Quantitative analysis of bioluminescent signals in preclinical imaging

Erwan Dereure

Thèse préparée à Télécom Paris et à l'Institut Pasteur de Paris
delivrée par Sorbonne Université

École Doctorale Informatique, Télécommunications, Électronique de Paris n°130
Unité d'Analyse d'Images Biologiques - UMR 3691

Quantitative analysis of bioluminescent signals in preclinical imaging

Analyse quantitative de signaux de bioluminescence en imagerie préclinique

Par Erwan DEREURE

Thèse de doctorat d'Informatique

Dirigée par Jean-Christophe OLIVO-MARIN et Elsa ANGELINI
et co-encadrée par Christophe KERVAZO

Présentée et soutenue publiquement le 12 avril 2024

Devant un jury composé de

Frédérique FROUIN	Chargée de Recherche (HDR), INSERM	Rapporteuse
Saïd MOUSSAOUI	Professeur, Centrale Nantes	Rapporteur
Bertrand TAVITIAN	Professeur, Université Paris-Cité	Examinateur
Caroline PETITJEAN	Professeure, Université de Rouen	Examinatrice
Anikitos GAROFALAKIS	Directeur de la R&D, Biospace Lab	Invité
Johanne SEGUIN	Ingénieure d'études, CNRS	Invitée
Christophe KERVAZO	Maître de conférences, Télécom Paris	Co-encadrant
Elsa ANGELINI	Professeure, Télécom Paris	Directrice de thèse
Jean-Christophe OLIVO-MARIN	Directeur de recherche, Institut Pasteur	Directeur de thèse

"You must strive to find your own voice. Because the longer you wait to begin, the less likely you are to find it at all." - John Keating

À mes proches.

Remerciements

Cette thèse est le résultat du travail que j'ai mené ces 3 dernières années, et, bien qu'il soit très personnel, il n'aurait pas pu être mené à bien sans tout le soutien que j'ai pu recevoir.

Je tenais tout d'abord à remercier mes encadrants. Jean-Christophe, merci pour m'avoir accueilli dans ce laboratoire, d'avoir accepté de diriger ma thèse, de tes apports scientifiques pertinents et de la liberté accordée dans nos travaux de recherche. Elsa, merci également d'avoir accepté de diriger ce projet de thèse, de t'être toujours montrée compréhensive ainsi que de m'avoir intégré au laboratoire de Télécom où j'ai notamment eu la chance de pouvoir enseigner. Christophe, merci tout d'abord d'avoir accepté de joindre ce projet en cours. Merci beaucoup d'avoir toujours été disponible pour prendre le temps de m'aider, parfois en me réexpliquant les choses de nombreuses fois ou en relisant pour la énième fois ce que j'écrivais.

Je tenais ensuite à remercier les membres de mon jury, et notamment mes rapporteurs Frédérique Frouin et Saïd Moussaoui, qui ont eu la lourde tâche de lire tout mon manuscrit. Merci pour vos commentaires plus que pertinents, et les différentes perspectives suggérées. Merci aussi à Bertrand Tavitian et Caroline Petitjean pour avoir accepté d'évaluer mon travail lors de la soutenance. Merci pour la discussion scientifique qui en a suivi.

Je tenais également à remercier Camille Kurtz, membre de mon comité de suivi, d'avoir toujours été très encourageant, de bon conseil et m'avoir permis de prendre un peu de confiance aux moments opportuns.

Cette thèse est un travail liant plusieurs équipes, et je souhaitais donc remercier mes partenaires. Tout d'abord, merci à l'entreprise Biospacelab d'avoir écrit ce projet ce thèse et notamment merci à Anikitos de m'avoir suivi hebdomadairement. Tu m'as toujours encouragé dans les travaux que je faisais, même lorsque ce n'était pas forcément dans ton domaine de prédilection. Ton impressionnante connaissance des différentes méthodes d'imagerie te permettait d'être d'une grande force de proposition pour tous les aspects applicatifs. Quelle coïncidence que tu habitais rue Simon Dereure quand j'ai commencé ma thèse. Merci aussi à Rémi, Mickaël et Robin pour toute votre assistance technique, notamment au début de ma thèse, et votre accueil dans les quelques journées que j'ai passées à vos côtés.

Ensuite, merci au laboratoire UTCBS d'avoir participé à ce projet et plus particulièrement à Johanne, toujours prête à m'aider à comprendre quelque chose, à valider l'utilité de mon travail, à essayer de nouvelles méthodes ou juste à discuter. Tu m'as souvent rassuré et a toujours veillé à me faire savoir que tu étais satisfaite de mon travail. Tu m'as également montré en détail les enjeux des protocoles expérimentaux, notamment en me faisant régulièrement assister à des expériences. Je ne pourrais assez te remercier pour le temps que tu as passé sur notre projet, pour l'attention que tu portais aux différentes pistes que j'explorais et pour tes encouragements. Merci également à Nathalie Mignet d'avoir supervisé notre travail. Merci à vous tous de m'avoir aidé à faire un travail de qualité pouvant avoir un impact bénéfique.

Durant cette thèse, j'étais intégré au sein de 2 laboratoires, dont j'aimerais maintenant remercier les membres. Merci d'abord à toute l'équipe BIA de l'Institut Pasteur, membres actuels comme passés : Aleix, Amandine, Daniel, Giacomo, Lucie, Maria, Marion. Plus particulièrement, merci à Thibault pour ta sincérité et ta bienveillance, que tu passes le plus clair de ton temps à cacher derrière de mauvaises blagues. J'ai cependant vite compris que, si on en faisait une encore plus mauvaise, tu devenais rapidement l'interlocuteur idéal pour nous expliquer les tenants et aboutissants du métier de chercheur, nous rassurer quant à la pertinence de nos travaux et nous aiguiller dans nos recherches (et nos programmes de course à pieds). Merci ensuite à Pascal pour le partage de ton savoir encyclopédique sur une quantité tellement diverse de sujets que je ne saurais pas par où commencer, et d'être malgré tout d'une grande humilité et sagesse dans ta quête continuelle d'apprentissage. Merci à Elisabeth d'être régulièrement d'excellente humeur et prête à rigoler, ainsi que de m'avoir fourni de la nourriture quelques matins où l'heure d'ouverture de la cantine était trop lointaine. Merci à Stéphane d'avoir été en permanence disponible pour nous aider sur Icy, disponibilité que tu as sans nul doute transmise à Thomas M. qui répond également toujours présent en cas de problème. Last but not least, thanks a lot to Suvadip and Ritu for having been the coolest mentors we could have had. I hope you'll live a happy and joyful life.

Dans ce laboratoire j'ai également pu travailler aux côtés d'une formidable équipe de doctorants que je voudrais remercier pour l'entraide, les moments de détente, les sessions de badminton le midi, les pauses sur le toit et les nombreux verres et camemberts farcis que nous avons consommés au Petit Pasteur. Tout d'abord, j'aimerais beaucoup remercier Alexandre et Diana de nous avoir accueillis, encadrés, encouragés et pris sous vos ailes au début de nos thèses. Alexandre, merci d'avoir été le grand frère qu'on rêve tous d'avoir, de m'avoir fait partager quelques footings, de nous avoir transmis ton savoir faire et l'habitude des verres au bar. Diana, camarade de soupirs et de ralages, merci pour ton humour et pour nous remonter le moral en permanence, tu es une force de la nature. Ensuite, un grand merci à Samuel et Yekta, constamment un peu en avance de nous, vous m'avez permis d'entrevoir sereinement les différentes échéances qui nous attendaient. Samuel, merci pour les discussions rugbystiques enflammées et pour avoir été notre modèle de persévérance et de rigueur. Yekta, notre amour commun pour les films et les mangas n'est rien comparé à ta gentillesse et ton dévouement. Merci ensuite à Benjamin, qui connait probablement la documentation de pytorch par coeur. Merci à Mounib, dont l'abnégation et le sacrifice forcent l'admiration, ainsi que la générosité vu les délicieuses pâtisseries que tu ramenais toujours de chez toi. Merci à Thomas T., biologiste mais qui nous a toujours écouté parler avec attention. Merci enfin à Thomas M., pas doctorant mais volontaire et féru de Uno.

Merci ensuite à l'équipe IMAGES de Télécom Paris. Je tenais à remercier tout d'abord les permanents pour avoir créé une équipe de cette qualité, où le niveau scientifique comme l'ambiance sont excellents : merci à Alasdair, Yann, Pietro, Isabelle et Florence. Merci en particulier à Saïd pour ta gentillesse, ton humour décapant, ton envie permanente d'aider et de partager tes innombrables connaissances, et les sorties vélo.

Merci ensuite à la très large équipe de doctorants et post-doctorants qui sont passés par ce labo : Zoé, Inès, Nicolas, Pierrick, Gwilherm, Rebeca, Emma, Edouard, Chloé, Marie, Raphaël A., Matthis, Alban, Camille, Rassim, Robin, Benoit, Carlo, Florence, Ali K., Ali M., Cristiano, Emanuele, Giammarco et Mateus. Dans le désordre, merci pour le Deep Learning Working Group (qui est un groupe incroyable pour monter en compétence en traitement d'images, merci aux organisateurs), les séances de course à pied le midi, les pauses cafés, les jeux de mots nuls, les verres à la Butte aux Cailles, le GRETSI, les travaux chez moi, le parc Astérix, les repas et les débordements au divan du monde. Malgré le long trajet, c'était toujours avec plaisir que je

venais à Télécom car je savais que je passerais ma journée avec vous.

En particulier, à l'intersection des 2, merci à Raphaël R. pour ton énergie inépuisable et de te rendre toujours diponible. C'est rare les gens sur qui on peut compter avec autant de certitude, que ce soit pour aider ou juste être présent. Merci également d'avoir été là pour les voyages en Colombie et en Grèce. Enfin, merci infiniment à Antoine H. d'avoir été là quasi-quotidiennement, à Télécom comme à Pasteur, pendant ces 3 ans. Au delà de tes compétences qui, honnêtement, m'impressionnent, merci d'avoir été là quand ça n'allait pas, pour tes encouragements, et pour nos discussions pleines de partages et de confidences. Cette thèse n'aurait pas été la même sans toi. Je suis très fier de te compter parmis mes amis. Merci d'être dans ma vie.

Au delà de l'aspect scientifique, ces laboratoires ne pourraient pas fonctionner sans des cellules administratives. J'ai eu le bonheur de tomber, dans chaque établissement, sur des personnes d'une gentillesse inouïe. Merci à Marie-Anne et Oliva côté Pasteur et Delphine côté Télécom. Vous simplifiez notre quotidien et rendez l'ambiance de travail beaucoup plus agréable en étant toujours aux petits soins, que ce soit pour de la gestion de problèmes ou de l'organisation d'événements. On ne vous remercie pas assez pour votre altruisme et votre aide de tous les instants donc merci beaucoup. Par la même occasion, j'aimerais remercier les employés des services de restauration de l'Institut Pasteur et de Télécom Paris pour nous assurer une nourriture copieuse quotidienne, et en particulier au lambescain de Télécom qui nous servait à chaque fois avec un grand sourire et une touche d'humour.

Cette thèse m'a permis d'en apprendre beaucoup scientifiquement, mais cet apprentissage n'a pas commencé là. J'aimerais donc remercier tous les enseignants que j'ai eu la chance de croiser dans ma scolarité qui m'ont fait aimer les sciences, et en particulier mes profs de prépa. Merci tout d'abord à Alexandre Marino, votre pédagogie et votre empathie nous ont permis de démarrer ce voyage scientifique exigeant avec beaucoup de sérénité. Vos apartés sur l'histoire des mathématiques étaient toujours captivants. Ensuite, merci à Véronique Chireux de m'avoir fait découvrir à quel point la physique pouvait être intéressante et merci infiniment pour votre dévouement infatigable à l'égard de vos élèves (même immobilisée après un accident). Merci de rester en contact avec nous et de prendre de nos nouvelles de temps en temps. Enfin, j'ai une pensée émue pour Michel Alessandri, disparu récemment, que je remercie énormément. Vous avez été un professeur passionné et passionnant qui a marqué beaucoup d'élèves, dont je fais partie. Force est de constater que vous avez transmis à bon nombre de gens votre goût pour les mathématiques. Votre gentillesse est un exemple que j'essaie de suivre dans ma vie. Merci pour tout.

Maintenant, j'aimerais remercier mes amis. Merci à mes amis d'école Pierre H., TanPhi, Gus, Pierre B., Tassaert, Renaud, Sechet, Emile, Tomasina, Latil et Kerleroux pour les vacances en Bourgogne, les raclettes, les randos, les soirées sur les quais, les matchs et les marmites. Merci à mes amis de prépa Thomas, Ivan, Aliénor, Nicolas, Paloma, Elias et Clémence pour les sessions de jeux de société et d'être des amis en or. Merci à mes amis de Montpellier Guilhem, Franck, Maxime, Adrien, Romain, Richard, Damien, François, Guillaume, Laurent, Fabrice et Thibault d'être toujours là malgré les années. En particulier, merci à Alexandre d'être tout le temps enthousiaste à l'idée de prendre un café à Paris. Merci à Matthieu, Luis et Corentin pour les ciné-concerts et les voyages. Merci à Rémi et Jonathan pour les sorties longues le dimanche. Merci à Julien d'être loin des yeux mais près du coeur. Merci à tous d'avoir toujours été là pour me changer les idées.

Merci plus particulièrement à mes colocataires Antoine G. et Marin, vous avez été ma

soupape de décompression, en permanence prêts à passer un bon moment à l'appart, regarder tout et n'importe quoi à la télé, manger des croustilles, rigoler, organiser des soirées couvre-chef ou juste discuter. Merci aussi énormément pour avoir essayé de m'aider dans cette fin de thèse, même si ce n'était pas forcément facile de trouver comment. Merci aussi à Virgile, notre 4ème colocataire par correspondance.

Merci maintenant à toute ma famille. Merci à ma mère, mon père, mon frère et ma soeur, mon beau-père, mon beau-frère, ma grand-mère, mes oncles et tantes et mes cousins. Merci notamment à Amaryllis et Gabriel, nos retrouvailles régulières étaient toujours plus que bienvenues. Merci également à mes grands-parents disparus ces dernières années, notamment merci à mon grand-père Michel avec qui j'aurais beaucoup aimé pouvoir parler de ce travail. Merci à tous de m'avoir toujours soutenu, surtout dans les moments les plus compliqués, et de m'avoir transmis le goût de l'apprentissage, de l'esprit critique et de l'humilité.

Merci enfin à Zoé pour ton soutien inconditionnel, j'espère pouvoir t'en rendre ne serait-ce qu'une infime partie.

Abstract

Abstract: Bioluminescence imaging (BLI) is an optical imaging technology in which a living organism or cell emits light through a biological substrate/enzyme reaction without any light excitation. This technology, used in preclinical oncology in order to quantify the tumor status in a non-invasive way, is still quite recent and for now biologists lack automated processing tools to improve the quantification of images. In addition, some experimental protocols require to extract the photon flux of multiple tumors on the same side of the animal. This can be difficult and can introduce errors and biases as BLI suffers from a lack of robustness because of a variability in vascularization, or hypoxic and necrotic zones within the tumors. In this work, we propose the use of Non-Negative Matrix Factorization to separate the photon flux of different tumors within the same bioluminescence image by leveraging the different pixel-wise temporal patterns. Such spatio-temporal unmixing yields several important challenges that we have tackled. In a first contribution, we use prior knowledge on the appearance of the tumors and show the importance of penalizing the norm of the wavelet coefficients corresponding to the sources estimated during the optimization process to obtain a high spatial consistency of unmixed tumors. In a second contribution we deal with strong heterogeneities within tumors corrupting the separation by presenting a dedicated pipeline for pre-aligning the photon flux of the different pixels. We show that the resulting method is capable of accurately extracting the photon flux of different tumors present within a single bioluminescence image. These algorithms were tested and validated on two real BLI datasets and on one synthetic dataset generated with a bioluminescence image simulator we designed and developed. In a third contribution, we propose a pharmacokinetics model to calibrate the tumor photon flux based on the bioluminescence signal emitted by a muscle. This allows us to extract meaningful physiological parameters from the image like substrate exchange rates. We show that these parameters represent significant features of the tumor state and can be used to improve the quantification of bioluminescence images.

Keywords: Bioluminescence imaging, blind source separation, non-negative matrix factorization, variability, preclinical oncology, quantification, pharmacokinetics model, reference region.

Résumé: L'imagerie par bioluminescence (BLI) est une technologie d'imagerie optique dans laquelle un organisme ou cellule vivant émet de la lumière à travers une réaction biologique substrat/enzyme sans aucune excitation lumineuse. Cette technologie, utilisée en oncologie préclinique afin de quantifier l'état des tumeurs de manière non invasive, est encore assez récente et, pour l'instant, les biologistes manquent d'outils de traitement automatisé pour améliorer la quantification des images. De plus, certains protocoles expérimentaux nécessitent l'extraction du flux de photons de plusieurs tumeurs situées sur le côté de l'animal. Cela peut être difficile et peut introduire des erreurs et des biais car la BLI souffre d'un manque de robustesse en raison d'une variabilité dans la vascularisation, ou des zones hypoxiques et nécrotiques au sein des tumeurs. Dans ce travail, nous proposons l'utilisation de la factorisation en matrices non négatives pour séparer le flux de photons de différentes tumeurs au sein de la même image de bioluminescence en tirant parti des différents patterns temporels pixel par pixel. Un tel démélange spatio-temporel présente plusieurs importants défis que nous avons relevés. Dans une première contribution, nous utilisons des connaissances préalables sur l'apparence des tumeurs et montrons l'importance de pénaliser la norme des coefficients d'ondelettes correspondant aux sources estimées pendant le processus d'optimisation afin d'obtenir une forte cohérence spatiale des tumeurs démêlées. Dans une deuxième contribution, nous traitons les fortes hétérogénéités au sein des tumeurs corrompant la séparation en présentant une chaîne de traitement dédiée pour pré-aligner le flux de photons des différents pixels. Nous montrons que la méthode résultante est capable d'extraire avec précision le flux de photons de différentes tumeurs présentes dans une seule image de bioluminescence. Ces algorithmes ont été testés et validés sur deux ensembles de données réelles de BLI et sur un ensemble de données synthétiques généré avec un simulateur d'image de bioluminescence que nous avons conçu et développé. Dans une troisième contribution, nous proposons un modèle de pharmacocinétique pour calibrer le flux de photons de la tumeur en fonction du signal de bioluminescence émis par un muscle. Cela nous permet d'extraire des paramètres physiologiques significatifs de l'image comme les taux d'échange de substrat. Nous montrons que ces paramètres représentent des caractéristiques significatives de l'état de la tumeur et peuvent être utilisés pour améliorer la quantification des images de bioluminescence.

Mots-clés: Imagerie par bioluminescence, séparation aveugle de sources, factorisation en matrice non négative, variabilité, oncologie préclinique, quantification, modèle pharmacocinétique, région de référence.

Contents

List of Figures

List of Tables

Chapter I

Introduction

I.1 Funding and Collaborative context

This PhD Project started in February 2021. It was funded by a grant from the PRPhD program of the Ile-de-France Region (PI: Jean-Christophe Olivo Marin and Lise Laumonier) and by the company Biospace Lab. It was carried out in collaboration with Biospace Lab (with Anikitos Garofalakis and Lise Laumonier as correspondants) and Nathalie Mignet and Johanne Seguin from the UTCBS laboratory (Université Paris-Cité, CNRS and Inserm), Faculty of Pharmaceutical and Biological Sciences at Université Paris-Cité. The biologists provided us with the datasets used in this thesis, acquired with a device developed by Biospace Lab.

I.2 Preclinical Imaging on small animals for oncology research

I.2.1 Introduction on preclinical imaging for oncology research

Cancer recorded more than 19 millions cases, and killed about 10 millons of people, according to the International Agency for Research on Cancer (IARC), around the world in 2020. Before a treatment is tested on humans, it is tested on animal subjects, especially mice or rats [1]. The transfer of knowledge and experiments from animals to human is called **preclinical research**. In this field of research, numerous technologies can help to visualize and quantify the impact of treatments on tumors, and one of them is imaging, therefore known as **preclinical imaging** [1].

There are 3 kinds of preclinical imaging : anatomical imaging, functional imaging and molecular imaging. Anatomical imaging targets specific anatomical structures with a contrast agent whereas functional imaging studies a physiological or chemical process that can happen anywhere in the organism. In functional imaging, contrast agents often target molecules, and we can talk in this context about molecular imaging [2, 3].

Preclinical can be divided in *ex vivo* and *in vivo*. First of all, *ex vivo* imaging is the analysis of tissues taken from a living subject. It can be applied on cells, tissues or histological sections [2]. Histology is the study of sections, between 5 and $200\mu m$ of thickness, of an organism with a microscope where objects of interest can be highlighted with a constrast agent to study physiological phenomena [4] (Figure I.1). Ex vivo imaging took great advantage from the breakthrough of microscopy, with confocal miscroscopy (that enables to study thick objects by isolating plans from the sample [5]) or FRET, Fluorescence Resonance Energy Transfer, which has a sensitivity up to a single molecule [3, 6].

Ex vivo imaging is widely developed, used, mastered and accurate [8]. However, it is invasive

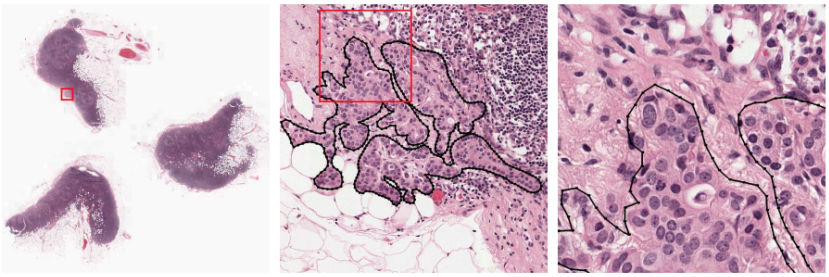

Figure I.1: Whole Slide Histopathology Image. Taken from the Camelyon17 dataset [7]. Delineated metastasis at increasing zoom levels from left to right.

and therefore the conditions of acquisition are completely different from the natural conditions of the studied subject and long-term longitudinal studies are proscribed [2]. Therefore, *in vivo* imaging, which refers to the study of a functioning living organism, is developed, to perform acquisitions in the most natural physiological conditions and long-term longitudinal studies. Some of the in vivo imaging modalities still suffers from a lack of knowledge of accuracy [9–11]. Longitudinal experiments is the approach used in the context of this thesis, requiring then in vivo imaging.

I.2.2 In vivo preclinical imaging

The use of many technologies can be designed for in vivo small animal imaging or derived from clinical imaging [2, 3] with several adaptation challenges in terms of resolution and accuracy related to the size difference between humans and small animals. There are several types of technologies used from localized to whole body imaging [12].

I.2.2.1 Localised imaging

An important part of localised preclinical imaging methods, with an axial resolution in the micrometre range [2], are adapted technologies from microscopy, like intravital microscopy [13]. However, they are limited to tissues of submillimetric thickness [13] and therefore are mainly used in dermatology [14] or ophtalmology [15]. Though, multiphotonic microscopy can be applied to neurology [16].

Ultrasound (US) imaging is considered between localised imaging and whole body imaging. Indeed, the acquired field of view is big enough to get images of a mouse foetus [2]. This modality is based on the measure of US waves propagated through the body [3], and their measure provides informations about the acoustics properties of the tissues/organs [17] as shown in Figure I.2. Moreover, Doppler Effect enables to image moving objects in the subject such as red blood cells, which provides informations on blood flow [17]. These constrast agents can also be modified to provide molecular information [18]. US imaging yields images with a rapid frame rate, with high spatial and temporal resolution, involves no radiation, while being low cost but is limited by its reproducibility and noise [3].

I.2.2.2 Whole body small animal imaging

Whole body images have a lower spatial resolution than localised images but they can yield macroscopic informations for small animals [2]. They enable for instance pharmacokinetics and biodistribution studies [19]. They can also be used to study live moving animals, which can be interesting to limit the impact of anesthesia both on the organism and the behavior of the

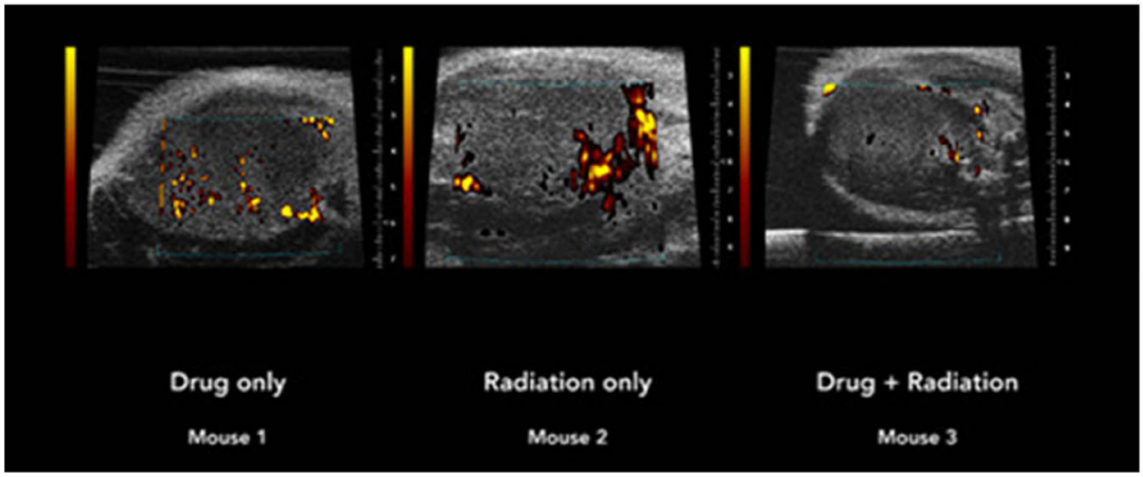

Figure I.2: US imaging from [2]. Visualization of tumoral vascularization in a mouse at 3 steps of treatment evolution.

Modality	**X-Ray Imaging**	**MRI**	**Radio-isotopic Imaging**	**Optical Imaging**
Type of images	Anatomical/Functional	Anatomical/Functional	Functional	Anatomical/Functional
Electromagnetic spectra	$40-100$ keV	$200-400$ MHz	$70-300$ keV in SPECT 511 keV in PET	$1.5-3$ eV ($400-800$ nm)
Contrast enhancement	Variation of tissues absorption coefficient	Density and environment of protons	Concentration of injected radiotracer	Activity of a luminescent protein or fluorophore excitation
Spatial and Depth resolutions	$(10\mu m)^3$ to $0.5\times0.5\times1$ mm^3, whole body	50μm to 1 mm, whole body	$1-2$ mm, whole body	$1-5$ mm, <5 cm
Temporal resolution	A minute	Few tenths of seconds to several minutes	Few seconds to a minute (in SPECT)	10 ms to a few seconds
Sensitivity	$1-5$ mmol/kg	$0.01-1$ mmol/L	$0.1-10$ pmol/L	10^{-10} to 10^{-8} μmol/kg
Contrast Agent	Iodine or barium sulfate	Gadolinium or iron oxide superparamagnetic	γ (SPECT) or positron (PET) emitters	Fluorophores or luminescent proteins
Cost	\$\$ – \$\$\$	\$\$\$	\$\$\$	\$ – \$\$

Table I.1: Recap table of the different whole body imaging modalities used in in vivo imaging, adapted from [2].

animal during the acquisition [20]. There are several imaging modalities that can be used in this context [21], which are presented next (Figure I.3). The characteristics of these different modalities are summarized in Table I.1.

a) X-Ray radiography and tomography

X-Rays modalities are based on the interaction of X-Rays with tissues in the organism. Each tissue has a specific absorption coefficient depending on its density and of the energy of the ray [3]. X-rays are used for $2D$ imaging (radiography) and $3D$ imaging (computerized tomography, CT) by rotating the scanner around the studied subject and reconstructing the tomographic image with the Radon transform [22]. Its preclinical equivalent is called microCT [2]. They differ in a few technical aspect such as the energy of X-Rays used (higher for preclinical imaging, 40 to $100eV$) to increase spatial resolution [23]. Both CT and microCT may need constrast agent like iodinated water-soluble compounds to increase contrast in some tissues [24]. This modality provides images with high resolution but the level of radiations involved is a limitation that should be taken into account. The safety precautions for the operator and the lack of molecular information are other limitating factors [3, 25]

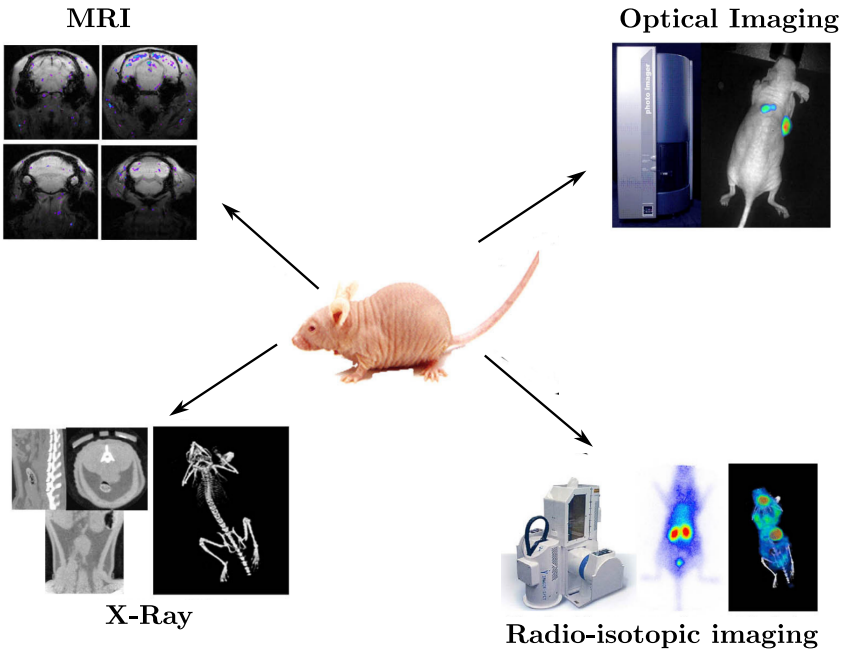

Figure I.3: Illustration of different whole body mouse imaging modalities (adapted from [2, 21]).

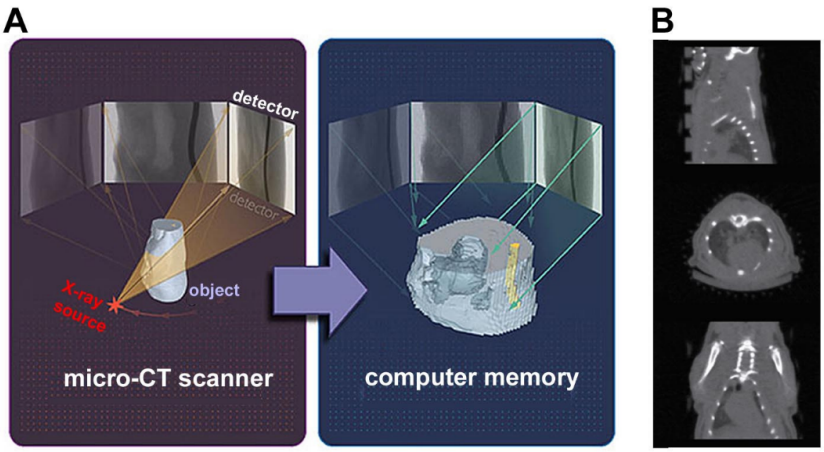

Figure I.4: Illustration of X-Ray imaging, taken from [2]. A: acquisition of multiple projections in microCT imaging and $3D$ reconstruction. B, from top to bottom: sagittal, axial and coronal slices of a rat.

b) Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is an imaging modality based on the Nuclear Magnetic Resonance (NMR) property of certain chemicals elements. It generates $2D$ or $3D$ images with better contrast in soft tissues than X-Rays based methods [2].

c) Radio-isotopic Imaging

Radio-isotopic Imaging, like Positon Emission Tomography (PET scan) or single Photon Emission Computerized Tomography (SPECT) is based on the emission of γ rays (with an energy greater than $100keV$) emitted by a radioactive radiotracer injected in the body [12]. For instance, $^{18}F - FDG$ (fluorine-18-labeled fluorodeoxyglucose) is a radioactive isotope of fluorine

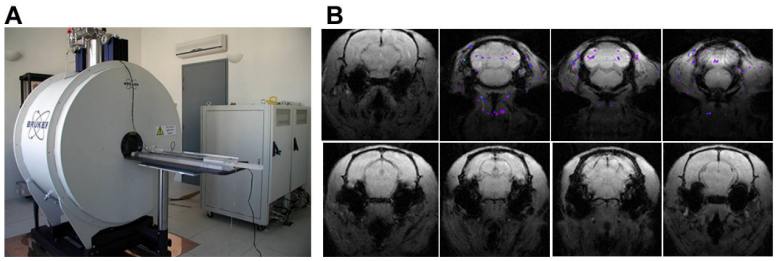

Figure I.5: Illustration of MRI preclinical small animal imaging. A: imaging device. B: MRI image of activity maps in coronal slice after visual stimulation. Taken from [2].

Small animal PET device **Combination PET/CT image**

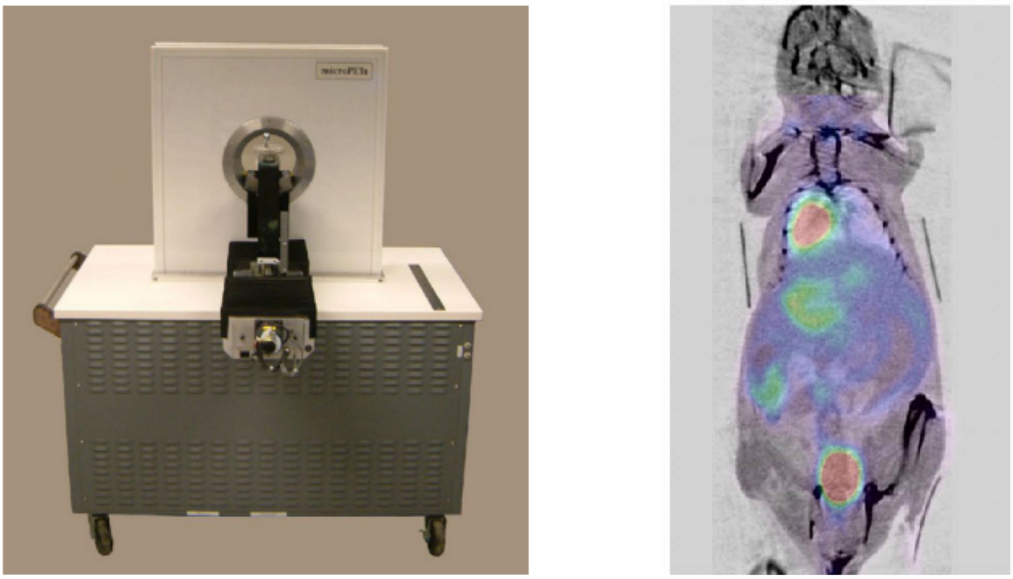

Figure I.6: Illustration of PET small animal imaging. Left: imaging device. Right: Combination of PET and CT images (adapted from [28]) projected in $2D$.

(^{18}F) [26] coupled with glucose [2], which is consumed abnormally by cancer cells (Figure I.6). PET employs a 360° detection gantry to detect the opposite rays emitted by the positron-electron annihilation process, providing a $3D$ image after reconstruction algorithms [27] whereas SPECT does not necessarily contain depth information [12] but it is possible with specific devices.

d) Optical Imaging

Optical imaging is based on light emission captured by a camera [29]. In molecular imaging, it has applications from *in vitro* microscopy imaging [30] to *in vivo* whole body small animal imaging [31].

Light photons interact with surrounding tissues, by being absorbed or scattered [32]. Only a small amout of photons, called ballistic, are directly captured by the imaging device. In optical imaging, we capture all photons arriving on the imaging device, whether they are ballistic of diffused [2]. In order to monitor tissues molecular activity, it is then common to use specific probes that have an affinity for a molecular target like a receptor or a protein. These probes are tagged with a luminescent probe that emits photons [3]. Molecular chemistry allowed the

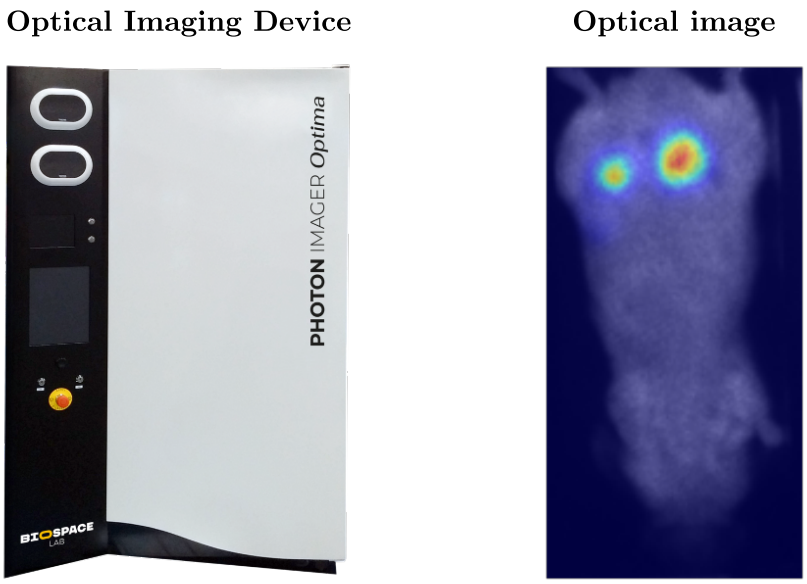

Figure I.7: Illustration of small animal optical preclinical imaging. Left: imaging device (PhotonImager OPTIMA, BiospaceLab, Nesles la Vallée, France). Right: superposition of bioluminescence image and natural image of animal.

labeling of specific molecules, leading to the development of molecular probes. It can for instance be a gene, coding for a luminescent protein, expressed by specific cells or tissues. In this case, light detection would provide information on this gene expression [12].

Due to the absorption and scattering of photons involved in optical imaging, accurate and efficient cameras are developed to capture low-intensity light [3]. We are going to briefly review these device later in the manuscript.

Optical imaging is a non-ionizing imaging method, on the contrary of PET scan, that therefore does not need any radiation protection and is not expensive. It is also non-invasive. Thus, there are cases where it is the most adapted method in the context of preclinical imaging [2], and is the one used in the context of this work. An example of optical image is shown in Figure I.7.

I.2.3 Light Emission Mechanisms

I.2.3.1 Energy Exchanges

Luminescence is the spontaneous emission of photons by an excited molecule, liberated when it returns to their stable state [12]. In optical imaging, these species are molecules [33].

There are different types of luminescences, depending on the origin of the excitation. For a light excitation, we talk about fluorescence (when the photon is emitted instantly, resulting in a fast emission of approximately $10^{-11}s$) [34] or phosphorescence (when the molecule transits by an intermediate state before emitting the photon, resulting in a slow emission during between a few microseconds to a few seconds) [2]. This energy transfers are illustrated in Figure I.8. In any case, the energy loss between excitation and emission causes the wavelength of the emitted photon to be higher than of the exciting photon, phenomenon known as Stokes shift [12]. For example when the excitation is due to an enzymatic reaction, the photon emission is called bioluminescence [34].

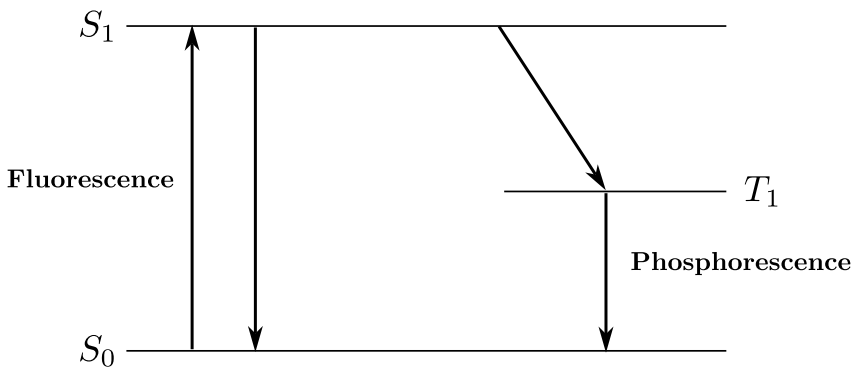

Figure I.8: Energy transfers, between a stable state S_0, an excited state S_1 and a transition state T_1, involved in luminescence.

I.2.3.2 Probes used in optical imaging

Different kind of probes are used in optical imaging. As in nuclear imaging, we can inject a luminescent probe that targets a specific phenomenon, these are called exogenes probes. But it is also possible to modify genetically the organisms to make them emit light in order to capture information, these are called endogenes.

a) Exogenous probes

There are different kinds of exogenous probes. The most well-known are organic fluorophores (such as boron-dipyrromethene *bodipy* [2]), which are organic molecules that can emit wavelengths in the whole visible spectrum [2]. They can also be easily modified to target tissues or physiological phenomena [35]. Their main drawbacks are a limited time stability, making it hard to do any long-exposures microscopic acquisitions [2]; and a limited Stokes Shift, challenging the filtering between exciting light and emitted light [36]. Some fluorophores, the cyanines, can emit in the red or near infrared, which are the preferred wavelength for optical *in-vivo* acquisitions because light can pass through efficiently [37] and has the lowest autofluorescence [12], but they are the ones having the most limited optical properties (low quantum efficiency : 5% to 40% in organic solvents) [2].

