi/>

validation de cette méthode présentés dans le chapitre 4, sont également fournis. Ces tests montrent les bonnes performances de RUDUR vis à vis des autres méthodes de l'état de l'art, sur des données synthétiques et cliniques. De légères variations dans le choix des régions d'intérêt n'influencent que très peu les résultats, ce qui indique que RUDUR est bien robuste à la sélection de ces régions. Ceci est rendu possible grâce à l'utilisation de l'information des distances aux régions d'intérêt.

Ces travaux ont donné lieu à deux publications dans une conférence nationale [Fil+15] et une conférence internationale [Fil+16], ainsi qu'à un article de journal [FDM17].

Perspectives

- Un modèle compartimental pourrait être intégré à la méthode de séparation afin de lier les courbes d'activité entre elles. L'intégration de ce modèle pourrait également permettre d'obtenir directement les cinétiques d'échange entre les compartiments conjointement avec les courbes d'activité et les cartes de présence.
- L'intégration des régions d'intérêt à l'aide d'une matrice des distances permet d'améliorer la robustesse de la séparation vis à vis du choix de ces régions. Il pourrait être intéressant d'examiner le comportement de la séparation lorsque ces régions d'intérêt sont choisies de manière automatique par un algorithme. De petites erreurs de ce dernier ne devrait que très peu perturber le résultat de la séparation. Ceci permettrait de rendre RUDUR automatique, à la sélection des constantes de pondération près.
- L'algorithme RUDUR est aisément transposable aux données 3D. En effet, dans sa version de base (2D), les images de la séquence sont vectorisées. Pour la version 3D, les volumes sont également vectorisés et l'algorithme est inchangé. Les voxels des séquences 3D sont aussi composés de mélanges, du fait du volume partiel et de l'imprécision des mesures. RUDUR peut donc être utilisé sur des données issues de TEMP ou de TEP, sous réserve que les régions d'intérêt 3D soient accessibles, afin de calculer par exemple le flux sanguin myocardique absolue [Shr+17].
- Il est envisageable d'appliquer RUDUR à des domaines autres que l'imagerie nucléaire. En particulier, il pourrait être intéressant de faire de la fusion de données hyperspectrales issues des instruments MUSE et HUBBLE. Ce dernier a une excellente résolution spatiale, mais une faible résolution spectrale. C'est l'inverse pour MUSE qui possède une faible résolution spatiale (car l'instrument est sur la terre) et une haute résolution spectrale [Bac+10]. À cause de cette faible résolution spatiale, il est nécessaire d'utiliser des méthodes de séparation de sources pour extraire de MUSE des spectres bien résolus des différentes galaxies du champ profond. Dans ce contexte, les régions d'intérêt de ces galaxies peuvent être obtenues avec les données spatialement bien résolues de HUBBLE, puis intégrées comme connaissance a priori dans RUDUR pour séparer les données issues de MUSE. Ceci permettrait d'obtenir les spectres des galaxies du champ profond à la résolution spectrale de MUSE, et d'en déduire des informations, par exemple la distance des galaxies étudiées.

Modèle compartimental et système d'équations différentielles du 6-DIG

Le modèle compartimental utilisé pour le 6-DIG a été présenté dans la figure 1.11. Les équations différentielles associées sont les suivantes :

$$\frac{dq_1(t)}{dt} = \frac{dI(t)}{dt} + (k_{12}q_2(t) + k_{13}q_3(t)) - (k_{01} + k_{21} + k_{31})q_1(t) \quad (\text{A.1})$$

$$\frac{dq_2(t)}{dt} = k_{12}q_1(t) - k_{21}q_2(t) \quad (\text{A.2})$$

$$\frac{dq_3(t)}{dt} = k_{13}q_1(t) - k_{31}q_3(t) \quad (\text{A.3})$$

où q_1 , q_2 et q_3 sont respectivement les quantités de traceurs dans le sang, le myocarde, et les autres tissus.

Calcul des dérivées partielles

Le détail des calculs des dérivées partielles des différents termes de f_{RUDUR} sont détaillées dans cet annexe. Les formules de dérivation matricielles qui simplifient grandement les calculs sont tout d'abord données.

B.1 Formules de dérivation matricielle

Dans les formules de dérivation matricielle suivante, X et C sont des matrices quelconques. La trace d'une matrice quelconque X est noté $Tr[X]$.

$$\frac{\partial Tr[XC]}{\partial X} = \frac{\partial Tr[CX]}{\partial X} = C^T \quad (\text{B.1})$$

$$\frac{\partial Tr[X^T CX]}{\partial X} = (C + C^T)X \quad (\text{B.2})$$

$$\frac{\partial Tr[XCX^T]}{\partial X} = X(C + C^T) \quad (\text{B.3})$$

B.2 Dérivation selon A

B.2.1 Terme de fidélité aux données

$$f_{\text{QWLS}}(A, F) = ||W(Y - AF)Q||_F^2 \quad (\text{B.4})$$

$$= Tr[(W(Y - AF)Q)(W(Y - AF)Q)^T] \quad (\text{B.5})$$

$$= Tr[WYQ^2Y^TW - 2AFQ^2Y^TW^2 + W^2AFQ^2F^TA^T] \quad (\text{B.6})$$

Ainsi, en dérivant par rapport à A et en utilisant les formules (B.1) et (B.3) on obtient :

$$\frac{\partial f_{\text{QWLS}}(A, F)}{\partial A} = -2W^2(Y - AF)Q^2F^T \quad (\text{B.7})$$

B.2.2 Terme de pénalisation des images factorielles

$$f_{\text{ROI}}(A) = \|D \circ D \circ A\|_{1,\mu} \quad (\text{B.8})$$

$$= \sum_{i=1}^N \sum_{k=1}^K \sqrt{D_{i,k}^4 A_{i,k}^2 + \mu^2} - \mu \quad (\text{B.9})$$

En dérivant selon $A_{i,k}$, on a :

$$\frac{\partial f_{\text{ROI}}(A)}{\partial A_{i,k}} = \frac{D_{i,k}^4 A_{i,k}}{\sqrt{D_{i,k}^4 A_{i,k}^2 + \mu^2}} \quad (\text{B.10})$$

En particulier, si $\mu = 0$ on a $\frac{\partial f_{\text{ROI}}(A)}{\partial A} = D \circ D$

B.3 Dérivation selon F

B.3.1 Terme de fidélité aux données

$$f_{\text{QWLS}}(A, F) = \|W(Y - AF)Q\|_F^2 \quad (\text{B.11})$$

$$= \text{Tr}[(W(Y - AF)Q)^T (W(Y - AF)Q)] \quad (\text{B.12})$$

$$= \text{Tr}[(Q^2 Y^T W^2 Y - 2Q^2 Y^T W^2 AF + F^T (A^T W^2 A) F Q^2)] \quad (\text{B.13})$$

Ainsi, en dérivant par rapport à F et en utilisant les formules (B.1) et (B.2) on obtient :

$$\frac{\partial f_{\text{QWLS}}(A, F)}{\partial F} = -2A^T W^2 (Y - AF) Q^2 \quad (\text{B.14})$$

B.3.2 Terme de régularisation - Tikhonov

$$f_{\text{Tik}}(F) = \|S^{0.5} \times \Gamma \times F^T\|_F^2 \quad (\text{B.15})$$

$$= \text{Tr}[(\Gamma F^T)^T S (\Gamma F^T)] \quad (\text{B.16})$$

$$= \text{Tr}[F(\Gamma^T S \Gamma) F^T] \quad (\text{B.17})$$

En dérivant selon F , on obtient alors :

$$\frac{df_{\text{Tik}}(F)}{dF} = 2F(\Gamma^T S \Gamma) \quad (\text{B.18})$$

Constantes de Lipschitz des dérivées partielles de f_{RUDUR}

Les constantes de Lipschitz sont calculées afin d'accélérer le processus d'optimisation décrit dans la section 3.3.7. Le détail de leur calcul est fourni dans la présente annexe.

C.1 Constante de Lipschitz de la dérivée partielle selon F

$$\frac{\partial f_{\text{RUDUR}}(A, F)}{\partial F} = -2A^T W^2 (Y - AF) Q^2 + 2\beta F (\Gamma^T S \Gamma) \quad (\text{C.1})$$

Ainsi,

$$\left\| \frac{\partial f_{\text{RUDUR}}(A, F)}{\partial F} - \frac{\partial f_{\text{RUDUR}}(A, F')}{\partial F} \right\|_F = \|2A^T W^2 A (F - F') Q^2 + 2\beta (F - F') (\Gamma^T S \Gamma)\|_F \quad (\text{C.2})$$

Par sous multiplicativité de la norme de Frobenius et inégalité triangulaire :

$$\left\| \frac{\partial f_{\text{RUDUR}}(A, F)}{\partial F} - \frac{\partial f_{\text{RUDUR}}(A, F')}{\partial F} \right\|_F \leq (\|2A^T W^2 A\|_F \|Q^2\|_F + 2\beta \|(\Gamma^T S \Gamma)\|_F) \|F - F'\|_F \quad (\text{C.3})$$

Ce qui donne donc une majoration de la constante de Lipschitz l_F de la dérivée partielle de f_{RUDUR} par rapport à F :

$$l_F \leq \|2A^T W^2 A\|_F \|Q^2\|_F + 2\beta \|(\Gamma^T S \Gamma)\|_F \quad (\text{C.4})$$

C.2 Constante de Lipschitz de la dérivée partielle selon A

En approximant la norme l_1 lisse (c'est à dire en choisissant $\mu = 0$) on a :

$$\frac{\partial f_{\text{RUDUR}}(A, F)}{\partial A_{i,:}} = D_{i,k}^2 - 2W_{i,i}^2 (Y_{i,:} - A_{i,:} F) Q^2 F^T \quad (\text{C.5})$$

Ainsi,

$$\left\| \frac{\partial f_{\text{RUDUR}}(A, F)}{\partial A_{i,:}} - \frac{\partial f_{\text{RUDUR}}(A, F')}{\partial A_{i,:}} \right\|_F = \|2W_{i,i}^2 (A_{i,:} - A'_{i,:}) F Q^2 F^T\|_F \quad (\text{C.6})$$

On obtient alors une majoration de la constante de Lipschitz l_i de la dérivée partielle de f_{RUDUR} par rapport à la $i^{\text{ème}}$ ligne de A :

$$l_{A,i} \leq 2W_{i,i}^2 \|FQ^2F^T\|_F \quad (\text{C.7})$$

Séquences d'images

Dans cet annexe, des exemples de séquences d'images de chaque jeu de données utilisé dans cette thèse sont affichés. Les séquences sont à lire de gauche à droite puis de haut en bas.

D.1 Données synthétiques simples, sans bruit de fond

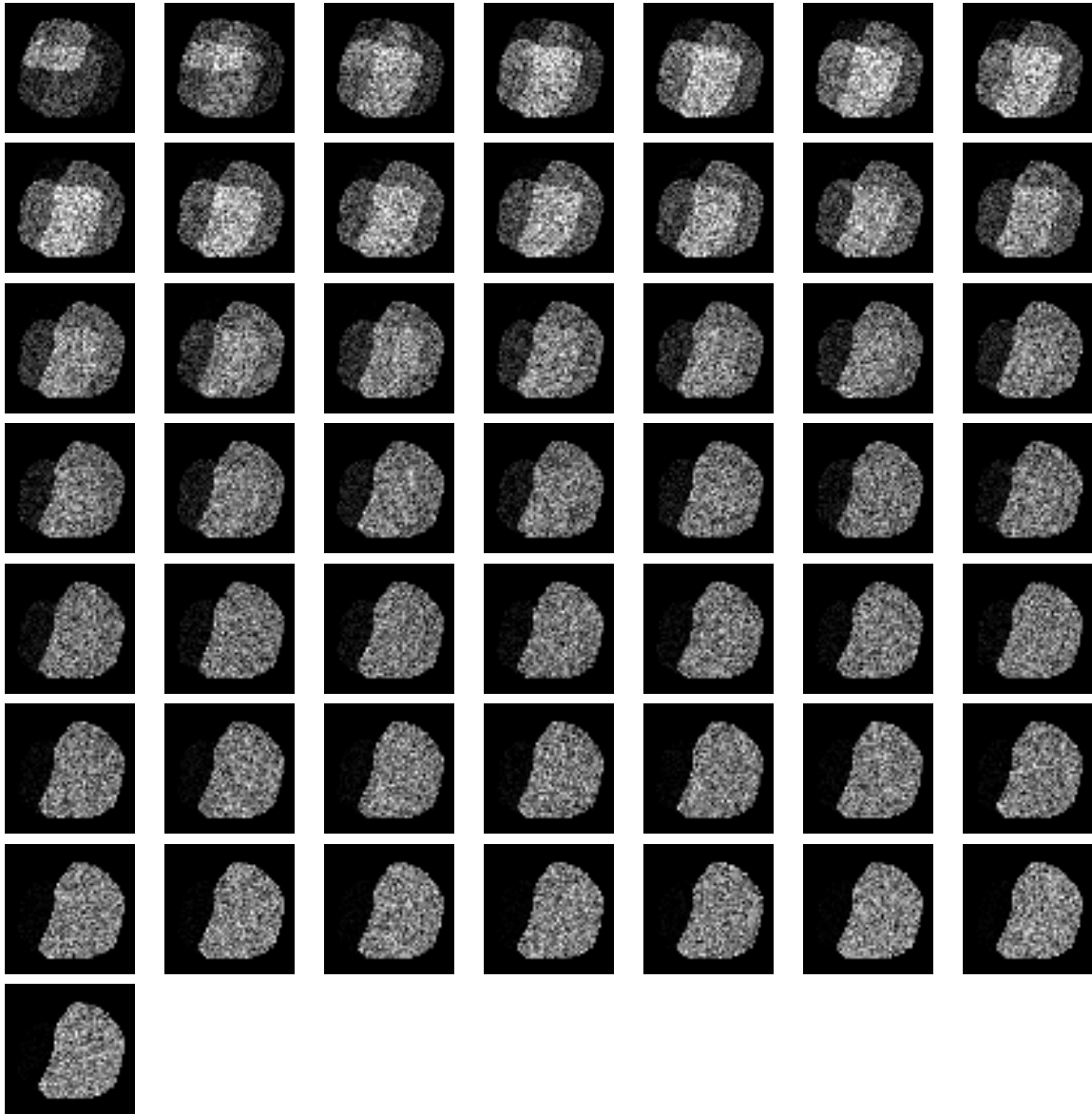


FIGURE D.1 – Exemple de séquence issue du jeu de données synthétiques simple ($\sigma = 0.5$), sans bruit de fond. Chaque séquence est composé de 50 images.

D.2 Données synthétiques simples, avec bruit de fond

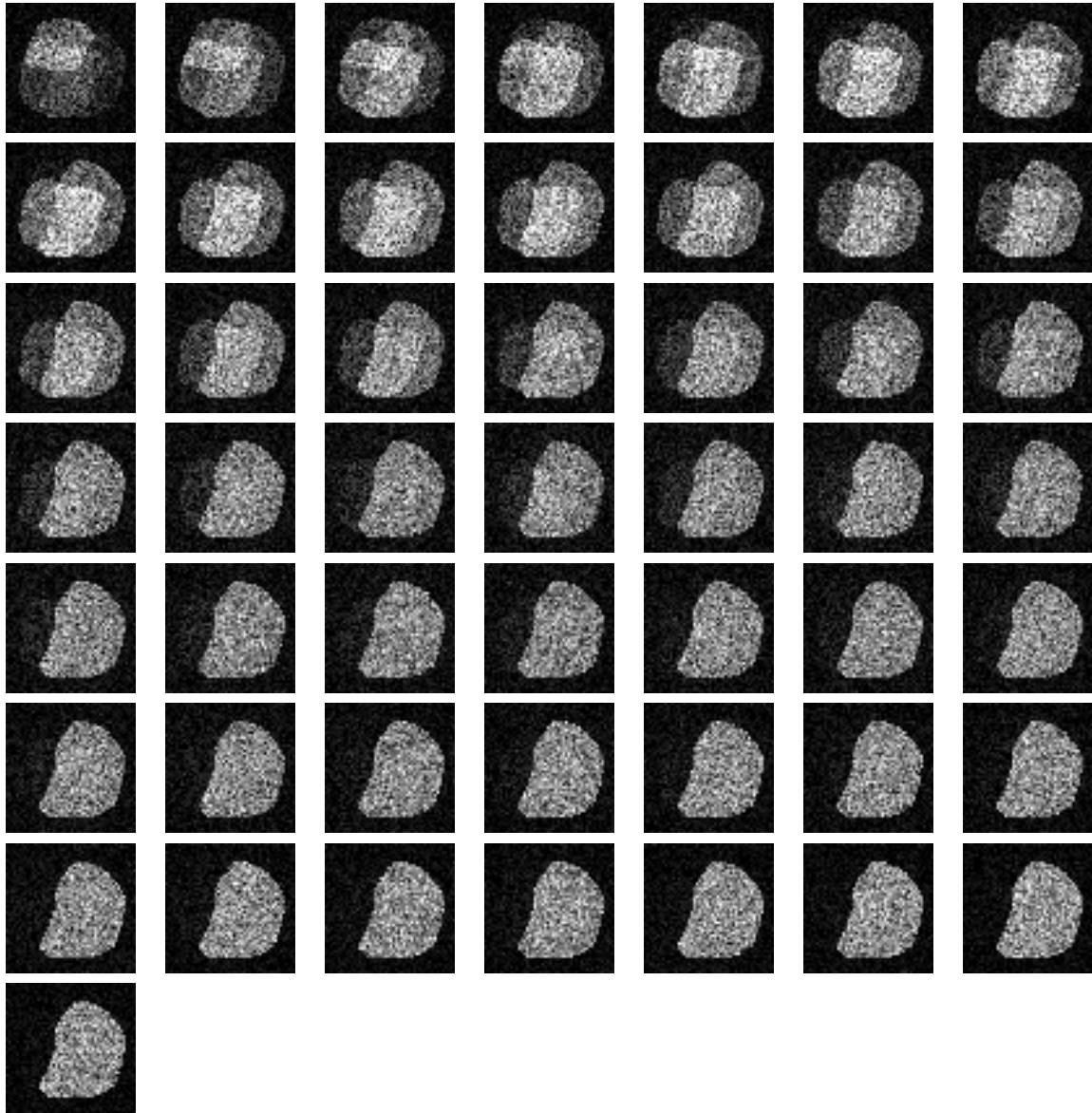


FIGURE D.2 – Exemple de séquence issue du jeu de données synthétiques simple ($\sigma = 0.5$), avec bruit de fond. Chaque séquence est composé de 50 images.

D.3 Rénographies synthétiques réalistes

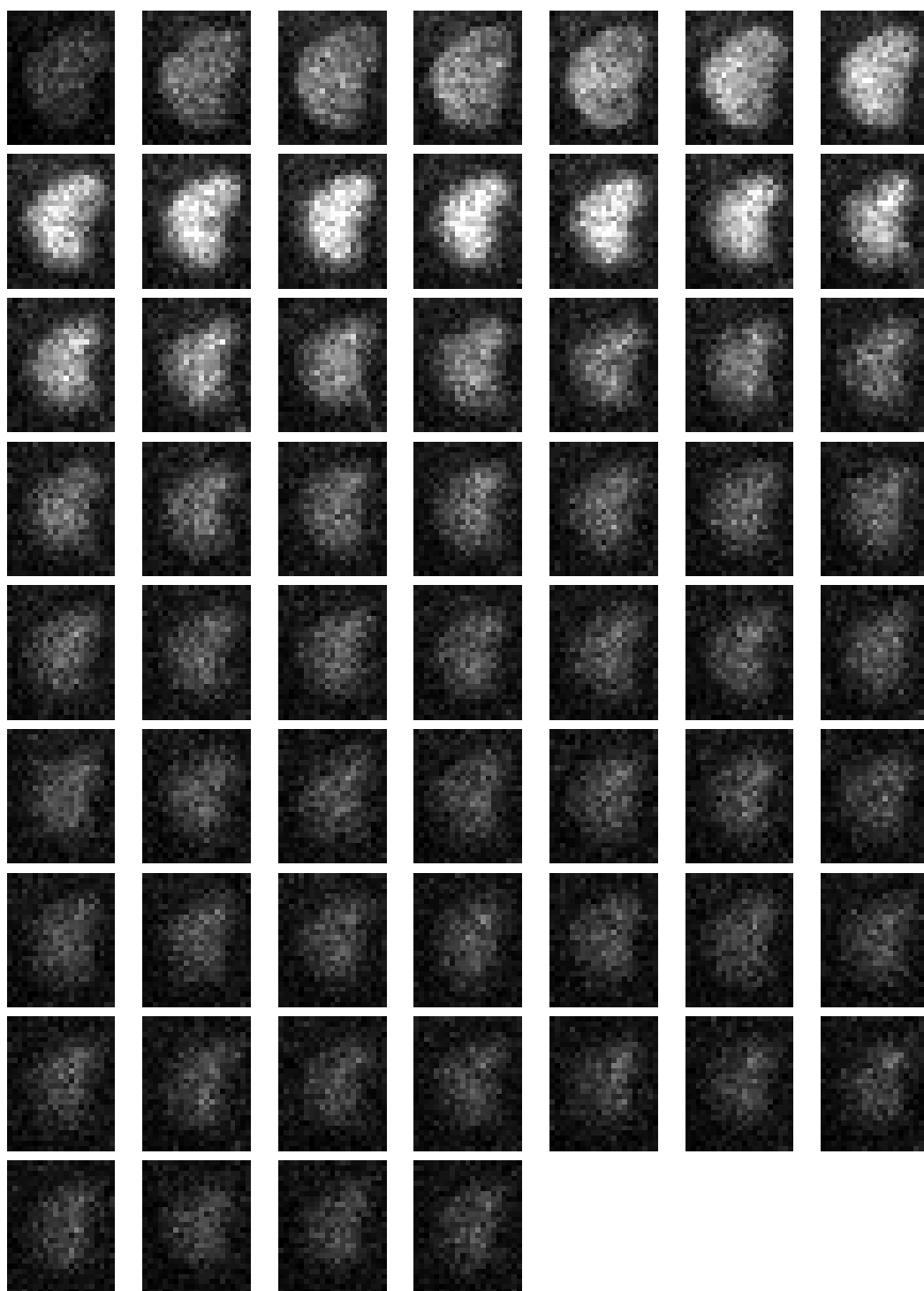


FIGURE D.3 – Exemple de séquence de rénographie synthétique réaliste. Chaque séquence est composée de 60 images.