It is not the case of the *quantum-dots*, which are inorganic cristals having tremendous optical properties. Their time stability is better than fluorophores and their emission time is longer [2], and they can emit also in the infrared [38]. At high concentration, they can be toxic but it has been shown that it is possible to use them in healthy concentrations for oncology [39]. They are however too big (often bigger than $10nm$) to be used for molecular imaging, because they perturb tissue properties or the associated measures [2].

Persistent luminescence nanoparticles have a very long emission time (a few hours [2]) and a high signal to noise ratio because they don't need a constant light excitation which makes the background of the image clear, so they could be interesting probes for optical imaging because they can emit in the red of near infrared [2]. The current challenges are mostly their sizes (they are also too big : $50-100nm$ [40]), and their complex functionalization, which are current fields of research [40].

b) Endogenous probes

It is also possible to directly modify the DNA of targeted cells by adding some part coding for a luminescent protein. In this set-up, we have a dual expression of the gene of interest

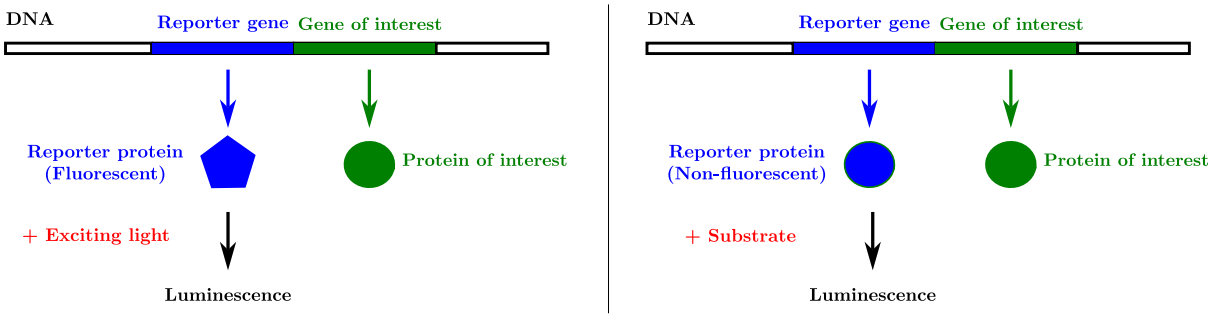

Figure I.9: Illustration of endogenous probes, adapted from [2]. Left: fluorescence probe, the gene of interest is linked to a reporter gene that codes for a fluorescent protein. Right: bioluminescence probe, the gene of interest is linked to a reporter gene that codes for a protein that produces light when excited with a substrate (luciferin).

and a reporter gene (coding for the light emission). The intensity of the signal provides some quantitative information on cellular phenomena [3]. In *in-vivo* experiments, both fluorescences and bioluminescence can be used as shown in [12]. In fluorescence, the most used protein is the Green Fluorescent Protein (GFP) [21], which generates a green signal when excited with a UV or blue light [12]. It has several applications in *in-vivo* experiments from tumors quantification [41] to cells dynamics [42].

On the other hand, bioluminescent proteins don't need any light excitation since the energy comes from an enzymatic reaction, where the associated enzyme is usually called luciferase [43]. In an environment that emits relatively little light it can be a useful tool to mark a specific area [44]. Mammalian tissues having little natural luminescence, Bioluminescence Imaging (BLI) appears to be a preferred imaging method for studying *in-vivo* phenomena [44]. It has therefore has applications in many fields of biomedical research [45] such as neuroscience [46] or oncology [47]. It also has many applications such as molecular evolutions and tumor quantification, which is the use case of the biologists involved in this thesis. Indeed, one of its main advantage is a very low background signal from surrounding tissue because there is no exciting light [11] .

I.2.3.3 Bioluminescence

So-called bioluminescence is the emission of photons by a living organism without the need of light excitation [48]. It is for instance the ability that enables fireflies [49] or glow-worm to produce light [50]. The usual mechanism involved is the creation of a high-energy molecule that produces a photon by returning to a stable state [33]. Most of bioluminescence phenomena are explained by a luciferin-luciferase reaction, and it had been a common assumption that it was the only explanation. Luciferin is the generic term that stands for an organic coumpound that provides the organism the energy necessary for photon emission by being oxidized [43], in presence of a specific enzyme called luciferase that catalyses the reaction, sometimes in the presence of others factors (Adenosine triphosphate, Magnesium ..) [33]. However, later studies stated that some proteins, like Aequorin, emit light without the presence of oxygen but with only other factors (calcium), thanks to an intramolecular reaction. This kind of protein cannot be fully described by the terms luciferin or luciferase, it has then be called in the litterature photoprotein [48].

Luciferin-luciferases reactions, summarized in Figure I.10, emit light in proportion of the amount of formed oxyluciferin [33], and so of the amout of luciferin [48]. On the other hand, reactions involving photoproteins emit light in proportion of the amount of so-called photopro-

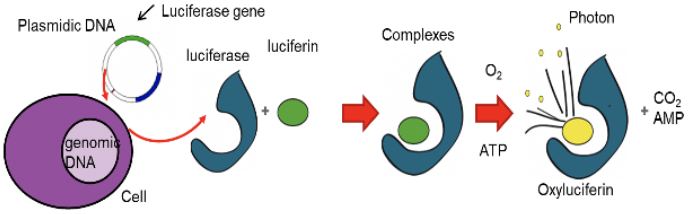

Figure I.10: Reaction of bioluminescence, adapted from [52].

teins [33], which are then mainly used to report the concentrations of C_a^{2+}, M_g^{2+} ... [51].

Bioluminescence Imaging is then a non-invasive technique that allows the monitoring of biological phenomena under the condition of producing the bioluminescence reaction. Even though both luciferin-luciferase [53] and photoproteins [54] based reactions have been used in BLI, in this work we will focus from now on luciferin-luciferase systems because it is the privileged set-up of the biologists we work with in this thesis. These systems can also emit in the near-infrared [55], which is the more suited window for in vivo imaging [56].

In this context, tumor cells are genetically modified to express luciferase. After an injection intravenous (IV) or intraperitoneal (IP) [57] of the luciferin, the emitted light intensity quantifies the tumor status and can be used to study longitudinally the evolution of a tumor in the context of experiments to asses the efficiency of cancer treatments [47].

I.2.4 Light detection

I.2.4.1 Possible Imaging Devices

Sensors involved in optical imaging, and especially in bioluminescence imaging, must respect some criteria [58]. First of all, in order to detected external light, they must be highly sensitive. Indeed, luminescent signals emitted by a bioluminescence reaction have intensities between 10^3 and $10^6 ph/s/cm^2/sr$. For fluorescence, it can be 2 to 3 orders of magnitude higher [2]. The maximum sensitivity of the human eye is at 555 nm, with a threhsold at about $1.4 \times 10^8 ph/s/cm^2/sr$ [59], which can be close to the the order of magnitude of fluorescence imaging.

However, *in vivo* experiments involve usually light in the red or near infrared, where the sensitivity of the human eye can drop by a factor 1000 [12]. In addition, we need to take into account diffused photons as explained in Figure I.2.2.2. Imaging device for optical imaging, and especially for bioluminescence imaging, must then contain a highly sensitive camera [12]. In optical imaging, we need also to have a high range of detection to be able to detect both very intense tumors and very low ones. This range is defined by the ratio between the lowest detectable signal (usually noise level) and the saturation threshold of the sensor. A high range of detection can be especially useful in the case of metastases, way less intense than the primary tumor [60].

In optical imaging, the most used types of sensors are the CMOS (Complementary Metal Oxyde Semi-conductor) technology bases sensors and the CCD (Charge Coupled Device) sensors [61]. Yet, the ratio signal to noise in the CMOS sensors is low, so is their sensitivity compared to CCDs [62, 63], which are therefore better suited for bioluminescence imaging [64, 65].

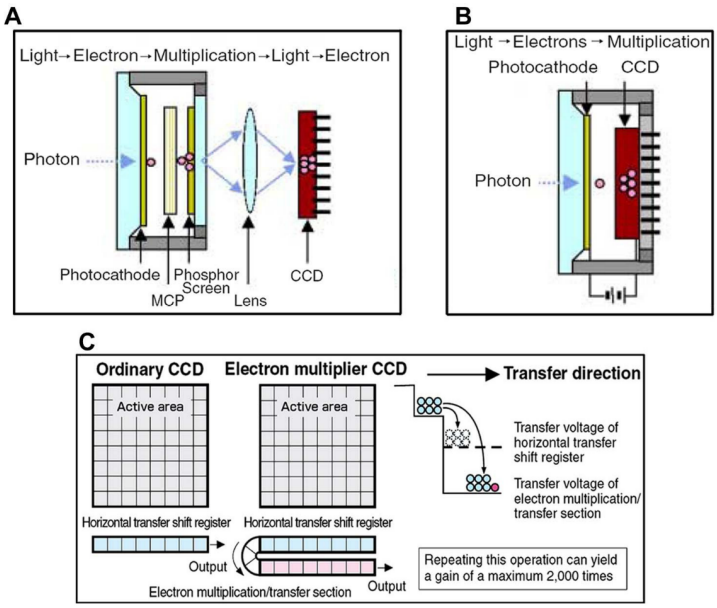

Figure I.11: How different kinds of CCDs work, taken from [2] from images of *Hamamatsu*. A: iCCD. B: EB-CCD. C:EM-CCD.

CCD sensors convert incident photons into electron-hole pairs to generate an image [33]. There are 2 kinds of CCD used in optical imaging :

- There are cooled CCD where the exposure time is increased (*integrated*) to improve the sensitivity of the sensor. Since the exposure time is linked to the spontaneous generated electrons in the device, cooling is applied on the sensor, sometimes at $-100°C$ to reduce that noise [66].
- Then, signal amplification CCD amplify the signal before it reaches the sensor [66]. Examples of signal amplification CCD include : Electron-Multiplication CCD (EM-CCD) [67] , EB-CCD (Electron Bombarded CCD) [68] and intensified CCD (iCCD) [69] (Figure I.11). EB-CCDs use a photocathode to fasten the electrons by applying a high voltage between the photocathode and the sensor, which results a good sensitivity and fast response time but is an expensive technology [68]. EM-CCDs use a gain register after the shift register and before the output. These sensors have the same sensitivity as cooled CCD and a better quantum efficiency but need a highly accurate cooling, at temperatures similar to cooled CCD, to avoid gain variations [67]. Finally, in iCCDs a photocathode converts photons to electrons, that are multiplied by a microchannel-plate (MCP), are converted back to electrons with a phosphor screen coupled to the CCD by a lens [70]. These cameras don't need a long exposure time. To increase sensitivity, they have a high signal to noise ratio for short acquisition times so they are adapted to photon counting, and don't need the same control as EM-CCDs even though the MCP needs to be cooled down at about $-30°C$ [20]. These last ones are thus preferred sensors for real time molecular imaging [2] and are the ones used in this thesis.

I.2.4.2 Planar Imaging Setups

There are 2 main set-ups used in optical imaging for *in vivo* experiments : $2D$ imaging and $3D$ tomographic methods [2]. Planar ($2D$) imaging methods are the most well-known ones

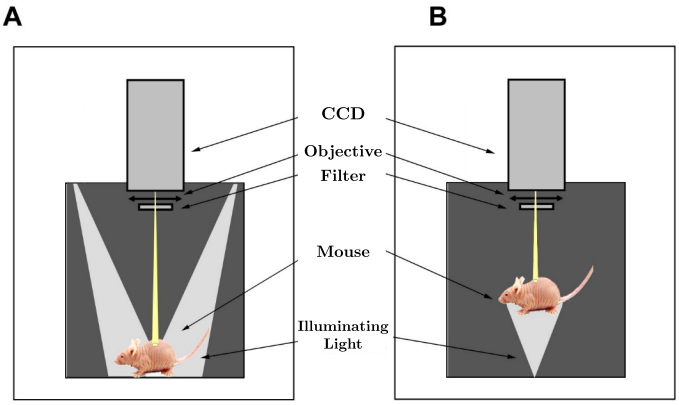

Figure I.12: Different types of planar devices for FLI, adapted from [2]. A: epi-illumination device. B: transillumination device.

because they are easy to set but cannot yield accurate quantification or deep location, that can be provided by tomography [71]. Planar imaging can be used for both bioluminescence and fluorescence [72,73] but tomography is still a more challenging problem in bioluminescence than fluorescence because of the absence of reference due to the absence of exciting light compare [74]. Bioluminesce Tomography therefore often needs prior informations from other modalities like CT or MRI [75].

The set-up of planar imaging is really simple, explaining its frequent use in optical imaging. It consists of a a light-tight enclosure where is placed the sensor having to acquire the luminescence emitted by the studied source. Fluorescence Imaging devices contain also the exciting light system. In this case, the light exciting system can be on the same side as the sensor with respect to the animal (epi-illumination) or on the other side (transillumination) [2], as shown in Figure I.12.

The orientation of the emitting subject can really impact the luminescent signal, therefore tomographic methods has been developed to improve quantification of deep tissues emitted light [76]. In fluorescence it is possible to do so thanks to the exciting light that acts as reference, by applying mathematical algorithms that describe photons propagation. Pulsed lasers can for instance enable to separate ballistic photons from others [2]. In bioluminescence, similarly to CT, it is possible to acquire $2D$ views from several sides of the animal, with for instance a mirror-based set-up [77] (Figure I.13) or a rotating device [78], to reconstruct the $3D$ image.

In this work, we used bioluminescence imaging with planar imaging for the sake of simplicity of set-up. The device used be described in the next subsection

I.2.4.3 Device used in this work: PhotonImager OPTIMA

This device has been developed by the company BiospaceLab (Nesles-la-Vallée, France), which we are working with in the context of this thesis. Its main competitive aspect comes from the real time imaging developed in this company. Even if we did not use it in this work because we focused on BLI, the PhotonImager OPTIMA has also a FLI mode. It is composed of an excitation light system and filters before the sensor [12].

The system uses an iCCD sensor and 2 objectives. One has a focal length of $50mm$, with

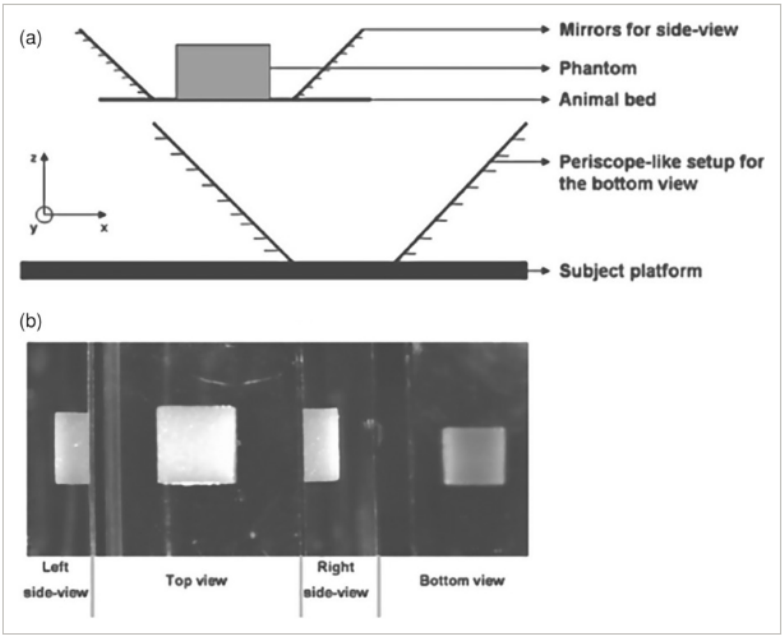

Figure I.13: Multiview set-up for bioluminescence tomography taken from [77, 78]. Top: schematic of the mirror-based set-up. Bottom: examples of image acquired by the CCD camera with this set-up.

an aperture between 1.2 and 15.6, the other has a focal length of $35mm$ and an aperture between 1.4 and 16.2. The sensor can acquire images at a frame rate of $11Hz$ and the gain of the microchannel-plate (MCP) is between 10^4 and 10^6 depending if recording fluorescence of bioluminescence images. The MCP is cooled at $-30°C$ to decrease thermal noise. The resolution of the iCCD is 1832×1377. This camera takes a picture of the animal before the acqusition as a preview to superpose the BLI on. The field of view is of $25cm \times 18cm$ and the distance between objective and object of $450mm$. Biologists can image up to 10 mice simultaneously.

Information of photons detected by the sensor is stored in a file with the time of arrival and the position [12]. In our experiments, we used this noisy raw data to be sure to control everything.

I.2.5 Typical Experiment

A typical BLI experiment ran by our collaborating biologists operates as follow : they inject luciferin into anesthetized mice, then place them in the imaging device [33]. The substrate is injected intraperitoneally as it is easier and faster than intravenous injections [79]. The sensor captures the signal emitted by the mice. This whole pipeline is illustrated in Figure I.14. After acquisition, biologists usually hand-draw a region of interest (ROI) around the studied tumor [80] in the Software associated with the imaging device (M3Vision from BiospaceLab for the PhotonImager OPTIMA system Figure I.15) which displays mainly the accumulation image (the integral of luminescent signal on all frames for all pixels) but also the signal frame by frame. The captured signal should contain quantitative information about the tumor, which we are going to study later.

In the following, since BLI of tumors is a major field of research for new cancer treatments [33], we are going to focus on bioluminescence imaging for tumor-bearing mice. Developments

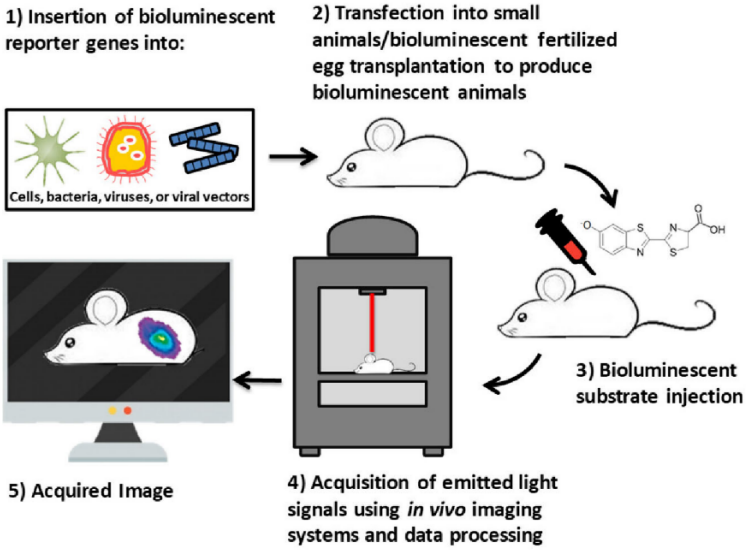

Figure I.14: A typical bioluminescence imaging acquisition for related experiment taken from [33].

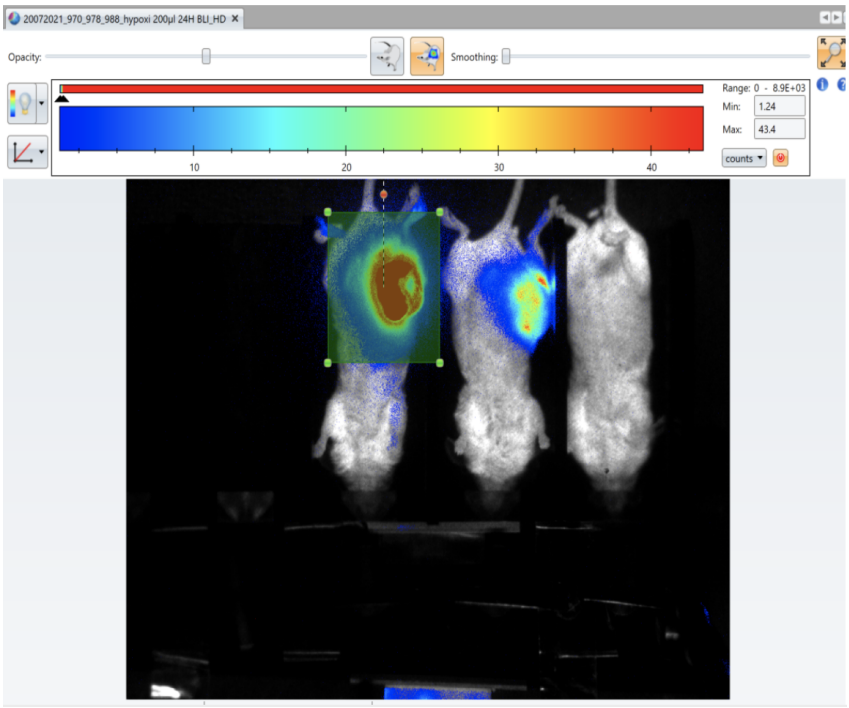

Figure I.15: An example of view for BLI on M3Vision.

in this field can include research on substrates [81], enzyme [82], or image analysis [83]. We will explain in next sections how and if image analysis tools can improve accuracy of BLI.

I.3 Bioluminescence Imaging

In the following, for the sake of display, sometimes we will show a $2D$ BLI. It represents the integral of photon counting on all frames, as in shown in Figure I.15.

I.3.1 Quantification of tumor state

In preclinical oncology research, it is common to use tumor volume to evaluate the tumor growth and estimate the treatment efficacy [53, 84, 85]. BLI is particularly useful for orthotopic tumor models, when measuring the tumor with a caliper is impossible due to its depth in the tissue [86] or for nonpalpable tumors [87]. The basic assumption is that the photon flux is correlated with the number of cancer cells [85, 88] and thus to the volume of the tumor [89]. It allows then to quantify the tumor burden of the animals without the need to euthanize them, even in deep tissues, and therefore to monitor the response of a subject to a cancer treatment [53, 88, 90] in the context of preclinical research. The bioluminescent signal can be quantified in several ways that we will explain in the following.

In that part, we will note $Y \in \mathbb{R}^{M\times N\times T}$ the BLI acquired by the device described earlier in a hand-drawn ROI around a specific tumor, where $Y(x, y, t)$ is the number of photons acquired by the pixel (x, y) at the frame t. The kinetics displayed from now on, as in Figure I.17, follow these rules: unless otherwise indicated, the time interval ranges from 0 to about 30 minutes after luciferin injection; and when a scale is specified, unless otherwise indicated, the unit is 8.9 seconds. In addition, unless otherwise indicated, the kinetics are presented after the pre-processing explained in Chapter II.

I.3.1.1 Average Photon Flux

It is quite common to use the average of photon flux over the acquisition period to quantify the Bioluminescent Signal (BLS) [53, 91–94] Equation I.1.

$$\text{Average Photon Flux} = \frac{\sum\limits_{x\leq M, y\leq N, t\leq T} Y(x, y, t)}{T} \tag{I.1}$$

In general, scientists using the luciferin/luciferase system to assess tumor growth inject the substrate and start imaging when the enzymatic reaction reaches the pre-determined plateau (about 10 minutes) [47,53,88,89,92,93,95] with acquisition times varying between a few seconds [88] and 10 minutes [47, 95]. This is motivated by the fact that the bioluminescence signal usually increases for about 10 or 15 minutes, then stagnates on a peak for 20 or 30 minutes before decreasing [72]. This procedure is summarized in Figure I.16.

However, using a fixed time window after substrate injection may be poor practice as it does not take into account the variability of BLS between tumors. Indeed, according to other many published works [57, 96–98], the temporal activity of the bioluminescence signal after substrate injection varies according to the tumor and the day as shown in Figure I.17, and therefore acquiring at a fixed time after injection, called a static acquisition, is not optimal for tumor quantification. Studying the entire time window may provide more accurate indicators of tumor state. Thus, some works use the average photon count on the whole bioluminescent kinetics [91, 99]. Nevertheless, according to [99], the optimal time point to begin the acquisition for a study using the average photon count over a static window is at 20 minutes after an IP injection because it is the moment where they found the correlation with the average photon count over the whole kinetics is the highest. However, this quantification method cannot differentiate between totally different kinetics with the same integral. A deeper analysis of the whole signal can provide more information.

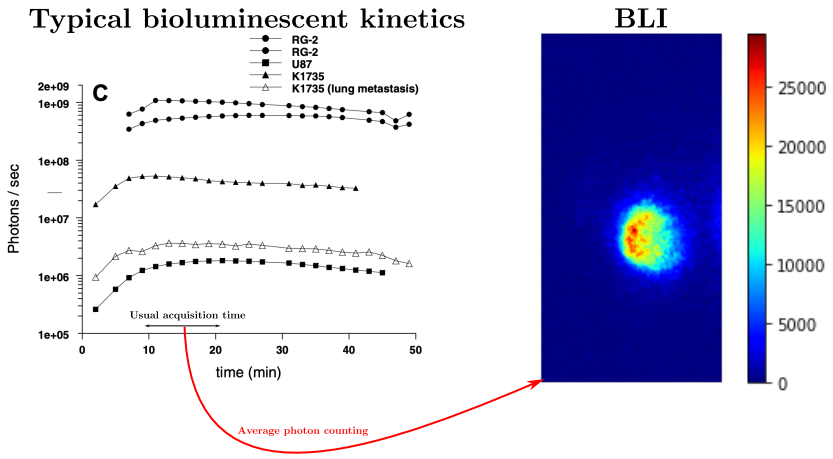

Figure I.16: Illustration of classical bioluminescent data (adapted from [72]). Left: Regular bioluminescent kinetics. Right: Accumulation image (integration on all frames of photon counting) from the usual acquisition time.

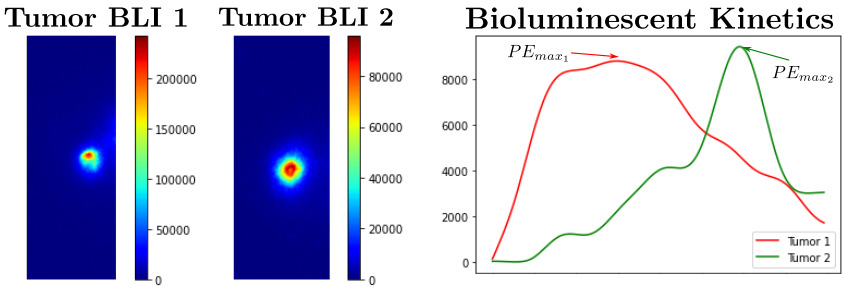

Figure I.17: Possible phase shift between kinetics. Left : BL Images. Right: Associated kinetics and their respective peak values.

I.3.1.2 Maximal Photon Emission

In [57], they estimate the maximum photon emission PE_{max} (see Figure I.17) of the tumor by calculating the 95th percentile on 5s time intervals over the 40 min acquisition, similarly to what is done in [94]. They also compute the average of photon counting during 1 minute at the average time-to-peak between the different BLIs, which should be even more accurate than the 20 minutes found earlier in [99]. This static measure is apparently less precise than PE_{max}, especially for IP injections.

$$PE_{max} = F_I^{-1}(0.95), \tag{I.2}$$

where F_I is the cumulative distribution function of I. However, this method is only a truncation and more information can be extracted when the entire acquisition is available.

I.3.1.3 Kinetic Modeling

Sim [100] and Dai's [84] work focuses on modelling the bioluminescence reaction in the context of an IP injection of luciferin. They use a multi-compartment pharmacokinetic model in which they study the uptake of luciferin before it reaches the luciferase-expressing area (the tumour tissue). They both made the assumptions that the rate of photon emitted from the tumor is proportional to the amount of substrate in the tumor area and that the reaction follows a

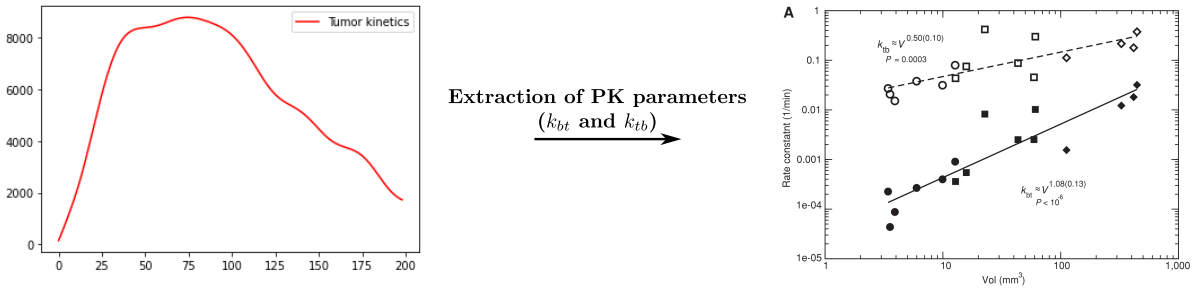

Figure I.18: Quantification of BLI with PK parameters, adapted from [100]. Left : Bioluminescent kinetic. Right : Logarithmic correlation between PK parameters and tumor volume.

Michaelis-Menten kinetics. [84] uses a 3 compartments model but [100] argues that the results are nearly identical with a 2 compartments model. This model allows them to obtain a temporal expression of the BLS in function of the pharmacokinetics parameters (PK parameters) that can be estimated from a real BLS by fitting the signal. They use one of these parameter, noted K, to quantity the tumor by using its logarithm, that is supposed to be proportional to the tumor volume (Figure I.18).

$$\text{PK quantification} = \log(K) \tag{I.3}$$

I.3.1.4 Other quantifications

In order to obtain an automatic unsupervised BLI analysis tool, [83] have built a 2-step algorithm. First, they use a K-Means algorithm to segment the image into 3 clusters: noise, background and signal (algorithm validated on synthetic data from Poisson distributions with different parameters). Then, they use the normalized BLI defined by the equation I.4 to quantify the signal in the same way as a standard bioluminescence image as displayed in Figure I.19. This normalization process allows them to account for auto luminescence and compare bioluminescence images from different devices.

$$NBLI = \frac{1}{T} \sum_{x,y,t} \frac{Y(x,y,t) - \mu(Background)}{\sigma(Noise)} \tag{I.4}$$

Moreover, in order to improve tumor volume monitoring and to standardize the quantification process of biolumienscence images, [47] proposes to use the luminoscore, defined as the sum of the photon fluxes acquired by the front and back views. This index thus requires though 2 times more acquisitions than the classical photon flux (Figure I.20).

$$\text{Luminoscore} = \frac{1}{T} \sum_{x,y,t} Y_{front}(x,y,t) + Y_{back}(x,y,t) \tag{I.5}$$

[101] proposes to calculate an estimate of the surface of the tumor, evaluated by counting the numbers of pixels belonging to the tumor after a segmentation computed with an Active Countour algorithm [102] on the accumulation image, defined by the integral of photon counting on all the frames. Before the segmantation, a pre-processing algorithm composed of a denoising and a deconvolution is applied to better segment the tumor.

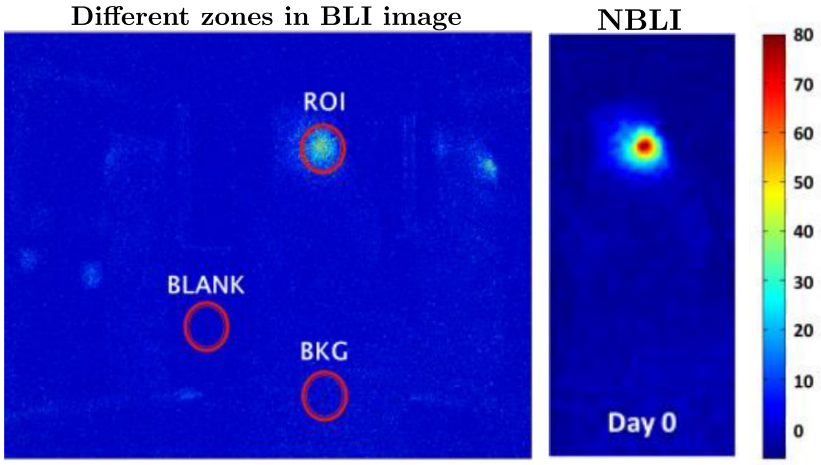

Figure I.19: Normalized BLI quantification method, adapted from [83]. Left: Different clusters in the accumulation image. Right : Associated NBLI.

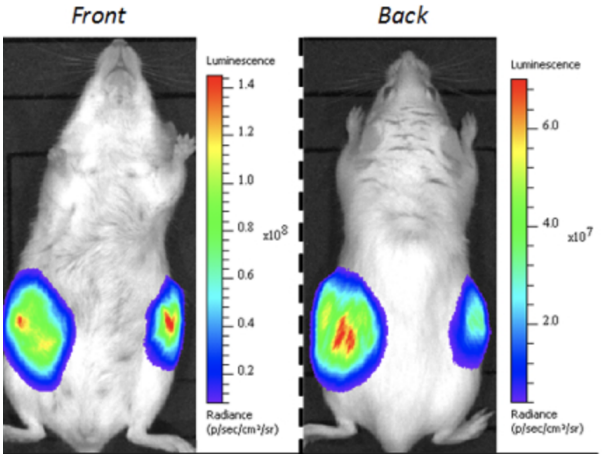

Figure I.20: Front and back view of a mouse bearing 2 tumors used to compute the luminoscore, taken from [47].

$$\text{Surface} = |\hat{I}|_0, \tag{I.6}$$

Where $I(x, y) = \sum_t Y(x, y, t)$ is the accumulation image and $\hat{}$ is the preprocessing.

I.3.1.5 Limitating factors and validity of quantification

To validate this process, biologists have used other modalities like caliper [85] or MRI [103] to measure the tumor volume. It showed that these quantification methods are not entirely reliable to determine accurately tumor volume, even if it can achieve satisfying results for big healthy tumors [94].

Indeed, Bioluminescence imaging suffers from factors that complicate signal quantification. First, the reaction requires the presence of molecules such as O_2 or ATP [72]. If one of these molecules is missing, the intensity of the emitted photon flux is not representative of the number of tumor cells [104]. This is particularly troublesome in hypoxic or necrotic zones appearing when tumors become too large [93]. Numerous works [57, 105–108] show the influence of these

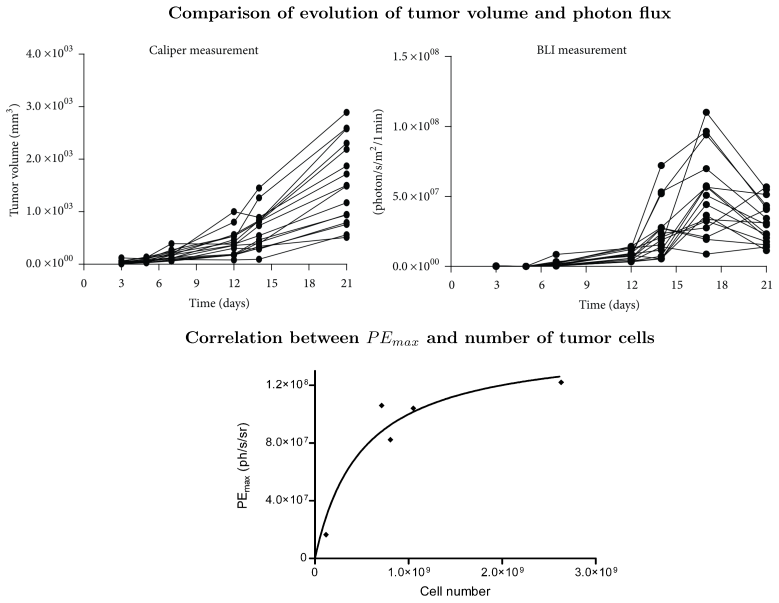

Figure I.21: Decrease of correlation between tumor volume and photon counting, adapted from [57, 106]. Top : comparison between evolution of tumor volume measured with a caliper and evolution of photon counting [106]. Bottom : Peak value of photon counting after IV administration in function of the histological cell count [57].

zones in the quantification of the bioluminescence signal. The correlation between bioluminescence signal and tumor volume, which grows almost exponentially as modeled in [109], drops as the tumor health status deteriorates and necrosis appears [47] (Figure I.21). This implies that bioluminescence imaging is no longer sufficient and that another modality (such as MRI) seems necessary to determine the volume of such a tumor. It appears necessary to develop methods to take these areas into account when quantifying the bioluminescence signal. The study brought by [110], extends a pre-existing model of photon flux from a tissue [111] by taking into account the oxygen dependence of the reaction between luciferin and luciferase. This study is a possible avenue to improve tumor quantification when the entire kinetics are available.

Moreover, the interactions between the emitted photons and the crossed tissues modify the intensity of the detected signal. The interactions between photons and their environment is defined by 2 phenomena : absorption and scattering, characterized by their respective associated coefficients μ_a and μ_s which depend on the wavelength of the photons. They are also both in homogeneous but it is possible to define an average μ_a from the tissues constitution. However, the scattering coefficient is also anisotropic, with an anisotropy factor noted g which defines the reduced scattering coefficient :

$$\mu_s' = (1 - g)\mu_s \tag{I.7}$$

A tumor closer to the surface of the skin then appears brighter than a tumor located at depth. In addition, the scattering of photons in the tissue can cause tumors to appear larger than they really are as is shown in Figure I.22 that shows how 2 different sources can emit the same signal. Indeed, according to [98], both the position of the animal during bioluminescence acquisition and the anesthetic used can be responsible for large variations in the resulting signal. However, [100] did not observe significant variance in their results by tilting the animals in the device, both in the signal and the resulting pharmacokinetic parameters.

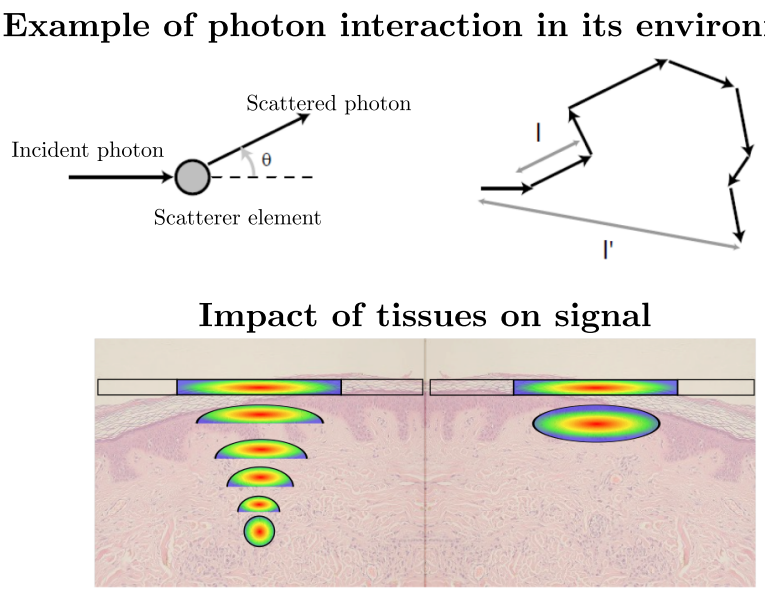

Figure I.22: Impact of tissue on photon emission, adapted from [12]. Top : explanation of scattering. Bottom : illustration of 2 very different sources that emit the same signal due to their position in the tissues and the associated scattering and absorption.

It is also important to note that, according to [72], the bioluminescence signal is often underestimated in larger tumors due to the increased optical density that reduces the amount of substrate available in the tumor core. However, BLI is often used in the context of longitudinal studies and not to determine the tumor burden of an isolated tumor. Normalizing the bioluminescence signal to the signal at the beginning of the treatment reduces inter-animal variation.

Another limitating factor of BLI is the variability in time-to-peak between different mice and tumors [57, 96–98]. [57] shows that there is less variance in the time-to-peak when proceeding with IV injections than with IP injections since there is no loss in the peritoneum, which could result in a better repeatability in the experiments and reducing the anesthesia of the animals. However, they are more difficult to perform and show the same gap of correlation between *in vivo* tumor volume and photon emission [57] (Figure I.21 bottom).

Therefore, there is a need to improve the way BLI is quantified to determine tumors status, both in terms of robustness and accuracy. Moreover, in order to quantify BLI, biologists have to manually draw a Region of Interest (ROI) around each tumor [80], which can be troublesome in some cases that we will discuss now.

I.3.2 Impact of number of tumors

First, it seems natural to ask whether the size of this ROI has an influence on the measured signal. According to [112] and [47], the BLI does not need excitation light therefore once the entire tumor is encompassed in the region, increasing the size of the ROI has no impact in terms of pure photon counting. It is an imaging technique more suitable for whole body imaging than fluorescence [47] and therefore taking a ROI that covers the entire animal, which can be easily done by a computer, should achieve accurate enough quantification. However in some experimental protocols the mice are bearing several tumors, by purpose or by metastases. In these cases, taking a ROI that covers the whole animal as proposed in [47] then mixes the

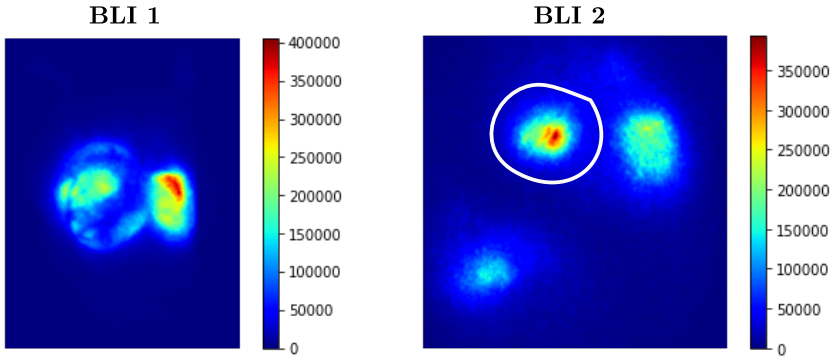

Figure I.23: BL Images where it's difficult to hand-draw ROI around the each tumor. Left: example with 2 very close tumors. Right: example with 2 tumors and one is in 2 parts, probably due to hypoxia or necrosis, making it difficult to attribute the white circle area to the right or left tumor.

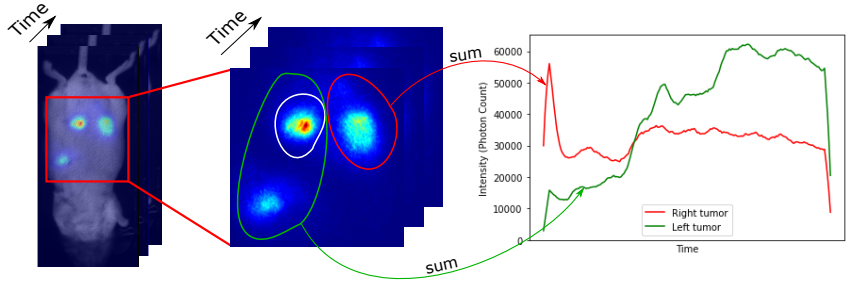

Figure I.24: Example of a BLI recording with 2 tumors that needs to be automated. Left: Superposition of BLI and mouse body. Middle: BLI photon count with expert hand-drawn ROIs (Green circled area: left tumor. Red circled area: right tumor. White circled area: zone difficult to attribute, belongs to the left tumor after close examination). Right: Average BLI photon accumulation time-series for each tumor.

signals from the different tumors and it then appears necessary to draw ROI around each tumor. It can get tricky, challenging, time-consuming and can introduce errors and biases in some cases like the ones presented in Figure I.23. In the first image, the tumors are too close and the border between them is unclear, so we can easily encompass pixels belonging to the wrong tumor when hand-drawing a ROI. In the second image, we have 3 luminous spots but we know from biologists who have the longitudinal information that there is only 2 tumors. It is then difficult to tell from this image whether the white circled area belonged to the right or left tumor. A close examination of the signal by biologists allowed us to determine that it belongs to the left tumor.

For now, there exists no dedicated pipeline that determine automatically the kinetics of each tumor located in a single Bioluminescence Image as illustrated in Figure III.12, which is then an interesting field of study.

I.3.3 Towards a more quantitative BLI by studying spatio-temporal patterns

In this work, we thus worked on 2 aspects :

- Separating the photon flux coming from each tumor automatically when the animal has

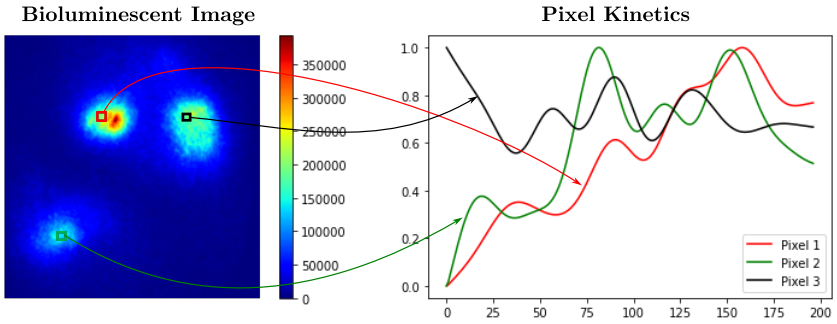

Figure I.25: Illustration of kinetic similarities between pixels of the same tumor. Left: BLI accumulation image. Right: pixel kinetics (after normalization).

several tumors

- Improving the quantification of the Bioluminescent Signal for preclinical imaging.

I.3.3.1 Blind source separation

To obtain the kinetics of the different tumors from a single BLI, we propose in this work to leverage the discrepancies between photon emission kinetics of different tumors [100]. More precisely, we made the assumption in our model that the different pixels belonging to only one tumor have almost the same temporal pattern (kinetics) as displayed in Figure I.25 that shows the normalized kinetics of 3 pixels in the image, unlike the pixels belonging to different tumors. Indeed, Figure I.25 shows that the kinetics of the 2 intense zones on the left are similar with a progressive increase, on the contrary of the pixel belonging to the intense zone on the right. Pixel-level kinetics seem therefore to be a relevant information to get the different tumor kinetics from a BLI.

To tackle such problem, many models can be adapted to these data. The most well-known is the spatial segmentation framework which classifies each pixel as belonging to a single tumor. However, when the tumors are very close, or superimposed if the point of view is not adapted, it may be difficult to attribute a pixel to a single tumor as the ones at the border between the 2 tumors of the first image of Figure I.23. The Blind Source Separation Framework (BSS) seems therefore to be adapted to that task, and is the one we are going to use in this work.

I.3.3.2 Quantitative analysis with reference region

The main limitations in bioluminescence imaging are the lack of repeatability of the experiments, the loss of signal due to necrosis and hypoxia and the interactions between photons and tissues, which are all valid perspectives to pursue. Before trying to model the impact of tissues on the kinetics, we first made experiments to quantify the possible variations occurring in the kinetics due to a slight tilt of the animal. We gathered acquisitions of some animals 3 times a day, during 40 minutes and compared the resulting kinetics for each tumor. In the same day, the tumor has not changed or grown and we imaged the animals in almost the same position so the influence of tissues should not be significant. The resulting difference between the associated kinetics was in fact very high and we could not find a particular pattern between the differences (that could have been due for instance to the hour of the experiments) as is shown in Figure I.26.

Hence, the quantification of tumor kinetics seemed to be a real bottleneck and it is the topic we focused on in terms of quantitative analysis of BLI. We tried to derive pharmacokinetics methods of quantification from other modalities such as PET [113] or DCE-MRI [114] based on

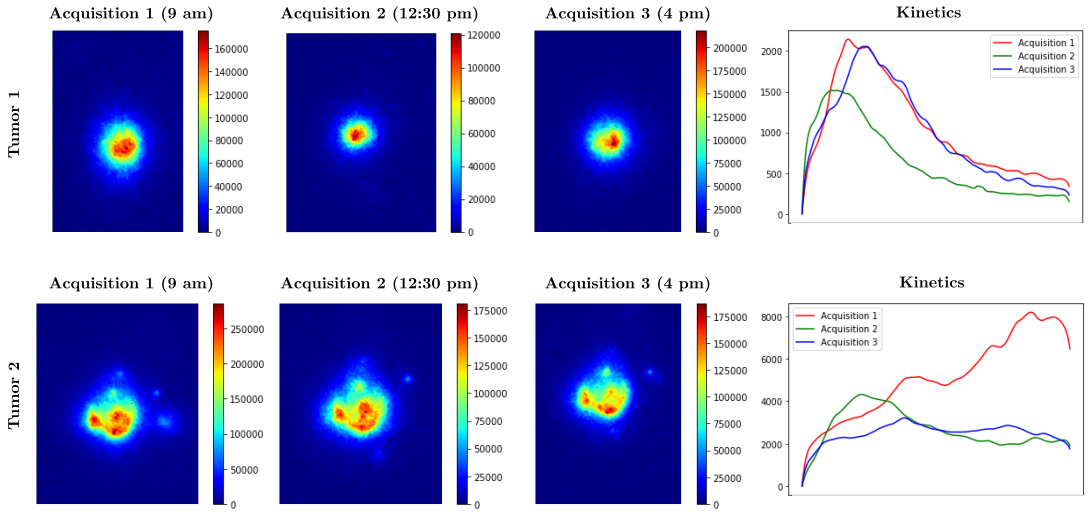

Figure I.26: Difference between kinetics of the same tumors at different times of the day. For each tumor, Left: BLI for each time stamp; Right: Associated kinetics.

the pharmacokinetics model of [100]. In particular, we focused on methods using signal coming from a Reference Region, such as usually the muscle, to better quantify the kinetics [115].

It should be noted that in this context the animals are supposed to be anesthetized thus motionless. However it is well known that physiological movement, such as the respiration, can complicate the analysis of such spatio-temporal data [116, 117]. In the scope of this thesis, we do not take this kind of movement into account but it represents a valid perspective for our work.

I.4 Manuscript outline, Collaborative Context and Publications

This thesis is a joint work between the Biological Analysis Unit of the Institut Pasteur and the IMAGES team of Télécom Paris, in collaboration with the company Biospace Lab and the UTCBS laboratory from the Université Paris-Cité, CNRS and Inserm.

The manuscript is structured as follows. We began with Chapter I - Introduction, where we provided general background information on pre-clinical imaging on small animals and the bioluminescence imaging technique.

We introduce in Chapter II - Bioluminescence Imaging Data the datasets we used in this thesis to design, test and validate our algorithms. We also present a synthetic tumor BLI generation framework that we designed. This simulator is the first to our knowledge able to generate spatio-temporal BLIs of tumors and will be a part of an article in preparation :

- Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J.-C. . Wavelet transform based non-negative matrix factorization for bioluminescence imaging with warping variability.

In Chapter III - Blind Source Separation on Bioluminescence Imaging, we dive into the algorithms we developed to separate the different tumors located within a single BLI. In par-

ticular, we adapted Blind Source Separation methods to our data, by adding prior information on tumor shape and heterogeneity. The associated publications are :

- Dereure, E., Angelini, E., Garofalakis, A., Seguin, J., Mignet, N., & Olivo-Marin, J. C. (2022). Factorisation Non-Négative de Matrice pour séparation de sources en Imagerie par Bioluminescence préclinique. In GRETSI'22: XXVIIIème Colloque Francophone de Traitement du Signal et des Images.
- Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J. C. (2024). Sparse Non-Negative Matrix Factorization for Preclinical Bioluminescent Imaging. In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI).
- Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J. C. (2024). Wavelet-based sparse non-negative matrix factorization for bioluminescent imaging unmixing. In 2024 IEEE 21st International Symposium on Biomedical Imaging (ISBI).

Some results will also be a part of an article in preparation:

- Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J.-C. . Wavelet transform based non-negative matrix factorization for bioluminescence imaging with warping variability.

Then, in Chapter IV - Quantitative analysis of BLI with pharmacokinetics model, we focus on improving the quantification of BLI by enhancing the calibration of an existing pharmacokinetics model.

This work is fully summarized in Chapter V - Conclusion and Perspectives, where we discuss our results and highlight interesting perspectives for future research directions.

Chapter II

Bioluminescence Imaging Data

II.1 Real BLI Datasets

In order to quantitatively analyze bioluminescent images, we worked alongside biologists, from the Université Paris Cité in the teams Plateformes mutualisées de l'institut du médicament (LIOPA) and UTCBS. These teams provided us with the datasets used in dedicated experiments for different problems tackled in this thesis.

All BL Images cohorts were acquired with an iCCD camera (PhotonImager OPTIMA, BiospaceLab, Nesles la Vallée, France). Luciferin potassium salt diluted in 10mM phosphate buffer was injected intraperitoneally (IP) at a rate of 2 mg per anesthetized mouse, which is in great excess compared to the amount of luciferase. Acquisitions were made 3 times a week from tumor implantation until the mice were sacrificed. Indeed, each mouse was sacrificed when the tumor volume, measured with a caliper, became too large (more than $1000mm^3$) or when the BLI signal became too strong (more than $5 \times 10^5 ph/s/cm^2/sr$). The frame rate of each acquisition is 89ms per frame and the pixel size is $0.12 \times 0.12mm^2$. In each frame, the pixel value represents the number of photons captured in that pixel.

II.1.1 Multiple-tumors bearing mice datasets

For these datasets, each BLI acquisition consists of a time series of 2D signals acquired over a fixed field of view containing several mice each bearing 2 tumors. We hand-drew ROIs around the 2 tumors of each mice as shown in Figure II.1. Each BLI consists then of a 2D+T series of about $500 \times 500 \times N_T$ pixels where N_T it the total number of frames, containing the signals of 2 tumors of each mouse. We evaluated our blind source separation algorithms, introduced in Chapter I and described in Chapter III on these datasets.

II.1.1.1 Short kinetics dataset

One dataset contains images coming from 6 different balb/cJRJ mice injected subcutaneously with a suspension of CT26-Luc cells ($2.5 \times 10^5/50\mu L$ culture medium). BLI acquisitions were initiated 8 min after substrate injection for a duration of about 7 min. A typical acquisition contains about $N_T \approx 4650$ time frames.

We considered only images where the 2 tumors emit signal and selected only data from acquisitions where the tumoral volume is of the same order of magnitude for the 2 tumors and where it is greater than $10mm^3$ to have enough signal. This selection of data places us in an ideal framework in the context of a preliminary analysis. These criteria led us to exclude 20

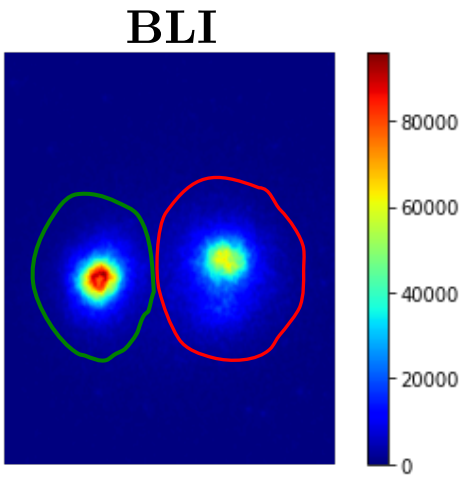

Figure II.1: Regions of Interest (ROIs) manually drawn around each tumor in a BLI.

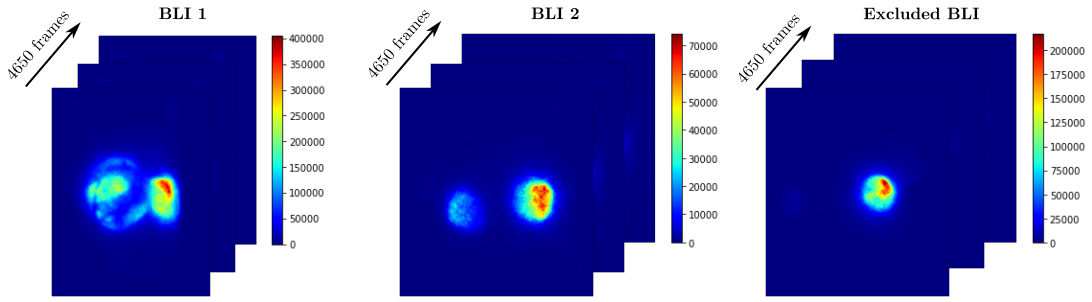

Figure II.2: Example of BLIs of mice bearing 2 tumors each. Left and Middle: BLI retained in our dataset; Right: BLI excluded from our dataset (the left tumor is approximately as voluminous as the right one according to the caliper measurement but does not emit signal, possibly due to an injection failure). Acquisitions of about 7 minutes ($N_T \approx 4650$ time frames). In this figure we illustrate the temporal aspect of BLI but the for the sake of visibility the frame displayed is the cumulative count of photon counting.

BLI. In total, with 6 mice we retained 15 BLI images and thus 30 tumors. Figure II.2 displays examples of such images, with an example of an excluded BLI (one of the tumors does not emit signal possibly due to an injection failure while being almost as voluminous as the other one according to the caliper measurement). In this type of figure we illustrate the temporal aspect of BLI but the for the sake of visibility the frame displayed is the cumulative photon counting.

II.1.1.2 Long kinetics dataset

A second dataset contains also images coming from 6 different balb/cJRJ mice injected subcutaneously with a suspension of CT26-Luc cells (4 with $10 \times 10^4/50\mu L$ cells, 1 with $8 \times 10^4/50\mu L$ cells and 1 with $5 \times 10^4/50\mu L$ cells). BLI acquisitions were initiated immediately after substrate injection for a duration of about 30 min (to capture the whole kinetic). A typical acquisition contains about $N_T \approx 20000$ time frames.

For this dataset, we considered only images where the 2 tumors emit signal and excluded 13 BLI. In total, with 6 mice we retained 18 BLI images and thus 36 tumors. Figure II.3 displays examples of such images, with an example of an excluded BLI (one of the tumors does not emit signal possibly due to an injection failure).

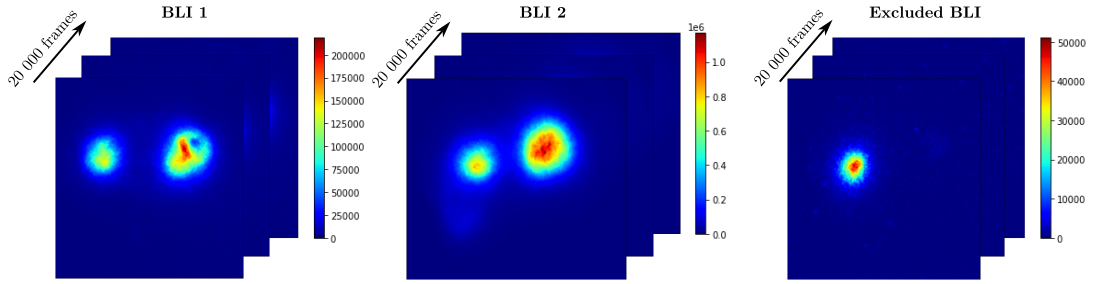

Figure II.3: Example of BLIs of mice bearing 2 tumors each. Left and Middle: BLI retained in our dataset; Right: BLI excluded from our dataset (the tumor on the right does not emit signal, possibly due to an injection failure). Acquisitions of about 30 minutes ($N_T \approx 20000$ time frames).

II.1.2 Tumor and muscle signal-emitting mice dataset

This BLI dataset is composed of images coming from 15 different BALB/c female mice (Janvier, St Genest de Lisle, France) aged of 6 weeks that were split into 3 groups of 5 mice ($5 \times 3 = 15$ mice). 2 groups were treated with different of Etoposide formulations, Etoposide Nanocrystals (noted ETO in the following), Etoposide/Prednisolone Nanocrystals (noted ETO/PRD in the following) and one group was used as control (noted CTR in the following). The first 2 groups received 6 injections of ETO formulations at 10 mg/kg. The untreated control group was used as comparison for tumor volume. Schedule: Murine carcinoma tumors CT26 were subcutaneous were implanted subcutaneously on day -3 using a 12-gauge trocar (38 mm) into the mouse flank previously disinfected with alcohol. Most of the mice bear 2 tumors but some bear only 1 because the other tumor has not grown.

In order to obtain a signal from a reference muscle during a BLI acquisition, which would yield a calibration of the arrival of the substrate, an internal control was used. It consists of an injection of $3\mu g$ of recombinant luciferase in $30\mu L$ of PBS into the right cranial tibial muscle of each mouse. This is supposed calibrate the signal coming from the tumor in the goal of improving its quantification [115] as we will explain more deeply in Chapter IV.

BLI acquisitions were initiated immediately after substrate injection for a duration of about 30 min (to capture the whole kinetic). A typical acquisition contains about $N_T \approx 20000$ time frames. For each acquisition, tumor volume was measured with a caliper by using the ellipsoid approximation of tumors shape [85].

We excluded 4 BLI for which we did not have correct annotations. In total, we have 77 BL images. For each BLI, ROIs were drawn around each tumor and reference muscle, leading up to a total of 147 tumors and their associated 77 muscles. 52 tumors belong to the group CTR, 44 to the group ETO and 51 to the group ETO/PRD.

We evaluated our tumor quantification algorithms, introduced in Chapter I described in Chapter IV on this dataset.

II.2 Modeling and Preprocessing of BLI

According to [101], only a small number of photons emitted by the source will reach the camera. It is possible to model, following their work, the image acquired by the iCCD camera Y as a mixed Poisson-Gaussian noise model described as:

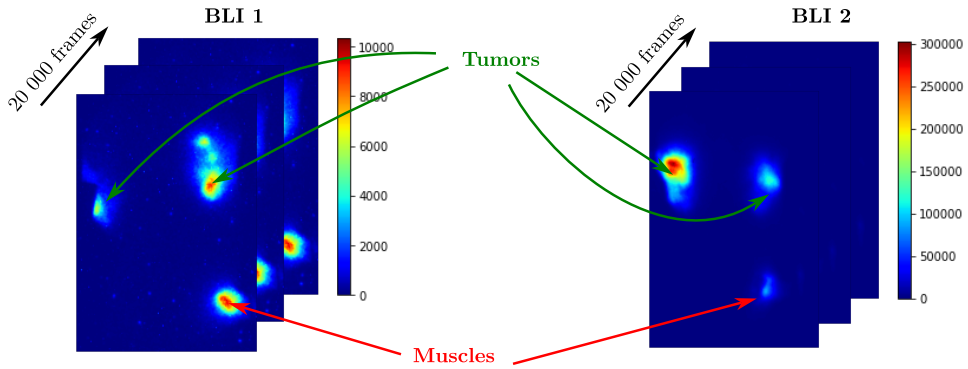

Figure II.4: Example of 2 BLIs of mice bearing 2 tumors each, with some signal emitted in the muscle thanks to a local luciferase injection. Acquisitions of about 30 minutes ($N_T \approx 20000$ time frames).

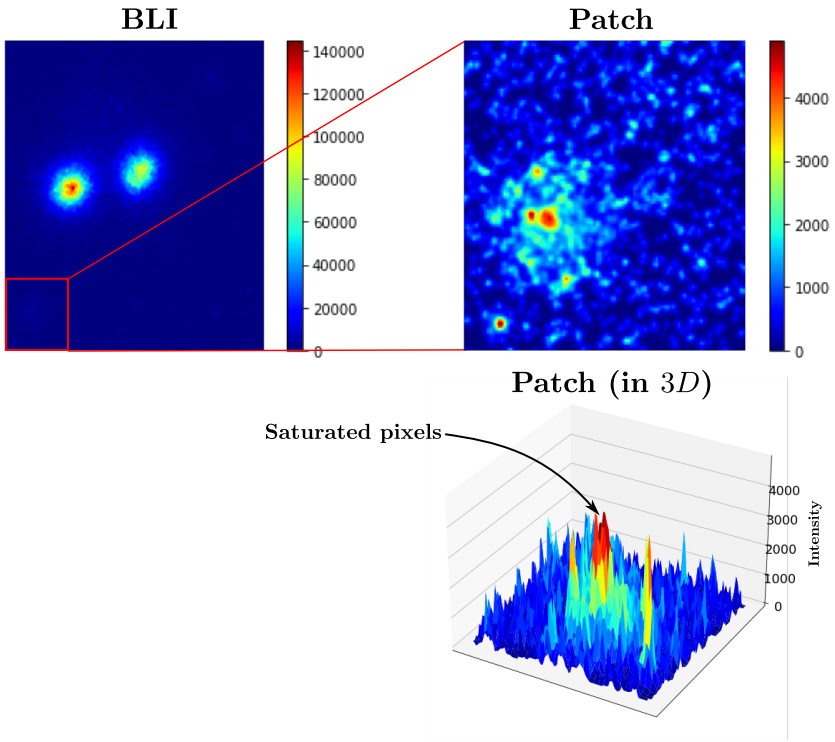

Figure II.5: Impact of additive noise on BLI. Left: BLI total cumulative count, Right: Patch in the Background (Top : visualization in $2D$, Bottom: visualization in $3D$).

$$Y(x,y,t) = P(O(x,y,t) * h(x,y,t)) + \sigma(x,y,t), \tag{II.1}$$

where $O(x,y,t)$ is the original image, P is a Poisson process, $h(x,y,t)$ is the Point Spread Function (PSF) that depends on the imaging device and $\sigma(x,y,t)$ is an additive noise component composed of a stationary white Gaussian noise due to camera read-out, quantization and dark current and an impulse noise due to cosmic radiation [83, 101, 118]. This last component is responsible for possible saturated pixels [47]. In the context of pixel-wise kinetic analysis, the raw data is too noisy and needs to be pre-processed to limit the risk of noisy detections as displayed in Figure II.5.

The signal is sampled with a sampling period noted $Te = 8.9s$ (100 time points) in order to

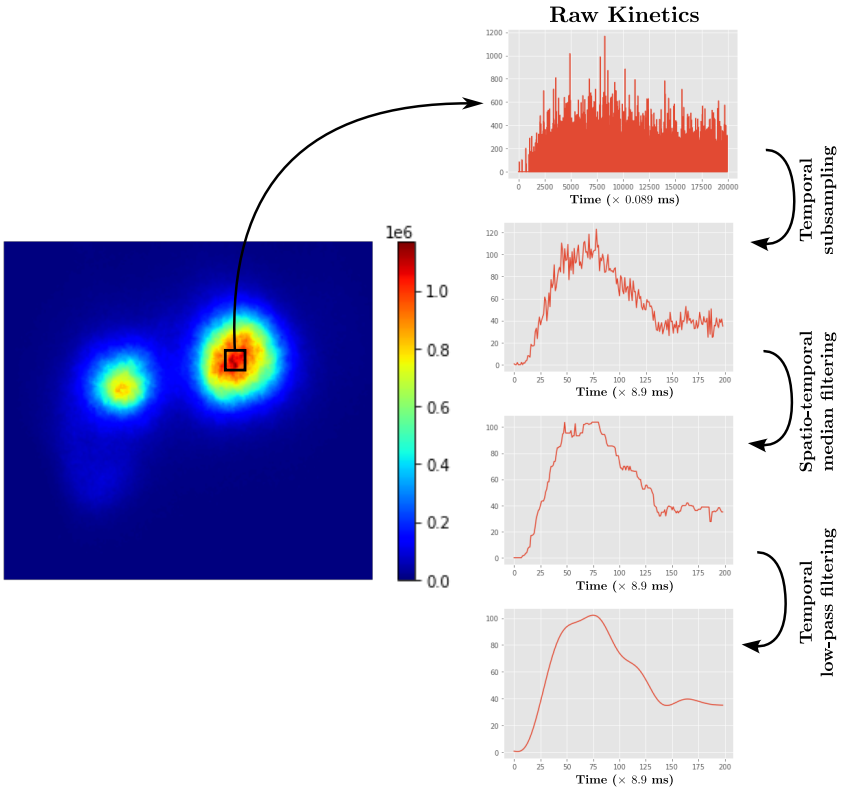

Figure II.6: Different pre-processing steps applied on raw BLI. Left: BLI, Right: Impact of the pre-processing on a pixel kinetic curve.

reduce the computation time, by averaging temporally, then smoothed spatio-temporally with a median filter of size 9×9 to follow the idea of [119]. Indeed, this filtering should smooth the image spatially and remove impulse noise. However, we can still observe non-relevant oscillations, that are cut thanks to a low-pass Butterworth filter of degree 10 and of cutoff frequency $f_c = \frac{1}{150} Hz$ for the short kinetics dataset and $f_c = \frac{1}{280} Hz$ for the long kinetics thers. These parameters have been set empirically and the difference between the cutoff frequency can be explained by the fact that in the case of short kinetics, if we smooth too much the signals, with a cutoff frequency too low, it becomes harder to distinguish between 2 tumors within the same image. For longer kinetics, relevant oscillations can be found at very low-frequency analysis therefore we can use a stricter low-pass filter. Figure II.6 illustrates the different pre-processing steps applied to BLI and their impact.

We then applied our algorithms to these pre-processed BLIs and evaluated them with approximated ground-truth as explained later in Chapter III. In order to to validate them with complete ground-truth, we then defined a synthetic BLI generation framework.

II.3 Proposed synthetic tumor BLI generation

We developed our own synthetic tumor BLI simulator to calibrate our various algorithms. One assumption made is that all pixels of a single tumors share a similar kinetic pattern as can be seen in Figure I.25 but with different orders of magnitude between pixels. In order to generate a BLI corresponding to a specific tumor, we first synthetize its Parametric Map (the intensity of each pixel), see subsubsection II.3.1.1 and its associated kinetic pattern, see subsubsection II.3.1.2. We then apply dedicated pixel-level degradations to mimick the different sources of noise and

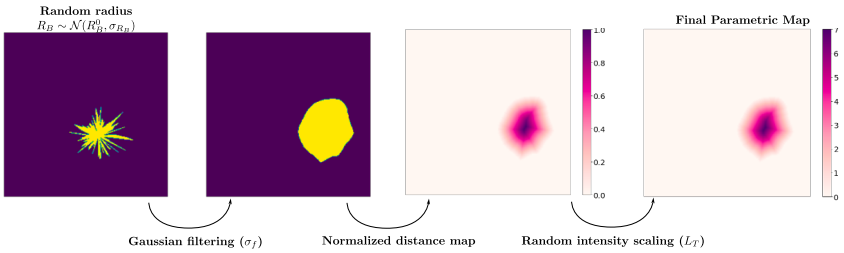

Figure II.7: Steps of generation of the parametric map of a generic tumor. For the sake of visualization, both masks are displayed after scaling the radius values by S_B. To generate this image we used $R_B^0 = 5, \sigma_{R_B} = 2.5, \sigma_f = 5$, $L_T = 7$ and $S_B = 50$.

heterogeneity, see subsection II.3.2. We finally explain how to simulate images with multiple tumors to generate a multiple-tumors synthetic BLI dataset in subsection II.3.3 by generating a Parametric Map and Kinetics for each tumor. This process allows us to handle the numbers of tumors in each image, and for each tumor its size, intensity, shape, position, kinetics and the associated perturbations.

II.3.1 Ideal tumor Bioluminescent Image generation

II.3.1.1 Parametric Maps generation

As seen in Figure II.2, Figure II.3 and Figure II.4, a tumor shape can be approximated by an almost circular form called in the following *blob*. All parameters are provided here for BLI of size 256×256 pixels.

Formally, a blob is therefore modeled as a disk whose radius R_B follows a normal distribution $R_B \sim \mathcal{N}(R_B^0, \sigma_{R_B}^2)$ and whose center is randomly placed on the image. We sample N_{R_B} values that are smoothed by a Gaussian filter of variance σ_f^2.

To obtain blobs of potentially very different sizes, these values are then multiplied by a scaling coefficient $\frac{S_B}{D_B}$ where S_B is randomly chosen as we will detail later, forced to be greater than 10 to matter, and D_B is the maximum of the horizontal and vertical diameters of the current blob.

In order to get the Parametric Map of the tumor, a normalized distance map is then computed on the blob, multiplied by a random intensity value $L_B \sim \mathcal{N}(L_B^0, \sigma_{L_B}^2)$ so that tumors have different intensities. A complete display of this process is shown in Figure II.7.

As we can see in Figure II.8, because of necrosis and hypoxia blocking the generation of photons, tumors can be composed of N_B blobs. Multiple blobs can be used to model a *disjoint* tumor, meaning in this work a tumor having hypoxic or necrotic areas, as the left one in Figure I.25. We used $N_B \in \{1, 2, 3\}$ to be exhaustive and representative of strongly degenerated tumors. N_B is a random variable that follows the distribution $\{0.6, 0.3, 0.1\}$ to mimick real data where severe deterioration is uncommon so tumors are very likely to be composed of a single blob, and very unlikely to be composed of 3 blobs. We limit the sizes (S_B) so that the more blobs a tumor contains, the smaller the blobs are in order to keep total tumor sizes of the same order of magnitude.

More precisely, the blob size S_B is in fact a random variable $S_B \sim 10 + \mathcal{U}(0, \frac{R_T}{\sqrt{N_B}})$ that uses $R_T \sim \mathcal{U}(30, \min(70, R_T^{sup}))$ where $R_T^{sup} = D - 10$ with D the distance between the center of the tumor, randomly chosen, and the closest border of the image.

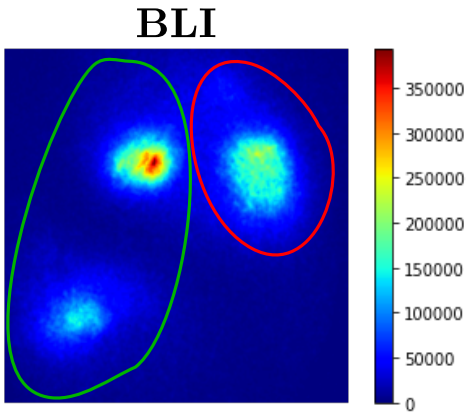

Figure II.8: Example of BLI with a disjoint tumor because of hypoxia or necrosis (the tumor on the left). Regions of Interest (ROIs) manually drawn around each tumor.

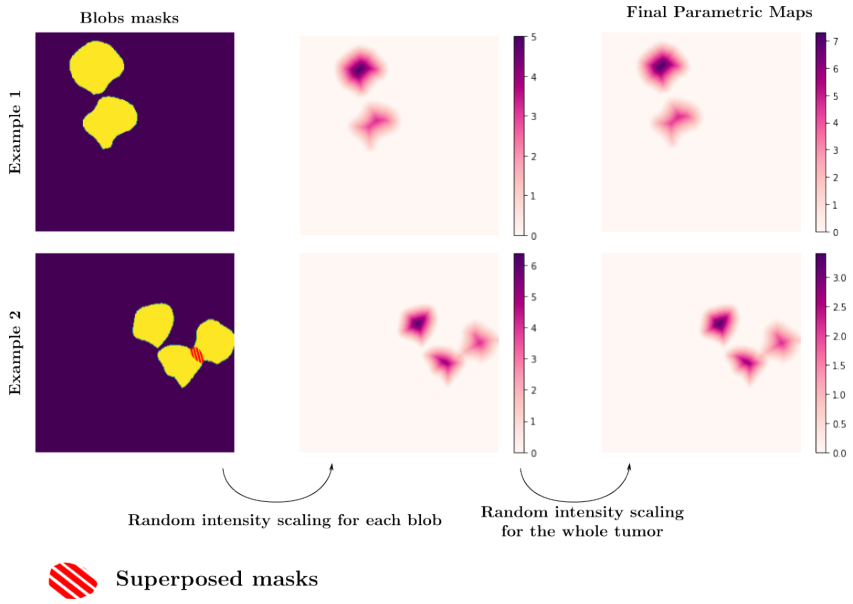

Figure II.9: 2 examples of Parametric Map generation for tumors composed of several blobs. The red hatched part corresponds to the superposition of the masks of different blobs. We used $R_B^0 = 5$, $\sigma_{R_B} = 2.5$, $N_{R_B} = 100$, $\sigma_f = 5$, $L_B^0 = 5$, $\sigma_{L_B} = 1.3$, $R_D = 50$, $L_T^0 = 5$ and $\sigma_{L_T} = 1.3$.

We also constrain the blobs of the same tumor to be near each other by keeping their centers in a disk with a random center and a radius R_D, while also avoiding the borders of the image. With these parameters different blobs may touch but not be completely superimposed, see Figure II.9 (Bottom).

The Parametric Map at this step is noted I, and the final Parametric Map $I^{(*)}$ is then computed as $I^{(*)} = L_T \frac{I}{\max(I)}$ where $L_T \sim \mathcal{N}(L_T^0, \sigma_{L_T})$ to be more realistic with tumors having possibly really different peak intensities.

Final examples of synthetic parametric maps $I^{(*)}$ are illustrated in Figure II.9. We used $R_B^0 = 5$, $\sigma_{R_B} = 2.5$, $N_{R_B} = 100$, $\sigma_f = 5$, $L_B^0 = 5$, $\sigma_{L_B} = 1.3$, $R_D = 50$, $L_T^0 = 5$ and $\sigma_{L_T} = 1.3$ in Figure II.9 and in all the following simulations.

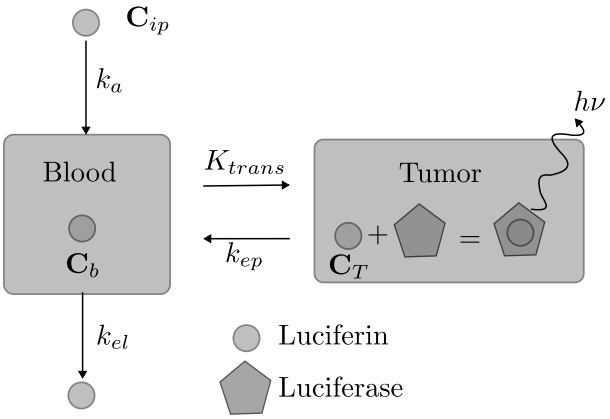

Figure II.10: BLI Pharmacokinetics model described in [100] that explains the emission of photons by the tumor from the interactions luciferin-luciferase.

II.3.1.2 Kinetics generation

In order to generate kinetics, we used the pharmacokinetic model from [100] that uses 4 parameters to describe the BL kinetics of a tumor as shown in Figure IV.1. We set the photon conversion factor and the initial concentration of luciferin injected to 1 to simplify equations. Moreover, in this work we consider this model valid even with necrotic or hypoxic tumors. Equations describing the model are the following:

$$\begin{aligned}
\frac{d\mathbf{C}_{ip}}{dt} &= -k_a\mathbf{C}_{ip} \\
\frac{d\mathbf{C}_b}{dt} &= -(K_{trans}+k_{el})\mathbf{C}_b + k_{ep}\mathbf{C}_T + k_a\mathbf{C}_{ip} \\
\frac{d\mathbf{C}_T}{dt} &= K_{trans}\mathbf{C}_b - k_{ep}\mathbf{C}_T,
\end{aligned} \tag{II.2}$$

Where $\mathbf{C}_{ip}, \mathbf{C}_b$ and $\mathbf{C}_T$ are the luciferin concentration respectively in peritoneum, blood and tumor. k_a is the rate constant describing the absorption of luciferin from the peritoneum into the blood and k_{el} is the elimination rate constant from the blood. K_{trans} and k_{el} are the rate constant describing respectively the transfer of luciferin from the blood to the tumor and from the tumor to the blood.

Since the concentration of substrate is negligible compared to the Michaelis constant of this enzymatic reaction, the photon flux $\frac{d\mathbf{P}_T}{dt}$ emitted by the tumor is proportional to the concentration of luciferin in it [100]:

$$\frac{d\mathbf{P}_T}{dt} \approx \lambda_T\mathbf{C}_T, \tag{II.3}$$

where λ_T is a constant. The explanation of this approximation is more elaborated in Chapter IV. For the simulations, for the sake of simplicity we used $\lambda_T = 1$.

Since most tumors are poorly vascularized we have $k_{el} >> k_{ep}, K_{trans}$ [100], and approximate $\frac{d\mathbf{C}_b}{dt} \approx -k_{el}\mathbf{C}_b + k_a\mathbf{C}_{ip}$. Thus:

$$\mathbf{C}_b(t) \approx k_a\int_0^t \mathbf{C}_{ip}(x)e^{-k_{el}(t-x)}dx \tag{II.4}$$

Table II.1: Parameters used to generate synthetic kinetics with model from [100]. $v_e = \frac{K_{trans}}{k_e}$

	Used values		
k_a (in min^{-1})	0.01	0.1	0.5
k_{el} (in min^{-1})	1	10	50
K_{trans} (in min^{-1})	0.01	0.1	0.35
v_e	0.01	0.1	0.5

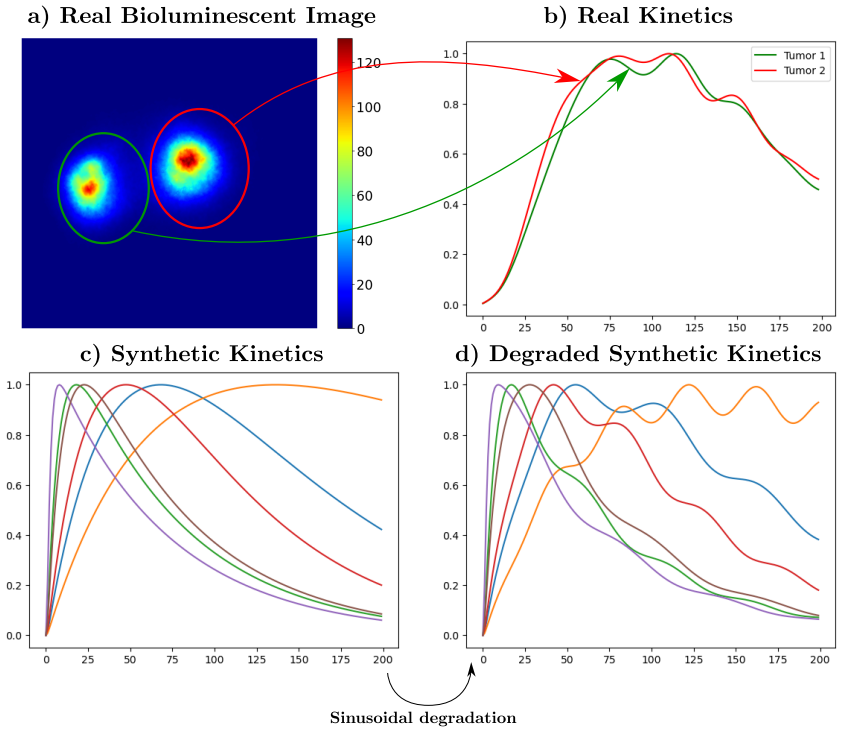

Figure II.11: Example of tumoral kinetics synthetically generated. In c) and d), colors correspond to different simulation parameters.

and

$$\mathbf{C}_T(t) = K_{trans} \int_0^t \mathbf{C}_b(x) e^{-k_{ep}(t-x)} dx \tag{II.5}$$

We use $\mathbf{C}_{ip}(0) = 1$, we then have $\mathbf{C}_{ip}(t) = e^{-k_a t}$, and therefore:

$$\mathbf{C}_b(t) \approx \frac{k_a}{k_{el} - k_a} (e^{-k_a t} - e^{-k_{el} t}) \tag{II.6}$$

We use this approximation to simulate the input functions (concentration of luciferin in blood) with (II.6) and then tumor kinetics with (II.5). We used 3 values for each pharmacokinetic parameter, to cover a wide distribution range, derived from [100] for input function and tumor kinetics, but also from [120] for tumor kinetics (to make the distribution even wider).