D.4 Rénographies cliniques

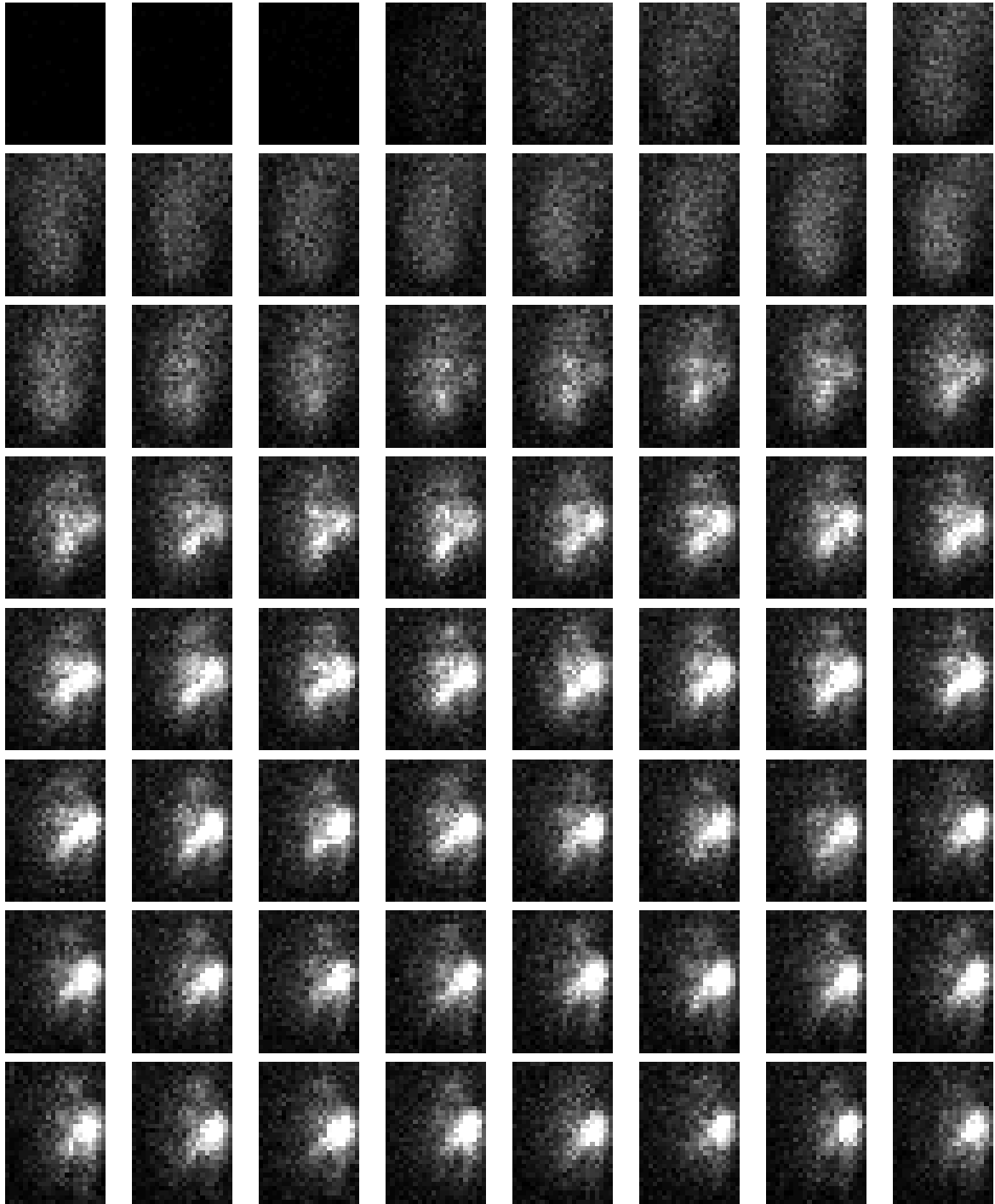


FIGURE D.4 – Exemple de séquence issue du jeu de rénographies clinique (rein droit). Seulement les 64 premières images sont affichés.

D.5 Rénographies cliniques

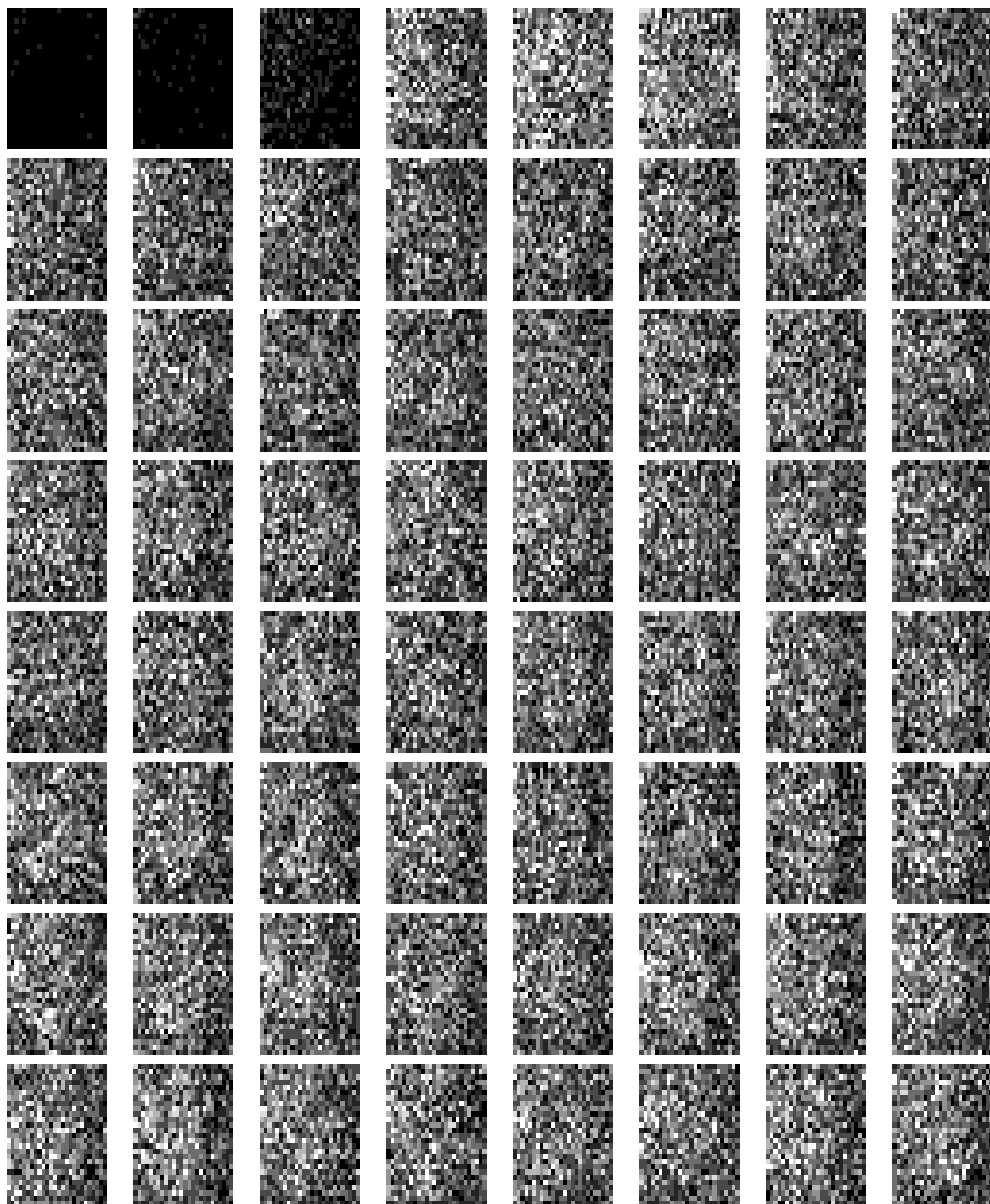


FIGURE D.5 – Exemple de séquence non retenue issue du jeu de rénographies clinique (rein gauche). Seulement les 64 premières images sont affichés.