More precisely, the parameters used are described in Table II.1.

This leads to a total of 81 possible combinations, with examples in Figure II.11(c).

We focused on long kinetics BLI and generated synthetic kinetics with a temporal dimension of size $T = 200$ with values spaced of $\frac{100 \times 0.089}{60}$ to be consistent with our real data (subsampled by a factor 100 from about $N_T = 20000$ time frames). However, Figure II.11(a-b) shows that

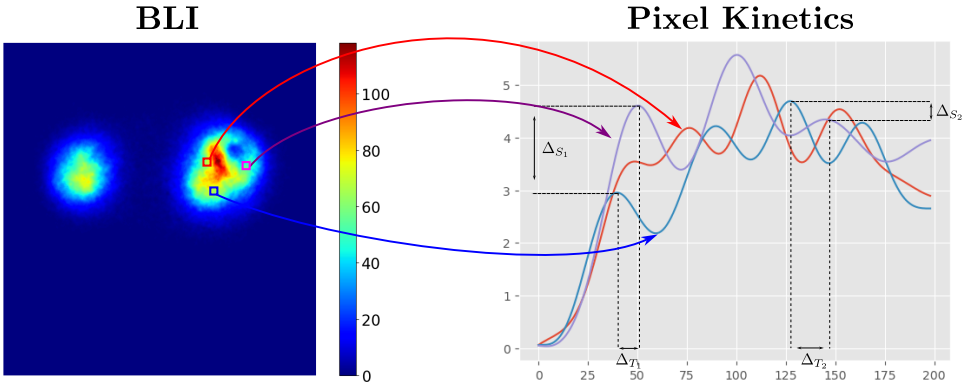

Figure II.12: Illustration of pixel-level kinetics variability for 3 random pixels within 1 tumor. Local maxima intensity ratio ($\Delta_{S_1} \neq \Delta_{S_2}$) and times of appearance ($\Delta_{T_1} \neq \Delta_{T_2}$) differ in scale and time.

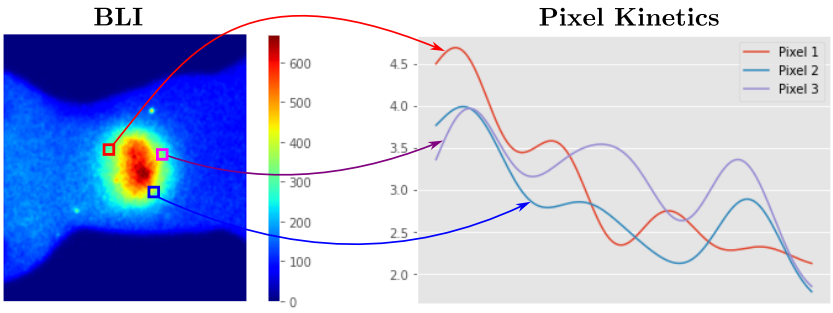

Figure II.13: Illustration of the variability of pixel-level kinetics for a homogeneous light source. Left: BlI, Right: kinetics of 3 pixels located randomly in the light source.

the real "mean" kinetics of tumors exhibit more oscillations than our initial synthetic curves. We therefore added a sinusoidal perturbation by multiplying the kinetics with $1+\frac{1}{15}cos(2\frac{\pi}{T_{pert}}t+\psi)$ where T_{pert} is a random variable uniformly sampled in the range $[\lfloor\frac{T}{5}\rfloor,\ldots,\lfloor\frac{T}{3}\rfloor]$ with $\lfloor\cdot\rfloor$ the floor operator and ψ is a random offset chosen uniformly in $[0,\ldots,100]$. We also set these parameters empirically. This generated the kinetics illustrated in Figure II.11(d).

Each kinetic is normalized according to the ℓ_2 norm to fully control tumoral intensity with the Parametric Map.

Usually, in BSS we take into account variations with respect to the linear model. We will then describe next how we deteriorated the generetic synthetic BLI model with pixel-level variability to make our framework more realistic.

II.3.2 Pixel-level perturbations

Figure II.12 shows the kinetics of pixels belonging to a single tumor, on a random set of 3 pixels. We can see in that figure that pixels located within the same tumor have the same general kinetics shape but the ratio between local maxima vary over time and the times of appearance of local maxima differ in a more complex fashion than can be explained by a simple time shift [121].

The origin of this variability is supposedly the imaging device, since we observed the same kind of variability for a homogeneous light source as displays Figure II.13.

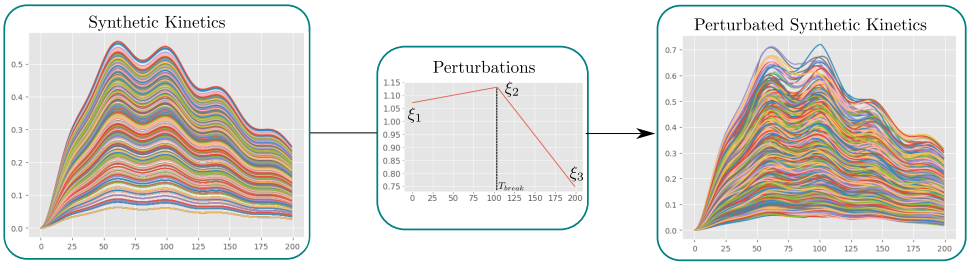

Figure II.14: Application of piece-wise affine variability on kinetics. One piece-wise affine function is generated for each kinetic.

II.3.2.1 Piece-Wise Affine Linear variability

To simulate the ratio variability between kinetic intensities, we used perturbations similar to [122] modeled as multiplicative piece-wise linear gain functions $\epsilon_i \in \mathbb{R}^T$ for each pixel i within the tumor, defined with 4 parameters: $\xi_i, i \in \{1,2,3\}$ and T_{break}:

$$\xi_i \sim \mathcal{U}(1 - \frac{c_{var}}{2}, 1 + \frac{c_{var}}{2}), i \in \{1,2,3\}, \tag{II.7}$$

$$T_{break} = \left\lfloor \frac{T}{2} + \left\lfloor \frac{T \times U}{10} \right\rfloor \right\rfloor, U \sim \mathcal{N}(0,1), \tag{II.8}$$

where T is the temporal dimension of the kinetics, $c_{var} > 0$ is a parameter set by the user that controls the amplitude of the perturbation. In our experiments, we used $c_{var} = 0.8$ because it empirically perturbed our kinetics enough to replicate real data variability. An example of such random piece-wise linear gain function ϵ_i is displayed in Figure II.14, and we generated one per pixel-level kinetic. The perturbated kinetics $\hat{\mathbf{y}}^i$ of each pixel are then the multiplication of the original kinetics and its associated gain function $\hat{\mathbf{y}}^i = \mathbf{y}^i \boldsymbol{\epsilon}_i$.

II.3.2.2 Time Warping variability

To simulate the variability in local maxima times of appearance, we desynchronized the synthetic kinetics with random elastic time warping functions generated with the `python-fdasrsf` package. These functions $\boldsymbol{\gamma}_i^{-1}$ are diffeomorphic deformations of the domain $[0,T]$. More precisely, we generated a random warping functions $\boldsymbol{\gamma}_i^{-1}$ for each pixel i within the tumor.

The procedure to generate a warping function is the following:

- draw $\alpha_i \sim \mathcal{N}(0, \sigma_\gamma^2)$.
- compute $\boldsymbol{v}_i(t) = \alpha_i(1 + \sqrt{2}(sin(2\pi t) + cos(2\pi t))$.
- compute $\boldsymbol{\psi}_i(t) = cos(||\boldsymbol{v}_i||_2 \sqrt{\Delta t}) + \frac{sin(||\boldsymbol{v}_i||_2)\boldsymbol{v}_i}{||\boldsymbol{v}_i||_2}$ where Δt is the the time step.
- Finally, with $\boldsymbol{\gamma}_i^0(t) = \int_0^t \boldsymbol{\psi}_i(t)^2 dt$. The final warping function is given by :

$$\boldsymbol{\gamma}_i^{-1}(t) = \frac{\boldsymbol{\gamma}_i^0(t) - \min_t(\boldsymbol{\gamma}_i^0)}{\max_t(\boldsymbol{\gamma}_i^0) - \min_t(\boldsymbol{\gamma}_i^0)} \tag{II.9}$$

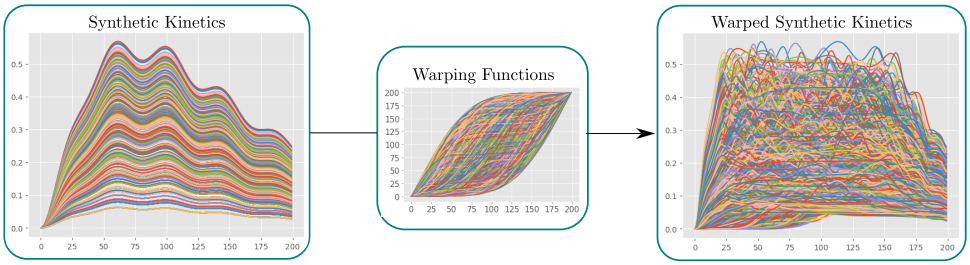

Figure II.15: Application of warping variability on kinetics.

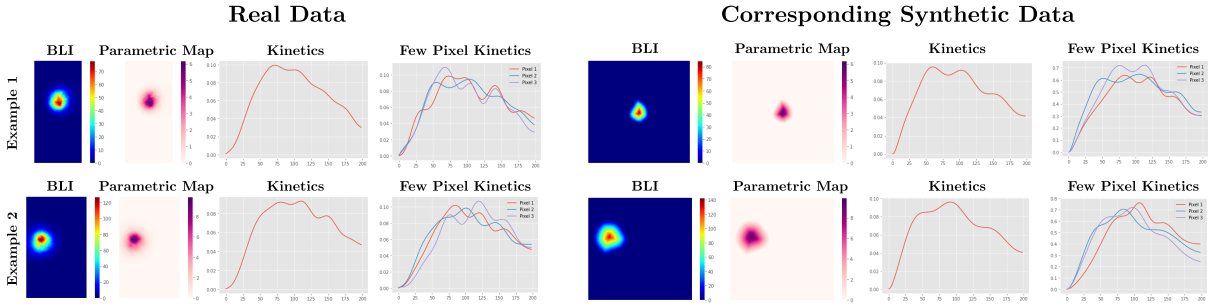

Figure II.16: Synthetic realistic BLI generation framework. Left: Real Data, Right: Synthetic Data generated by the framework supposed to be similar to the Real Data. For each image from left to right : BLIs, Parametric Maps, Tumor Kinetics and a the kinetics of a few pixels located randomly in the image.

We used a standard deviation of $\sigma_\gamma = 0.3$. Examples of such random warping functions $\boldsymbol{\gamma}_i^{-1}$ are displayed in Figure II.15, and we generated one per pixel. The perturbated kinetics $\hat{\mathbf{y}}^i$ of each pixel are then the composition of the original kinetics and its associated warping function $\hat{\mathbf{y}}^i(t) = \mathbf{y}^i(\boldsymbol{\gamma}_i^{-1}(t))$.

This complete framework is able to generate realistic BLI of tumors, as illustrated in Figure II.16.

We here made the choice to generate pre-processed BLI and it could be a possible improvement to our method to generate raw BLI by leveraging (II.1).

II.3.3 Generated multiple-tumors synthetic BLI dataset

We generated images containing several tumors to have synthetic data similar to subsection II.1.1. To generate a BLI of J tumors, the process is the following:

- Generate J parametric maps, noted $\mathbf{A}^{(*)}$ after vectorization and concatenation. It should be noted that in order to be realistic with what is observed (Figure I.23) we constrained parametric maps of different tumors not to overlap completely as shown in Figure II.17.
- Select randomly J combinations of kinetic parameters that should belong to a single animal (so with constant k_a and k_{el}). Each kinetic is then normalized according to the ℓ_2 norm to be consistent. After concatenation, we note $\mathbf{S}^{(*)}$ the retained kinetics.

We then obtained the final image $\mathbf{Y}$ by applying the perturbations defined in subsection II.3.2 to the product $\mathbf{A}^{(*)}\mathbf{S}^{(*)}$. Examples of such synthetic BLI of several tumors, with the parametric map of each tumor, are displayed in Figure II.18.

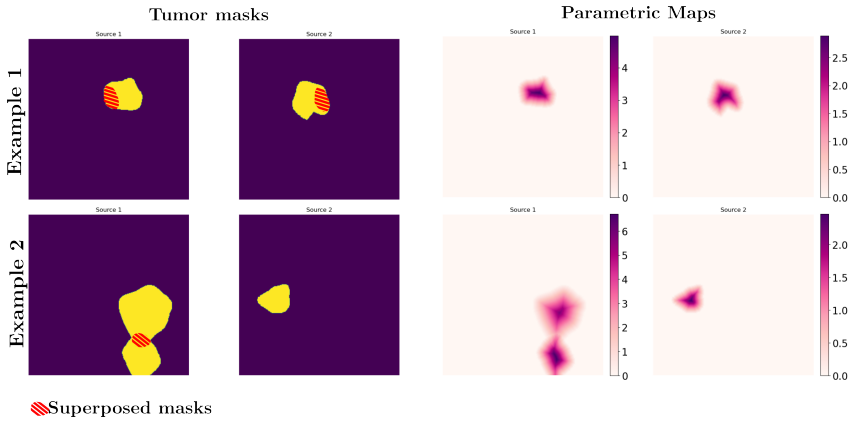

Figure II.17: Example of synthetic parametric maps for images with 2 tumors. Top: Case with touching tumors. Bottom: Case where one tumor has 2 overlapping blobs.

This framework allowed us to generate a synthetic BLI dataset of 24 BLI. We in fact have 4 datasets of 24 images corresponding to the 4 types of perturbation : None, time warping (subsubsection II.3.2.2), piece-wise linear gain (subsubsection II.3.2.1), time warping + piece-wise linear gain. In this dataset, there are 8 BLI with 2 tumors, 7 BLI with 3 tumors and 9 BLI with 4 tumors. These datasets have been used to accurately evaluate our BSS algorithms described in Chapter III.

Further more, the 81 synthetic kinetics generated (subsubsection II.3.1.2) helped us to validate our BLI quantification method as will be explained in Chapter IV.

This BLI generation framework will be part of an article in preparation:

Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J.-C. . Wavelet transform based non-negative matrix factorization for bioluminescence imaging with warping variability.

It should be noted that the preprocessing process and associated hyperparameters described in section II.2 were chosen because they seemed to provide empirically satisfactory results and had not been studied in depth. Improving them and analyzing their influence on the perturbations described in subsection II.3.2 represents an interesting perspective for our work, for example by optimizing a spatio-temporal SVD filtering [123].

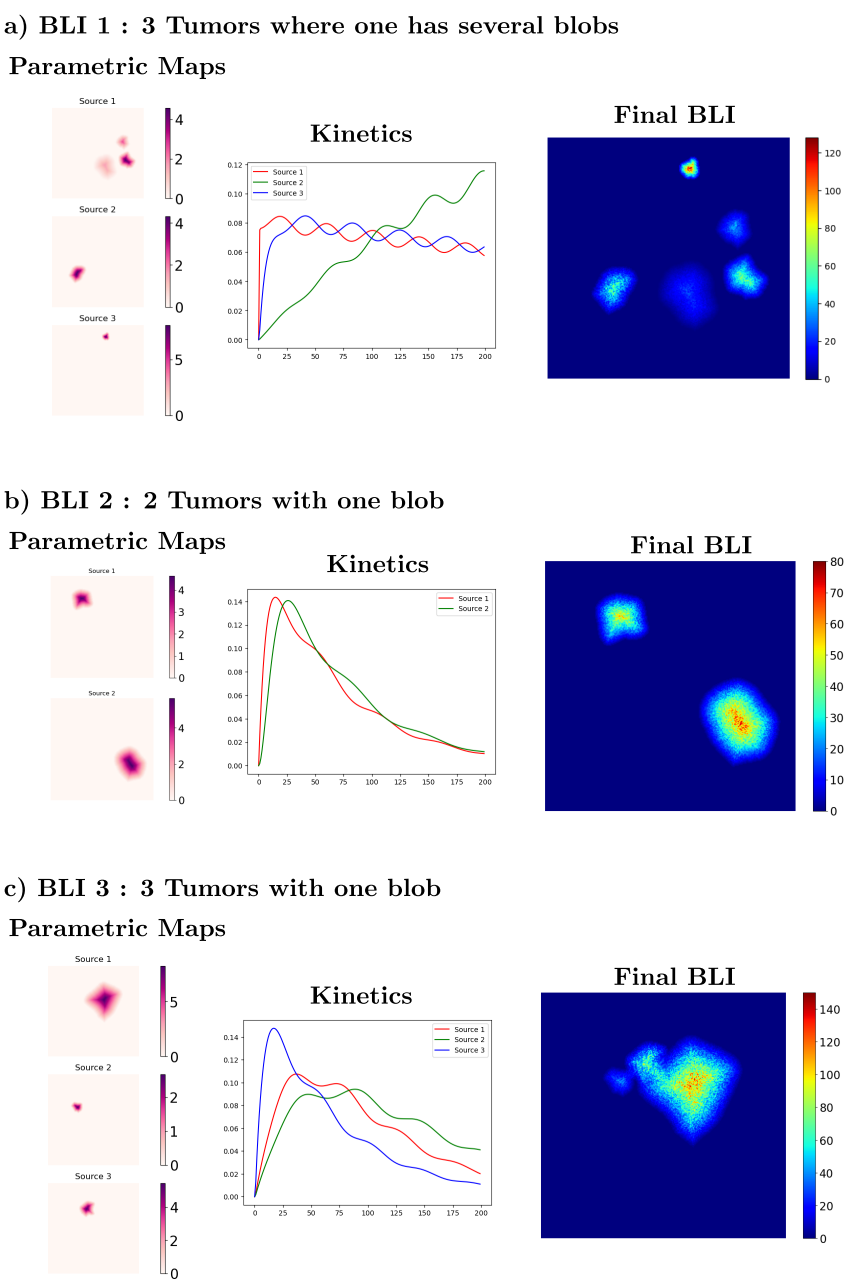

Figure II.18: Examples of 3 generated BLIs. Top: BLI with 3 tumors where one is composed of several blobs. Middle: BLI with 2 tumors with one blob per tumor. Bottom: BLI with 3 tumors with one blob per tumor. For each example, Left: Parametric Maps; Middle: Kinetics and Right: Final BLI.

Chapter III

Blind Source Separation on Bioluminescence Imaging

III.1 State of the art on Blind Source Separation

III.1.1 Introduction and formalism

Blind source separation methods are algorithms designed to separate mixed signals. Theys are widely used in the remote sensing community, such as in Hyperspectral Imaging [122, 124] or Astronomy [125], but also in medical imaging with PET applications [126, 127] or DCE-MRI [128]. Typically, data are described using a linear mixture model (LLM), which is solved using various optimization methods [129, 130].

III.1.1.1 Examples of Blind Source separation problems

a) Audio: cocktail party problem

The most well known BSS problem is the cocktail party problem. During a party, everyone is speaking at the same time. If there are m microphones, the goal of Blind Source Separations is to recover the sounds $\mathbf{S}_1, \ldots, \mathbf{S}_n$ of the n speakers present in the room from the recordings $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ of the microphones. Each sound $\mathbf{Y}_i$ is a mixing of the different sounds $\mathbf{S}_j$, weighted by different physic properties such their respective distance to the microphone.

b) Biomedical Imaging: PET Images

Dynamic Positron Emission Tomography (PET) is a functional imaging technique that quantifies biological processes thanks to a radioactive tracer injected intravenously. This tracer, after fixation on tissues, emits a positron that produces gamma photons that are recorded by the imaging device [131]. The time-activity curves (TACs) of each tissue represent the concentration of the tracer in that tissue [132]. The TACs of each voxel represent a mixture of the TACs of each tissue present in that voxel, and it can then be useful to unmix these TACs thanks to a Blind Source Separation method [127, 133]. An illustration of this kind of signal is provided in Figure III.1, where the signal evolution of each voxel is shown schematically.

c) Astronomical data

Astronomical data, such as the *Chandra* [1] dataset, is composed of pictures of the sky taken at

[1] http://chandra .harvard .edu/

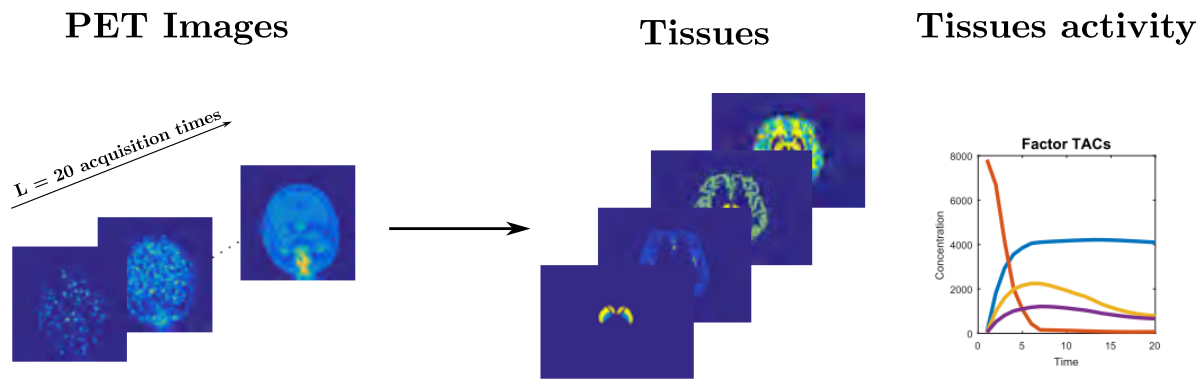

Figure III.1: Illustration of Dynamic PET unmixing (adapted from [131]). All images are $2D$ slices for the sake of simplicity and visualization but they represent $3D$ images. Left: PET scan acquisitions (with $L = 20$ acquisitions), each image corresponds to the observation at a given time spot. Middle: Tissues weights maps, Right: Tissues TACs (with $L = 20$ time points). The TAC of a pixel corresponds to a linear mixture of the TACs of the different tissues, weighted by the values of the weight maps at this pixel.

different wavelengths. More precisely, as can be seen on Figure III.2, the signal at each pixel does not result from the physical emission of a single object, but from a mixture of elementary emissions (the different sources) as can be seen in Figure III.2. BSS methods can help to unmix these signals in order to study the physical properties of each source in these *supernovae* remnants [125].

d) Earth monitoring through hyperspectral imaging

Such as for astronomical data, Hyperspectral Imaging is composed of pictures acquired at different wavelength. It is often used for Earth monitoring [122]. Examples of such datasets are the AVIRIS datasets[2] or more recently the EnMAP dataset[3]. The unmixing provides information about the different materials involved in the images generation, that can be identified by comparison with visual results [135]. The outputs of the unmixing are the *endmembers*, the spectral signature of each material, and the *abundances*, the spatial distribution of each material in the image as illustrated in Figure III.3.

III.1.1.2 Instantaneous Linear Model

In blind source separation, the data is usually described with a Linear Mixing Model (LMM) [134]. It means that the P observation $\mathbf{y}^1, \ldots, \mathbf{y}^P$ is a linear combination of the J sources $\mathbf{s}^1, \ldots, \mathbf{s}^J$:

$$\mathbf{y}^i = \sum_{j=1}^{J} \mathbf{a}_j^i \mathbf{s}^j, \tag{III.1}$$

where the coefficients $\mathbf{a}_j^i$ are the mixing weight coefficients. In matrix form, it yields:

[2]The Moffett and Cuprite images are available at `http://www.ehu.es/ccwintco/` and `http://aviris.jpl.nasa.gov/`

[3]The EnMAP Data Access Portal is available at `https://www.enmap.org/data_access/`

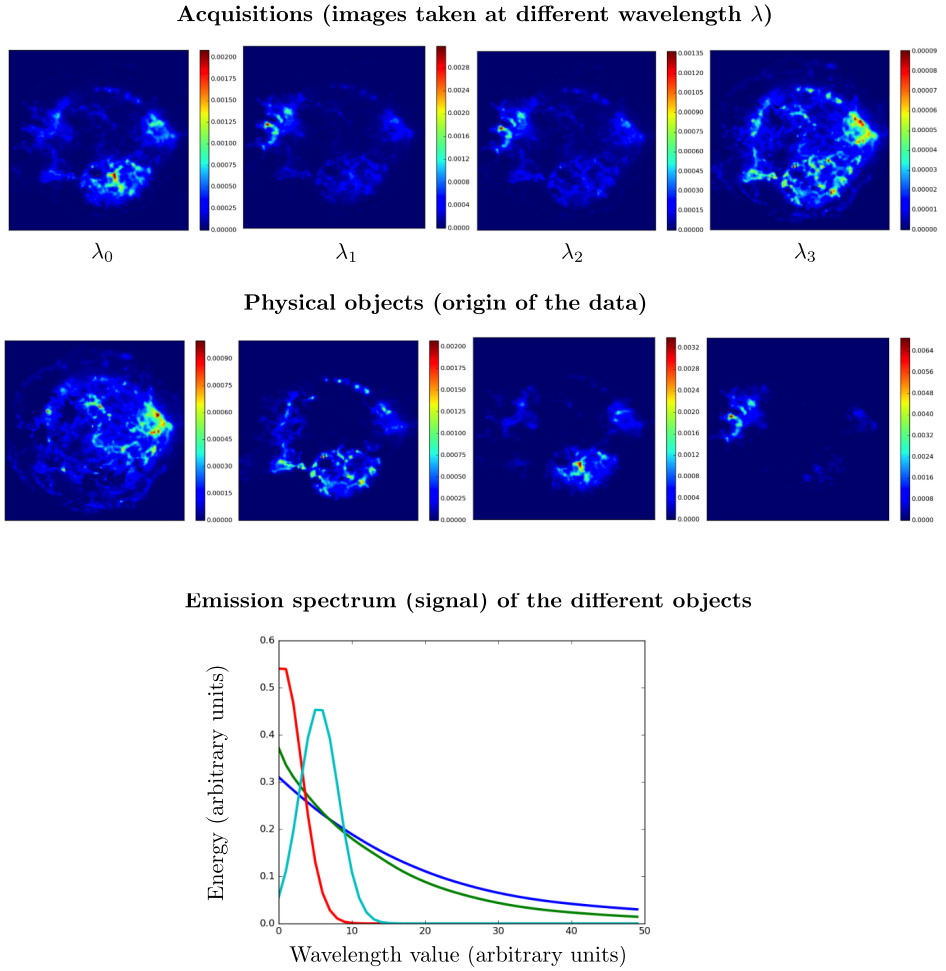

Figure III.2: Illustration of Chandra dataset unmixing, adapted from [134]. Top: Example of images taken at 4 different wavelengths, Middle: spatial emission sources of each physical object, Bottom: emission spectra of each elementary emission. At a given wavelength, the acquired image is a linear mixture of the different objects, whose weights correspond to the emission spectrum of the various objects at this wavelength .

$$\mathbf{Y} = \mathbf{AS} + \mathbf{N}, \tag{III.2}$$

where the observations $\mathbf{y}^i \in \mathbb{R}^T$ are grouped in a matrix $\mathbf{Y} \in \mathbb{R}^{P\times T}$, $\mathbf{A} \in \mathbb{R}^{P\times J}$ and $\mathbf{S} \in \mathbb{R}^{J\times T}$. The $\mathbf{N} \in \mathbb{R}^{P\times T}$ matrix accounts for the additive noise and model discrepancies. In this work, for any matrix $\mathbf{M}$ we will note $\mathbf{m}_j$ its column j and $\mathbf{m}^j$ its line j. In the following, we will assume $\mathbf{A}$ and $\mathbf{S}$ to be full rank matrices.

III.1.1.3 Classical approaches

In BSS, the problem of finding the number of sources J might be very hard, especially when the level of noise is unknown. In the scope of this work, we suppose its value known. However, it can be an interesting topic to pursue [137], for instance with criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [138].

The first idea is to use a Singular Value Decomposition (SVD) [139] which decomposes the matrix $\mathbf{Y}$ into a product $\mathbf{UDV}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{D}$ is a pseudo-

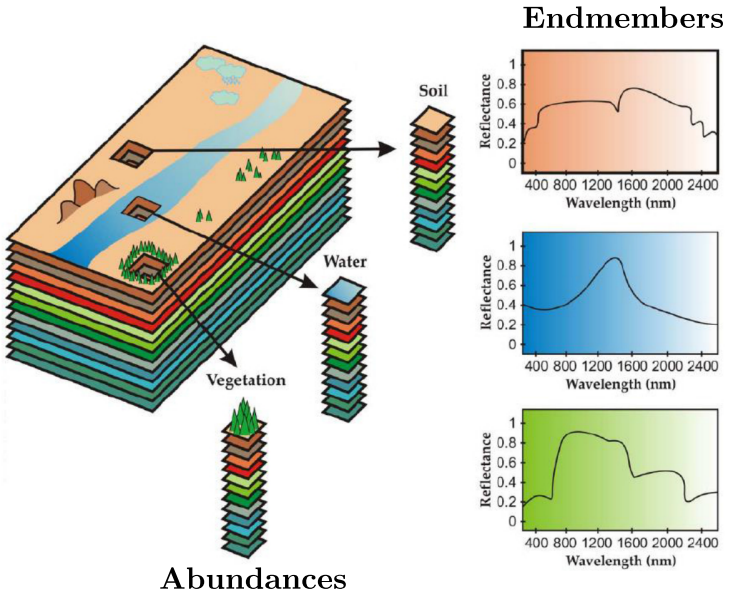

Figure III.3: Illustration of Hyperspectral Unmixing, taken from [136]. The elementary emission (endmember) of each material is given on the right. We can visualize in each pixel the abundances, meaning the weight of each material in it .

diagonal matrix containing the singular values of $\mathbf{Y}$ in descending order. According to the Eckart-Young theorem, the best approximation possible of $\mathbf{Y}$ of rank J is given by the matrix Equation III.3 :

$$\mathbf{U}_{:,1:J}\mathbf{D}_{1:J,1:J}\mathbf{V}^T_{:,1:J}, \tag{III.3}$$

.

where $\mathbf{X}_{1:J}$ is the subset of the first J elements of any vector $\mathbf{X}$. The matrices $\mathbf{A} = \mathbf{U}_{:,1:J}$ and $\mathbf{S} = \mathbf{D}_{1:J,1:J}\mathbf{V}^T_{:,1:J}$ provide therefore a factorization of the matrix $\mathbf{Y}$.

SVD can also lead to a fast implementation of Principal Component Analysis (PCA) [140], by computing the SVD after standardization and centering of the data, and retrieving the principal components in the same way as Equation III.3, and then provide the requested decomposition of $\mathbf{Y}$ after adding of the mean and multiplication by the former standard deviations. Our data is not orthogonal, therefore SVD and PCA are not the most suited possible methods.

Factor Analysis of Dynamic Structures (FADS), introduced by [141] and later improved by [142], is a BSS method to analyse Dynamic Structures in nuclear medecine. The idea of the method is to compute an *oblique* analysis on the first principal components found after a PCA, that produces physiologically meaningful components, under positivity constraint [143]. The solution yielded is however not unique [144], so the output may be somehow different from the ground-truth.

Independent Component Analysis (ICA) is a statistical method used for BSS that assumes that at least one of the sources is non Gaussian, as is explained in the extensive review on the subject [145]. The goal is to make the estimated sources independent. A practical way is to maximize the non-Gaussiannity of the sources. The philosophy behind this strategy is the central limit theorem, which states that a sum of independent random variables tends to

a Gaussian distribution. After centering and standardization of the data, ICA unmixes the signal $\mathbf{Y}$ into a product $\mathbf{AS}$ through the minimization of a cost function usually based on a non-Gaussianity measure [146, 147].

ICA is vastly used in BSS but it suffers from the presence of noise [148] and mutual independence of the sources may be an invalid hypothesis in our case, as with Planck Data [149] or Hyperspectral data with the sum-to-one constraint on abundance maps [150]. Indeed, since in our case the data is bioluminescent signals coming from tumors within the same animal, they can be correlated with spatial support not necessarily disjoint.

Finally, Non-Negative Matrix Factorization (NMF) is a BSS method, wildly used in Hyperspectral [151] but also in medical imaging [127, 133], usually solved using optimization algorithms, that enforces all coefficients involved to be non-negative [129]. These optimization processes will be more detailed in the following.

III.1.1.4 Deep Learning approaches

Such unmixing can also be computed with Deep Learning approaches. Some of them are black-box methods, for instance based on autoencoders [152]. Others are based on optimization principles, by using unrolling that writes the optimization problem as a neural network [153] or Deep Image Prior that substitutes the regularizer of an optimization problem (described later in subsubsection III.1.3.2) by a neural network [154].

There is no previous work on Biolominescent Imaging unmixing and we showed in Chapter II a huge variability in bioluminescence kinetics, making it possibly hard to extract a pattern. Therefore, we used in this work more classical approaches than Deep Learning. This makes the associated challenges more explainable as a first work on the subject. Since our data is composed only of positive signals (photon counting) and not exactly independant (several tumors of the same animal), we decided to use NMF.

III.1.2 Non-Negative Matrix Factorization

III.1.2.1 Formulation

In general, trying to find the matrices $\mathbf{A}$ and $\mathbf{S}$ from the sole knowledge of $\mathbf{Y}$ admits an infinite number of solutions. It is then mandatory to add prior information in this kind of unmixing problem. NMF imposes then a non-negativity constraint on $\mathbf{A}$ and $\mathbf{S}$:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} D(\mathbf{Y},\mathbf{AS}) + i_{\geq 0}(\mathbf{A}) + i_{\geq 0}(\mathbf{S}) \tag{III.4}$$

where $i_C(.)$ is the indicator function of a set C defined in (III.5) and D is a distance metric between data $\mathbf{Y}$ and the approximation $\mathbf{AS}$.

$$i_C(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{else} \end{cases} \tag{III.5}$$

The choice of this distance is very important because it can be adapted to the problem being addressed. Its choice is based on the model suited for the studied data. The used divergences are separable, meaning that $\forall \mathbf{X}, \mathbf{Y}, D(\mathbf{X},\mathbf{Y}) = \sum_{i,j} d(\mathbf{X}_{i,j}, \mathbf{Y}_{i,j})$. The main divergence metrics used in the literature are the following:

- The **Euclidian distance**: $d(x,y) = (y-x)^2$

  This distance has been introduced for NMF since the beginning [130, 155]. It can be seen as the maximum likelihood when the sources are corrupted with a Gaussian noise. This, and the simplicity of the associated optimization algorithms, makes it the one most often chosen metric in NMF algorithms [156, 157].

- The **Kullback-Leibler divergence** : $d(x,y) = x\log(\frac{x}{y}) - x + y$

  This divergence, used to quantity the difference between 2 probability distributions, has been introduced in NMF in order to construct a probabilistic model of image generation [158]. It can also be seen as a maximum likelihood when the sources are corrupted with a Poisson noise [159].

- The **Itakura-Saito divergence** : $d(x,y) = \frac{x}{y} - \log(\frac{x}{y}) - 1$

  This divergence, used to quantify the difference between 2 spectra, has been introduced in NMF for the decomposition of audio data [160]. It can indeed be seen as a maximum likelihood when the sources are corrupted with a Gamma noise [160].

In our case, according to [118], the bioluminescent images are corrupted with a Poisson and a Gaussian noise. However, the pre-processing used really smoothes the images so the sources are corrupted with a noise having a small energy, which can be approximated as Gaussian. We therefore decided to use the Euclidian distance because of its easiness to set up (quadratic, regular and its Lipschitzian gradient simplifies the task of solving optimization problems).

The NMF formulation used in this work is then:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + i_{\geq 0}(\mathbf{A}) + i_{\geq 0}(\mathbf{S}) \tag{III.6}$$

The multiplication by $\frac{1}{2}$ used to simplify the equations of the optimization algorithms described later.

III.1.2.2 Basic NMF algorithms

Now that the problem has been correctly formulated, we need to choose between the different minimization algorithms for this type of multi-convex optimization problem. Indeed, the product **AS** makes it non-convex. Altough, the sub-problem on A

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + i_{\geq 0}(\mathbf{A}) \tag{III.7}$$

and the sub-problem on S

$$\underset{\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + i_{\geq 0}(\mathbf{S}) \tag{III.8}$$

are both convex which makes them easier to solve. This is why many algorithms alternate between a minimization on **A** with a fixed **S**, and a minimization on **S** with a fixed **A**. In an

algorithm, one can choose to either solve these sub-problems completely or just to decrease the cost function a little bit. Moreover, we can choose to either minimize these cost functions by taking into account the non-negativity constraint or to apply the constraint afterwards with a projection on the positive space.

a) Global minimization without constraint

In this algorithm, each sub-problem is solved without taking the non-negativity constraint into account, and then the result is projected on the non-negative space with the operator $[.]_+ = max(., 0)$ [155]. The minimization is done with a least-squares algorithm. This algorithm is called projected Alternating Least Squares (pALS), and is fully described in Algorithm 1

Algorithm 1 pALS algorithm

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0$
  $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
  $\mathbf{S} \leftarrow \mathbf{S}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{A}^{(k+1)} \leftarrow [Y\mathbf{S}^{(k)T}(\mathbf{S}^{(k)}\mathbf{S}^{(k)T})^{-1}]_+$
    $\mathbf{S}^{(k+1)} \leftarrow [(\mathbf{A}^{(k+1)T}\mathbf{A}^{(k+1)})^{-1}\mathbf{A}^{(k+1)T}\mathbf{Y}]_+$
  end for
  Return $\mathbf{A}, \mathbf{S}$

b) Partial minimization without constraint

This time, each sub-problem is not fully solved at each iteration, but the cost function is decreased with a gradient descent with a well defined Lipschitz parameter [161]. With the projection on the non-negative space, we then have a projected gradient step at each iteration for each sub-problem. The full algorithm is described in Algorithm 2.

Algorithm 2 Projected Gradient Algorithm

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0, \gamma_A, \gamma_S$
  $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
  $\mathbf{S} \leftarrow \mathbf{S}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{A}^{(k+1)} \leftarrow [\mathbf{A}^{(k)} - \gamma_A^{(k)}(\mathbf{Y} - \mathbf{A}^{(k)}\mathbf{S}^{(k)})\mathbf{S}^{(k)T}]_+$
    $\mathbf{S}^{(k+1)} \leftarrow [\mathbf{S}^{(k)} - \gamma_S^{(k)}\mathbf{A}^{(k)T}(\mathbf{Y} - \mathbf{A}^{(k)}\mathbf{S}^{(k)})]_+$
  end for
  Return $\mathbf{A}, \mathbf{S}$

c) Partial minimization with constraint

This method decreases the cost functions by taking into account the non-negativity constraint, thanks to a carefully set term-by-term matrix multiplication. This algorithm, described in Algorithm 3, is called multiplicative update [130].

Algorithm 3 Multiplicative Update Algorithm

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0$
  $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
  $\mathbf{S} \leftarrow \mathbf{S}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{A}^{(k+1)} \leftarrow \mathbf{A}^{(k)} \odot \frac{\mathbf{Y}\mathbf{S}^{(k)T}}{\mathbf{A}^{(k)}\mathbf{S}^{(k)}\mathbf{S}^{(k)T}}$
    $\mathbf{S}^{(k+1)} \leftarrow \mathbf{S}^{(k)} \odot \frac{\mathbf{A}^{(k+1)T}\mathbf{Y}}{\mathbf{A}^{(k+1)T}\mathbf{A}^{(+1)}\mathbf{S}^{(k)}}$
  end for
  Return $\mathbf{A}, \mathbf{S}$

Where $\frac{\cdot}{\cdot}$ and $\odot$ are term-by-term operations.

d) Global minimization with constraint

Finally, it is also possible to solve exactly each sub-problem at each iteration by taking into account the non-negativity constraint. As described in Algorithm 4, each sub-problem is solved by using the forward-backward splitting algorithm [162] described later. In practice, each sub-problem is in fact solved by repeating the projected gradient steps described in Algorithm 2 .