D.6 Données synthétiques simples, avec source de faible intensité

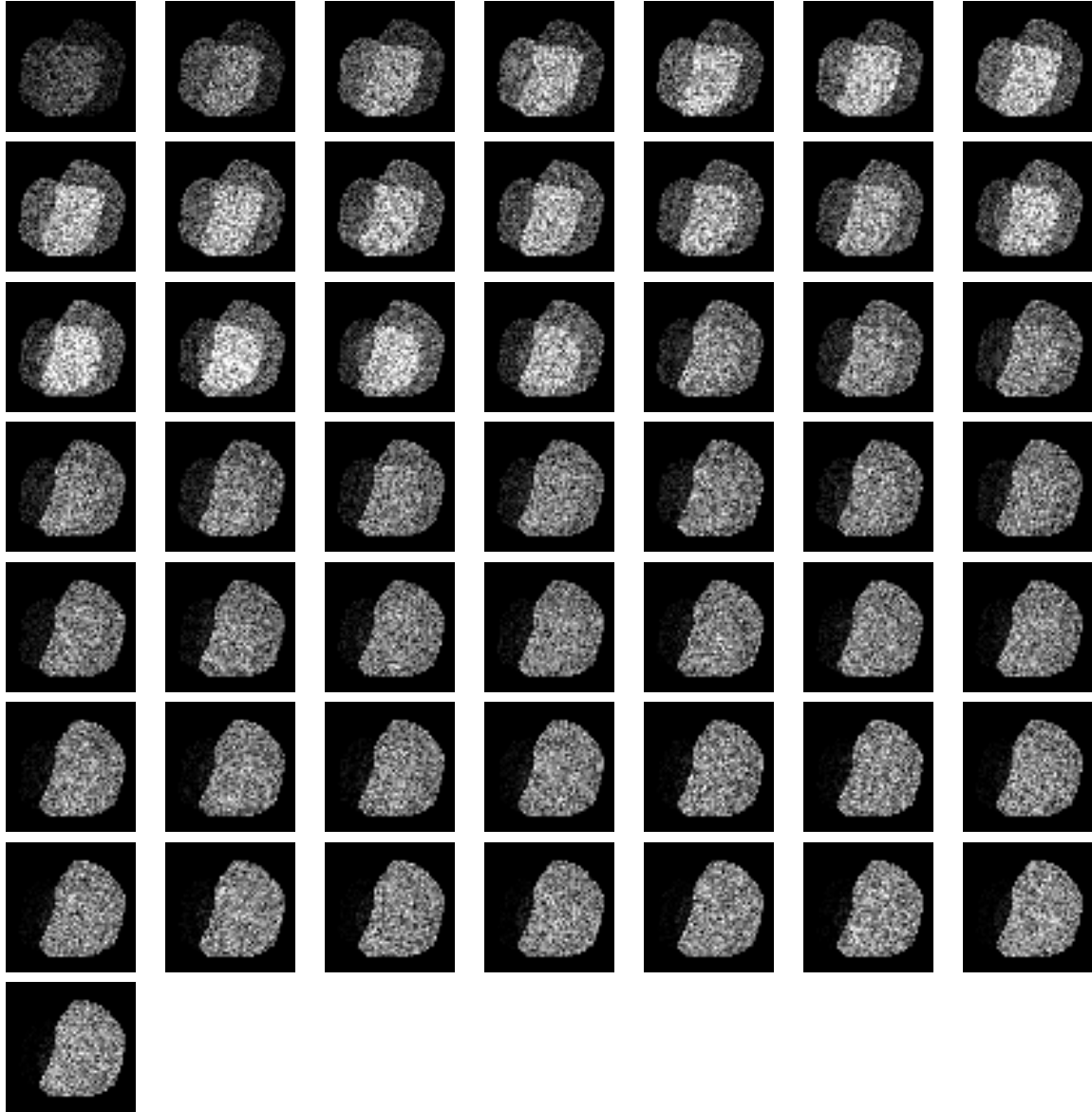


FIGURE D.6 – Exemple de séquence issue du jeu de données synthétiques simple ($\sigma = 0.5$), avec source de faible intensité. Chaque séquence est composé de 50 images.

D.7 Simulation d'une séquence 6-DIG

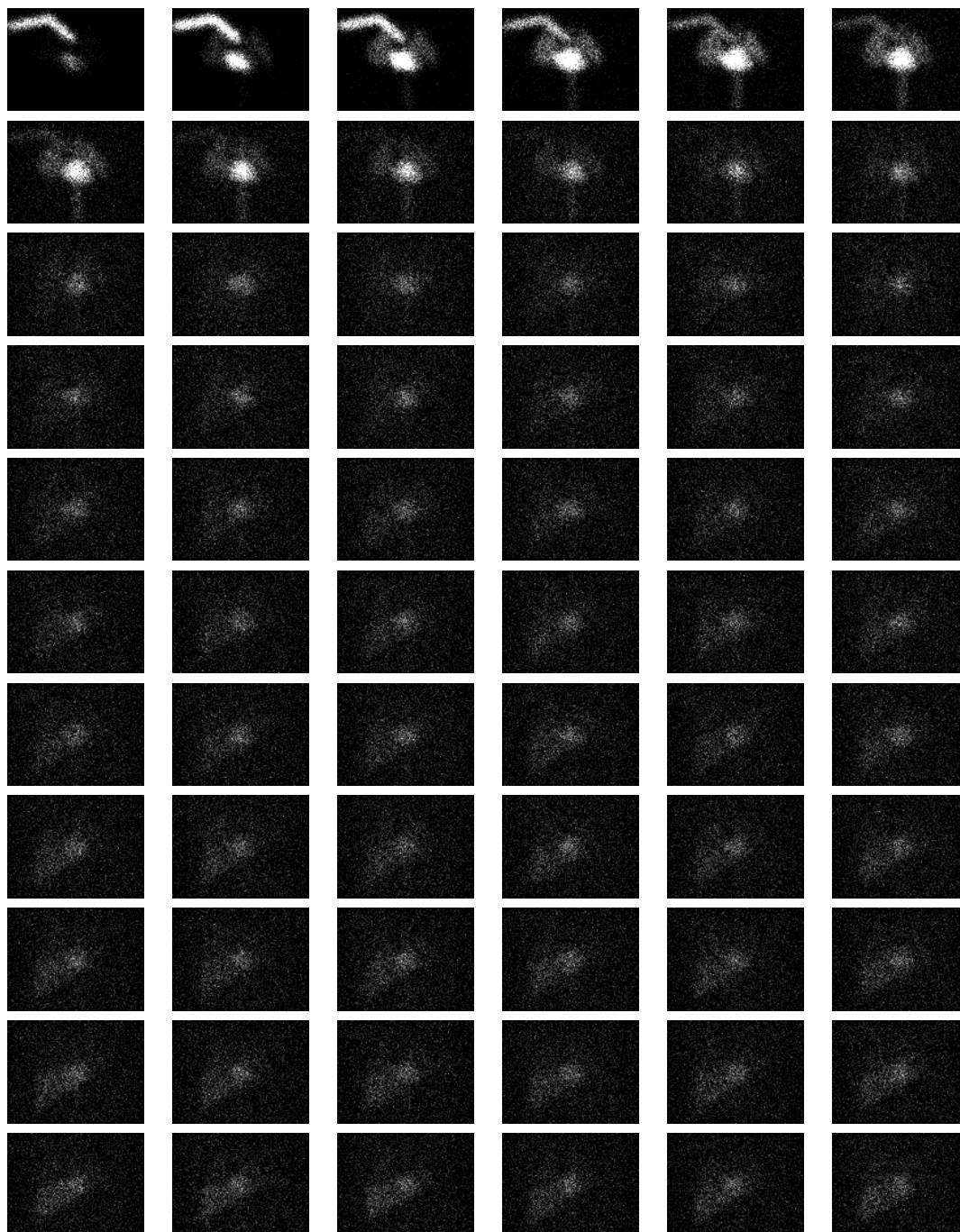


FIGURE D.7 – Exemple de séquence issue de la simulation d'une étude avec le 6-DIG. Seulement les 66 premières images sont affichées.

D.8 Exemple de séquence 6-DIG (phase basale)

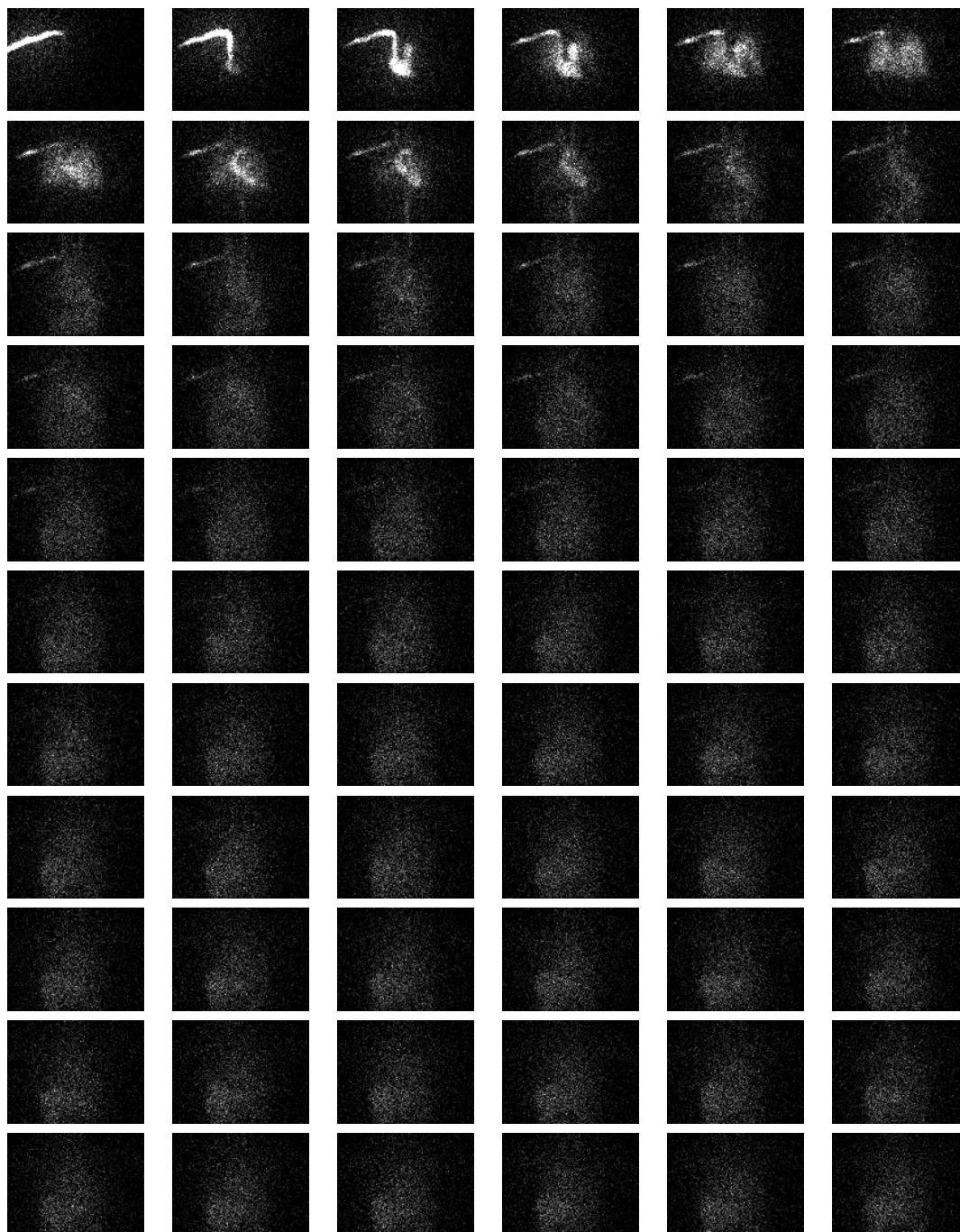


FIGURE D.8 – Exemple de séquence scintigraphique utilisant du 6-DIG (phase basale). Seulement les 66 premières images sont affichées.