Algorithm 4 Global minimization with constraint algorithm

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0$
  $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
  $\mathbf{S} \leftarrow \mathbf{S}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{A}^{(k+1)} \leftarrow \text{argmin}_{\mathbf{A} \in \mathbb{R}^{P \times J}} \frac{1}{2}||\mathbf{Y} - \mathbf{A}\mathbf{S}^{(k)}||_F^2 + i_{\geq 0}(\mathbf{A})$
    $\mathbf{S}^{(k+1)} \leftarrow \text{argmin}_{\mathbf{S} \in \mathbb{R}^{J \times T}} \frac{1}{2}||\mathbf{Y} - \mathbf{A}^{(k)}\mathbf{S}||_F^2 + i_{\geq 0}(\mathbf{S})$
  end for
  Return $\mathbf{A}, \mathbf{S}$

e) Hierarchical ALS (HALS)

It is also possible to optimize the ALS algorithm by avoiding to compute such big matrix multiplications. Indeed, Cichocki et al. [163] proposes to minimize the sub-problems source-by-source. Meaning that we are now going to alternate between the sub-problem in $\mathbf{A}$ and $\mathbf{S}$ inside each source. More precisely, we define $\mathbf{Y}^{(j)} = \mathbf{Y} - \sum_{i \neq j} \mathbf{a}_j \mathbf{s}^j$

We have:

$||\mathbf{Y} - \mathbf{A}\mathbf{S}||_F^2 = ||\mathbf{Y}^{(j)} - \mathbf{a}_j \mathbf{s}^j||_F^2$ and that $\forall j$

The goal of the algorithm HALS is then to minimize alternatively the set of cost functions $D^j_{\mathbf{A}}(\mathbf{a}_j)$ and $D^j_S(\mathbf{s}^j)$ for every j, because each row of $\mathbf{A}$ only affects the corresponding row of the product $\mathbf{AS}$ [129]:

$D^j_{\mathbf{A}}(\mathbf{a}_j) = \frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j \mathbf{s}^j||_F^2$ for a fixed $\mathbf{s}^j$ to update $\mathbf{a}_j$

$D^j_S(\mathbf{s}^j) = \frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j \mathbf{s}^j||_F^2$ for a fixed $\mathbf{a}_j$ to update $\mathbf{s}^j$

To do that, they are using the ALS algorithm [155] to get the update rules for $\mathbf{a}_j$ and $\mathbf{s}^j$ by equating the gradient to zero and enforcing non-negativity:

$\mathbf{s}^j \leftarrow \frac{1}{a_j^T \mathbf{a}_j}[\mathbf{a}_j^T \mathbf{Y}^{(j)} \mathbf{a}_j]_+ = \frac{1}{\mathbf{a}_j^T \mathbf{a}_j}[\mathbf{Y}^{(j)T} \mathbf{a}_j]_+^T$ and $\mathbf{a}_j \leftarrow \frac{1}{\mathbf{s}^j \mathbf{s}^{jT}}[\mathbf{Y}^{(j)} \mathbf{s}^{jT}]_+$ where $[.]_+$ refers to taking only the positive coefficients.

However, we can rewrite these update rules. Let's take $\mathbf{a}_j$ for instance

$\mathbf{Y}^{(j)}\mathbf{s}^j = (\mathbf{Y} - \mathbf{AS} + \mathbf{a}_j\mathbf{s}^j)\mathbf{s}^{jT} = \mathbf{Y}\mathbf{s}^{jT} - \mathbf{A}S\mathbf{s}^{jT} + \mathbf{a}_j\mathbf{s}^j\mathbf{s}^{jT} = [\mathbf{Y}S^T]_j - [\mathbf{ASS}^T]_j + \mathbf{a}_j\mathbf{s}^j\mathbf{s}^{jT}$

So $\frac{1}{\mathbf{s}^j\mathbf{s}^{jT}}[\mathbf{Y}^{(j)}\mathbf{s}^{jT}]_+ = [\mathbf{a}_j - \frac{1}{\mathbf{s}^j\mathbf{s}^{jT}}([\mathbf{ASS}^T]_j) - [\mathbf{YS}^T]_j)]_+$ is the new update rule for $\mathbf{a}_j$.

Symmetrically, by doing relevant transpositions :

$\mathbf{s}^j \leftarrow [\mathbf{s}^{jT} - \frac{1}{\mathbf{a}_j^T\mathbf{a}_j}([\mathbf{S}^T\mathbf{A}^T\mathbf{A}]_j - [\mathbf{Y}^T\mathbf{A}]_j)]_+^T$ is the new update rule for $\mathbf{s}^j$.

Finally, the updates rules for a_j and s^j are the following :

$$\mathbf{a}_j \leftarrow \left[\mathbf{a}_j - \frac{1}{\mathbf{s}^{\mathbf{s}} s^{jT}}\left(\mathbf{A}[\mathbf{SS}^T]_j - [\mathbf{YS}^T]_j\right)\right]_+, \tag{III.9}$$

$$\mathbf{s}^j \leftarrow \left[\mathbf{s}^{jT} - \frac{1}{\mathbf{a}_j^T\mathbf{a}_j}\left(\mathbf{S}^T[\mathbf{A}^T\mathbf{A}]_j - [\mathbf{Y}^T\mathbf{A}]_j\right)\right]_+^T, \tag{III.10}$$

These update rules are of the same form as projected gradient Algorithm 2 with a Lipschitz constant equal to the hessian of the cost function.

The Fast-HALS algorithm is then described in Algorithm 5.

Algorithm 5 Fast HALS algorithm

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0$
  $\mathbf{A} \leftarrow \mathbf{A}^{(0)}$
  $\mathbf{S} \leftarrow \mathbf{S}^{(0)}$
  for $k \leq N_{iter}$ **do**
    for $j \leq J$ **do**
      Update $\mathbf{a}_j$ with Equation III.9
      Update $\mathbf{s}^j$ with Equation III.10
    end for
  end for
  Return $\mathbf{A}, \mathbf{S}$

III.1.2.3 Application on Bioluminescent Imaging

a) Linear Mixing Model for BLI

For BLI, we denote as $\mathbf{y}^i \in \mathbb{R}^T$ the kinetics observed in the i-th pixel with $i \in \{1, \ldots, P\}$ with P the number of pixels. The goal of the method is to find the kinetics of each tumor and their associated spatial distribution. Depending on the distance between tumors, these spatial distributions may not be completely disjoint. We then assume that the kinetics observed in a pixel $\mathbf{y}^i$ is a linear combination of individual tumor kinetics. Mathematically, written in a matrix form, it yields the mixing model:

$$\mathbf{Y} = \mathbf{AS} + \mathbf{N}, \tag{III.11}$$

where each line of $\mathbf{S} \in \mathbb{R}^{J \times T}$ corresponds to the kinetics of a single tumor (with J the number of tumors), $\mathbf{A} \in \mathbb{R}^{P \times J}$ are the linear mixing weights (the parametric maps of the tumors) and $\mathbf{Y} \in \mathbb{R}^{P \times T}$ is the observed signal. The $\mathbf{a}_j^i$ coefficient of $\mathbf{A}$ corresponds to the loading of the j-th tumor in the i-th pixel. The $\mathbf{N} \in \mathbb{R}^{P \times T}$ matrix accounts for the additive noise and model discrepancies.

The Fast-HALS algorithm Algorithm 5 has been tested on the 2 datasets of BLI described in Chapter II containing multiple-tumors bearing mice (one with short kinetics and one with long kinetics). We chose this algorithm as a first step because it seemed a good trade off between performance and ease of use [125, 164]. All the following BSS algorithms have been also evaluated on these 2 datasets.

It should be noted that the algorithms described in subsubsection III.1.2.2 do not ensure the unicity of the solution because for any $\lambda > 0$, $\mathbf{AS} = (\frac{1}{\lambda}\mathbf{A})(\lambda\mathbf{S}) = (\lambda\mathbf{A})(\frac{1}{\lambda}\mathbf{S})$. To enforce it, we therefore constrained the lines of $\mathbf{S}$ to be in the ℓ_2 ball.

In these datasets, the tumors consist globally of pure-pixels (meaning pixels belonging to only one source). For each tumor j, we hand-drew a small bounding box around tumor it, as displayed in Figure III.12 , noted Ω_j, and we thus defined a ground truth for $\mathbf{A}$ and for $\mathbf{S}$ on real images:

- For each source j and for each time frame t we use as ground truth $\mathbf{s}^{j(*)}(t) = \frac{\text{mean}_{i \in \Omega_j} \mathbf{y}^i(t)}{||\text{mean}_{i \in \Omega_j} \mathbf{y}^i||_2}$
- For each source j and for each pixel $i \in \Omega_j$, $\mathbf{a}_j^{i(*)} = \underset{t}{\text{median}}(\frac{\mathbf{y}^i(t)}{\mathbf{s}^{j(*)}(t)})$

It should be noted that this definition of ground-truth is not perfect because it supposes that all the pixels only belong to one tumor, which is wrong in practice because there are pixels at the intersection (see Chapter I). However, we did not find any better way to define ground-truth for real data for the evaluation of our unmixing algorithms.

b) Metrics used

We denote the observed ground truth photon cumulated total count in a tumor as $\mathbf{H}^{(*)} \in \mathbb{R}^{J,T}$. With these notations, $\mathbf{H}^{(*)}(j,t) = \sum_{i \in \Omega_j} Y(i,t)$ and $\mathbf{H}^{(*)}(j,t)$ is the total BL signal emitted by the tumor j at the frame t.

NMF separation yields $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_J]$ and $\mathbf{S} = [\mathbf{s}^1, \ldots, \mathbf{s}^J]^T$. For each source j we reconstructed the whole kinetic defined by $\mathbf{a}_j$ and $\mathbf{s}^j$: $\mathbf{H}^j = (\sum_{i \leq P} \mathbf{a}_j^i)\mathbf{s}^j \in \mathbb{R}^T$.

We have several metrics to compare $\mathbf{H}^{(*)}$ and $\mathbf{H}$ to evaluate reconstruction quality on kinetics:

- Pearson correlation coefficient (noted *corr*) that returns a number between -1 and 1
- The Normalized Mean Squared Error ($NMSE$), defined in Equation III.12

$$NMSE(\mathbf{H}, \mathbf{H}^{(*)}) = \frac{||\mathbf{H} - \mathbf{H}^{(*)}||_F^2}{||\mathbf{H}^{(*)}||_F^2} \tag{III.12}$$

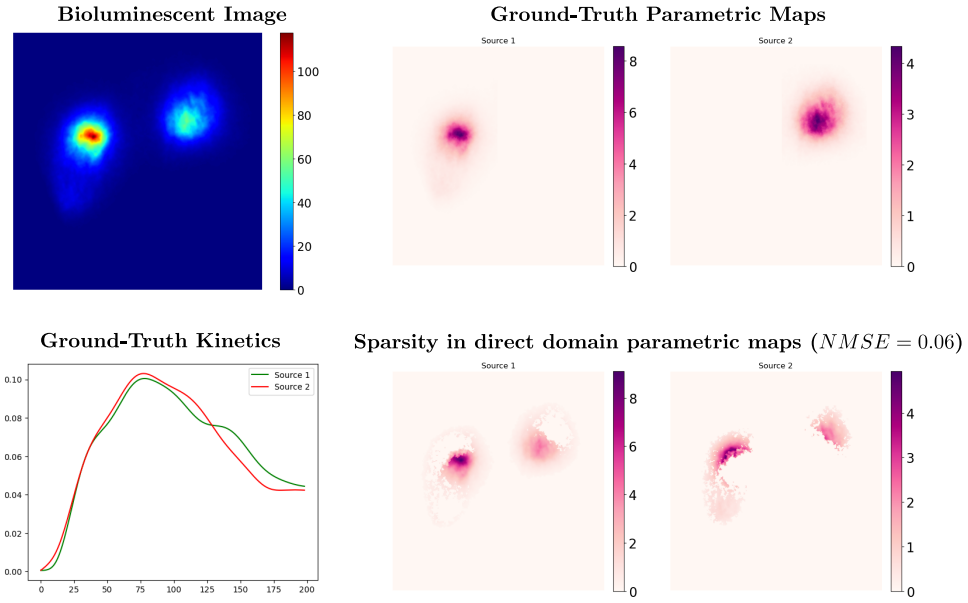

Figure III.4: Example of a BLI poorly unmixed but with a good $NMSE$.

However, this metric does not evaluate how well spatially separated are the tumors. Indeed, this error can be really low if the tumors kinetics are very similar and the algorithm does not manage to separate them (see Figure III.4).

Using the ROI-based ground-truths for $\mathbf{A}$ and for $\mathbf{S}$, noted $\mathbf{A}^{(*)}$ and $\mathbf{S}^{(*)}$, we defined the following metrics to better assess the tumor separation quality:

$$NMSE_A = \frac{||\mathbf{A} - \mathbf{A}^{(*)}||_F^2}{||\mathbf{A}^{(*)}||_F^2} \tag{III.13}$$

$$NMSE_S = \frac{||\mathbf{S} - \mathbf{S}^{(*)}||_F^2}{||\mathbf{S}^{(*)}||_F^2} \tag{III.14}$$

And a global metric to take into account both $\mathbf{A}$ and $\mathbf{S}$

$$NMSE_{tot} = \frac{NMSE_A + NMSE_S}{2} \tag{III.15}$$

It is important to note that all these metrics are computed for each tumor j in each image. The final metric is then obtained by taking the average over all tumors.

We initialized $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$ with NNDSVD [165] and set the maximum number of iterations to 1000. Preliminary tests with NMF using the Fast-HALS algorithm led to the quantitative results reported on Figure III.5 and Table III.1. These results were communicated in 2 conference articles [166, 167].

We note that, if we look only the correlation metric for the global signals, this algorithm suggests a pretty good fit with the NMF decomposition. However, the more precise $NMSE_A$ and $NMSE_S$ metrics show that there is room for a lot of improvement.

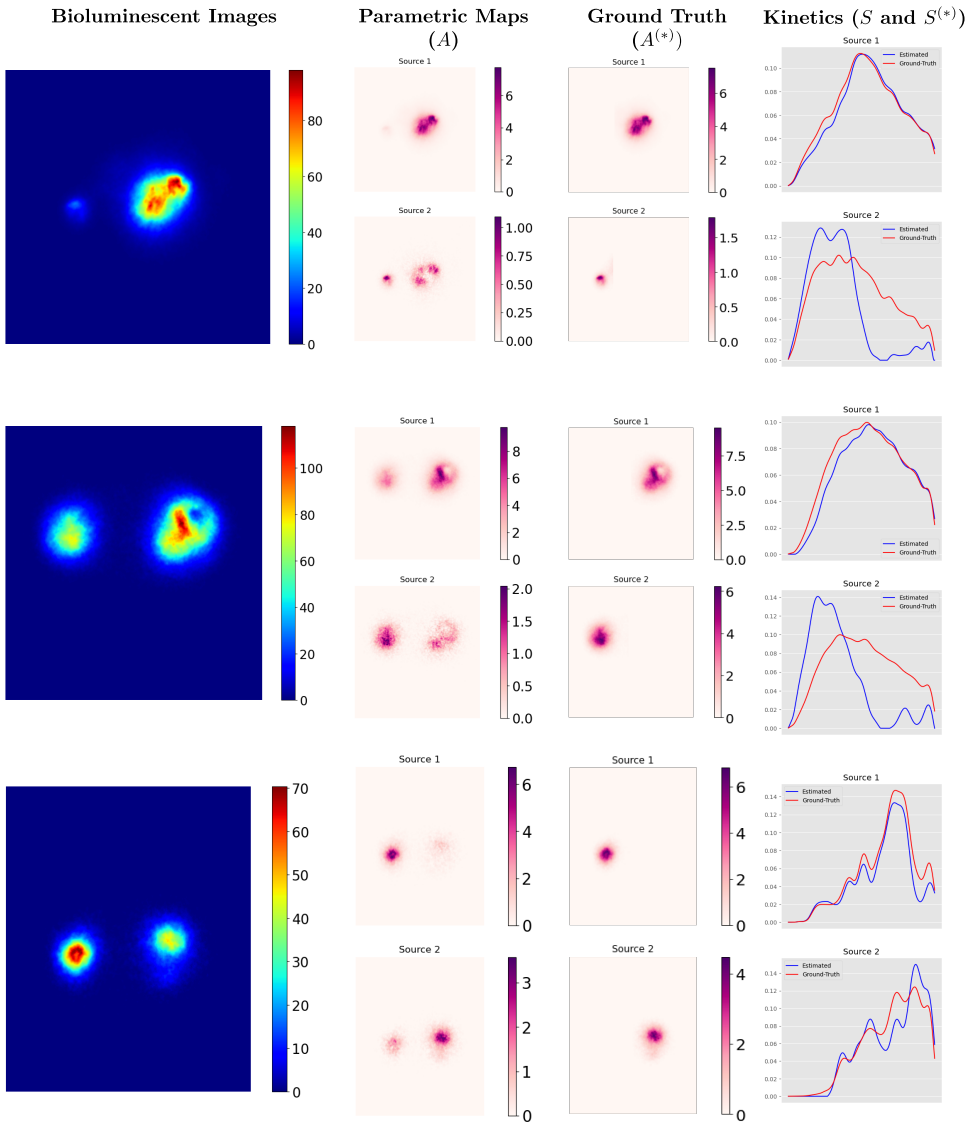

Figure III.5: Visual results (Parametric Maps and Kinetics) of the Fast-HALS algorithm on 3 real BLI with long kinetics.

	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Short kinetics	0.59 ± 0.93	0.17 ± 0.16	0.38 ± 0.63	0.57 ± 1.3	0.65 ± 0.42
Long kinetics	0.45 ± 0.39	0.16 ± 0.20	0.31 ± 0.35	0.30 ± 0.45	0.76 ± 0.33

Table III.1: Quantitative Results of Fast-HALS Algorithm on 2 real BLI datasets.

As the visual results Figure III.5 show spatial mixtures of sources, the non-negativity constraint is thus not sufficient to obtain a satisfying unmixing of BLI. We proposed next to force the parametric maps $\mathbf{A}$ to be more *sparse*. In the following part, we detail the concept of sparsity and how it can be applied to our problem.

III.1.3 Sparsity and Sparse BSS

Sparsity is a concept that has been vastly used over the last few decades in inverse problems [168, 169] such as image denoising [170], sparse coding [171], tomographic reconstruction [172] and in Blind Source Separation (BSS) [173]. Constraining the desired solution to be sparse helps

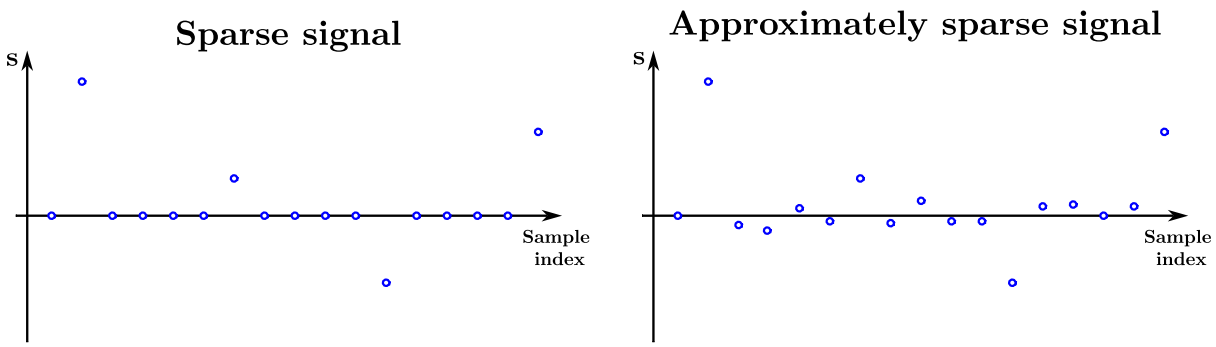

Figure III.6: Illustration of notions of signal sparsity.

to better-pose initially ill-conditionned inverse problems, possibly corrupted with observation noise [174]. In BSS, it is used to promote small weight components or low-energy sources to become exactly zero. We now briefly present the concept of sparsity and how it can be applied in the context of blind source separation, especially for Non-Negative Matrix Factorization.

III.1.3.1 Sparsity

a) Definition

A signal is said to be "sparse" when it has only very few non-zeros coefficients. Formally, we write $||.||_0$, the ℓ_0 pseudo-norm, which is the number of non-zeros coefficients of a signal $\mathbf{s} \in \mathbb{R}^S$. The signal $\mathbf{s}$ is sparse if $||\mathbf{s}||_0 = k$ with $k << S$ (left part of Figure III.6).

In practice, it is not straightforward to use the ℓ_0 norm as a real indicator of sparsity in optimization problems because the ℓ_0 norm is non-convex. The ℓ_1 norm $||s||_1 = \sum_i |s_i|, \forall s \in \mathbb{R}^S$ is the convex norm closest to ℓ_0. It has also been proven that solutions sparse for a problem with ℓ_1 can be solutions of that problem with ℓ_0 when they are very sparse [175]. The ℓ_1 is therefore used in practice when looking for sparse solutions. This metric also allows us to define the notion of approximately sparse signals, meaning signals having just a few coefficients having a high amplitude, as is illustrated in the right part of Figure III.6.

b) Sparsity in a transformed domain

The data itself (*ie* in the *direct* domain) is often not sparse but it can be represented by only a few large coefficients after a transformation $\mathbf{\Phi}^T$ in an appropriate domain. The transformed signal is then noted $\mathbf{\Phi}^T\mathbf{s} \in \mathbb{R}^t$. This is the case for instance of sinusoidal signals that are obviously not sparse in space/time but are sparse in the Fourier domain [176]. Wavelet basis functions are able to sparsify polynomial by parts signals, which are a good approximation of most natural signals [177]. We briefly introduce the wavelet theory. Interested readers can refer to the book of Mallat [177].

A wavelet basis is a set of functions $\mathbf{\Phi}_1^T, \ldots, \mathbf{\Phi}_n^T$, which each are dilated and shifted version of a basic function. More precisely, a function $\mathbf{\Phi}^T$ represents a pair of analysis function respectively responsible of encoding the approximation and the details of the signal $\mathbf{s}$ at a given location and scale (dilation). Changing the location (*i.e.* shifting the function) allows to get informations at different places of the signal and the scale is responsible for the size of the signal support studied. In fact, we restrict ourselves to a finite subset of all the possible shifts and scales (also called *levels* of analysis). Several wavelet basis functions exist, for different kind of geometric or morphological features of the data, like the Haar wavelet [178], the Daubechies wavelets [179] or the startlet [180]. Figure III.7 illustrates the output of a wavelet tranform.

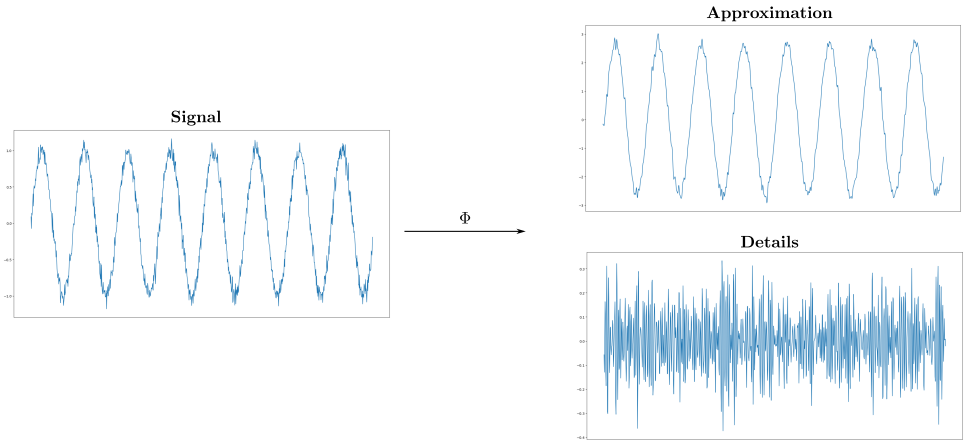

Figure III.7: Illustration of wavelet transform $\mathbf{\Phi}^T$. Left: original signal. Right: approximation and details coefficient at a given scale. The wavelet used here is a Daubechies-3 [181].

Wavelet basis functions can be orthogonal, where an orthogonal transform $\mathbf{\Phi}^T$ is designed such as $\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{Id}$. The orthogonality is a convenient property when enforcing sparsity in an optimization problem. Usually, Wavelet Transforms are *decimated*, meaning that the signal is downsampled at each level of decomposition. It is however possible to maintain the original signal length during the analysis, thanks to *undecimated* (or *stationary*, *redundant*) Wavelet Transforms.

Figure III.8 displays the histogram of the coefficients of a natural image (taken from [182]), before and after a Daubechies-3 Wavelet Transform [181]. We can see that the histogram of the transformed image has a larger proportion of very low coefficients, which means that it is sparser in the wavelet domain than in the direct domain.

In the following, we will see how we can enforce the solution of an inverse problem such as BSS to be sparse, in the direct or a transformed domain.

III.1.3.2 General Problem (Regularization)

Let's consider an inverse problem, defined by the minimization of the following cost function:

$$\underset{\mathbf{x}\in\mathbb{R}^s}{\operatorname{argmin}}\, d(\mathbf{Y},\mathbf{x}) + \mathcal{J}(\mathbf{x}), \tag{III.16}$$

where $d(.,.)$ is the data fidelity term that links the data to the approximation model we are trying to infer and such that d is differentiable with respect to $\mathbf{x}$. It is also supposed to have Lipschitz gradients and to be convex.

$\mathcal{J}$ is called a *regularization* term, that enforces the solution $\mathbf{x}$ to have some desired properties. For instance, for NMF, we have $J(\mathbf{x}) = i_{\geq 0}$ to force solutions to be non-negative. We can also force the solution $\mathbf{x}$ to be sparse in some domain characterized by a transform $\mathbf{\Phi}^T$ by using $J(\mathbf{x}) = ||\mathbf{\Lambda} \odot \mathbf{\Phi}^T\mathbf{x}||_1$ where $\mathbf{\Lambda} \in \mathbb{R}^t$ (the same shape as $\mathbf{\Phi}^T\mathbf{x}$) is the coefficient-wise weight matrix associated to the regularization and $\odot$ is the Hadamard product. The higher and less sparse the Λ, the more sparse the solution. Regularization is a concept that can be extended to many applications. For instance in image denoising, we can use the Total-Variation (TV) regularization, $TV(\mathbf{x}) = ||\nabla\mathbf{x}||_1$, to force the solution to conserve image edges while enforcing piece-wise smooth values in $\mathbf{x}$ [183].

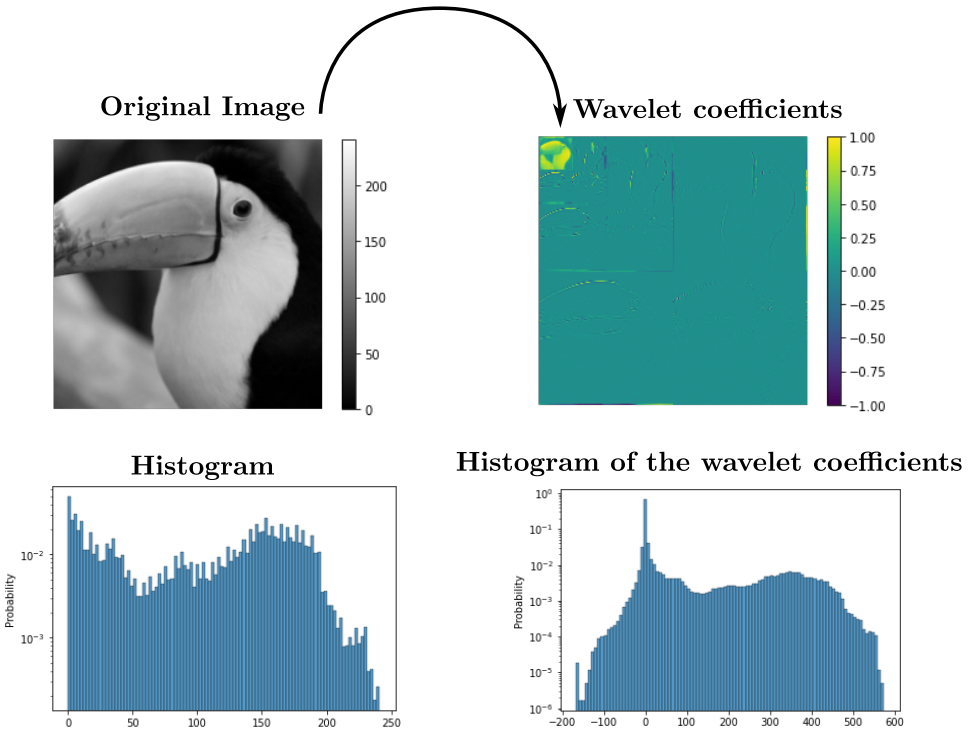

Figure III.8: Illustration of sparsity in a transformed domain. Top left : original image in the spatial domain taken from [182]. Top right : representation of the detail coefficients in the wavelet (Daubechies-3 [181]) domain at scales $1, 2$ and approximation and details coefficients at scale 3. Bottom left: histogram of signal in the spatial domain. Bottom right: histogram of the wavelet coefficients : there are more 0 or very low coefficients, meaning the data can be considered as approximately sparse in the wavelet domain.

III.1.3.3 Sparsity prior for blind source separation

For BSS, we are using in this work the Euclidian distance for d. The appropriate inverse problem we solve, if we consider classical BSS (*ie* without the non-negativity constraint), with sparsity constraints on the weights $\mathbf{A}$ is the following:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + ||\Lambda \odot \mathbf{\Phi}^T\mathbf{A}||_1 + i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S}) \tag{III.17}$$

To prevent degenerated solutions where $||\mathbf{A}|| \to 0$ and $||\mathbf{S}|| \to \infty$, the columns of $\mathbf{S}$ are constrained to be in the ℓ_2 ball, with the appropriate regularization function $i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S})$.

In the following section, we explain how we solved such inverse non-convex problem for our use-case on BLI data.

III.2 Optimization frameworks for Blind Source Separation

In general terms, in order to perform BSS we minimize the following cost function:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} d(\mathbf{Y},\mathbf{AS}) + \mathcal{J}(\mathbf{A}) + \mathcal{G}(\mathbf{S}), \tag{III.18}$$

where d is a metric supposed to be differentiable with a Lipschitz gradient (such as the Euclidian distance) and $\mathcal{J}, \mathcal{G}$ are the regularization functions respectively on $\mathbf{A}$ and $\mathbf{S}$.

Such a minimization can be difficult to perform for several reasons:

- d is supposed to be differentiable but it is not the case of $\mathcal{J}$ and $\mathcal{G}$ (for example we can consider a sparsity prior with the ℓ_1 norm or the non-negativity). It is then necessary to use dedicated tools, such as proximal algorithms that we will explain next, for this kind of non-smooth optimization problems.
- The problem is non-convex in $\mathbf{A}$ and $\mathbf{S}$ but it becomes convex when one of the variables is fixed. It is thus *multi-convex* [184], requiring then the use of specific algorithms to find optimal solutions.

III.2.1 Proximal operators and proximal algorithms

Proximal algorithms enable to minimize convex non-smooth optimization problems, when they respect few regularity properties, by using what is called a *proximal* operator. First, we are going to present proximal operators and then how it can be applied to solve optimization problems.

III.2.1.1 Proximal operators

a) Definition

A function $g : \mathbb{R}^t \to \mathbb{R} \cup \pm\infty$ is said to be *proximable* when it is a proper lower semicontinuous convex function. The *proximal* operator of g is defined by:

$$\text{prox}_g(\mathbf{x}) = \underset{\mathbf{y}\in\mathbb{R}^t}{\text{argmin}}\, g(\mathbf{y}) + \frac{1}{2}||\mathbf{x}-\mathbf{y}||_2^2 \tag{III.19}$$

This function is strongly convex and not infinite everywhere, it then has a unique minimizer. To better grasp the concept of proximal operators, we can provide some possible interpretations:

- It can first be seen as a local minimizer of the function g. Indeed, we aim at minimizing the sum of the function g and the Euclidian distance that penalizes the high distances between the current point $\mathbf{x}$ and $\mathbf{y}$. More precisely, if $\mathbf{x}_m$ is a minimum of g, then $\text{prox}_g(\mathbf{x}_m) = \mathbf{x}_m$. This is called the fixed-point property, which is the base of the proximal optimization algorithms explained later.
- Then, let's suppose that g is differentiable. Differential calculus yields:

  $$\text{prox}_g(\mathbf{x}) = \mathbf{x} - \nabla_g(\text{prox}_g(\mathbf{x})) \tag{III.20}$$

  We can see that the proximal operator on $\mathbf{x}$ looks like a gradient descent step, except that the gradient is computed on $\text{prox}_g(\mathbf{x})$. Nevertheless, a gradient step is also a way to minimize a function therefore Equation III.20 provides a natural interpretation on the process of proximal algorithms.
- Finally, let's consider g being the indicator function of a convex set C, noted i_C (see (III.5)):

Then the proximal operator of g is the Euclidian projection on the set C:

$$\text{prox}_g(\mathbf{x}) = \Pi_C(\mathbf{x}) = \underset{\mathbf{y}\in C}{\text{argmin}} \frac{1}{2}||\mathbf{x}-\mathbf{y}||_2^2 \tag{III.21}$$

This interpretation enable us to see the proximal operator as a generalization of the projection.

This last property is very useful in optimization problems, since minimizing the following problem:

$$\underset{\mathbf{y}\in C}{\text{argmin}}\, f(\mathbf{y}) \tag{III.22}$$

is equivalent to minimizing the regularized following problem :

$$\underset{\mathbf{y}}{\text{argmin}}\, f(\mathbf{y}) + i_C(\mathbf{y}) \tag{III.23}$$

It can however seem costly to compute the proximal operator with a minimization scheme. In fact, lots of useful functions in optimization problems have an explicit proximal operator, which helps to use them in practical algorithms. We will see next some classical proximal operators used in this work. A more complete list can be found in [185].

b) Proximal operators used in this work

We will focus here on the constraint $\mathcal{J}$ and $\mathcal{G}$ used in this work.

- Constraints $\mathcal{J}$ on the parametric maps $\mathbf{A}$:
  - *Sparsity in a transformed domain:* This constraint can be expressed as (subsubsection III.1.3.2):

$$\mathcal{J}(\mathbf{A}) = ||\mathbf{\Lambda} \odot \mathbf{\Phi}^T A||_1, \tag{III.24}$$

where $\mathbf{\Phi}^T$ is a transform, supposed to be orthogonal in the following, into a domain in which $\mathbf{A}$ can be sparsely represented. $\mathbf{\Lambda}$ are the coefficient-wise weights associated to the constraint. The proximal operator of $\mathcal{J}$ has an explicit form:

$$\text{prox}_{\mathcal{J}}(\mathbf{A}) = \mathbf{\Phi} Soft_{\mathbf{\Lambda}}\left(\mathbf{\Phi}^T\mathbf{A}\right), \tag{III.25}$$

where

$$Soft_{\mathbf{\Lambda}}(\mathbf{X}) = sign(\mathbf{X})\left[|\mathbf{X}| - \mathbf{\Lambda}\right]_+, \tag{III.26}$$

is the soft-thresholding operator applied element-wise and $[\cdot]_+$ is the projector on the non-negative orthant:

$$[x]_+ = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{else} \end{cases} \tag{III.27}$$

Equation III.28 explains that we first transform the data, then use the soft-thresholding to sparsify it, and then retransform the data in the original domain.

In particular, if $\mathbf{\Phi} = \mathbf{Id}$, meaning that the sparsity is applied in the direct domain, we have :

$$\text{prox}_{\mathcal{J}}(\mathbf{A}) = Soft_{\mathbf{\Lambda}}(\mathbf{A}) \tag{III.28}$$

– *Non-Negativity in the direct domain*: all the coefficients in $\mathbf{A}$ must be non-negative. That can be expressed as a constraint to be in the non-negative orthant $\{\mathbf{A} \in \mathbb{R}^{P,J}; \forall i \leq P; \forall j \leq J; \mathbf{A}^i_j \geq 0\}$:

$$\mathcal{J}(\mathbf{A}) = i_{\geq 0}(\mathbf{A}) \tag{III.29}$$

Since the non-negative orthant is a convex set, the proximal operator of $\mathcal{J}$ is the projection on it, noted $[\cdot]_+$. It means identity for non-negative coefficients and 0 for the negative ones.

$$\text{prox}_{\mathcal{G}}(\mathbf{A}) = [\mathbf{A}]_+ \tag{III.30}$$

- Constraints $\mathcal{G}$ on the kinetics $\mathbf{S}$:

  – *Constraint to lie on the ℓ_2 hypersphere*: In order to avoid degenerated solutions where $\mathbf{A} \to \infty$ and $\mathbf{S} \to 0$, we constrain the columns of $\mathbf{S}$ to be in the ℓ_2 ball $\mathcal{O} = \{\forall j \leq J, ||\mathbf{s}^j||_2^2 \leq 1\}$ (subsubsection III.1.3.3).

  The appropriate regularization function is then the indicator function of this convex set $\mathcal{O}$:

$$\mathcal{G}(\mathbf{S}) = i_{\mathcal{O}}(\mathbf{S}), \tag{III.31}$$

and its proximal operator is the projection on C:

$$\text{prox}_{\mathcal{G}}(\mathbf{S}) = \Pi_{\mathcal{O}}(\mathbf{S}), \tag{III.32}$$

where $\Pi_{\mathcal{O}}$ is the projection on the ℓ_2 unitary ball $\Pi_{||.||_2 \leq 1}$ of every column $\mathbf{s}^j$ of $\mathbf{S}$:

$$\Pi_{||.||_2 \leq 1}(\mathbf{s}^j) = \begin{cases} \mathbf{s}^j / ||\mathbf{s}^j||_2 & \text{if } ||\mathbf{s}^j||_2 \geq 1 \\ \mathbf{s}^j & \text{else} \end{cases} \tag{III.33}$$

  – *Non-Negativity in the direct domain*: Same as for $\mathbf{A}$

$$\mathcal{G}(\mathbf{S}) = i_{\geq 0}(\mathbf{S}), \tag{III.34}$$

$$\text{prox}_{\mathcal{G}}(\mathbf{S}) = [\mathbf{S}]_+ \tag{III.35}$$

III.2.1.2 Proximal algorithms

Now that all these proximal operators are defined, we are going to explain how to use them in the context of minimization of a convex non-smooth cost function. We will therefore now present different proximal algorithms.

a) Proximal Point Algorithm

Let $f : \mathbb{R}^t \to \mathbb{R}$ be a proximable function. The problem is to find an optimum of:

$$\underset{\mathbf{y}}{\operatorname{argmin}}\, f(\mathbf{y}) \tag{III.36}$$

The most simple algorithm to perform this task is the proximal point algorithm (Algorithm 6).

Algorithm 6 Proximal Point Algorithm

Require: $N_{iter} > 0$, $\lambda > 0$
  $\mathbf{x} \leftarrow \mathbf{x}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{x}^{(k+1)} = \text{prox}_{\lambda f}(\mathbf{x}^{(k)})$
  end for
  Return $\mathbf{x}$

If f has a minimum, the algorithm converges to it. However, this algorithm is only useful in the context of functions f difficult to minimize but where f plus a quadratic term is easy to minimize. To perform optimization on more advanced problems, other algorithms have been designed.

b) Forward Backward Splitting Algorithm

Now let's consider the problem

$$\underset{\mathbf{y}}{\operatorname{argmin}}\, f(\mathbf{y}) + \mathcal{J}(\mathbf{y}), \tag{III.37}$$

where $f : \mathbb{R}^t \to \mathbb{R}$ is strictly convex, differentiable with a Lipschitz gradient and $\mathcal{J} : \mathbb{R}^t \to \mathbb{R} \cup \pm\infty$ convex and proximable.

The Forward-Backward Splitting Algorithm (FBS [162]) uses the differentiability of f to split the problem in 2 terms, as explained in Algorithm 7. It indeed alternates between a forward gradient step and a backward proximal step.

Algorithm 7 Forward-Backward Splitting Algorithm

Require: $N_{iter} > 0$, $\nu > 0$
  $\mathbf{x} \leftarrow \mathbf{x}^{(0)}$
  for $k \leq N_{iter}$ **do**
    $\mathbf{x}^{(k+1)} = \text{prox}_{\nu\mathcal{J}}(\mathbf{x}^{(k)}) - \nu\nabla f(\mathbf{x}^{(k)})))$
  end for
  Return $\mathbf{x}$

In this algorithm, ν is a step size. If ∇f is L-Lipschitzian, this algorithm converges to a global minimum of $f+\mathcal{J}$ provided that $\nu \in \left]0, \frac{2}{L}\right[$. This algorithm can also be accelerated using a linear combination of previous estimates in a variant known as the FISTA algorithm [169].

In a practical setting, the problem (III.17) at a fixed $\mathbf{S}$ is an instance of this kind of minimization problem where we can use the FBS algorithm. Indeed, in the problem expressed in

(III.38), $\frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2$ is differentiable and $||\Lambda \odot \mathbf{\Phi}^T\mathbf{A}||_1$ is proximable with the proximal given in (III.28).