Bibliographie

- [Ans48] Francis J ANSCOMBE. “The transformation of Poisson, binomial and negative-binomial data”. In : *Biometrika* 35.3/4 (1948), p. 246–254 (cf. p. 52).
- [ASJ86] John B ADAMS, Milton O SMITH et Paul E JOHNSON. “Spectral mixture modeling : A new analysis of rock and soil types at the Viking Lander 1 site”. In : *Journal of Geophysical Research : Solid Earth* 91.B8 (1986), p. 8098–8112 (cf. p. 49).
- [Aur89] André AURENGO. “Analyse factorielle des séquences d’images en médecine nucléaire”. Thèse de doct. Paris 11, 1989 (cf. p. 50).
- [Bac+10] R BACON et al. “The MUSE second-generation VLT instrument”. In : *SPIE Astronomical Telescopes+ Instrumentation*. International Society for Optics et Photonics. 2010, p. 773508–773508 (cf. p. 25, 135).
- [Bac+15] Roland BACON et al. “The MUSE 3D view of the hubble deep field south”. In : *Astronomy & Astrophysics* 575 (2015), A75 (cf. p. 25).
- [Bai+05] Dale L BAILEY et al. *Positron emission tomography*. Springer, 2005 (cf. p. 9).
- [Bar+75] DC BARBER et al. “Principal components : a new approach to the analysis of gastric emptying”. In : *Dynamic studies with radioisotopes in medicine 1974*. 1975 (cf. p. 48).
- [Bar80] DC BARBER. “The use of principal components in the quantitative analysis of gamma camera dynamic studies”. In : *Physics in medicine and biology* 25.2 (1980), p. 283 (cf. p. 16, 29, 35, 48, 88).
- [Baz+80] JP BAZIN et al. “Factor analysis of dynamic scintigraphic data as a modeling method. An application to the detection of metastases”. In : *Information processing in medical imaging, Paris, 2-6 July 1979*. 1980 (cf. p. 48).
- [BD+12] José M BIOUCAS-DIAS et al. “Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches”. In : *IEEE journal of selected topics in applied earth observations and remote sensing* 5.2 (2012), p. 354–379 (cf. p. 24, 27, 68).
- [BD09] José M BIOUCAS-DIAS. “A variable splitting augmented Lagrangian approach to linear spectral unmixing”. In : *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing, 2009. WHISPERS’09. First Workshop on*. IEEE. 2009, p. 1–4 (cf. p. 47).
- [Ben+93] H BENALI et al. “A statistical model for the determination of the optimal metric in factor analysis of medical image sequences (FAMIS)”. In : *Physics in medicine and biology* 38.8 (1993), p. 1065 (cf. p. 50, 53, 70).
- [Ben+94a] Habib BENALI et al. “CAMIS : clustering algorithm for medical image sequences using a mutual nearest neighbor criterion”. In : *Mathematical Methods in Medical Imaging III. Proc. SPIE* 2299 (1994), p. 336–347 (cf. p. 50).

- [Ben+94b] Habib BENALI et al. “Foundations of factor analysis of medical image sequences : a unified approach and some practical implications”. In : *Image and Vision Computing* 12.6 (1994), p. 375–385 (cf. p. 50, 53, 71, 87, 88, 92).
- [Ben92] J-P BENZÉCRI. *Correspondence analysis handbook*. Marcel Dekker, 1992 (cf. p. 48, 50, 53).
- [Ber+04] Mark BERMAN et al. “ICE : A statistical approach to identifying endmembers in hyperspectral images”. In : *IEEE transactions on Geoscience and Remote Sensing* 42.10 (2004), p. 2085–2095 (cf. p. 47).
- [Ber+07] Michael W BERRY et al. “Algorithms and applications for approximate nonnegative matrix factorization”. In : *Computational statistics & data analysis* 52.1 (2007), p. 155–173 (cf. p. 44, 53, 82).
- [Ber+10] DS BERMAN et al. “SPECT/PET myocardial perfusion imaging versus coronary CT angiography in patients with known or suspected CAD.” In : *The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN)[and] the International Association of Radiopharmacology (IAR),[and] Section of the Society of...* 54.2 (2010), p. 177–200 (cf. p. 10).
- [Ber+99] H BERGMANN et al. “Improved automatic separation of renal parenchyma and pelvis in dynamic renal scintigraphy using fuzzy regions of interest”. In : *European Journal of Nuclear Medicine and Molecular Imaging* 26.8 (1999), p. 837–843 (cf. p. 51).
- [BGL13] Gustav BROLIN, Katarina Sjögren GLEISNER et Michael LJUNGBERG. “Dynamic 99mTc-MAG3 renography : images for quality control obtained by combining pharmacokinetic modelling, an anthropomorphic computer phantom and Monte Carlo simulated scintillation camera imaging”. In : *Physics in medicine and biology* 58.10 (2013), p. 3145 (cf. p. 100).
- [BKG95] Joseph W BOARDMAN, Fred A KRUSE et Robert O GREEN. “Mapping target signatures via partial unmixing of AVIRIS data”. In : (1995) (cf. p. 45).
- [BPDP70] JP BAZIN, C PARMENTIER et R DI PAOLA. “Analyse des compartiments et analyse factorielle dans l’étude de la cinétique du fer”. In : *dynamic studies with radioisotopes in medicine* (1970), p. 489 (cf. p. 48).
- [BR13] Gilles BARONE-ROCHETTE. “Etudes biologiques et cliniques du 6-Déoxy-6-Iodo-D-Glucose : Traceur du transport du glucose marqué à l’iode 123 et marqueur de l’insulinorésistance chez l’homme”. Thèse de doct. Université Grenoble Alpes, 2013 (cf. p. 17, 19, 20).
- [Bre13] Richard P BRENT. *Algorithms for minimization without derivatives*. Courier Corporation, 2013 (cf. p. 55).
- [Bri+07] Arnaud BRIAT et al. “In vivo assessment of cardiac insulin resistance by nuclear probes using an iodinated tracer of glucose transport”. In : *European journal of nuclear medicine and molecular imaging* 34.11 (2007), p. 1756–1764 (cf. p. 19).

- [Bro97] Rasmus BRO. “PARAFAC. Tutorial and applications”. In : *Chemometrics and intelligent laboratory systems* 38.2 (1997), p. 149–171 (cf. p. 25, 58, 61).
- [Buv+93] J BUVAT et al. “Target apex-seeking in factor analysis of medical image sequences”. In : *Physics in medicine and biology* 38.1 (1993), p. 123 (cf. p. 49, 50).
- [Cec+15] Diego CECCHIN et al. “Analytical and experimental FWHM of a gamma camera : theoretical and practical issues”. In : *PeerJ* 3 (2015), e722 (cf. p. 13).
- [CFC15] Jeremy COHEN, Rodrigo Cabral FARIAS et Pierre COMON. “Fast decomposition of large nonnegative tensors”. In : *IEEE Signal Processing Letters* 22.7 (2015), p. 862–866 (cf. p. 58, 61).
- [CGK08] Meltem CAGLAR, Gonca Kara GEDIK et Erdem KARABULUT. “Differential renal function estimation by dynamic renal scintigraphy : influence of background definition and radiopharmaceutical”. In : *Nuclear medicine communications* 29.11 (2008), p. 1002–1005 (cf. p. 48).
- [Cha+06] C-I CHANG et al. “A new growing method for simplex-based endmember extraction algorithm”. In : *IEEE Transactions on Geoscience and Remote Sensing* 44.10 (2006), p. 2804–2819 (cf. p. 46).
- [Cha+11] Tsung-Han CHAN et al. “A simplex volume maximization framework for hyperspectral endmember extraction”. In : *IEEE Transactions on Geoscience and Remote Sensing* 49.11 (2011), p. 4177–4193 (cf. p. 46).
- [Che53] E Colin CHERRY. “Some experiments on the recognition of speech, with one and with two ears”. In : *The Journal of the acoustical society of America* 25.5 (1953), p. 975–979 (cf. p. 24).
- [CLDA09] Pierre COMON, Xavier LUCIANI et André LF DE ALMEIDA. “Tensor decompositions, alternating least squares and other tales”. In : *Journal of chemometrics* 23.7-8 (2009), p. 393–405 (cf. p. 58, 61).
- [Coc+03] Emmanuel COCHE et al. “Diagnosis of acute pulmonary embolism in outpatients : comparison of thin-collimation multi-detector row spiral CT and planar ventilation-perfusion scintigraphy”. In : *Radiology* 229.3 (2003), p. 757–765 (cf. p. 10).
- [Com14] Pierre COMON. “Tensors : a brief introduction”. In : *IEEE Signal Processing Magazine* 31.3 (2014), p. 44–53 (cf. p. 59, 61).
- [Com94] Pierre COMON. “Independent component analysis, a new concept ?” In : *Signal processing* 36.3 (1994), p. 287–314 (cf. p. 29).
- [Cos85] PS COSGRIFF. “Region of interest, ROI, analysis-Confidence in derived parameters”. In : *Nuclear medicine communications* 6.5 (1985), p. 305–309 (cf. p. 15, 48).
- [CP09] Andrzej CICHOCKI et Anh-Huy PHAN. “Fast local algorithms for large scale nonnegative matrix and tensor factorizations”. In : *IEICE transactions on fundamentals of electronics, communications and computer sciences* 92.3 (2009), p. 708–721 (cf. p. 44).

- [Cra94] Maurice D CRAIG. “Minimum-volume transforms for remotely sensed data”. In : *IEEE Transactions on Geoscience and Remote Sensing* 32.3 (1994), p. 542–552 (cf. p. 47, 49).
- [Dab+07] Kostadin DABOV et al. “Image denoising by sparse 3-D transform-domain collaborative filtering”. In : *IEEE Transactions on image processing* 16.8 (2007), p. 2080–2095 (cf. p. 52).
- [DD10] Brian D’ALESSANDRO et Atam P DHAWAN. “Depth-dependent hemoglobin analysis from multispectral transillumination images”. In : *IEEE Transactions on Biomedical Engineering* 57.10 (2010), p. 2568–2571 (cf. p. 25).
- [DDL13a] Ignat DOMANOV et Lieven DE LATHAUWER. “On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I : Basic results and uniqueness of one factor matrix”. In : *SIAM Journal on Matrix Analysis and Applications* 34.3 (2013), p. 855–875 (cf. p. 62).
- [DDL13b] Ignat DOMANOV et Lieven DE LATHAUWER. “On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II : Uniqueness of the overall decomposition”. In : *SIAM Journal on Matrix Analysis and Applications* 34.3 (2013), p. 876–903 (cf. p. 62).
- [Dev08] Karthik DEVARAJAN. “Nonnegative matrix factorization : an analytical and interpretive tool in computational biology”. In : *PLoS computational biology* 4.7 (2008), e1000029 (cf. p. 74).
- [Dha+09] Atam P DHAWAN et al. “Multispectral optical imaging of skin-lesions for detection of malignant melanomas”. In : *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE. 2009, p. 5352–5355 (cf. p. 25).
- [DLDMV95] Lieven DE LATHAUWER, Bart DE MOOR et Joos VANDEWALLE. “Fetal electrocardiogram extraction by source subspace separation”. In : *IEEE SP/Athos Workshop on Higher-Order Statistics*. 1995, p. 134 (cf. p. 25).
- [Dow+10] Nicholas DOWSON et al. “Joint factor and kinetic analysis of dynamic FDOPA PET scans of brain cancer patients”. In : *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010* (2010), p. 185–192 (cf. p. 50, 51).
- [DP+82] R DI PAOLA et al. “Handling of dynamic sequences in nuclear medicine”. In : *IEEE Transactions on Nuclear Science* 29.4 (1982), p. 1310–1321 (cf. p. 25, 34, 38, 48).
- [DS04] David DONOHO et Victoria STODDEN. “When does non-negative matrix factorization give a correct decomposition into parts?” In : *Advances in neural information processing systems*. 2004, p. 1141–1148 (cf. p. 38, 41).
- [EF+05] Georges EL FAKHRI et al. “Quantitative dynamic cardiac 82Rb PET using generalized factor and compartment analyses”. In : *Journal of nuclear medicine* 46.8 (2005), p. 1264–1271 (cf. p. 49).