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + ||\Lambda \odot \mathbf{\Phi}^T\mathbf{A}||_1 \tag{III.38}$$

This algorithm is very efficient in optimization problems having an explicit proximal operator for $\mathcal{J}$. When this is not the case but $\mathcal{J}$ is the sum of operators having an explicit proximal, then it is common to use the Generalized Forward Backward Splitting Algorithm explained next.

c) Generalized Forward Backward Splitting Algorithm

If the optimization problem we want to solve is:

$$\underset{\mathbf{y}}{\operatorname{argmin}}\, f(\mathbf{y}) + \sum_i^n \mathcal{J}_i(\mathbf{y}), \tag{III.39}$$

where $f : \mathbb{R}^t \to \mathbb{R}$ is strictly convex, differentiable with a Lipschitz gradient and $\forall 1 \leq i \leq n, \mathcal{J} : \mathbb{R}^t \to \mathbb{R}\cup\pm\infty$ is convex and proximable. If $\sum_i^n \mathcal{J}_i(\mathbf{y})$ does not have an explicit proximable but each of the $\mathcal{J}$ has one, we can use the Generalized Forward Backward Splitting Algorithm (GFBS [186]) detailed in Algorithm 8.

Algorithm 8 Generalized Forward-Backward Splitting Algorithm

Require: $N_{iter} > 0$, $\mu > 0$, $\gamma > 0$, ω_i such as $\sum_i^n \omega_i = 1$
  $\mathbf{x} \leftarrow \mathbf{x}^{(0)}$
  for $k \leq N_{iter}$ **do**
    for $i \leq n$ **do**
      $\mathbf{z}^{(k+1,i)} = \mathbf{z}^{(k,i)} + \mu \text{prox}_{\frac{\gamma}{\omega_i}\mathcal{J}_i}(2\mathbf{x}^{(k)} - \mathbf{z}^{(k,i)} - \gamma\nabla f(\mathbf{x}^{(k)})) - \mu\mathbf{x}^{(k)}$
    end for
    $\mathbf{x}^{(k+1)} = \sum_i^n \omega_i \mathbf{z}^{(k+1,i)}$
  end for
  Return $\mathbf{x}$

In this algorithm, the ω_i are the weights of each constraint $\mathcal{J}_i$ in the optimization and γ is a step size. If ∇f is L-Lipschitzian, this algorithm converges to a global minimum of $f + \sum_i^n \mathcal{J}$ provided that $\gamma \in \left]0, \frac{2}{L}\right[$ and $\nu \in \left]0, \min\left(\frac{3}{2}, \frac{1+\frac{2}{L\gamma}}{2}\right)\right[$.

For instance, the following problem:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + ||\Lambda \odot \Phi\mathbf{A}||_1 + i_{\geq 0}(\mathbf{A}), \tag{III.40}$$

with a fixed $\mathbf{S}$ is an instance of this kind of minimization problem where we can use the GFBS algorithm. Indeed, $\frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2$ is differentiable and $\mathcal{J}_1(\mathbf{A}) = ||\Lambda\odot\Phi\mathbf{A}||_1$ and $\mathcal{J}_2(\mathbf{A}) = i_{\geq 0}(\mathbf{A})$ are both proximable with explicit proximal operators but we don't know any explicit proximal of the sum $\mathcal{J}_1 + \mathcal{J}_2$.

For optimization problems where we don't have any explicit proximal of the constraints $\mathcal{J}$ or when f is not differentiable anymore, others well-known algorithms exist such as the Chambolle-Pock algorithm [187] or the Douglas-Rachford algorithm [188]. We can also use a subrouting to compute the proximal, but it is poorly efficient.

Now that we detailed the proximal algorithms that can help us to solve non-smooth optimization problem, we will now explain how we can tackle multi-convex problems.

III.2.2 Multi-Convex Optimization for BSS

III.2.2.1 Problem overview

In this part, we aim at minimizing a cost function like the following:

$$\underset{\mathbf{U}_1\in\mathbb{R}^{P_1\times J_1},\ldots,\mathbf{U}_n\in\mathbb{R}^{P_n\times J_n}}{\operatorname{argmin}} f(\mathbf{U}_1,\ldots,\mathbf{U}_n)+\sum_{i}^{n}\mathcal{J}_i(\mathbf{U}_i), \tag{III.41}$$

where:

- the function f is supposed to be block multi-convex. It means that, for all $1\leq i\leq n$, f is a convex function of $\mathbf{U}_i$ if all the others blocks are fixed. We will note in the following f_i the corresponding convex function $f_i(\mathbf{U})=f(\mathbf{U}_1,\ldots,\mathbf{U},\ldots,\mathbf{U}_n)$ for any $\mathbf{U}\in\mathbb{R}^{P_i,T_i}$ and for any arbitrary $\mathbf{U}_1\in\mathbb{R}^{P_1\times J_1},\ldots,\mathbf{U}_{i-1}\in\mathbb{R}^{P_{i-1}\times J_{i-1}},\mathbf{U}_{i+1}\in\mathbb{R}^{P_{i+1}\times J_{i+1}},\ldots,\mathbf{U}_n\in\mathbb{R}^{P_n\times J_n}$. In the following, we will suppose the f_i to be differentiable with a Lipschitzian gradient (of Lipschitian coefficient L_i).
- for all $1\leq i\leq n$, the constraint functions $\mathcal{J}_i:\mathbb{R}^{P_i,T_i}\to\mathbb{R}$ are supposed to be proximable.

It appears that the BSS problem is a specific case of such a multi-convex problem with $n=2,\mathbf{U}_1=\mathbf{A}$ and $\mathbf{U}_2=\mathbf{S}$.

All the usual algorithms for multi-convex optimization, such as the Block-Coordinate Descent (BCD) [189] or the Proximal Block Coordinate (PBC) [190] are part of the Gauss-Seidel iteration scheme. It means that they perform a alternated minimization over each block $\mathbf{U}_1,\ldots,\mathbf{U}_n$. These algorithms rely on a strong mathematical background and the framework under which we work on respects the convergence conditions of the different algorithms. Since it's not the core of our work, we admit that property in this document. As in [125], in this work we used the Proximal Alternating Linearized Minimization (PALM) algorithm that is a computationally cheap algorithm with a convergence assurance if some conditions are respected.

III.2.2.2 Proximal Alternating Linearized Minimization

The Proximal Alternating Linearized Minimization (PALM) Algorithm has been introduced in [191] and has a lot of applications in matrix factorization and BSS [192–194]. In order to solve the optimization problem efficiently, it minimizes a relaxed version of the problem where f is linearized close to the previous estimate:

$$\underset{\mathbf{U}_i\in\mathbb{R}^{P_i\times J_i}}{\operatorname{argmin}} f_i^{(k)}(\mathbf{U}_i^{(k)})+\langle\nabla f_i^{(k)}(\mathbf{U}_i^{(k)})|\mathbf{U}_i-\mathbf{U}_i^{(k)}\rangle+\mathcal{J}_i(\mathbf{U}_i)+\frac{L_i}{2}||\mathbf{U}_i-\mathbf{U}_i^{(k)}||^2, \tag{III.42}$$

where L_i is a constant, which we will take equal as the Lipschitzian constant of $\nabla f_i^{(k)}$ in the following. Now, the optimization on $\mathbf{U}_i$ does not depend on $f_i^{(k)}$ anymore, which makes it less constrained.

Moreover, we can note that, since some terms do not depend on $\mathbf{U}_i$:

$$\begin{aligned}
&\underset{\mathbf{U}_i \in \mathbb{R}^{P_i \times J_i}}{\operatorname{argmin}} f_i^{(k)}(\mathbf{U}_i^{(k)}) + \langle \nabla f_i^{(k)}(\mathbf{U}_i^{(k)}) | \mathbf{U}_i - \mathbf{U}_i^{(k)} \rangle + \mathcal{J}_i(\mathbf{U}_i) + \frac{L_i}{2} ||\mathbf{U}_i - \mathbf{U}_i^{(k)}||^2 \\
&= \underset{\mathbf{U}_i \in \mathbb{R}^{P_i \times J_i}}{\operatorname{argmin}} \langle \nabla f_i^{(k)}(\mathbf{U}_i^{(k)}) | \mathbf{U}_i - \mathbf{U}_i^{(k)} \rangle + \mathcal{J}_i(\mathbf{U}_i) + \frac{L_i}{2} ||\mathbf{U}_i - \mathbf{U}_i^{(k)}||^2 \\
&= \underset{\mathbf{U}_i \in \mathbb{R}^{P_i \times J_i}}{\operatorname{argmin}} \langle \frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)}) | \mathbf{U}_i - \mathbf{U}_i^{(k)} \rangle + \frac{1}{L_i} \mathcal{J}_i(\mathbf{U}_i) + \frac{1}{2} ||\mathbf{U}_i - \mathbf{U}_i^{(k)}||^2 \\
&= \underset{\mathbf{U}_i \in \mathbb{R}^{P_i \times J_i}}{\operatorname{argmin}} \frac{1}{L_i} \mathcal{J}_i(\mathbf{U}_i) + \frac{1}{2} ||\mathbf{U}_i - \mathbf{U}_i^{(k)} + \frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)})||^2 - ||\frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)})||^2 \\
&= \underset{\mathbf{U}_i \in \mathbb{R}^{P_i \times J_i}}{\operatorname{argmin}} \frac{1}{L_i} \mathcal{J}_i(\mathbf{U}_i) + \frac{1}{2} ||\mathbf{U}_i - (\mathbf{U}_i^{(k)} - \frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)}))||^2
\end{aligned} \tag{III.43}$$

Therefore, by definition of the proximal operator, this update can be rewritten using the proximal of $\mathcal{J}_i$:

$$\operatorname{prox}_{\frac{\mathcal{J}_i}{L_i}}(\mathbf{U}_i^{(k)} - \frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)})) \tag{III.44}$$

Therefore, each update is in fact one gradient step on the smooth part followed by proximal step on the non-smooth part, exactly like a step of the FBS algorithm. Then, if the proximal operators of the $\mathcal{J}_i$ are explicit the PALM algorithm, fully described in Algorithm 9, is very efficient.

Algorithm 9 Proximal Alternating Linearized Minimization Algorithm

Require: $N_{iter} > 0$
  $\mathbf{U}_i \leftarrow \mathbf{U}_i^{(0)}$ for all $1 \leq i \leq n$
  for $k \leq N_{iter}$ **do**
    for $i \leq n$ **do**
      $\mathbf{U}_i^{(k+1)} = \operatorname{prox}_{\frac{\mathcal{J}_i}{L_i}}(\mathbf{U}_i^{(k)} - \frac{1}{L_i} \nabla f_i^{(k)}(\mathbf{U}_i^{(k)}))$
    end for
  end for
  Return $\mathbf{U}_1, \ldots, \mathbf{U}_n$

The projected gradient described in Algorithm 2 is in fact a specific case of the PALM algorithm with the constraints $\mathcal{J}_i$ being the non-negativity constraints $i_{\geq 0}$

There are some possible extensions to the PALM algorithm [195,196] that we will not study in the scope of this work but that can represent interesting perspectives by accelerating the algorithm.

In the following, we will explain how we used the PALM algorithm in the context of BSS for BLI.

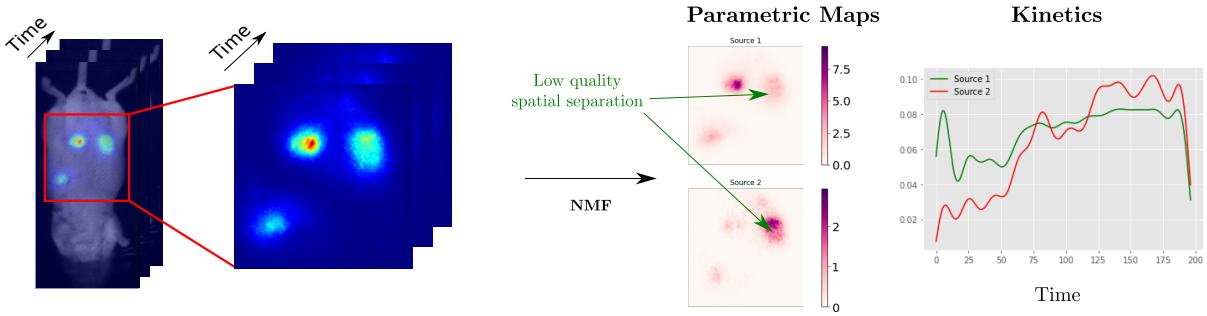

Figure III.9: Illustration of NMF on BLI and its current issues. Left: superposition of BLI and mouse body. Middle: BLI photon count. Right: Results of NMF (parametric maps and kinetics). The parametric maps are corrupted by each other.

III.3 Proposed approach : Sparsity in direct domain

Trying to find $\mathbf{A}$ and $\mathbf{S}$ from the sole knowledge of $\mathbf{Y}$ using (III.11) admits unfortunately an infinity of spurious solutions which calls for additional prior assumptions. We saw in subsubsection III.1.2.3 that classical NMF algorithms, like HALS, mixes spatially the sources as is illustrated in Figure III.9 where the Parametric Maps of the 2 tumors are corrupted by each other. Forcing the Parametric Maps to be more sparse can therefore remove a lot of non-zeros coefficients and result in a better spatial separation of the tumors. To do that, we added a ℓ_1 sparsity constraint on $\mathbf{A}$ to separate the tumors. To prevent degenerated solutions where $||\mathbf{A}|| \to 0$ and $||\mathbf{S}|| \to \infty$ due to the sparsity, the columns of S are constrained to be in the ℓ_2 ball. We aimed then at minimizing:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{AS}||_F^2 + ||\mathbf{\Lambda}\odot\mathbf{A}||_1 + i_{\geq 0}(\mathbf{A}) + i_{\geq 0}(\mathbf{S}) + i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S}), \quad \text{(III.45)}$$

where $\mathbf{\Lambda} \in \mathbb{R}^{P\times K}$ are sparsity weights applied coefficients by coefficients.

Optimization of (III.45) is a non-smooth non-convex problem, calling for the dedicated optimization tools, presented before. In addition, it is highly sensitive to hyperparameters, discussed in section subsection III.3.2.

III.3.1 Efficient iterative algorithm under sparsity constraint

First, and similarly to the Fast-HALS [163] algorithm, in this work the update is done component-by-component in order to be more computationally efficient. By reusing the notations of subsubsection III.1.2.3 $\mathbf{Y}^{(j)} = \mathbf{Y} - \sum_{i\neq j}\mathbf{a}_j\mathbf{s}^j$, we have:

For all $1 \leq j \leq J$, $||\mathbf{Y}-\mathbf{AS}||_F^2 = ||\mathbf{Y}^{(j)} - \mathbf{a}_j\mathbf{s}^j||_F^2$.

The cost function of (III.45) can be expressed, for all j, by:

$$\frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j\mathbf{s}^j||_F^2 + \sum_j ||\mathbf{\Lambda}_j\odot\mathbf{a}_j||_1 + \sum_j i_{\geq 0}(\mathbf{s}_j) + \sum_j i_{\geq 0}(\mathbf{a}_j) + i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S}) \quad \text{(III.46)}$$

We therefore minimized alternatively the set of cost functions $D_A^j(a_j)$ and $D_S^j(s_j)$ for every j [129]:

$$D_{\mathbf{A}}^j(\mathbf{a}_j) = \frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j\mathbf{s}^j||_F^2 + ||\mathbf{\Lambda}_j \odot \mathbf{a}_j||_1 + i_{\geq 0}(\mathbf{a}_j) \tag{III.47}$$

and

$$D_{\mathbf{S}}^j(\mathbf{s}^j) = \frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j\mathbf{s}^j||_F^2 + i_{\{||\cdot||_2^2 \leq 1\}}(\mathbf{s}^j) + i_{\geq 0}(\mathbf{s}_j) \tag{III.48}$$

These cost functions are not differentiable but the regularizations are proximable with an explicit proximal operator.

Indeed, if we write $\mathcal{J}_j(\mathbf{a}_j) = ||\mathbf{\Lambda}_j \odot \mathbf{a}_j||_1 + i_{\geq 0}(\mathbf{a}_j)$ then

$$\text{prox}_{\mathcal{J}_j}(\mathbf{a}_j) = [Soft_{\mathbf{\Lambda}}(\mathbf{a}_j)]_+ \tag{III.49}$$

Similarly, with $\mathcal{G}_j(\mathbf{s}^j) = i_{\{||\cdot||_2^2 \leq 1\}}(\mathbf{s}^j) + i_{\geq 0}(\mathbf{s}_j)$

$$\text{prox}_{\mathcal{G}_j}(\mathbf{s}^j) = [\Pi_{||\cdot||_2}(\mathbf{s}^j)]_+ \tag{III.50}$$

In order to minimize these blocks at each iteration, we then used the Proximal Alternating Linearized Minimization (PALM) algorithm described earlier.

In the following, we will note $f_j(\mathbf{a}_j, \mathbf{s}^j) = \frac{1}{2}||\mathbf{Y}^{(j)} - \mathbf{a}_j\mathbf{s}^j||_F^2$

For the update of $\mathbf{a}_j$, we have:

$$\nabla_{f_j}(\mathbf{a}_j) = \left(\mathbf{a}_j\mathbf{s}^j - \mathbf{Y}^{(j)}\right)\mathbf{s}^{jT} \tag{III.51}$$

∇_{f_j} is L-Lipschitzian with $L = \mathbf{s}^j\mathbf{s}^{j^T}$.

Therefore, one PALM step for $\mathbf{a}_j$ is:

$$\begin{aligned} Va_j &= \left[Soft_{\frac{\mathbf{\Lambda}_j}{\mathbf{s}^j\mathbf{s}^{j^T}}}\left(\mathbf{a}_j - \frac{1}{\mathbf{s}^j\mathbf{s}^{j^T}}\left(\left(\mathbf{a}_j\mathbf{s}^j - \mathbf{Y}^{(j)}\right)\mathbf{s}^{j^T}\right)\right)\right]_+ \\ &= \left[\mathbf{a}_j - \frac{1}{\mathbf{s}^j\mathbf{s}^{j^T}}\left(\left(\mathbf{a}_j\mathbf{s}^j - \mathbf{Y}^{(j)}\right)\mathbf{s}^{j^T} + \mathbf{\Lambda}_j\right)\right]_+, \end{aligned} \tag{III.52}$$

thanks to an proof by exhaustion. Indeed, let's suppose $x \in \mathbb{R}$ and $\lambda > 0$:

- if $x \leq 0$: $[Soft_\lambda(x)]_+ = \left[sign(x)\left[|x| - \lambda\right]_+\right]_+ = 0$ since $x \leq 0$ and $\left[x - \lambda\right]_+ = 0$
- if $x \geq \lambda$: $[Soft_\lambda(x)]_+ = \left[sign(x)\left[|x| - \lambda\right]_+\right]_+ = \left[\left[x - \lambda\right]_+\right]_+ = \left[x - \lambda\right]_+$

- if $0 \leq x \leq \lambda$: $[Soft_\lambda(x)]_+ = \Big[sign(x)\left[|x| - \lambda\right]_+\Big]_+ = [sign(x) \times 0]_+ = 0$ and $[x - \lambda]_+ = 0$

Moreover, since $(\mathbf{a}_j\mathbf{s}^j - \mathbf{Y}^{(j)})s^{jT} = (\mathbf{AS} - \mathbf{Y})\mathbf{s}^{jT} = \mathbf{AS}s^{jT} - \mathbf{Y}\mathbf{s}^{jT} = \mathbf{A}[\mathbf{SS}^T]_j - [\mathbf{YS}^T]_j$ we can rewrite the update rule (Equation III.52) as:

$$\mathbf{a}_j = \left[\mathbf{a}_j - \frac{1}{\mathbf{s}^j\mathbf{s}^{jT}}\left(\mathbf{A}[\mathbf{SS}^T]_j - [\mathbf{YS}^T]_j + \mathbf{\Lambda}_j\right)\right]_+ \tag{III.53}$$

Symmetrically, for $\mathbf{s}^j$:

$$\nabla_{f_j}(\mathbf{s}^j) = \mathbf{a}_j^T\left(\mathbf{a}_j\mathbf{s}^j - \mathbf{Y}^{(j)}\right) \tag{III.54}$$

∇_{f_j} is L-Lipschitzian with $L = \mathbf{a}_j^T\mathbf{a}_j$.

Therefore, one PALM step for $\mathbf{s}^j$ is:

$$\mathbf{s}^j = \left[\Pi_{||.||_2\leq 1}\left(\mathbf{s}^{jT} - \frac{1}{\mathbf{a}_j^T\mathbf{a}_j}\left(\mathbf{S}^T[\mathbf{A}^T\mathbf{A}]_j - [\mathbf{Y}^T\mathbf{A}]_j\right)\right)\right]_+^T, \tag{III.55}$$

These updates rules lead to the following PALM algorithm:

Algorithm 10 $PALM(\mathbf{Y}, \mathbf{A}^{(0)}, \mathbf{S}^{(0)})$

While the stopping criterion Δ_k has not reached the threshold τ, iterate over (k):
1− Update each column j of $\mathbf{A}$:

$$\mathbf{a}_j^{(k+1)} = \left[\mathbf{a}_j^{(k)} - \frac{1}{\mathbf{s}^{j(k)}\mathbf{s}^{jT(k)}}\left(\mathbf{A}[\mathbf{SS}^T]_j - [\mathbf{YS}^T]_j + \mathbf{\Lambda}_j\right)\right]_+ \tag{III.56}$$

2− Update each line j of $\mathbf{S}$:

$$\mathbf{s}^{j(k+1)} = \left[\Pi_{||.||_2\leq 1}\left(\mathbf{s}^{jT(k)} - \frac{1}{\mathbf{a}_j^{T(k)}\mathbf{a}_j^{(k)}}\left(\mathbf{S}^T[\mathbf{A}^T\mathbf{A}]_j - [\mathbf{Y}^T\mathbf{A}]_j\right)\right)\right]_+^T, \tag{III.57}$$

Return $\mathbf{A}, \mathbf{S}$

In Algorithm 10 we defined a stopping criterion Δ_k defined at each iteration (k) that stops the algorithm if it exceeds a threshold τ. Now, we need to compute the sparsity threshold $\mathbf{\Lambda}$.

III.3.2 Adaptive sparsity threshold management

Sparsity hyperparameters Λ in (III.45) used to control the Soft-Thresholding in (III.53) are critical to obtain satisfying separations [125]. We therefore set the threshold Λ automatically. Here, we choose to set the sparsity parameters as:

$$\mathbf{\Lambda} = \mathbf{W}_\mathbf{A}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_J\end{bmatrix}, \tag{III.58}$$

where λ_j is the parameter associated with the j-th source and $\mathbf{W_A}$ is a $P \times J$ matrix used for ℓ_1 reweighting, discussed next.

The first step is to determine the λ_j values. In this part, we assume there is no reweighting involved, meaning that Λ_j is equal to λ_j. Inspired by both [197] and [164], we set λ_j to large values over the first iterations to focus the initial separation on the most important features. We then used a lower λ_j on subsequent iterations so that features associated with smaller details are incorporated. More precisely, with p a gain factor, we set:

- For the first iteration: $\lambda_j = p \times ||(\mathbf{AS} - \mathbf{Y})\mathbf{S}^T||_\infty$ for all $j \in \{1, \ldots, J\}$ sources;
- Over the following iterations (k), for all $1 \leq j \leq J$ we assume that after many iterations, the algorithm has converged to $\mathbf{a}_j^{(\infty)}$ and $\mathbf{s}^{j(\infty)}$ close to the true values $\mathbf{a}_j^{(*)}$ and $\mathbf{s}^{j(*)}$. If we assume that there is no non-negativity constraint for the sake of simplicity, then by considering a fixed point argument (by developing and considering $\mathbf{a}_j^{(\infty)} \approx \mathbf{a}_j^{(*)}$ and $\mathbf{s}^{j(\infty)} \approx \mathbf{s}^{j(*)}$):

$$\begin{aligned}
\mathbf{a}_j^{(\infty)} &= Soft_{\frac{\lambda_j}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}}} \left(\mathbf{a}_j^{(k)} - \frac{1}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \left(\mathbf{a}_j^{(\infty)} \mathbf{s}^{j(\infty)} - \mathbf{Y}^{(j)} \right) \mathbf{s}^{j^T(\infty)} \right) \\
&= Soft_{\frac{\lambda_j}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}}} \left(\frac{1}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \mathbf{Y}^{(j)} \mathbf{s}^{j^T(\infty)} \right) \\
&= Soft_{\frac{\lambda_j}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}}} \left(\frac{1}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \left(\mathbf{a}_j^{(*)} \mathbf{s}^{j(*)} + \mathbf{N} \right) \mathbf{s}^{j^T(\infty)} \right) \\
&\approx Soft_{\frac{\lambda_j}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}}} \left(\mathbf{a}_j^{(*)} + \frac{\mathbf{N}\mathbf{s}^{j^T(\infty)}}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \right),
\end{aligned} \tag{III.59}$$

where $\mathbf{N}\mathbf{s}^{j^T(\infty)}$ is a correlated Gaussian noise (since we assumed that $\mathbf{N}$ was Gaussian). Then, the sources before thresholding are equal to the true ones $\mathbf{a}_j^{(*)}$ up to an additive Gaussian noise. In order to retrieve the real values of such denoising problems, the usual method is to set a threshold at $p\sigma_j$ where σ_j is the standard deviation of the implicated noise and $p \in \mathbb{R}^+$. Indeed, if $p = 3$, the probability that a coefficient of $\mathbf{a}_j^{(*)}$ greater than $3\sigma_j$ is noise is about 0.4%. In practice, we do not know σ_j but it can fortunately be approximated using a Median Absolute Deviation (MAD), defined as :

$$\forall \mathbf{u} \in \mathbb{R}^T \text{MAD}(\mathbf{u}) = \text{median}_{t \leq T} |u_t - \text{median}_{i \leq T}(u_i)| \tag{III.60}$$

We have that $\sigma_j^{(k)} \approx 1.48 \times \text{MAD}(\mathbf{a}_j^{(k)})$ is the approximation of the standard deviation of the Gaussian noise at iteration (k). Indeed, the MAD is robust to a sparse contamination since it would only change a few values, the median would then not be much affected [125] meaning that $\text{MAD}\left(\mathbf{a}_j^{(*)} + \frac{\mathbf{N}\mathbf{s}^{j^T(\infty)}}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \right) \approx \text{MAD}\left(\frac{\mathbf{N}\mathbf{s}^{j^T(\infty)}}{\mathbf{s}^{j(\infty)}\mathbf{s}^{j^T(\infty)}} \right)$ since $\mathbf{a}_j^{(*)}$ is supposed to be sparse.

If the sources were well separated, setting $p = 3$ to remove 99% of the noise. However, we initialize the algorithm with NNDSVD [165], which yields a first imperfect separation. Due to this imperfect separation, each source is initially partially contaminated by the other sources. Because of the possible intensity imbalance between the sources, the less intense sources are likely

to be more strongly contaminated by the most intense ones at initialization. To counterbalance this initialization, we wanted to more strongly regularize the less intense sources.

Specifically, we used a gain factor p_j for each source that depends on the MAD of the estimated source in the initialization $p_j = 3 \times \frac{\max_t \text{MAD}(\mathbf{a}_t^{(0)})}{\text{MAD}(\mathbf{a}_j^{(0)})}$. Thus, if the MAD are comparable in the case of balanced sources, we obtain $p_j \approx 3$ which is consistent with a Gaussian noise. In the following, we will call such source dependent p_j an *adaptive gain factor*, in contrast to a *fixed gain factor* ($p_j = 3$).

According to [125], this strategy of threshold tuning alone does not work well for PALM because using a ℓ_1-norm in Eq. (III.45) instead of the ℓ_0 pseudo-norm to enforce sparsity introduces a bias in the estimates. Since we are doing just 1 proximal gradient step, the parameters are computed from sources with high interferences and noise. To limit such bias, a ℓ_1 reweighting scheme [198] is used here, through the introduction of the $\mathbf{W_A}$ matrix. The idea is then to adjust the matrix $\mathbf{W_A}$ to threshold more low intensity pixels, which are more likely to correspond to interferences, with the following formula:

$$\mathbf{W}^i_{\mathbf{A}_j} = \frac{\epsilon}{\epsilon + \frac{|\mathbf{a}^i_j|}{||\mathbf{a}_j||_\infty}} \tag{III.61}$$

Where ϵ is a small constant and $\mathbf{W}^i_{\mathbf{A}_j}$ is the coefficient of the i-th line and j-th column of $\mathbf{W_A}$ at iteration (k).

III.3.3 Experiments

We tested the PALM algorithm with the different setups (with or without reweighting and with or without an adaptive gain factor) on the real dataset with long kinetics presented in Chapter II. As in subsubsection III.1.2.3, $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$ are initialized with NNDSVD [165] and the maximum number of iterations is set to 1000. Δ_k is the stability measure defined as the minimum cosine distance in the evolution of each $\mathbf{s}^j$ or $\mathbf{a}_j$, $\tau = 1 - 4 \times 10^{-11}$ and $\epsilon = 10^{-3}$ (from [125]).

$$\begin{aligned}\Delta_k = \min\Big(&\min_{j \leq J} \Big\langle \frac{\mathbf{a}_j^{(k)}}{||\mathbf{a}_j^{(k)}||_F}, \frac{\mathbf{a}_j^{(k-1)}}{||\mathbf{a}_j^{(k-1)}||_F} \Big\rangle, \\ &\min_{j \leq J} \Big\langle \frac{\mathbf{s}^{j(k)}}{||\mathbf{s}^{j(k)}||_F}, \frac{\mathbf{s}^{j(k-1)}}{||\mathbf{s}^{j(k-1)}||_F} \Big\rangle \Big)\end{aligned} \tag{III.62}$$

We set $J = 2$ since there are 2 tumors by image. We reported in Table III.2 metrics for the different algorithmic refinements and a comparison with popular blind-source separation methods: PCA and FastICA [199].

We illustrate on Figure III.10 the impact of the sparsity options on the separation of 2 BLI cases with different kinetics as illustrated on few frames. The top BLI case corresponds to the case presented in Figure III.12 for which separating the tumors based on total photon count alone is challenging. The bottom BLI case has one tumor much larger and more intense than the other one.

In addition to these results, we also evaluated this obtained PALM algorithm (with Adaptive Gain Factor and Reweighting) with all the metrics defined previously on the 2 real BLI datasets

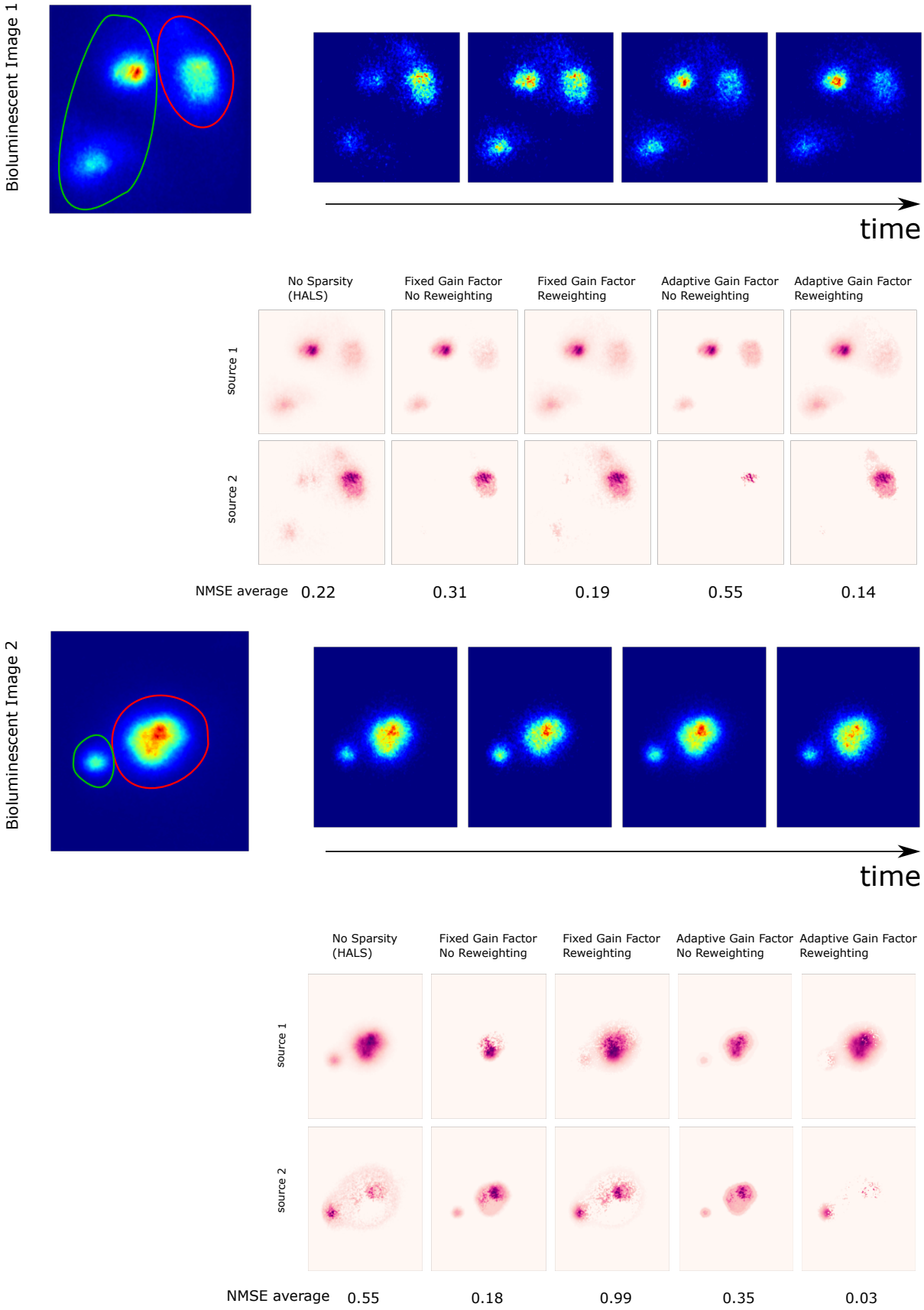

Figure III.10: Impact of Sparsity Parameters on 2 BLI Images. For both BLI : Top Left: global visualization of the data with the accumulation image (integral on all frames); Top Right: : selection of some frames; Bottom: Parametric maps $\mathbf{a}_1$ and $\mathbf{a}_2$ for 5 setups of PALM.

Table III.2: Ablation study on the proposed PALM algorithm with adaptive sparsity constraints. Average NMSE values and standard deviation are reported over the whole dataset (18 BLI).

				Fixed Gain Factor		Adaptive Gain Factor	
	PCA	FastICA	No Sparsity (HALS)	No Reweighting	Reweighting	No Reweighting	Reweighting
$NMSE$	4.9 ± 11	4.9 ± 11	0.34 ± 0.65	0.57 ± 0.35	0.38 ± 1.52	0.69 ± 1.06	$\mathbf{0.08 \pm 0.12}$

	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
HALS	0.59 ± 0.83	0.17 ± 0.16	0.38 ± 0.63	0.57 ± 1.3	0.65 ± 0.42
PALM	$\mathbf{0.58 \pm 0.82}$	$\mathbf{0.06 \pm 0.15}$	$\mathbf{0.32 \pm 0.65}$	$\mathbf{0.32 \pm 0.65}$	$\mathbf{0.84 \pm 0.29}$

Table III.3: Quantitative Results of Fast-HALS Algorithm and the proposed PALM algorithm on the real BLI dataset with short kinetics.

	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
HALS	0.45 ± 0.39	0.16 ± 0.20	0.31 ± 0.35	0.30 ± 0.45	0.76 ± 0.33
PALM	$\mathbf{0.44 \pm 0.59}$	$\mathbf{0.01 \pm 0.05}$	$\mathbf{0.23 \pm 0.47}$	$\mathbf{0.08 \pm 0.12}$	$\mathbf{0.96 \pm 0.14}$

Table III.4: Quantitative Results of Fast-HALS Algorithm and the proposed PALM algorithm on the real BLI dataset with long kinetics.

(short and long kinetics), and on the synthetic datasets described in Chapter II (with the different kinds of variability). The results are presented in Table III.3, Table III.4, Table III.5 and Table III.6. Visual results on 2 synthetic BLI taken from the synthetic dataset without variability are shown in Figure III.11.

III.3.4 Discussion

According to Table III.2 we can see that it is the combination of adaptive gain factor and sparsity reweighting that allows to completely outperform the basic algorithms HALS, PCA and FastICA in terms of average values and standard deviation $NMSE$. We can also see that PCA and FastICA do not seem suited for this task.

In Figure III.10, we can visualize the impacts of the different parts of the algorithm on the separations. On those examples, biologists confirmed that the 2 left spots on the top BLI case correspond to the same tumor and that the right spot is a second tumor. Without the reweighting scheme on this image, the parametric maps become too sparse and we miss information. Adding the reweighting scheme allows to focus on most important pixels and combining it with the adaptive gain factor allows here only a marginal improvement. However, the bottom case shows a much better separation with our proposed setup with 2 tumors differing in both size and intensity. In particular, with reweighting but without the adaptive gain factor, the second source still mixes the 2 tumors.

In this work, we focused on the iterative part for the optimization of (III.45) with the setting of sparsity hyperparameters. However, this problem can be sensitive to initialization, we therefore also tried to refine $\mathbf{S}^{(0)}$ by using physiological knowledge based on the work of

Table III.5: Quantitative Results of Fast-HALS Algorithm and the proposed PALM algorithm on synthetic data without time warping perturbations.

	Variability : None					Variability : Piece-wise Affine				
Algorithm	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
HALS	0.49 ± 0.40	0.20 ± 0.25	0.35 ± 0.36	0.48 ± 0.68	0.66 ± 0.45	$\mathbf{0.58 \pm 0.46}$	0.30 ± 0.38	0.44 ± 0.45	0.76 ± 1.7	0.59 ± 0.49
PALM	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.03 \pm 0.08}$	$\mathbf{0.17 \pm 0.28}$	$\mathbf{0.22 \pm 0.26}$	$\mathbf{0.89 \pm 0.27}$	0.58 ± 0.49	$\mathbf{0.04 \pm 0.09}$	$\mathbf{0.31 \pm 0.45}$	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.83 \pm 0.31}$

Table III.6: Quantitative Results of Fast-HALS Algorithm and the proposed PALM algorithm on synthetic data with time warping perturbations.

	Variability : Warping					Variability : Warping + Piece-wise Affine				
Algorithm	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
HALS	1.19 ± 1.1	0.43 ± 0.39	0.81 ± 0.93	2.4 ± 6.1	0.47 ± 0.46	1.26 ± 1.2	0.45 ± 0.38	0.86 ± 99	2.6 ± 6.8	0.44 ± 0.49
PALM	$\mathbf{1.03 \pm 0.76}$	$\mathbf{0.29 \pm 0.42}$	$\mathbf{0.66 \pm 0.72}$	$\mathbf{0.68 \pm 1.1}$	$\mathbf{0.54 \pm 0.49}$	$\mathbf{1.05 \pm 0.77}$	$\mathbf{0.32 \pm 0.45}$	$\mathbf{0.69 \pm 0.72}$	$\mathbf{0.66 \pm 0.92}$	$\mathbf{0.49 \pm 0.51}$

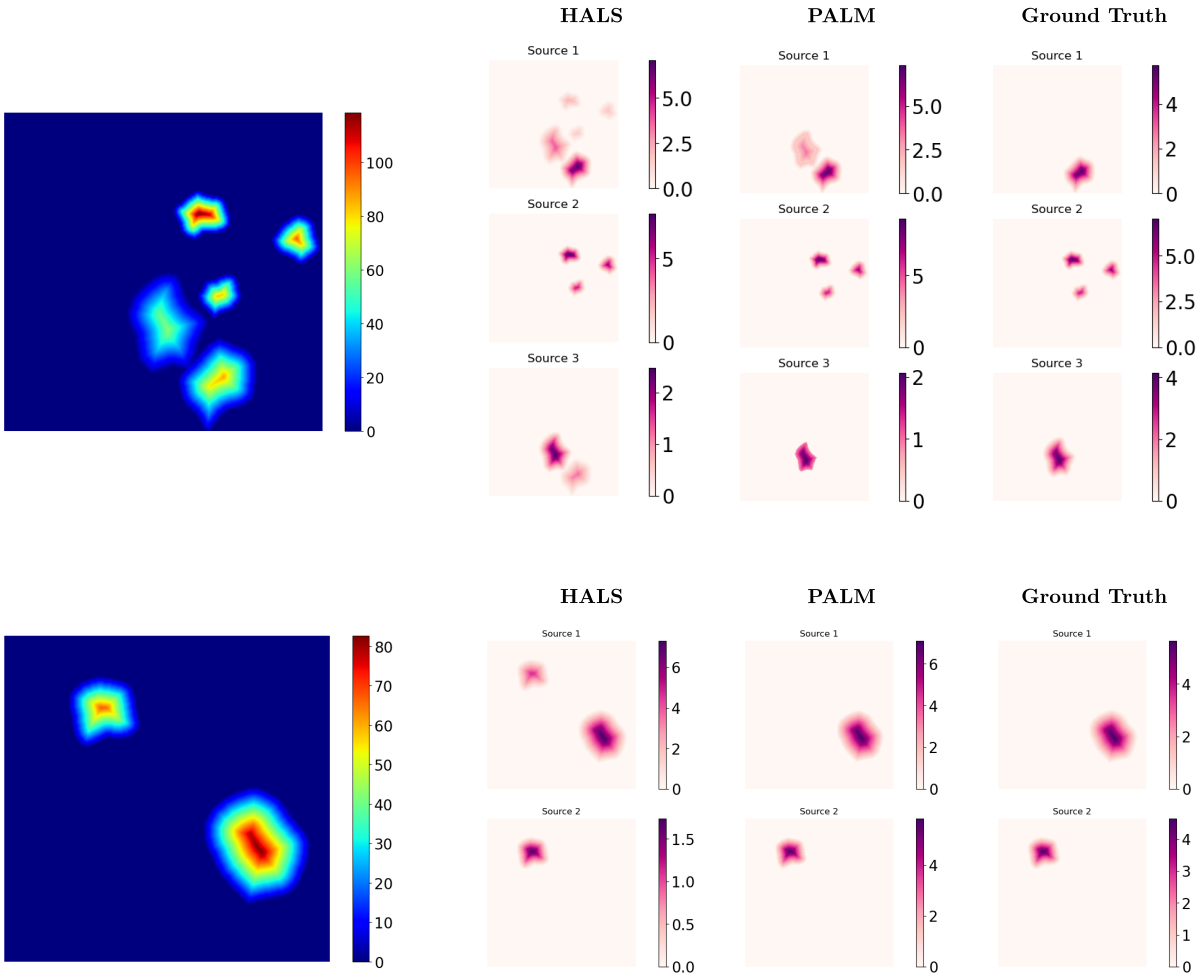

Figure III.11: Visual results of NMF on 2 synthetic BLI for each algorithm (PALM and HALS). The results are taken from the dataset without variability.

Sim *et al.* but the improvement was not significant. These results were communicated in one conference article [167].

Table III.3, Table III.4, Table III.5 and Table III.6 show that for almost all datasets and all metrics, the PALM algorithm with a sparsity constraint is more able to separate tumors than HALS, as illustrated in Figure III.11. However, it appears in these tables that there seems to be a lot of possible improvement in the context of separating real BLIs. One possible improvement is that for now, we did not put any spatial constraint on the Parametric Maps, which can often lead to isolated artifact pixels as in the case of the bottom image in Figure III.10. As was explained in subsubsection III.1.3.1, the Parametric Maps are maybe more sparse in another domain. In the next section, we will see how enforcing them to be sparse in an adapted domain can lead to a better spatial consistency.

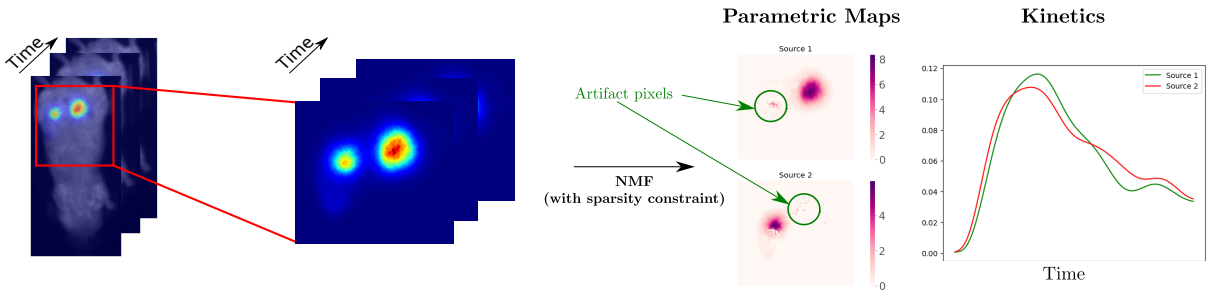

Figure III.12: Illustration of NMF with sparsity constraint on BLI and its current issues. Left: superposition of BLI and mouse body. Middle: BLI photon count. Right: Results of NMF (parametric maps and kinetics). The parametric maps are corrupted by isolated artifact pixels.

III.4 Proposed approach : Sparsity in transformed domain

III.4.1 Adaptation of the algorithm for sparsity in the wavelet domain

Enforcing the sparsity in the direct domain can generate isolated artifact pixels in the parametric maps as is illustrated in Figure III.12.

In this part, we propose to reduce these artifacts by using a spatial regularization via ℓ_1 sparsity constraint in a transform domain. This can also be done by adding a regularization on the gradient of the parametric maps. For instance, it is possible to minimize the ℓ_1 norm of the gradient (TV) [200] or the ℓ_2 norm [122].

The constraint in [167] does not take into account the spatial coherence of the signals which deteriorates the separation results. We then adapted the work of [167] to a sparsity constraint in the wavelet domain, which can leverage that prior knowledge of spatial coherence. Sparsity has been shown to improve source separation by providing more contrast between them [164, 167]. However, in the previous work, the sparsity is enforced only on the direct domain while the source can be even sparser in a transformed domain. For instance, a sinusoidal function is not sparse in the direct domain but is very sparse in the Fourier domain. In the context of images, complex geometrical structures can be sparse in transformed domains thanks to a multi-scale signal analysis. According to [177], the wavelet transform makes sparse piece-wise polynomial signals, which are a good approximation of most of natural signals. For this reason, constraining sparsity in the wavelet domain is widely used in blind source separation [125, 197, 201] and we will work with this transform in this work. The optimization problem is then:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{A}\mathbf{S}||_F^2 + ||\mathbf{\Lambda}\odot\mathbf{\Phi}^T\mathbf{A}||_1 + i_{\geq 0}(\mathbf{A}) + i_{\geq 0}(\mathbf{S}) + i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S}), \quad \text{(III.63)}$$

where $\mathbf{\Phi}^T$ is the wavelet transform, chosen to be undecimated and orthogonal in the following and $i_C(.)$ is the indicator function of a set C. $\frac{1}{2}||\mathbf{Y}-\mathbf{A}\mathbf{S}||_F^2$ is the data fidelity term, $||\mathbf{\Lambda}\odot\mathbf{\Phi}^T\mathbf{A}||_1$ corresponds to the sparsity in the wavelet domain, $i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S})$ is the constraining of the lines of S in the ℓ_2 unitary ball to avoid degenerated solutions where $||\mathbf{A}||\to 0$ and $||\mathbf{S}||\to\infty$, and the $i_{\geq 0}(\cdot)$ are the non-negativity terms.

For computational reasons, we decided then to formulate the problem in synthesis, and not in analysis as in (III.63), because sparsity and data fidelity terms can then be expressed in the same domain, facilitating the implementation as it would enable us to use the same code

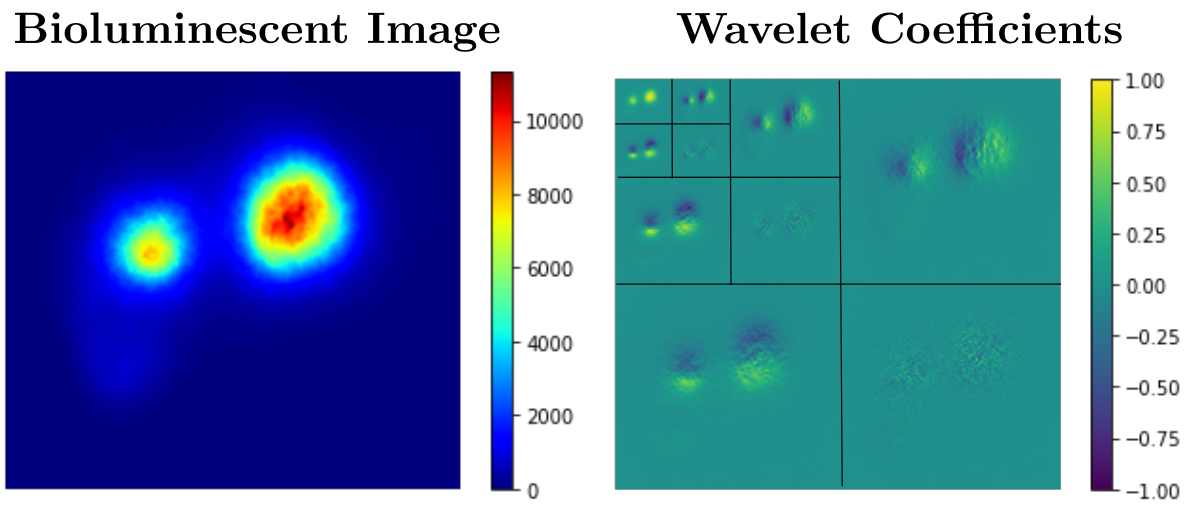

Figure III.13: Illustration of wavelet coefficients of a BLI. Left : Accumulation image of a Bioluminescent Image (integral of photon counting). Right : associated Haar Wavelet [177] coefficients, each (details for 3 scales and approximation of last scale). Each map of coefficients is normalized between -1 and 1 for clarity of display.

foundations as in [167]. Indeed, since $\mathbf{\Phi}^T$ is assumed to be an orthogonal transform, minimizing $||\mathbf{Y} - \mathbf{AS}||$ is equivalent to minimizing $||\mathbf{\Phi}^T\mathbf{Y} - \mathbf{\Phi}^T(\mathbf{AS})|| = ||\mathbf{\Phi}^T\mathbf{Y} - (\mathbf{\Phi}^T\mathbf{A})\mathbf{S}||$ according to the Parseval theorem. Let's note $\mathbf{X}_{\mathbf{\Phi}^T} = \mathbf{\Phi}^T\mathbf{X}$ in the following for any signal $\mathbf{X}$. Figure III.13 displays a visual illustration of these wavelet coefficients for a given BLI. With this formulation, (III.63) becomes:

$$\underset{\mathbf{A}_{\mathbf{\Phi}^T}\in\mathbb{R}^{P_{\mathbf{\Phi}^T}\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}_{\mathbf{\Phi}^T} - \mathbf{A}_{\mathbf{\Phi}^T}S||_F^2 + ||\mathbf{\Lambda}\odot\mathbf{A}_{\mathbf{\Phi}^T}||_1 + i_{\geq 0}(\mathbf{\Phi}\mathbf{A}_{\mathbf{\Phi}^T}) + i_{\geq 0}(\mathbf{S}) + i_{\{\forall j\leq J, ||\mathbf{s}^j||_2^2\leq 1\}}(\mathbf{S}), \tag{III.64}$$

where $P_{\mathbf{\Phi}^T} = P \times K \times 4$, with K the number of scales. A big optimization difficulty is that the sparsity and non-negativity terms are not expressed in the same domain. In order to solve this optimization problem, one can use the Generalized Forward-Backward Splitting Algorithm [186] that treats the proximals one by one or the Chambolle and Pock algorithm [187] that uses the dual problem, as specified in [201]. As an alternative, we can approximate the global proximal by composing the proximal operators of both constraints [202]. The resulting algorithm, although being an approximation, is computationally cheaper and does not introduce further hyperparameters to tune. The proximal of $\mathcal{G}(\mathbf{A}_{\mathbf{\Phi}^T}) = ||\mathbf{\Lambda}\odot\mathbf{A}_{\mathbf{\Phi}^T}||_1 + i_{\geq 0}(\mathbf{\Phi}\mathbf{A}_{\mathbf{\Phi}^T})$ can then be approximated by :

$$prox_{\mathcal{G}(.)}(\mathbf{A}_{\mathbf{\Phi}^T}) = Soft_{\mathbf{\Lambda}}\left(\mathbf{\Phi}^T\left[\mathbf{\Phi}\mathbf{A}_{\mathbf{\Phi}^T}\right]_+\right), \tag{III.65}$$

Since our signals are in fact stacks of vectorized 2D images, after logical successive reshapes to apply the wavelet transform on the 2D images and rasterize the output, for a signal $\mathbf{X}\in\mathbb{R}^{P\times T}$, the corresponding $\mathbf{X}_{\mathbf{\Phi}^T}$ in the wavelet domain is of size $(M\times N\times K\times 4)\times T$ where K is the numbers of scales used (1 dimension for the approximation coefficients and 3 for the details coefficients).

Similarly to [167], the cost function can then be optimized using the PALM algorithm. It is very important to handle the non-negativity before the soft-thresholding as the other way around can generate features that would have been softened by the sparsity. The PALM algorithm alternates between a proximal step on $\mathbf{A}_{\mathbf{\Phi}^T}$ and a proximal step on $\mathbf{S}$. In this part, similarly to the sparsity in the direct domain, the update is done component-by-component: we alternate between the update of the columns $\mathbf{a}_{\mathbf{\Phi}_j^T}$ of $\mathbf{A}_{\mathbf{\Phi}^T}$ and the lines $\mathbf{s}^j$ of $\mathbf{S}$. Based on this modeling, we propose the following Algorithm 12 to retrieve the matrices $\mathbf{A}^{(*)}$ and $\mathbf{S}^{(*)}$ from $\mathbf{Y}$. We can reasonably expect the proposed algorithm to converge to a critical point of (III.63) but we have no proof of convergence here since we have used an approximated proximal in (III.65) [191]. It is an iterative algorithm with a stopping criterion Δ_k defined at each iteration (k) that stops the algorithm if it exceeds a threshold τ.

These proximal steps yield the following update rules:

$$\mathbf{a}_{\mathbf{\Phi}_j^T} = Soft_{\frac{\Lambda}{\mathbf{s}^j\mathbf{s}^{jT}}}\left(\mathbf{\Phi}^T\left[\mathbf{\Phi}\left(\mathbf{a}_{\mathbf{\Phi}_j^T} - \frac{1}{\mathbf{s}^j\mathbf{s}^{jT}}\left(\nabla_{\mathbf{a}_{\mathbf{\Phi}_j^T}}\right)\right)\right]_+\right), \tag{III.66}$$

where $\nabla_{\mathbf{a}_{\mathbf{\Phi}_j^T}} = \mathbf{A}_{\mathbf{\Phi}^T}[\mathbf{S}\mathbf{S}^T]_j - [\mathbf{Y}\mathbf{S}^T]_j$

And:

$$\mathbf{s}^j = \left[\Pi_{||.||_2\leq 1}\left(\mathbf{s}^{jT} - \frac{1}{\mathbf{a}_{\mathbf{\Phi}_j^T}^T\mathbf{a}_{\mathbf{\Phi}_j^T}}\left(\nabla_{\mathbf{s}^j}\right)\right)\right]_+^T, \tag{III.67}$$

where $\nabla_{\mathbf{s}^j} = \mathbf{S}^T[\mathbf{A}_{\mathbf{\Phi}^T}^T\mathbf{A}_{\mathbf{\Phi}^T}]_j - [\mathbf{Y}^T\mathbf{A}_{\mathbf{\Phi}^T}]_j$

Algorithm 11 $PALM(\mathbf{Y}, \mathbf{A}^{(0)}, \mathbf{S}^{(0)})$

$\mathbf{A}_{\mathbf{\Phi}^T}^{(0)} = \mathbf{\Phi}^T\mathbf{A}^{(0)} \in \mathbb{R}^{(M\times N\times K\times 4)\times T}$ with K number of scales, M and N spatial dimensions.
While the stopping criterion Δ_k has not reached the threshold τ, iterate over (k):
1− Update each column j of $\mathbf{A}_{\mathbf{\Phi}^T}$:

$$\mathbf{a}_{\mathbf{\Phi}_j^T}^{(k+1)} = Soft_{\frac{\mathbf{\Lambda}_j}{\mathbf{s}^{j(k)}\mathbf{s}^{jT(k)}}}\left(\mathbf{\Phi}^T\left[\mathbf{\Phi}\left(\mathbf{a}_{\mathbf{\Phi}_j^T}^{(k)} - \frac{1}{\mathbf{s}^{j(k)}\mathbf{s}^{jT(k)}}\left(\mathbf{A}_{\mathbf{\Phi}^T}[\mathbf{S}\mathbf{S}^T]_j - [\mathbf{Y}\mathbf{S}^T]_j\right)\right)\right]_+\right) \tag{III.68}$$

2− Update each line j of $\mathbf{S}$:

$$\mathbf{s}^{j(k+1)} = \left[\Pi_{||.||_2\leq 1}\left(\mathbf{s}^{jT(k)} - \frac{1}{\mathbf{a}_{\mathbf{\Phi}_j^T}^{T(k)}\mathbf{a}_{\mathbf{\Phi}_j^T}^{(k)}}\left(\mathbf{S}^T[\mathbf{A}_{\mathbf{\Phi}^T}^T\mathbf{A}_{\mathbf{\Phi}^T}]_j - [\mathbf{Y}^T\mathbf{A}_{\mathbf{\Phi}^T}]_j\right)\right)\right]_+^T, \tag{III.69}$$

$\mathbf{A} = \mathbf{\Phi}\mathbf{A}_{\mathbf{\Phi}^T}$.
Return $\mathbf{A}, \mathbf{S}$

To set the parameters $\mathbf{\Lambda}$, we used the same strategy as in [167] based on a fixed-point argument, with an adaptive gain factor and a reweighting on the Median Absolute Deviation esstimator (MAD) on the non-zero values of $\mathbf{a}_{\mathbf{\Phi}^T j}^{(k)}$ that remove the noise contaminating the parametric maps. This provides a spatial regularization that does not need any hyperparameter tuning. To do that, we need the noise in each scale to be comparable with each other. Indeed,

Sparsity Domain	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct [167]	0.58 ± 0.82	$\mathbf{0.06 \pm 0.15}$	0.32 ± 0.65	$\mathbf{0.32 \pm 0.65}$	$\mathbf{0.84 \pm 0.29}$
Wavelet	$\mathbf{0.48 \pm 0.70}$	0.10 ± 0.19	$\mathbf{0.29 \pm 0.55}$	0.33 ± 0.57	0.78 ± 0.37

Table III.7: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains on the real BLI dataset with short kinetics.

Sparsity Domain	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct [167]	0.44 ± 0.59	$\mathbf{0.01 \pm 0.05}$	0.23 ± 0.47	$\mathbf{0.08 \pm 0.12}$	$\mathbf{0.96 \pm 0.14}$
Wavelet	$\mathbf{0.31 \pm 0.37}$	0.05 ± 0.12	$\mathbf{0.18 \pm 0.31}$	0.14 ± 0.18	0.91 ± 0.22

Table III.8: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains on the real BLI dataset with long kinetics.

Table III.9: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains on synthetic data without time warping perturbations.

	Variability : None					Variability : Piece-wise Affine				
Sparsity Domain	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.03 \pm 0.08}$	$\mathbf{0.17 \pm 0.28}$	$\mathbf{0.22 \pm 0.26}$	$\mathbf{0.89 \pm 0.27}$	0.58 ± 0.49	$\mathbf{0.04 \pm 0.09}$	$\mathbf{0.31 \pm 0.45}$	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.83 \pm 0.31}$
Wavelet	0.39 ± 0.35	0.10 ± 0.21	0.25 ± 0.33	0.37 ± 0.41	0.78 ± 0.40	$\mathbf{0.48 \pm 0.34}$	0.17 ± 0.34	0.32 ± 0.37	0.37 ± 0.33	0.70 ± 0.45

the MAD operator is computed over all the scales and all the types of coefficients in the same time. This was done by computing the standard deviation of a Gaussian white noise of variance 1 over all the scales and all the types of coefficients. When performing a wavelet transform, we then divided the results of each map of coefficients in each scale by the associated standard deviation value. A Gaussian white noise of variance 1 is then preserved over all the scales and all the types of coefficients, as illustrated in Figure III.14. The noise levels are then comparable between all scales and types of coefficients.

III.4.2 Experiments

We chose the Haar Wavelet transform [177,203] which is orthogonal and generalizes to different scales the smoothness constraint defined in [122] that enforces neighborings pixels to have the same value. As in section III.3, matrices $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$ are initialized with NNDSVD [165]. The maximum number of iterations is set to 1000, Δ_k is the stability measure defined as the minimum cosine distance in the evolution of each $\mathbf{s}^j$ or $\mathbf{a}_j$ and $\tau = 1 - 4 \times 10^{-11}$. We reported the results of the algorithm on real data with short kinetics in Table III.7, on real data with long kinetics in Table III.8 and visual examples on 2 real BLIs in Figure III.15 illustrating typical behavior of the algorithm. This algorithm has also be tested on the synthetic datasets described in Chapter II (with the different kinds of variability). The associated results are presented in Table III.9 and Table III.10. Visual results on 3 synthetic BLI taken from the synthetic dataset are displayed in Figure III.11.

III.4.3 Discussion

Table III.7 and Table III.8 show that source sparsity in the wavelet domain performs better in terms of spatial separation of parametric maps, measured by $NMSE_A$, which is the desired

Table III.10: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains on synthetic data with time warping perturbations.

	Variability : Warping					Variability : Warping + Piece-wise Affine				
Sparsity Domain	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	1.03 ± 0.76	0.29 ± 0.42	0.66 ± 0.72	0.68 ± 1.1	0.54 ± 0.49	1.05 ± 0.77	$\mathbf{0.32 \pm 0.45}$	0.69 ± 0.72	0.66 ± 0.92	$\mathbf{0.49 \pm 0.51}$
Wavelet	$\mathbf{0.77 \pm 0.67}$	$\mathbf{0.29 \pm 0.38}$	$\mathbf{0.53 \pm 0.59}$	$\mathbf{0.66 \pm 0.99}$	$\mathbf{0.52 \pm 0.53}$	$\mathbf{0.77 \pm 0.65}$	0.34 ± 0.44	$\mathbf{0.56 \pm 0.59}$	$\mathbf{0.65 \pm 0.88}$	0.49 ± 0.53

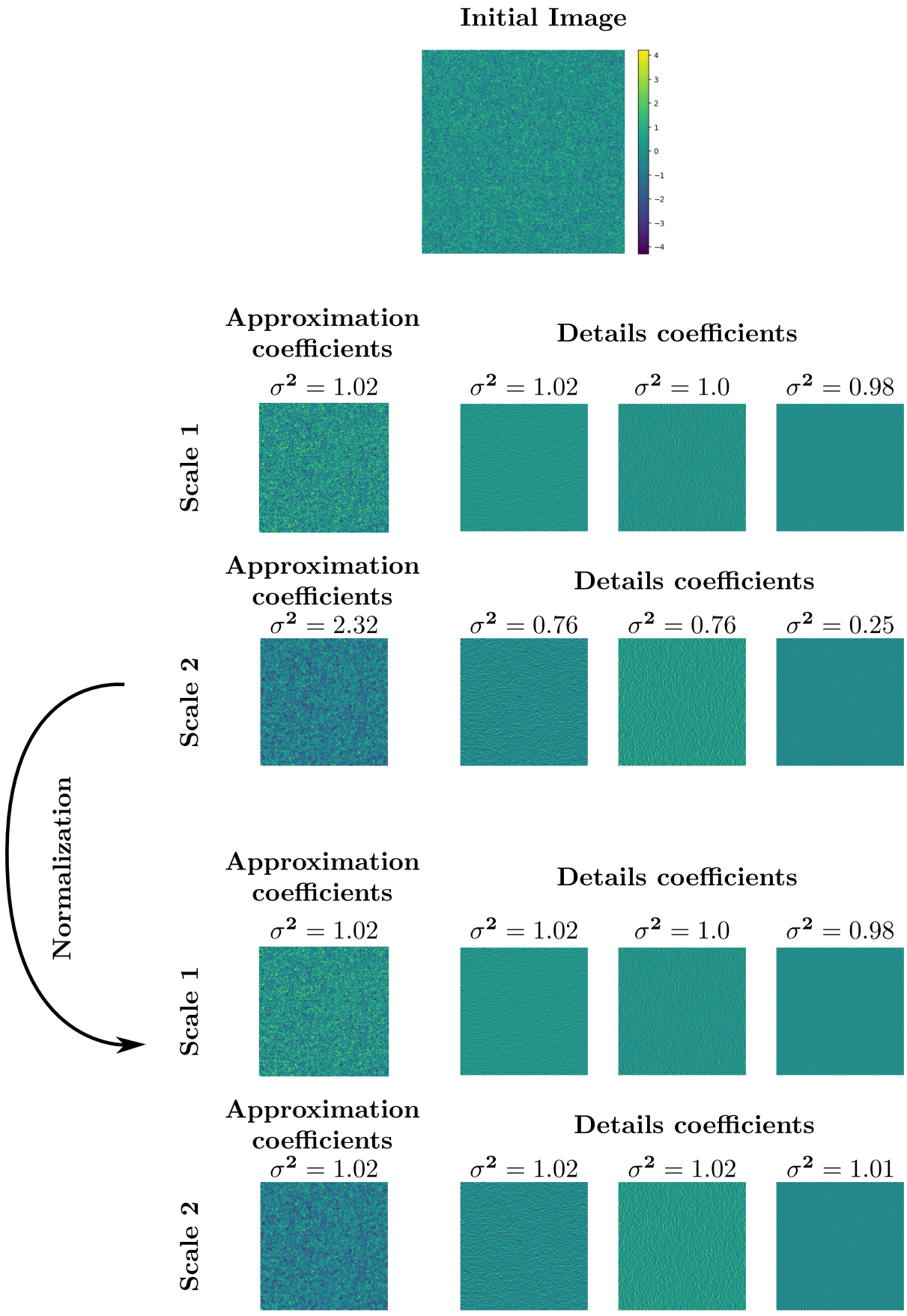

Figure III.14: Normalization process to keep the noise values comparable between scales for a Haar Wavelet Transform. The input image is a Gaussian white noise of standard deviation $\sigma = 1$.

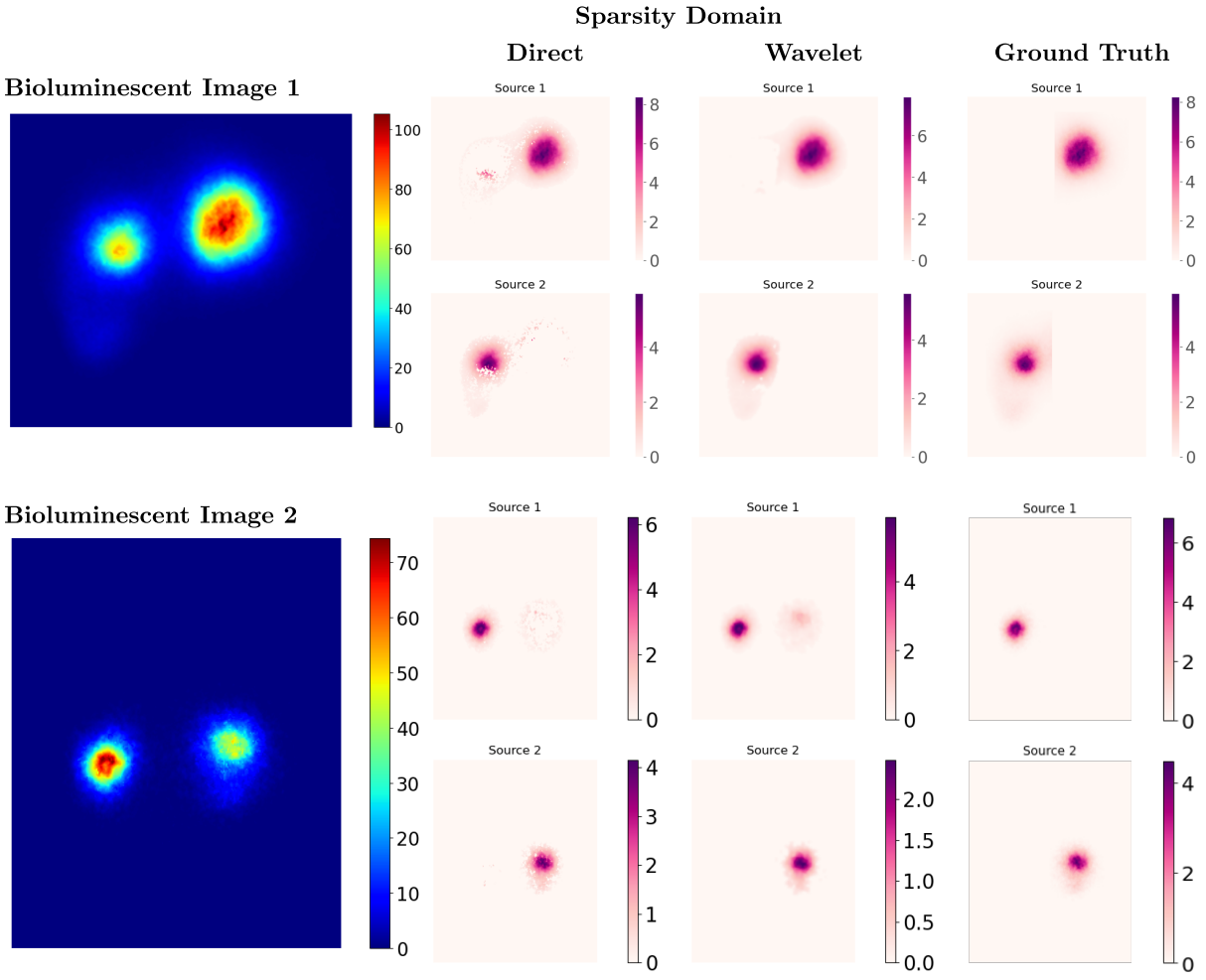

Figure III.15: Visual results for each NMF algorithm on 2 real BLIs. For each BLI, Left : accumulation image (integral on all frames of photon counting); Right : parametric maps.

behavior. Indeed, the top image in Figure III.15 illustrates that it can remove isolated artifact pixels.

It is also interesting to note that the new algorithm seems to downgrade a little bit the kinetics separation performances, measured by $NMSE_S$, which can be explained because it focuses more on the spatial separation (with the spatial constraint) in the optimization problem at the expense of a slightly less accurate temporal estimate. This would be an intriguing problem to pursue. $NMSE_{tot}$ shows however that here the improvement of the spatial separation is enough to compensate for this deterioration.

Finally, enforcing the sources to be sparse in the transformed domain may also decrease the initial imperfect $NMSE$ and *corr* metrics, but that is explained by tumors previously badly separated but so similar both in terms of kinetics shape and intensity that it misled the metrics as illustrated in Figure III.4. As such, the degradation of these metrics should not be taken into account.

Table III.9 and Table III.10 show that source sparsity in the wavelet domain performs better in terms of spatial separation of parametric maps, evaluated by $NMSE_A$, when there is perturbation involved. This is illustrated in Figure III.16 where we can see that the algorithm smoothes the isolated pixels in the parametric maps generated by the previous algorithm, that

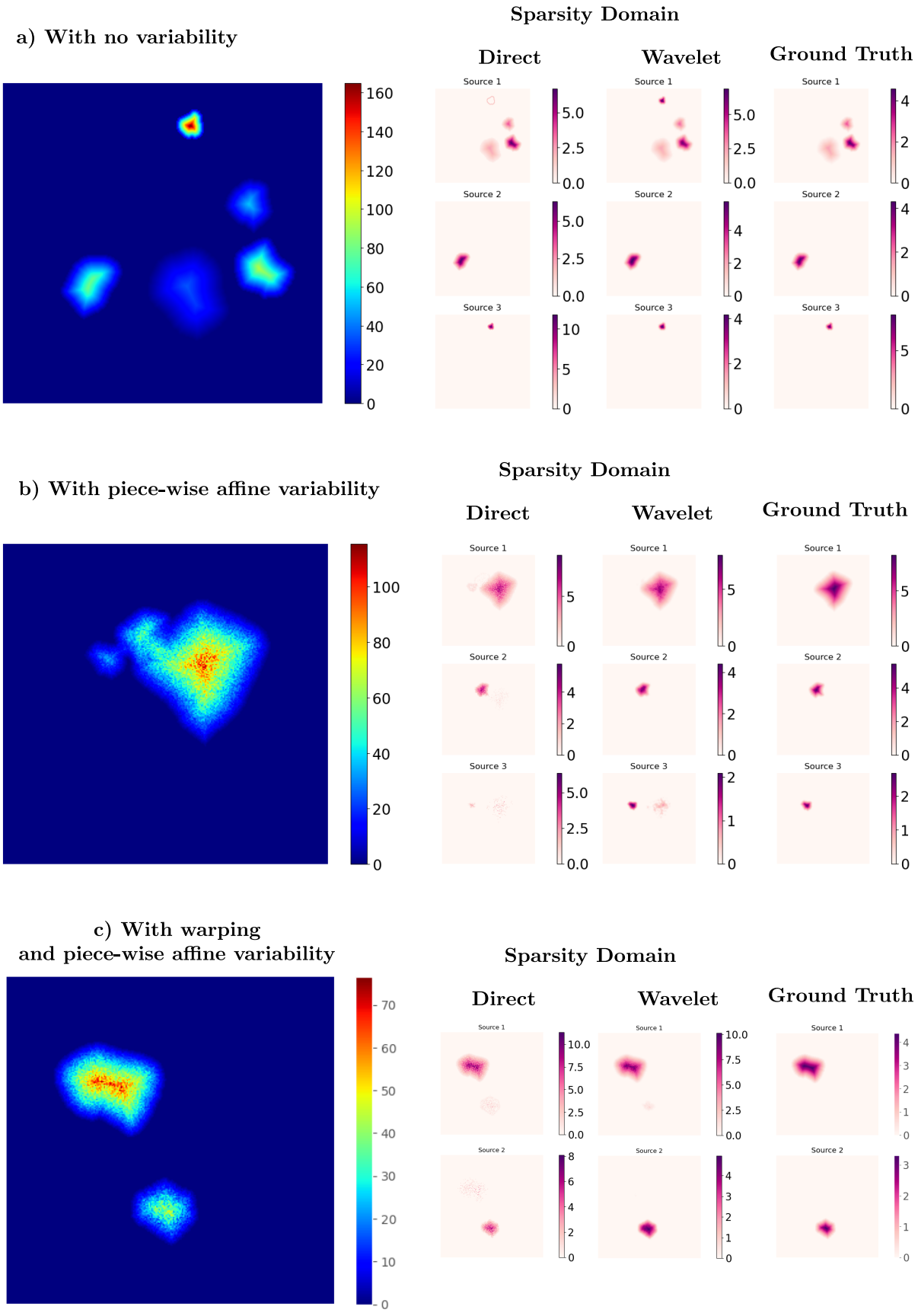

Figure III.16: Visual results for each NMF algorithm on 2 real BLIs. For each BLI, Left : accumulation image (integral on all frames of photon counting); Right : parametric maps.

constrained the parametric maps to be sparse in the direct domain and thus did not force any spatial consistency in the parametric maps. We can also see in this figure that enforcing the parametric maps to be sparse in the wavelet domain downgrades the performance of the algorithm if there is no perturbation, a phenomenon confirmed by Table III.9. These results will be communicated in one accepted conference article:

Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J. C. (2024). Wavelet-based sparse non-negative matrix factorization for bioluminescent imaging unmixing. In 2024 IEEE 21st International Symposium on Biomedical Imaging (ISBI).

However, in some cases like the BLI in the bottom of Figure III.15 and the ones with variability in Figure III.16 (middle and bottom), the optimization algorithm with this spatial constraint seems to sometimes create *halos* around isolated pixels. It may be caused by the choice of regularization parameters and the spatial variability of the kinetics inside a tumor due to its heterogeneity [204]. In the next part, we will see how it is possible to take these variability into account in order to improve the separation.

III.5 Pixel-level kinetics variability

III.5.1 Introduction on variability

In section III.4 and section III.4, we made the assumption that a linear mixing model was sufficient to describe the data we were handling. However, similarly to what is called spectral variability in Hyperspectral Imaging [122], physiological and technical constraints can introduce non-linearities in the kinetics observed in pixels from the same source (tumor). This phenomena can be observed in Figure III.17 where are displayed the spatial average and standard deviation curves from two tumors in one BLI. We can see that the kinetics of the pixels belonging to a single tumor can vary locally. Such variability can be explained by small local heterogeneities in the tumors, spatially-variable tissue attenuation phenomenons, and slight synchronization issues in the imaging device. According to Chapter II, this is mostly due in our case to the imaging device. In order to improve the separation, it is then possible to into account this variability in the unmixing process.

Pixel-level variability can be modeled by introducing a source-modulating function f_i in the mixing model for each pixel i:

$$\mathbf{y}^i = \sum_{j=1}^{J} \mathbf{a}_j^i f_i(\mathbf{s}^j) + \mathbf{n}^i \tag{III.70}$$

In [205], the authors assume in their work that the variability is only resulting from scaling factors. They used pixel-level linear scaling factor functions $f_i(\mathbf{s}^j) = \boldsymbol{\lambda}_j^i \mathbf{s}^j$ to handle this spectral variability while keeping the usual sum-to-one inherent to hyperspectral data. The corresponding optimization problem can be expressed by:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\boldsymbol{\Lambda}\in\mathbb{R}^{P\times J},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y} - (\boldsymbol{\Lambda}\odot\mathbf{A})\mathbf{S}||_F^2 + \mathcal{J}(\mathbf{A}) + \mathcal{G}(\mathbf{S}) + \mathcal{H}(\boldsymbol{\Lambda}), \tag{III.71}$$

where $\boldsymbol{\Lambda}$ is the matrix representing the scaling factors, computed with an optimization algorithm, and $\mathcal{H}$ is an appropriated regularization applied to this matrix. The assumption of

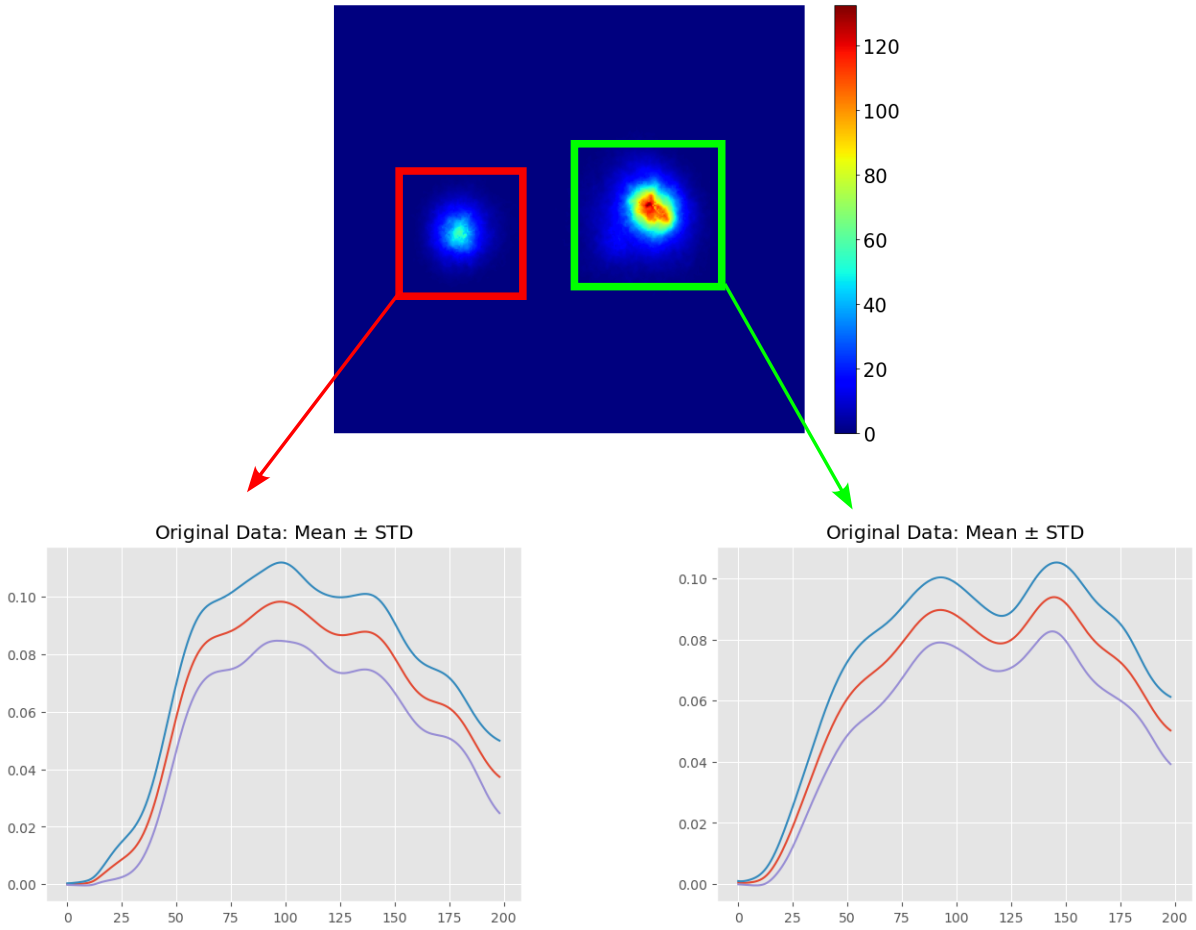

Figure III.17: Example of spatial average and standard deviation in kinetics of 2 tumors within on BLI. The kinetics of a given tumor may vary locally.

a variability caused only by scaling factors is too restrictive for most problems though.

To capture the variability in a more agnostic way, [122] introduced an additive perturbation in the modulating function $f_i(\mathbf{s}^j) = \mathbf{s}^j + \boldsymbol{\delta}_j^i$ inspired by the Taylor expression. The resulting model, that does not require explicit prior knowledge of the variability, can be written as:

$$\begin{aligned} \mathbf{Y} &= \mathbf{AS} + \left[(\mathbf{a}^1\boldsymbol{\delta}^{(1)})^T|\ldots|(\mathbf{a}^P\boldsymbol{\delta}^{(P)})^T\right]^T + \mathbf{N} \\ &= \mathbf{AS} + \boldsymbol{\Delta} + \mathbf{N}, \end{aligned} \tag{III.72}$$

where for each pixel i, $\boldsymbol{\delta}^{(i)} \in \mathbb{R}^{J\times T}$ is the matrix containing the perturbations associated this pixel and $\boldsymbol{\Delta} = \left[(\mathbf{a}^1\boldsymbol{\delta}^{(1)})^T|\ldots|(\mathbf{a}^P\boldsymbol{\delta}^{(P)})^T\right]^T$.