- [Eri+89] Johan ERIKSSON et al. "Early metabolic defects in persons at increased risk for non-insulin-dependent diabetes mellitus". In : *New England Journal of Medicine* 321.6 (1989), p. 337–343 (cf. p. 16, 18).
- [ES+06] Einat EVEN-SAPIR et al. "The detection of bone metastases in patients with high-risk prostate cancer : 99mTc-MDP Planar bone scintigraphy, single-and multi-field-of-view SPECT, 18F-fluoride PET, and 18F-fluoride PET/CT". In : *Journal of Nuclear Medicine* 47.2 (2006), p. 287–297 (cf. p. 7).
- [Far+04] Dario FARINA et al. "Blind separation of linear instantaneous mixtures of nonstationary surface myoelectric signals". In : *IEEE Transactions on Biomedical Engineering* 51.9 (2004), p. 1555–1567 (cf. p. 25).
- [Faz+09] Reza FAZEL et al. "Exposure to low-dose ionizing radiation from medical imaging procedures". In : *New England Journal of Medicine* 361.9 (2009), p. 849–857 (cf. p. 7).
- [FBDP89] Frederique FROUIN, Jean-Pierre BAZIN et Robert DI PAOLA. "Image sequence processing using factor analysis and compartmental modelling". In : *1989 Intl Congress on Optical Science and Engineering*. International Society for Optics et Photonics. 1989, p. 37–45 (cf. p. 50).
- [FDM17] Marc FILIPPI, Michel DESVIGNES et Eric MOISAN. "Robust Unmixing of Dynamic Sequences Using Regions of Interest (RUDUR)". In : *IEEE transactions on medical imaging* (2017) (cf. p. 51, 70, 73, 77, 81, 135).
- [Fil+15] M FILIPPI et al. "A priori spatiaux et analyse factorielle de séquences scintigraphiques". In : *GRETSI*. 2015 (cf. p. 70, 72–74, 135).
- [Fil+16] Marc FILIPPI et al. "Factor Analysis of Dynamic Sequence with Spatial Prior for 2D Cardiac Spect Sequences Analysis". In : *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer. 2016, p. 228–237 (cf. p. 70, 73, 77, 135).
- [Fre+01] Wendy L FREEDMAN et al. "Final results from the Hubble Space Telescope key project to measure the Hubble constant". In : *The Astrophysical Journal* 553.1 (2001), p. 47 (cf. p. 25).
- [Fro+92] Frédérique FROUIN et al. "FAMIS : a software package for functional feature extraction from biomedical multidimensional images". In : *Computerized medical imaging and graphics* 16.2 (1992), p. 81–91 (cf. p. 51).
- [Fro+93] Frédérique FROUIN et al. "Extraction of functional volumes from medical dynamic volumetric data sets". In : *Computerized medical imaging and graphics* 17.4-5 (1993), p. 397–404 (cf. p. 51).
- [Fro+99] F FROUIN et al. "Spatial regularization applied to factor analysis of medical image sequences (FAMIS)". In : *Physics in medicine and biology* 44.9 (1999), p. 2289 (cf. p. 50).

- [Gal+78] Brian M GALLAGHER et al. “Metabolic trapping as a principle of radiopharmaceutical design : some factors responsible for the biodistribution of [18F] 2-deoxy-2-fluoro-D-glucose”. In : *Journal of Nuclear Medicine* 19.10 (1978), p. 1154–1161 (cf. p. 7, 10, 19).
- [GGC01] Roger N GUNN, Steve R GUNN et Vincent J CUNNINGHAM. “Positron emission tomography compartmental models”. In : *Journal of Cerebral Blood Flow & Metabolism* 21.6 (2001), p. 635–652 (cf. p. 13).
- [Gho+10] Nina GHOSH et al. “Assessment of myocardial ischaemia and viability : role of positron emission tomography”. In : *European heart journal* 31.24 (2010), p. 2984–2995 (cf. p. 10).
- [GL06] Rémi GRIBONVAL et Sylvain LESAGE. “A survey of sparse component analysis for blind source separation : principles, perspectives, and new challenges”. In : *ESANN’06 proceedings-14th European Symposium on Artificial Neural Networks*. d-side publi. 2006, p. 323–330 (cf. p. 29).
- [GPS11] Isky GORDON, Amy PIEPSZ et Rune SIXT. “Guidelines for standard and diuretic renogram in children”. In : *European journal of nuclear medicine and molecular imaging* 38.6 (2011), p. 1175 (cf. p. 15, 35, 103, 105).
- [GRH04] John GRUNINGER, Anthony J RATKOWSKI et Michael L HOKE. *The sequential maximum angle convex cone (SMACC) endmember model*. Rapp. tech. SPECTRAL SCIENCES INC BURLINGTON MA, 2004 (cf. p. 46).
- [Gul+10] Grant T GULLBERG et al. “Dynamic single photon emission computed tomography—basic principles and cardiac applications”. In : *Physics in medicine and biology* 55.20 (2010), R111 (cf. p. 6, 8).
- [HA84] Jeanny HÉRAULT et Bernard ANS. “Réseau de neurones à synapses modifiables : Décodage de messages sensoriels composites par apprentissage non supervisé et permanent”. In : *Comptes rendus des séances de l’Académie des sciences. Série 3, Sciences de la vie* 299.13 (1984), p. 525–528 (cf. p. 49).
- [HBB11] Brian F HUTTON, Irene BUVAT et Freek J BEEKMAN. “Review and current status of SPECT scatter correction”. In : *Physics in medicine and biology* 56.14 (2011), R85 (cf. p. 13).
- [Hen+97] Christelle HENRY et al. “[123I]-6-deoxy-6-iodo-d-glucose (6DIG) : A potential tracer of glucose transport”. In : *Nuclear medicine and biology* 24.6 (1997), p. 527–534 (cf. p. 7, 19).
- [Hen87] Ronald C HENRY. “Current factor analysis receptor models are ill-posed”. In : *Atmospheric Environment* (1967) 21.8 (1987), p. 1815–1820 (cf. p. 39).
- [Hit27] Frank L HITCHCOCK. “The expression of a tensor or a polyadic as a sum of products”. In : *Studies in Applied Mathematics* 6.1-4 (1927), p. 164–189 (cf. p. 58).
- [HKK98] Toshihiko HARA, Noboru KOSAKA et Hiroichi KISHI. “PET imaging of prostate cancer using carbon-11-choline”. In : *The Journal of Nuclear Medicine* 39.6 (1998), p. 990 (cf. p. 10).

- [Hou80] AS HOUSTON. “Classification of dynamic function studies in nuclear medicine. Does features extraction help”. In : *Information processing in medical imaging, Paris, 2-6 July 1979*. 1980 (cf. p. 48).
- [Hou84] AS HOUSTON. “The effect of apex-finding errors on factor images obtained from factor analysis and oblique transformation (nuclear medicine)”. In : *Physics in medicine and biology* 29.9 (1984), p. 1109 (cf. p. 39, 41, 49).
- [Hoy02] Patrik O HOYER. “Non-negative sparse coding”. In : *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*. IEEE. 2002, p. 557–565 (cf. p. 43).
- [Hoy04] Patrik O HOYER. “Non-negative matrix factorization with sparseness constraints”. In : *Journal of machine learning research* 5.Nov (2004), p. 1457–1469 (cf. p. 43).
- [HS06] Matthias HEILER et Christoph SCHNÖRR. “Learning sparse representations by non-negative matrix factorization and sequential cone programming”. In : *Journal of Machine Learning Research* 7.Jul (2006), p. 1385–1407 (cf. p. 44).
- [HS14] Kejun HUANG et Nicholas D SIDIROPOULOS. “Putting nonnegative matrix factorization to the test : a tutorial derivation of pertinent Cramer—Rao bounds and performance benchmarking”. In : *IEEE Signal Processing Magazine* 31.3 (2014), p. 76–86 (cf. p. 28).
- [HSS14] Kejun HUANG, Nicholas D SIDIROPOULOS et Ananthram SWAMI. “Non-negative matrix factorization revisited : Uniqueness and algorithm for symmetric decomposition”. In : *IEEE Transactions on Signal Processing* 62.1 (2014), p. 211–224 (cf. p. 41, 42).
- [Jol02] Ian JOLLIFFE. *Principal component analysis*. Wiley Online Library, 2002 (cf. p. 29).
- [JT00] Christian JUTTEN et Anisse TALEB. “Source separation : from dusk till dawn”. In : *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA2000)*. 2000, p. 15–26 (cf. p. 49).
- [KAS07] Amar KACHENOURA, Laurent ALBERA et Lotfi SENHADJI. “Séparation aveugle de sources en ingénierie biomédicale”. In : *IRBM* 28.1 (2007), p. 20–34 (cf. p. 25, 29).
- [KBA+08] R KLEIN, RS BEANLANDS, A ADLER et al. “Model-based factor analysis of dynamic sequences of cardiac positron emission tomography”. In : *Nuclear Science Symposium Conference Record, 2008. NSS’08. IEEE*. IEEE. 2008, p. 5198–5202 (cf. p. 50, 51).
- [Ken+75] RW KENNY et al. “Deconvolution analysis of the scintillation camera renogram”. In : *The British journal of radiology* 48.570 (1975), p. 481–486 (cf. p. 48).
- [Kie00] Henk AL KIERS. “Towards a standardized notation and terminology in multi-way analysis”. In : *Journal of chemometrics* 14.3 (2000), p. 105–122 (cf. p. 60).

- [Kle+07] Ran KLEIN et al. “A minimal factor overlap method for resolving ambiguity in factor analysis of dynamic cardiac PET”. In : *Nuclear Science Symposium Conference Record, 2007. NSS’07. IEEE*. T. 5. IEEE. 2007, p. 3268–3272 (cf. p. 49).
- [KM02] Nirmal KESHA et John F MUSTARD. “Spectral unmixing”. In : *IEEE signal processing magazine* 19.1 (2002), p. 44–57 (cf. p. 24, 25).
- [KMPL81] G KLOSTER, C MÜLLER-PLATZ et P LAUFER. “3-[11C]-methyl-D-glucose, a potential agent for regional cerebral glucose utilization studies : Synthesis, chromatography and tissue distribution in mice”. In : *Journal of Labelled Compounds and Radiopharmaceuticals* 18.6 (1981), p. 855–863 (cf. p. 7, 19).
- [KP11] Jingu KIM et Haesun PARK. “Fast nonnegative matrix factorization : An active-set-like method and comparisons”. In : *SIAM Journal on Scientific Computing* 33.6 (2011), p. 3261–3281 (cf. p. 44).
- [Kra+93] W KRAMER et al. “Follow-up after kidney transplantation. Sequential functional scintigraphy with technetium-99m-DTPA or technetium-99m-MAG3”. In : *Der Urologe. Aug. A* 32.2 (1993), p. 115–120 (cf. p. 10).
- [Kru77] Joseph B KRUSKAL. “Three-way arrays : rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics”. In : *Linear algebra and its applications* 18.2 (1977), p. 95–138 (cf. p. 62).
- [Lau+08] Hans LAURBERG et al. “Theorems on positive data : On the uniqueness of NMF”. In : *Computational intelligence and neuroscience* 2008 (2008) (cf. p. 41, 42).
- [LBD08] Jun LI et José M BIOUCAS-DIAS. “Minimum volume simplex analysis : A fast algorithm to unmix hyperspectral data”. In : *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*. T. 3. IEEE. 2008, p. III–250 (cf. p. 47).
- [LC09] Lek-Heng LIM et Pierre COMON. “Nonnegative approximations of nonnegative tensors”. In : *Journal of chemometrics* 23.7-8 (2009), p. 432–441 (cf. p. 59, 62).
- [Lin+15] Chia-Hsiang LIN et al. “Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing : The no-pure-pixel case”. In : *IEEE Transactions on Geoscience and Remote Sensing* 53.10 (2015), p. 5530–5546 (cf. p. 46).
- [Lin07] Chih-Jen LIN. “Projected gradient methods for nonnegative matrix factorization”. In : *Neural computation* 19.10 (2007), p. 2756–2779 (cf. p. 44).
- [Lju12] Michael LJUNGBERG. “The SIMIND Monte Carlo program”. In : *Monte Carlo calculations in nuclear medicine : Applications in diagnostic imaging* 20126027 (2012) (cf. p. 100).
- [LS01] Daniel D LEE et H Sebastian SEUNG. “Algorithms for non-negative matrix factorization”. In : *Advances in neural information processing systems*. 2001, p. 556–562 (cf. p. 29, 44, 49, 61, 74).

- [LZL03] Weixiang LIU, Nanning ZHENG et Xiaofeng LU. “Non-negative matrix factorization for visual coding”. In : *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP’03). 2003 IEEE International Conference on*. T. 3. IEEE. 2003, p. III–293 (cf. p. 43).
- [Ma+15] Wing-Kin MA et al. “When can the minimum volume enclosing simplex identify the endmembers correctly when there is no pure pixel ?” In : *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2015 7th Workshop on*. IEEE. 2015 (cf. p. 46).
- [MAG93] B MOURATIDIS, JM ASH et DL GILDAY. “Comparison of planar and SPECT 99Tcm-DMSA scintigraphy for the detection of renal cortical defects in children.” In : *Nuclear medicine communications* 14.2 (1993), p. 82–86 (cf. p. 10).
- [Meg+11] Inès MEGANEM et al. “Extraction of stellar spectra from dense fields in hyperspectral MUSE data cubes using non-negative matrix factorization”. In : *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*. IEEE. 2011, p. 1–4 (cf. p. 25).
- [MF11] Markku MAKITALO et Alessandro FOI. “Optimal inversion of the Anscombe transformation in low-count Poisson image denoising”. In : *IEEE Transactions on Image Processing* 20.1 (2011), p. 99–109 (cf. p. 52).
- [Mig99] O MIGNECO. “Application du recalage d’images en médecine nucléaire”. In : *Revue de l’Acomen* 5 (1999), p. 180 (cf. p. 32).
- [MQ07] Lidan MIAO et Hairong QI. “Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization”. In : *IEEE Transactions on Geoscience and Remote Sensing* 45.3 (2007), p. 765–777 (cf. p. 47).
- [MS11] Sören MATTSSON et Marcus SÖDERBERG. “Radiation dose management in CT, SPECT/CT and PET/CT techniques”. In : *Radiation protection dosimetry* 147.1-2 (2011), p. 13–21 (cf. p. 12).
- [NB85] KS NIJRAN et DC BARBER. “Towards automatic analysis of dynamic radionuclide studies using principal-components factor analysis”. In : *Physics in medicine and biology* 30.12 (1985), p. 1315 (cf. p. 48, 49).
- [NB86] KS NIJRAN et DC BARBER. “Factor analysis of dynamic function studies using a priori physiological information (nuclear medicine)”. In : *Physics in medicine and biology* 31.10 (1986), p. 1107 (cf. p. 49).
- [NB88] KS NIJRAN et DC BARBER. “The importance of constraints in factor analysis of dynamic studies”. In : *Information processing in medical imaging*. Springer. 1988, p. 521–529 (cf. p. 30, 32, 50, 71, 72).
- [NC74] DB NASH et JE CONEL. “Spectral reflectance systematics for mixtures of powdered hypersthene, labradorite, and ilmenite”. In : *Journal of Geophysical Research* 79.11 (1974), p. 1615–1621 (cf. p. 27).

- [ND05] José MP NASCIMENTO et José MB DIAS. “Vertex component analysis : A fast algorithm to unmix hyperspectral data”. In : *IEEE transactions on Geoscience and Remote Sensing* 43.4 (2005), p. 898–910 (cf. p. 46).
- [Niu+11] Xiaofeng NIU et al. “Effects of motion, attenuation, and scatter corrections on gated cardiac SPECT reconstruction”. In : *Medical physics* 38.12 (2011), p. 6571–6584 (cf. p. 13).
- [NSK89] MASAHIKO NAKAMURA, YUTAKA SUZUKI et SHIN KOBAYASHI. “A method for recovering physiological components from dynamic radionuclide images using the maximum entropy principle : a numerical investigation”. In : *IEEE transactions on biomedical engineering* 36.9 (1989), p. 906–917 (cf. p. 49).
- [OF96] Bruno A OLSHAUSEN et David J FIELD. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”. In : *Nature* 381.6583 (1996), p. 607 (cf. p. 43).
- [Ond15] Tichý ONDŘEJ. “Bayesian Blind Source Separation in Dynamic Medical Imaging”. In : *CTU v Praze, (Praha 2015)* (2015) (cf. p. 57).
- [Opp78] BE OPPENHEIM. “Identification of renograms by factor analysis”. In : *Information Processing in Medical Imaging*. Eds. Brill and Price. ORNL/BTIC-2, Springfield (1978), p. 481–503 (cf. p. 48).
- [Oue+12] Wendyam SB OUEDRAOGO et al. “Geometrical method using simplicial cones for overdetermined nonnegative blind source separation : application to real PET images”. In : *International Conference on Latent Variable Analysis and Signal Separation*. Springer. 2012, p. 494–501 (cf. p. 47).
- [Ped+08] Michael Syskind PEDERSEN et al. “Convolutional blind source separation methods”. In : *Springer Handbook of Speech Processing*. Springer, 2008, p. 1065–1094 (cf. p. 24, 29).
- [Per+03] P PERRET et al. “Assessment of insulin sensitivity in vivo in control and diabetic mice with a radioactive tracer of glucose transport : [125I]-6-deoxy-6-iodo-D-glucose”. In : *Diabetes/metabolism research and reviews* 19.4 (2003), p. 306–312 (cf. p. 19).
- [Per+07] Pascale PERRET et al. “Assessment of insulin resistance in fructose-fed rats with 125I-6-deoxy-6-iodo-D-glucose, a new tracer of glucose transport”. In : *European journal of nuclear medicine and molecular imaging* 34.5 (2007), p. 734–744 (cf. p. 14, 16, 19, 20).
- [Per+11] P PERRET et al. “Nuclear Imaging of Glucose Transport/Metabolism—An Interesting Tool to Screen Insulin Resistance, Refine Diagnosis of Type 2 Diabetes, Understand Disease Mechanisms, and/or Evaluate New Therapies”. In : *Medical Complications of Type 2 Diabetes*. InTech, 2011 (cf. p. 19).
- [Plu03] Mark D PLUMBLEY. “Algorithms for nonnegative independent component analysis”. In : *IEEE Transactions on Neural Networks* 14.3 (2003), p. 534–543 (cf. p. 44).

- [PN04] Asit K PAUL et Hani A NABI. “Gated myocardial perfusion SPECT : basic principles, technical aspects, and clinical applications”. In : *Journal of nuclear medicine technology* 32.4 (2004), p. 179–187 (cf. p. 32).
- [PR69] Elijah POLAK et Gerard RIBIERE. “Note sur la convergence de méthodes de directions conjuguées”. In : *Revue française d’informatique et de recherche opérationnelle. Série rouge* 3.16 (1969), p. 35–43 (cf. p. 85).
- [PT94] Pentti PAATERO et Unto TAPPER. “Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values”. In : *Environmetrics* 5.2 (1994), p. 111–126 (cf. p. 29, 49).
- [QCL16a] Yang QI, Pierre COMON et Lek-Heng LIM. “Semialgebraic geometry of nonnegative tensor rank”. In : *SIAM Journal on Matrix Analysis and Applications* 37.4 (2016), p. 1556–1580 (cf. p. 58, 62).
- [QCL16b] Yang QI, Pierre COMON et Lek-Heng LIM. “Uniqueness of nonnegative tensor approximations”. In : *IEEE Transactions on Information Theory* 62.4 (2016), p. 2170–2183 (cf. p. 58, 62).
- [Qui+06] Anthony QUINN et al. *The variational Bayes method in signal processing*. Springer Science & Business Media, 2006 (cf. p. 51, 56).
- [Rag+93] Michael RAGOSTA et al. “Quantitative planar rest-redistribution 201Tl imaging in detection of myocardial viability and prediction of improvement in left ventricular function after coronary bypass surgery in patients with severely depressed left ventricular function.” In : *Circulation* 87.5 (1993), p. 1630–1641 (cf. p. 10).
- [Rea88] Gerald M REAVEN. “Role of insulin resistance in human disease”. In : *Diabetes* 37.12 (1988), p. 1595–1607 (cf. p. 18).
- [Rep+15] Audrey REPETTI et al. “Euclid in a Taxicab : Sparse Blind Deconvolution with Smoothed l1/l2 Regularization”. In : *IEEE Signal Processing Letters* 22.5 (2015), p. 539–543 (cf. p. 86).
- [Roc15] Ralph Tyrell ROCKAFELLAR. *Convex analysis*. Princeton university press, 2015 (cf. p. 41).
- [ROF92] Leonid I RUDIN, Stanley OSHER et Emad FATEMI. “Nonlinear total variation based noise removal algorithms”. In : *Physica D : Nonlinear Phenomena* 60.1-4 (1992), p. 259–268 (cf. p. 44, 53, 82).
- [Ron+75] PM RONAI et al. “Technetium-99m-pyridoxylideneglutamate : a new hepatobiliary radiopharmaceutical. II. Clinical aspects”. In : *Journal of Nuclear Medicine* 16.8 (1975), p. 728–737 (cf. p. 48).
- [SBB07] Marine SORET, Stephen L BACHARACH et Irene BUVAT. “Partial-volume effect in PET tumor imaging”. In : *Journal of Nuclear Medicine* 48.6 (2007), p. 932–945 (cf. p. 16).
- [SBG05] Age SMILDE, Rasmus BRO et Paul GELADI. *Multi-way analysis : applications in the chemical sciences*. John Wiley & Sons, 2005 (cf. p. 25).
- [Sch+77] P SCHMIDLIN et al. “Evaluation of dynamic studies by means of factor analysis”. In : *Medical radionuclide imaging*. 1977 (cf. p. 48).

- [Sch77] P SCHMIDLIN. “Factor analysis of sequence scintigrams”. In : *Fifth International Conference on Information Processing in Medical Imaging*. Oak Ridge National Laboratory, Oak Ridge, TN. 1977 (cf. p. 48).
- [SDBG99] A SITEK, EVR DI BELLA et GT GULLBERG. “Factor analysis of dynamic structures in dynamic SPECT imaging using maximum entropy”. In : *IEEE Transactions on Nuclear Science* 46.6 (1999), p. 2227–2232 (cf. p. 49).
- [Seg+10] WP SEGARS et al. “4D XCAT phantom for multimodality imaging research”. In : *Medical physics* 37.9 (2010), p. 4902–4915 (cf. p. 100).
- [SGH02] Arkadiusz SITEK, Grant T GULLBERG et Ronald H HUESMAN. “Correction for ambiguous solutions in factor analysis using a penalized least squares objective”. In : *IEEE transactions on medical imaging* 21.3 (2002), p. 216–225 (cf. p. 49, 54, 92).
- [SH05] Amnon SHASHUA et Tamir HAZAN. “Non-negative tensor factorization with applications to statistics and computer vision”. In : *Proceedings of the 22nd international conference on Machine learning*. ACM. 2005, p. 792–799 (cf. p. 58, 61).
- [Sho+73] MD SHORT et al. “Gamma-camera renography using 123I-hippuran”. In : *The British journal of radiology* 46.544 (1973), p. 289–294 (cf. p. 48).
- [Shr+17] Uttam SHRESTHA et al. “Measurement of absolute myocardial blood flow in humans using dynamic cardiac SPECT and 99m Tc-tetrofosmin : Method and validation”. In : *Journal of Nuclear Cardiology* 24.1 (2017), p. 268–277 (cf. p. 135).
- [SI89] J SHEN et GW ISRAËL. “A receptor model using a specific non-negative transformation technique for ambient aerosol”. In : *Atmospheric Environment* (1967) 23.10 (1989), p. 2289–2298 (cf. p. 47).
- [SMB98] Jean-Luc STARCK, Fionn D MURTAGH et Albert BIJAOU. *Image processing and data analysis : the multiscale approach*. Cambridge University Press, 1998 (cf. p. 52).
- [SR85] J SOLOMON et B ROCK. “Imaging spectrometry for earth remote sensing”. In : *Science* 228.4704 (1985), p. 1147–1152 (cf. p. 24).
- [SVBDL13] Laurent SORBER, Marc VAN BAREL et Lieven DE LATHAUWER. “Optimization-based algorithms for tensor decompositions : Canonical polyadic decomposition, decomposition in rank-($L_r, L_r, 1$) terms, and a new generalization”. In : *SIAM Journal on Optimization* 23.2 (2013), p. 695–720 (cf. p. 58, 61).
- [TF09] Vincent YF TAN et Cédric FÉVOTTE. “Automatic relevance determination in nonnegative matrix factorization”. In : *SPARS’09-Signal Processing with Adaptive Sparse Structured Representations*. 2009 (cf. p. 56).
- [Tik63] Andrey TIKHONOV. “Solution of incorrectly formulated problems and the regularization method”. In : *Soviet Meth. Dokl.* 4 (1963), p. 1035–1038 (cf. p. 44, 53, 82).

- [TŠ15] Ondrej TICHÝ et Václav ŠMÍDL. “Bayesian blind separation and deconvolution of dynamic image sequences using sparsity priors”. In : *IEEE transactions on medical imaging* 34.1 (2015), p. 258–266 (cf. p. 51, 57, 92, 93).
- [Vav09] Stephen A VAVASIS. “On the complexity of nonnegative matrix factorization”. In : *SIAM Journal on Optimization* 20.3 (2009), p. 1364–1377 (cf. p. 42).
- [VD+91] M VAN DAELE et al. “A new vertex-finding algorithm for the oblique rotation step in factor analysis”. In : *Physics in Medicine and Biology* 36.1 (1991), p. 77 (cf. p. 49).
- [Vid00] Poul VIDEBECH. “PET measurements of brain glucose metabolism and blood flow in major depressive disorder : a critical review”. In : *Acta Psychiatrica Scandinavica* 101.1 (2000), p. 11–20 (cf. p. 11).
- [Vig97] Ricardo Nuno VIGÁRIO. “Extraction of ocular artefacts from EEG using independent component analysis”. In : *Electroencephalography and clinical neurophysiology* 103.3 (1997), p. 395–404 (cf. p. 25).
- [Win99] Michael E WINTER. “N-FINDR : An algorithm for fast autonomous spectral end-member determination in hyperspectral data”. In : *SPIE’s International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics et Photonics. 1999, p. 266–275 (cf. p. 46).
- [Wis+98] John Oersted WISBECK et al. “Application of ica in the separation of breathing artifacts in ecg signals”. In : (1998) (cf. p. 25).
- [ZFS08] Bo ZHANG, Jalal M FADILI et Jean-Luc STARCK. “Wavelets, ridgelets, and curvelets for Poisson noise removal”. In : *IEEE Transactions on Image Processing* 17.7 (2008), p. 1093–1108 (cf. p. 52).
- [Zha+08] Qiang ZHANG et al. “Tensor methods for hyperspectral data analysis : a space object material identification study”. In : *JOSA A* 25.12 (2008), p. 3001–3012 (cf. p. 58, 61).
- [ŠKZ91] M ŠÁMAL, Miroslav KÁRNÝ et Daniel ZAHÁLKA. “Confirmatory aspects in factor analysis of image sequences”. In : *Biennial International Conference on Information Processing in Medical Imaging*. Springer. 1991, p. 397–407 (cf. p. 70).
- [ŠT12] Václav ŠMÍDL et Ondřej TICHÝ. “Automatic regions of interest in factor analysis for dynamic medical imaging”. In : *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE. 2012, p. 158–161 (cf. p. 51, 56, 67, 68, 92).
- [ŠTŠ11] Václav ŠMÍDL, Ondrej TICHÝ et Martin ŠÁMAL. “Factor analysis of scintigraphic image sequences with integrated convolution model of factor curves”. In : *Proceedings of the second international conference on Computational Bioscience. IASTED*. 2011 (cf. p. 51).
- [Šám+86] Martin ŠÁMAL et al. “Enhancement of physiological factors in factor analysis of dynamic studies”. In : *European journal of nuclear medicine* 12.5-6 (1986), p. 280–283 (cf. p. 51).

- [Šám+87] M ŠÁMAL et al. "Rotation to simple structure in factor analysis of dynamic radionuclide studies". In : *Physics in medicine and biology* 32.3 (1987), p. 371 (cf. p. 49).
- [Šám+88] Martin ŠÁMAL et al. "The reality and meaning of physiological factors". In : *Information Processing in Medical Imaging*. Springer. 1988, p. 499–519 (cf. p. 49).
- [Šám+89] M ŠÁMAL et al. "On the existence of an unambiguous solution in factor analysis of dynamic studies". In : *Physics in medicine and biology* 34.2 (1989), p. 223 (cf. p. 42, 50, 72).
- [Šám+98] Martin ŠÁMAL et al. "Relative renal uptake and transit time measurements using functional factor images and fuzzy regions of interest". In : *European Journal of Nuclear Medicine and Molecular Imaging* 25.1 (1998), p. 48–54 (cf. p. 51, 67, 70).

Résumé — En imagerie nucléaire (scintigraphie, TEMP, TEP), les diagnostics sont fréquemment faits à l'aide des courbes d'activité temporelles des différents organes et tissus étudiés. Ces courbes représentent l'évolution de la distribution d'un traceur radioactif injecté dans le patient. Leur obtention est compliquée par la superposition des organes et des tissus dans les séquences d'images 2D, et il convient donc de séparer les différentes contributions présentes dans les pixels. Le problème de séparation de sources sous-jacent étant sous-déterminé, nous proposons d'y faire face dans cette thèse en exploitant différentes connaissances a priori d'ordre spatial et temporel sur les sources. Les principales connaissances intégrées ici sont les régions d'intérêt (ROI) des sources qui apportent des informations spatiales riches. Contrairement aux travaux antérieurs qui ont une approche binaire, nous intégrons cette connaissance de manière robuste à la méthode de séparation, afin que cette dernière ne soit pas sensible aux variations inter et intra-utilisateurs dans la sélection des ROI. La méthode de séparation générique proposée prend la forme d'une fonctionnelle à minimiser, constituée d'un terme d'attache aux données ainsi que de pénalisations et de relâchements de contraintes exprimant les connaissances a priori. L'étude sur des images de synthèse montrent les bons résultats de notre approche par rapport à l'état de l'art. Deux applications, l'une sur les reins, l'autre sur le cœur illustrent les résultats sur des données cliniques réelles.

Mots clés : Séparation de sources, Imagerie nucléaire, Factorisation en matrices non-négatives, Analyse factorielle, Connaissances a priori, Région d'intérêt

Abstract — In nuclear imaging (scintigraphy, SPECT, PET), diagnostics are often made with time activity curves (TAC) of organs and tissues. These TACs represent the dynamic evolution of tracer distribution inside patient's body. Extraction of TACs can be complicated by overlapping in the 2D image sequences, hence source separation methods must be used in order to extract TAC properly. However, the underlying separation problem is underdetermined. We propose to overcome this difficulty by adding some spatial and temporal prior knowledge about sources on the separation process. The main knowledge used in this work is region of interest (ROI) of organs and tissues. Unlike state of the art methods, ROI are integrated in a robust way in our method, in order to face user-dependancy in their selection. The proposed method is generic and minimize an objective function composed with a data fidelity criterion, penalizations and relaxations expressing prior knowledge. Results on synthetic datasets show the efficiency of the proposed method compare to state of the art methods. Two clinical applications on the kidney and on the heart are also adressed.

Keywords : Source separation, Nuclear imaging, NOn-negative matrix factorization, Factor analysis, Prior knowledge, Region of interest