The optimization problem to solve is then:

$$\underset{\mathbf{A}\in\mathbb{R}^{P\times J},\boldsymbol{\Delta}\in\mathbb{R}^{P\times J\times T},\mathbf{S}\in\mathbb{R}^{J\times T}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y} - \mathbf{AS} - \boldsymbol{\Delta}||_F^2 + \mathcal{J}(\mathbf{A}) + \mathcal{G}(\mathbf{S}) + \mathcal{H}(\boldsymbol{\Delta}), \tag{III.73}$$

where $\mathcal{H}$ is an appropriate regularization applied to the perturbation matrix. For instance, in [122] and [127] they penalized the total energy of the matrix $\mathcal{H} = ||.||_F^2$.

This method is data-agnostic, meaning that it can be applied to any kind of data. However, if we understand the variability involved in our data, we can integrate that prior knowledge in the model in a more appropriate way to improve the unmixing.

Figure II.12 displays a random selection of kinetics belonging to a single tumor. We can observe that the kinetics differ in both peaks heights and locations in a more complex fashion, which can be explained by a time shift [121], as was presented in Chapter II. We will next see how we dealt with that particular variability in our data.

III.5.2 Proposed approach : Pre-alignment of kinetics

In this part, based on the observed variability in our data Figure II.12, we proposed to model the variability f_i as pixel-level elastic time warping functions $\boldsymbol{\gamma}_i^{-1}$. These functions $\boldsymbol{\gamma}_i^{-1}$ are diffeomorphic deformations of the domain $[0, T]$. In order to improve our unmixing, we then have to find the optimal $\boldsymbol{\gamma}_i^{-1}$ for each pixel i.

Handling that in the optimization problem would require to find an analytical formulation of the projection into the space Γ of diffeomophisms of $[0, T]$. We therefore instead decided to use an alignment algorithm to register the pure pixels (*i.e.* pixels belonging to only one tumor) to the kinetics of the tumor they belong before the unmixing algorithm as a pre-processing step.

A first step is then to find pure pixels. The data we observe consists essentially in tumors located next to each other. Some pixels at the borders of the tumor can belong to several tumors but the others can reasonably be assumed to be pure pixels.

We thus propose to make groups of pure-pixels on the accumulation image with an unsupervised algorithm. Indeed, if we keep only the groups composed of more than 30 pixels (the cardinality threshold commonly used to consider a relevant distribution in statistics), the average kinetics of each group would be close to the kinetic of the associated tumor since there would only be pure-pixels in these groups. The algorithm can then register the kinetics of each pixel in each group to the average kinetics of the group. Formally, we have to find an algorithm that extracts $\Omega_1, \Omega_2, \ldots, \Omega_G$ from the BLI image I. Ω_g is the mask corresponding to the group g of pure-pixels. These groups will be called *superpixels* in the following.

Once we have those superpixels $(\Omega_g)_{g \geq G}$, for all g, for all i in Ω_g, we aim at finding $\boldsymbol{\gamma}_i$ as the solution of the alignment algorithm (described below) between $\mathbf{y}^i$ and $\boldsymbol{\mu}_g$ where $\boldsymbol{\mu}_g$ is the average kinetics of the group g.

The framework we propose tackles 2 issues : finding the pure pixels, and then the kinetics registration algorithm.

III.5.2.1 Pure Pixels Extraction

There are several superpixels algorithms wildly used in the litterature (SLIC [206] or Felzenszwalb's method [207]) but they only use the intensity as information and result in whether too precise or too rough segmentations. Indeed, it can be tricky to get superpixels composed of only pure-pixels belonging to the same tumor as can be seen in Figure III.18, where the segmentations of the accumulation image end up with superpixels composed of pixels belonging to both tumors.

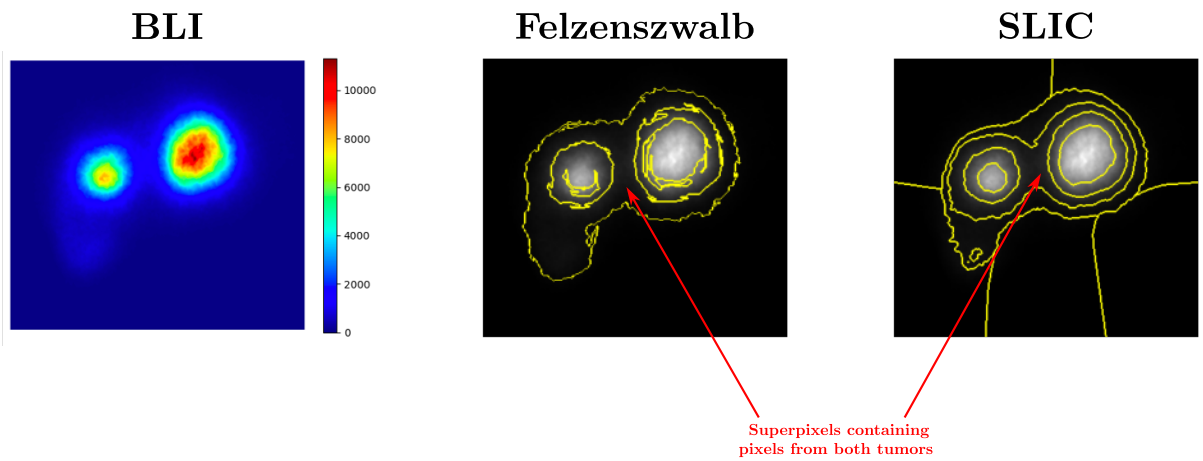

Figure III.18: Superpixels segmentation with SLIC and Felzenszwalb's method on a BLI. Some superpixels contain pixels belonging to the 2 tumors.

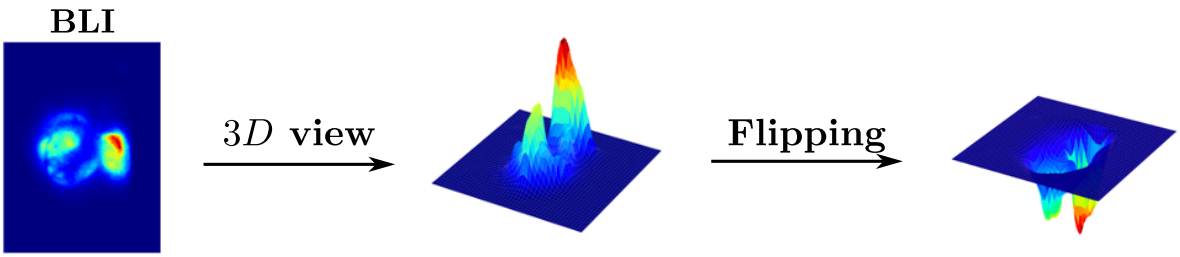

Figure III.19: Representation of a BLI by a juxtaposition of basins, which justifies the use of the watershed algorithm.

However, it can be noted that our images look like a juxtaposition of "mountains", where the pixels belonging to a specific mountain belong to the same tumor. What's interesting is that, as illustrated in Figure III.21, a single tumor can be represented by several mountains (not just one) as detailed in Chapter II. If we take the opposite of our images, these mountains become basins (Figure III.19) and it becomes logical to use the watershed algorithm [208] to delineate these basins.

The first step of this algorithm, illustrated in Figure III.20 is to find the bottom of the basins, hence the peaks of the image. To do that, we compare the original image with its convolution with a Maximum Filter of size N_m, as illustrated in Figure III.21. We then assign a different label to each peak, located where the maximum image is equal to the original image. We then consider the opposite of the accumulation image and initialize a heap of pixels with all the found bottoms, with their corresponding values. Then, the algorithm is the following :

While the heap is not empty:

- Select the lowest item i in the heap
- For all its neighbors n:
  - if n is already labeled, continue
  - else, label it with the same label as i and add it to the heap

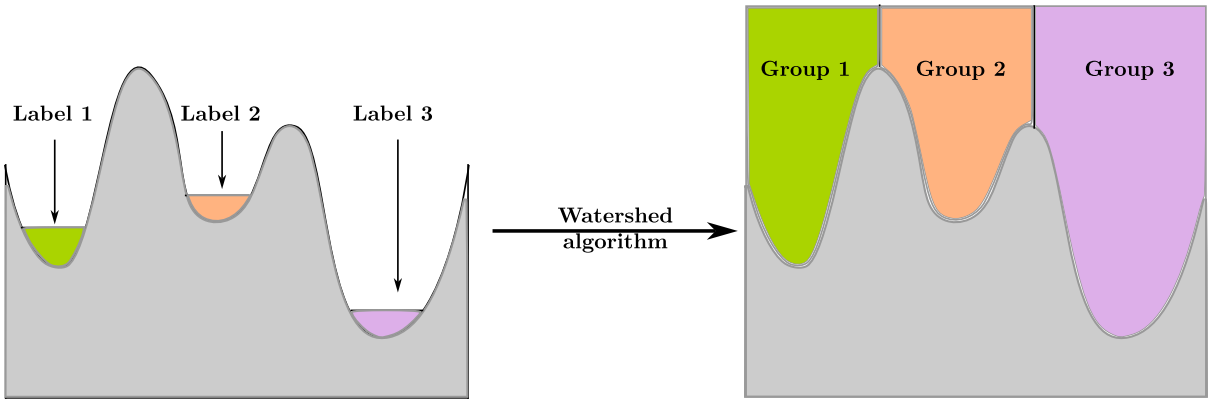

Figure III.20: Illustration of the watershed algorithm. It groups together points belonging to the same basins.

After taking only the most intense pixels in each superpixel, in order to be sure that only pure pixels remain, this algorithm outputs the G masks $\Omega_1, \Omega_2, \ldots, \Omega_G$ described above with G the number of peaks found. In practice, for each Ω_g, we kept only the pixels whose intensity were bigger than 10% of the maximum value

In fact, this watershed algorithm is applied to the whole image $\mathbf{I}$ restricted to pixels whose intensity were greater than a threshold set at MAD($\mathbf{I}$) to keep only relevant pixels that make a difference in the algorithm.

III.5.2.2 Alignment algorithm

To register the kinetics, we used the algorithm of [209]. In this section we will then explain its principle and assets. To align the geometric features of $\mathbf{f}_1$ and $\mathbf{f}_2$, where $\mathbf{f}_1$ and $\mathbf{f}_2$ are 2 continuous functions of $[0, T]$, the algorithm seeks a diffeomorphism noted $\boldsymbol{\gamma}$ which acts as a time warping. Let's note Γ the space of orientation-preserving diffeomorphisms of $[0, T]$ to itself. The problem is the to find a good metric d to find $\text{argmin}_{\gamma \in \Gamma}\, d(\mathbf{f}_1, \mathbf{f}_2(\boldsymbol{\gamma}))$. In the following, we will call that the alignment of $\mathbf{f}_2$ to $\mathbf{f}_1$. It is then necessary to use a good metric d. This metric has to be:

- Invariant to a composition by a warping function $\boldsymbol{\gamma}$: $d(\mathbf{f}_1, \mathbf{f}_2) = d(\mathbf{f}_1 \circ \boldsymbol{\gamma}, \mathbf{f}_2 \circ \boldsymbol{\gamma})$
- Symmetric: $d(\mathbf{f}_1, \mathbf{f}_2) = d(\mathbf{f}_2, \mathbf{f}_1)$
- Respectful of the Triangle inequality: $d(\mathbf{f}_1, \mathbf{f}_3) \leq d(\mathbf{f}_1, \mathbf{f}_2) + d(\mathbf{f}_2, \mathbf{f}_3)$ for any $\mathbf{f}_3$

The invariance property assure that it is not possible to increase the similarity between $\mathbf{f}_1$ and $\mathbf{f}_2$ by warping them simultaneously to by the same warping function $\boldsymbol{\gamma}$ [210]. Finally, the Symmetry property and Triangle inequality are necessary to define a proper distance.

Moreover, the symmetry property makes sure that the alignment of $\mathbf{f}_1$ to $\mathbf{f}_2$ is the same as the alignment of $\mathbf{f}_2$ to $\mathbf{f}_1$ [210]:

$$inf_{\gamma \in \Gamma} d(\mathbf{f}_1, \mathbf{f}_2 \circ \boldsymbol{\gamma}) = inf_{\gamma \in \Gamma} d(\mathbf{f}_1 \circ \boldsymbol{\gamma}, \mathbf{f}_2) \tag{III.74}$$

Many warping algorithms use the regular ℓ_2-distance, but it does not satisfy important properties for functions registration [209]. A metric that satisfy all these properties is the Fisher-

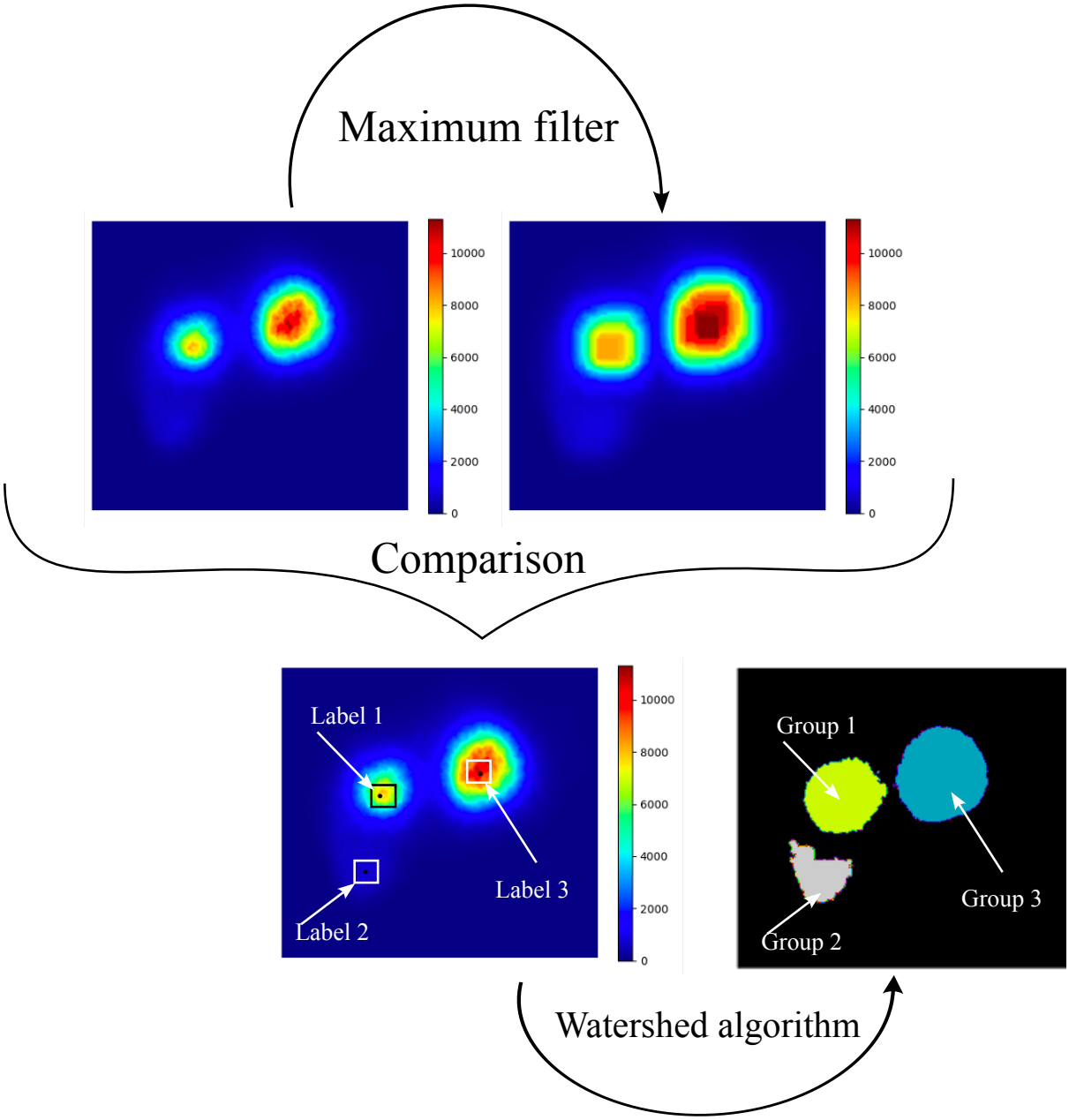

Figure III.21: Watershed Algorithm used to create the superpixels used for the alignment procedure

Rao Riemannian metric d_{FR} [211]. This metric is difficult to use in practice, but according to [209], under the Square Root Velocity Framework (SRVF) transform $\mathbf{q}$ of $\mathbf{f}$:

$$\mathbf{q}(t) = \frac{\mathbf{f}'(t)}{\sqrt{|\mathbf{f}'(t)|}}, \tag{III.75}$$

where $\mathbf{f}'$ is the derivative of $\mathbf{f}$, it is equivalent to the ℓ_2 norm:

$$d_{FR}(\mathbf{f}_1, \mathbf{f}_2) = ||\mathbf{q}_1 - \mathbf{q}_2||_2 \tag{III.76}$$

The optimization problem to align $\mathbf{f}_2$ to $\mathbf{f}_1$ is then reformulated as:

$$\underset{\gamma\in\Gamma}{\operatorname{argmin}}\, d_{FR}(\mathbf{f}_1, \mathbf{f}_2 \circ \boldsymbol{\gamma}) = \underset{\gamma\in\Gamma}{\operatorname{argmin}}\, ||\mathbf{q}_1 - (\mathbf{q}_2 \circ \boldsymbol{\gamma})\sqrt{\boldsymbol{\gamma}'}||_2 \tag{III.77}$$

The minimization of such an optimization problem is not straightforward without the explicit proximal operator of i_Γ. In [209] they then used a Dynamic Programming Algorithm to find an optimal solution.

In our context, we align all n pixels in Ω_g regions individually with the following iterative procedure:

- initialize the mean $\boldsymbol{\mu}$ as the average of the kinetics $\{\mathbf{q}_i\}_{i\in\Omega_g}$ where $\mathbf{q}_i$ is the SRVF transform of $\mathbf{y}^i$
- align each $\mathbf{q}_i$ to $\boldsymbol{\mu}$ by solving $\operatorname{argmin}_{\gamma_i\in\Gamma} ||\boldsymbol{\mu} - (\mathbf{q}_i \circ \boldsymbol{\gamma}_i)\sqrt{\boldsymbol{\gamma}_i'}||_2$ and set $\tilde{\mathbf{q}}_i = \mathbf{q}_i \circ \boldsymbol{\gamma}_i$. This minimization is solved using a Dynamic Programming Algorithm.
- Set $\tilde{\mathbf{y}}^i = \mathbf{y}^i \circ \boldsymbol{\gamma}_i$

Figure III.22 illustrates the effect of the alignment on kinetics from on BLI region with their corresponding time warping transforms $\boldsymbol{\gamma}_i$. This alignment pipeline, summarized in Figure III.23, outputs the matrix $\tilde{\mathbf{Y}}$ where the kinetics of all the pixels within each group Ω_g are aligned. The Linear Model defined in (III.11) becomes

$$\tilde{\mathbf{Y}} = \mathbf{AS} + \mathbf{N}, \tag{III.78}$$

That BSS problem is then solved on $\tilde{\mathbf{Y}}$ using the PALM algorithm section III.4. The complete pipeline is summarized in Algorithm 12.

Algorithm 12 Pseudo-code for automated BLI tumor separation with constrained NMF on aligned kinetics

Require: $\mathbf{Y} \geq 0$, $J > 0$, $N_{iter} > 0$, $\tau > 0$
  if Alignment **then**
    $\Omega_1, \ldots, \Omega_G \leftarrow$ Watershed Algorithm ($\mathbf{Y}$)
    for $g \leq G$ **do**
      for $i \in \Omega_g$ **do**
        $argmin_{\gamma_i\in\Gamma}||\boldsymbol{\mu} - (\mathbf{q}_i \circ \boldsymbol{\gamma}_i)\sqrt{\boldsymbol{\gamma}_i'}||_2$
        $\tilde{\mathbf{y}}^i = \mathbf{y}^i \circ \boldsymbol{\gamma}_i$
      end for
    end for
  end if
  Initialize $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$ from $\tilde{\mathbf{Y}}$
  $\mathbf{A}, \mathbf{S} = \text{PALM}(\tilde{\mathbf{Y}}, \mathbf{A}^{(0)}, \mathbf{S}^{(0)})$ Algorithm 11
  Return $\mathbf{A}, \mathbf{S}$

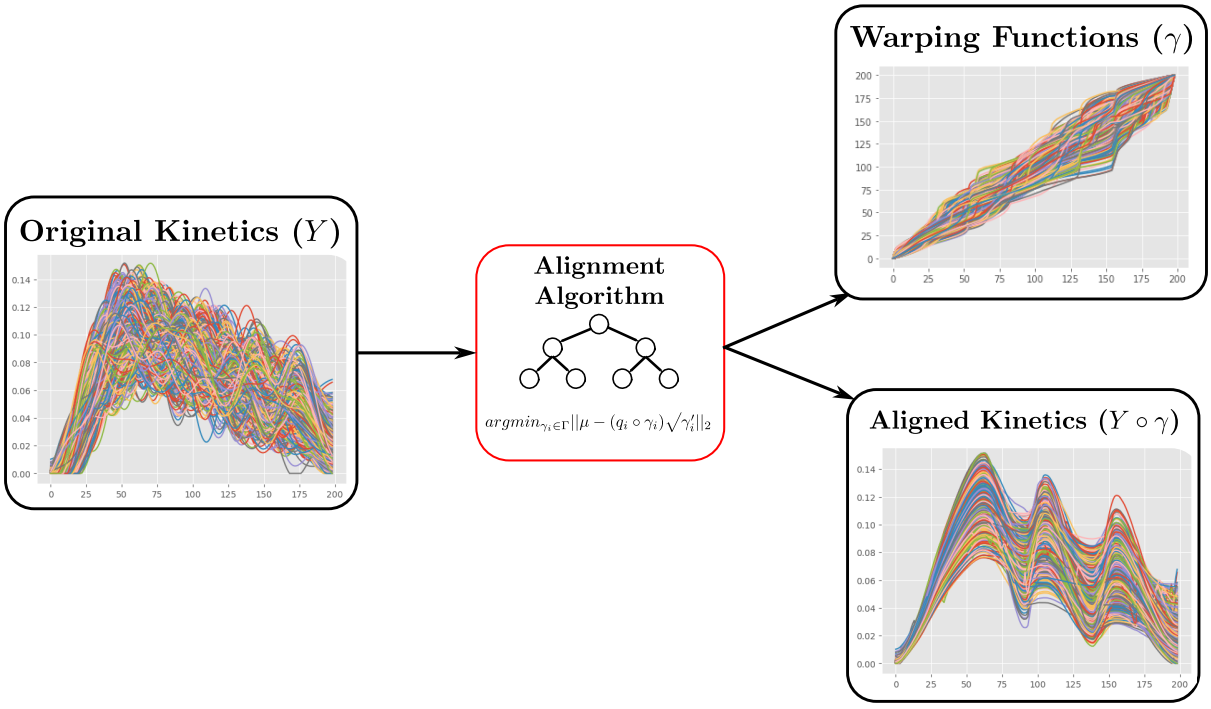

Figure III.22: Temporal alignment of kinetics. Inputs are original kinetics within a homogeneous region and outputs include the warping functions and the aligned kinetics.

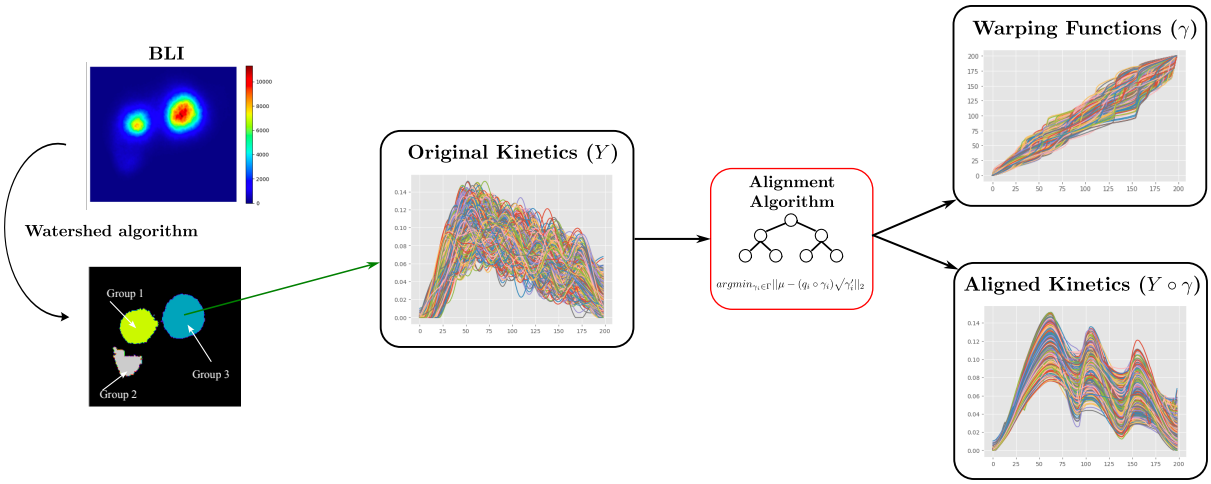

Figure III.23: Alignment procedure pipeline as a pre-processing for BLI unmixing. First step: make groups of pure-pixels. Second step: align kinetics inside each one of these groups.

III.5.3 Experiments & Results

We used the same hyperparameters as in section III.3 and section III.4, meaning an initialization of the matrices $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$ with NNDSVD [165]. The maximum number of iterations is set to 1000, Δ_k is the stability measure defined as the minimum cosine distance in the evolution of each $\mathbf{s}^j$ or $\mathbf{a}_j$ and $\tau = 1 - 4 \times 10^{-11}$ and we used the Haar Wavelet Transform. To be fair, we also tested the alignment pre-processing with the direct domain sparsity PALM algorithm Algorithm 10.

We reported the results of the algorithm on real data with short kinetics in Table III.11, on real data with long kinetics in Table III.12 and visual examples on 2 real BLIs in Figure III.24

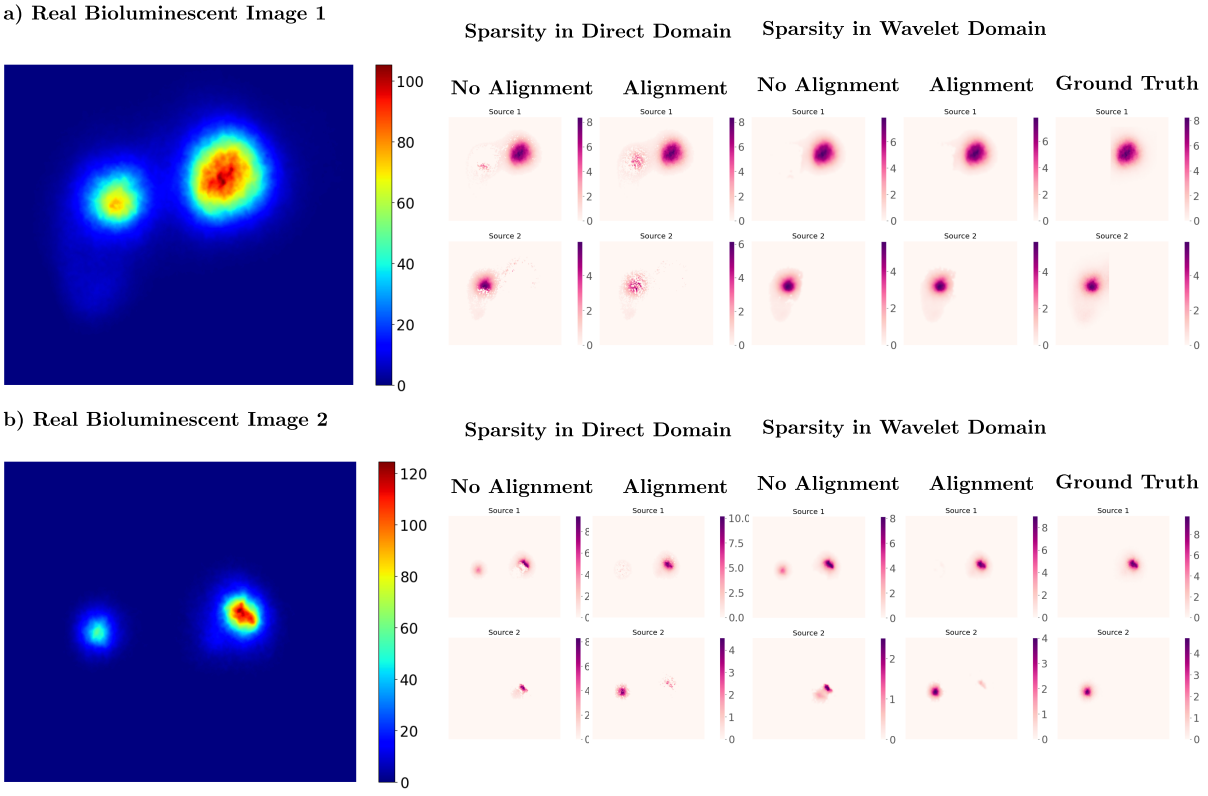

Figure III.24: Results of the different PALM algorithms illustrating their behavior on 2 examples of real BLI with and without the alignement pre-processing.

Table III.11: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on the real BLI dataset with short kinetics.

Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	0.58 ± 0.82	0.06 ± 0.15	0.32 ± 0.65	0.32 ± 0.65	0.84 ± 0.29
Direct	Yes	0.47 ± 0.59	$\mathbf{0.04 \pm 0.11}$	0.26 ± 0.47	$\mathbf{0.24 \pm 0.42}$	$\mathbf{0.92 \pm 0.12}$
Wavelet	No	0.48 ± 0.70	0.10 ± 0.19	0.29 ± 0.55	0.33 ± 0.57	0.78 ± 0.37
Wavelet	Yes	$\mathbf{0.39 \pm 0.45}$	0.09 ± 0.19	$\mathbf{0.24 \pm 0.38}$	0.26 ± 0.35	0.82 ± 0.38

Table III.12: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on the real BLI dataset with long kinetics.

Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	0.44 ± 0.59	$\mathbf{0.01 \pm 0.05}$	0.23 ± 0.47	$\mathbf{0.08 \pm 0.12}$	$\mathbf{0.96 \pm 0.14}$
Direct	Yes	0.38 ± 0.57	0.02 ± 0.09	0.20 ± 0.45	0.12 ± 0.18	0.96 ± 0.15
Wavelet	No	0.31 ± 0.37	0.05 ± 0.12	0.18 ± 0.31	0.14 ± 0.18	0.91 ± 0.22
Wavelet	Yes	$\mathbf{0.24 \pm 0.35}$	0.06 ± 0.20	$\mathbf{0.15 \pm 0.30}$	0.18 ± 0.25	0.91 ± 0.26

displaying typical behavior of the algorithm. This algorithm has also be tested on the synthetic datasets described in Chapter II (with the different kinds of variability). The associated results are presented in Table III.13 and Table III.14. Visual results on 3 synthetic BLI taken from the synthetic dataset are shown in Figure III.25.

Table III.13: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on synthetic data without time warping perturbations.

		Variability : None					Variability : Piece-wise Affine				
Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.03 \pm 0.08}$	$\mathbf{0.17 \pm 0.28}$	$\mathbf{0.22 \pm 0.26}$	$\mathbf{0.89 \pm 0.27}$	0.58 ± 0.49	$\mathbf{0.04 \pm 0.09}$	$\mathbf{0.31 \pm 0.45}$	$\mathbf{0.30 \pm 0.33}$	$\mathbf{0.83 \pm 0.31}$
Direct	Yes	0.39 ± 0.43	0.15 ± 0.16	0.22 ± 0.37	0.27 ± 0.32	0.85 ± 0.34	0.61 ± 0.56	0.05 ± 0.11	0.33 ± 0.50	0.35 ± 0.41	0.82 ± 0.36
Wavelet	No	0.39 ± 0.35	0.10 ± 0.21	0.25 ± 0.33	0.37 ± 0.41	0.78 ± 0.40	$\mathbf{0.48 \pm 0.34}$	0.17 ± 0.34	0.32 ± 0.37	0.37 ± 0.33	0.70 ± 0.45
Wavelet	Yes	0.45 ± 0.39	0.12 ± 0.23	0.29 ± 0.36	0.42 ± 0.43	0.75 ± 0.43	0.51 ± 0.38	0.15 ± 0.27	0.33 ± 0.37	0.40 ± 0.38	0.70 ± 0.44

Table III.14: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on synthetic data with time warping perturbations.

		Variability : Time Warping					Variability : Time warping + Piece-wise Affine				
Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	1.03 ± 0.76	0.29 ± 0.42	0.66 ± 0.72	0.68 ± 1.1	0.54 ± 0.49	1.05 ± 0.77	0.32 ± 0.45	0.69 ± 0.72	0.66 ± 0.92	0.49 ± 0.51
Direct	Yes	0.53 ± 0.44	$\mathbf{0.14 \pm 0.23}$	$\mathbf{0.33 \pm 0.40}$	$\mathbf{0.38 \pm 0.35}$	$\mathbf{0.83 \pm 0.23}$	0.70 ± 0.52	$\mathbf{0.17 \pm 0.27}$	0.43 ± 0.50	$\mathbf{0.41 \pm 0.39}$	$\mathbf{0.72 \pm 0.36}$
Wavelet	No	0.77 ± 0.67	0.29 ± 0.38	0.53 ± 0.59	0.66 ± 0.99	0.52 ± 0.53	0.77 ± 0.65	0.34 ± 0.44	0.56 ± 0.59	0.65 ± 0.88	0.49 ± 0.53
Wavelet	Yes	$\mathbf{0.51 \pm 0.37}$	0.18 ± 0.24	0.35 ± 0.35	0.46 ± 0.42	0.70 ± 0.37	$\mathbf{0.57 \pm 0.42}$	0.19 ± 0.24	$\mathbf{0.38 \pm 0.39}$	0.42 ± 0.34	0.67 ± 0.41

III.5.4 Discussion

Quantitative results on synthetic data (Table III.13 and Table III.14) lead to the following observations:

- Adding perturbations to the kinetics downgrades the performances of every algorithm. This phenomenon seems natural as the linear model becomes less relevant with such data. Our baseline algorithm [167] is the least robust to degraded kinetics.

- Adding the alignment process (subsubsection III.5.2.1 and subsubsection III.5.2.2) has no positive influence on source separation (and can even be detrimental and downgrade a little bit both the spatial and kinetics separation performances) when there is no time warping perturbations involved (Figure III.25 a-c) but leads to better separations with time warping perturbations (Figure III.25 b-d).

- Source sparsity in the wavelet domain performs always better than sparsity in the direct domain in terms of spatial separation of sources when there is any kind of piece-wise affine perturbation involved (Figure III.25 b-c-d).

Finally, since the synthetic data that should better model the real data are the ones degraded with both time warping and piece-wise affine gain, enforcing the sparsity in the transformed domain and adding the alignment pre-processing should better separate tumors in real BLI.

Results on real data (Table III.11 and Table III.12) confirm that hypothesis. We illustrate those results visually in Figure III.24 on several examples. Indeed, without alignment, tumors are sometimes not well separated and the spatial constraint induced by the sparsity in the wavelet domain is not able to compensate. With temporal alignment, only isolated erroneous pixels remain that can be corrected by the sparsity in wavelet domain. The separation is however not perfect, probably due to the hyperparameters and to other variability than the time-warping one involved. [122] and [127] uses the PLMM to take into account the piece-wise affine variability in the algorithm, which could be an interesting problem to pursue. It may also be due to the fact that we just use a pre-processing to deal with kinetics variability. Handling them directly in the optimization process could provide more precise separations.

Table III.15: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on synthetic data containing 2 sources without time warping perturbations.

		Variability : None					Variability : Piece-wise Affine				
Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	$\mathbf{0.12 \pm 0.27}$	0.02 ± 0.08	$\mathbf{0.07 \pm 0.21}$	$\mathbf{0.07 \pm 0.16}$	0.90 ± 0.32	0.30 ± 0.39	$\mathbf{0.01 \pm 0.06}$	0.15 ± 0.31	$\mathbf{0.10 \pm 0.20}$	$\mathbf{0.85 \pm 0.39}$
Direct	Yes	0.18 ± 0.44	0.02 ± 0.10	0.1 ± 0.33	0.11 ± 0.39	0.89 ± 0.33	0.38 ± 0.62	0.03 ± 0.11	0.21 ± 0.48	0.16 ± 0.42	0.82 ± 0.41
Wavelet	No	0.15 ± 0.27	$\mathbf{0.01 \pm 0.05}$	0.08 ± 0.21	0.11 ± 0.19	$\mathbf{0.90 \pm 0.31}$	$\mathbf{0.23 \pm 0.33}$	0.04 ± 0.1	$\mathbf{0.14 \pm 0.26}$	0.13 ± 0.2	0.78 ± 0.47
Wavelet	Yes	0.21 ± 0.40	0.02 ± 0.09	0.11 ± 0.30	0.15 ± 0.36	0.89 ± 0.33	0.34 ± 0.5	0.05 ± 0.09	0.19 ± 0.39	0.22 ± 0.39	0.76 ± 0.47

Table III.16: Quantitative Results of the proposed PALM algorithms with sparsity constraints applied on direct and wavelet domains, with and without the alignment pre-processing, on synthetic data containing 2 sources with time warping perturbations.

		Variability : Time Warping					Variability : Time warping + Piece-wise Affine				
Sparsity Domain	Alignment	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$	$NMSE_A$	$NMSE_S$	$NMSE_{tot}$	$NMSE$	$corr$
Direct	No	0.85 ± 1.47	0.19 ± 0.38	0.52 ± 1.12	0.57 ± 1.58	0.73 ± 0.39	1.01 ± 1.52	0.18 ± 0.38	0.6 ± 1.18	0.53 ± 1.53	0.65 ± 0.46
Direct	Yes	0.70 ± 1.45	0.18 ± 0.38	0.44 ± 1.09	0.99 ± 3.15	0.71 ± 0.40	0.53 ± 0.63	$\mathbf{0.10 \pm 0.15}$	0.32 ± 0.51	$\mathbf{0.29 \pm 0.44}$	$\mathbf{0.75 \pm 0.38}$
Wavelet	No	0.33 ± 0.49	0.09 ± 0.14	0.21 ± 0.38	$\mathbf{0.22 \pm 0.30}$	0.83 ± 0.28	0.80 ± 1.48	0.22 ± 0.39	0.51 ± 1.12	0.99 ± 3.05	0.57 ± 0.52
Wavelet	Yes	$\mathbf{0.30 \pm 0.50}$	$\mathbf{0.08 \pm 0.10}$	$\mathbf{0.19 \pm 0.38}$	0.40 ± 0.99	$\mathbf{0.84 \pm 0.28}$	$\mathbf{0.42 \pm 0.59}$	0.11 ± 0.13	$\mathbf{0.27 \pm 0.46}$	0.44 ± 1.01	0.72 ± 0.45

In addition, it appears that the separation performance is lower with synthetic data than with real BLI images. To explain that, we should first note that having more sources degrade the separation performance. Indeed, we generated a synthetic dataset by enforcing $J = 2$ sources in each images and tested our algorithms on it. The results are presented in Table III.15 and Table III.16. Except for a few cases, we can observe the separation metrics are better than with a random number of sources greater than 2. They are however still not as good as the ones from the real dataset. It can have several causes :

- The variability can be stronger in synthetic data.
- The kinetics synthetically generated are maybe too close to each other, resulting to a stronger correlation between the synthetic kinetics than between the real ones and thus to more difficult cases for separation.

Morover, as usual, the metrics $NMSE$ and $corr$ are less precise to assess of the quality of the separation and so their respective potential decrease should not be taken into account. We can finally note that the final algorithm seems to downgrade a little bit the kinetics separation performances. It can be explained because with the proposed framework we only align the kinetics of the pixels we know for sure to be pure. As is illustrated in Figure III.21, the kinetics of a lot of pixels are not aligned (the ones on the borders of the superpixels) which corrupts the obtained source kinetics. This would be an intriguing issue to tacke in future works. A possible idea in that perspective would be to integrate the alignment in the optimization pipeline to cover all pixels.

Overall, this alignment pre-processing step resulted in enhanced separations of tumors but could probably be improved by being a part of the optimization process. To overcome the limitation of not knowing the proximal operator of the indicator function of the space of diffeomorphisms, we could restrict ourselves to a small family of diffeomorphisms such as polynoms with strictly positive coefficients. Another interesting idea would be to use the Wasserstein barycenter instead of the usual average $\boldsymbol{\mu}$ to calibrate the alignment.

These results will be part of an article in preparation:

Dereure, E., Kervazo, C., Seguin, J., Garofalakis, A., Mignet, N., Angelini, E., & Olivo-Marin, J.-C. . Wavelet transform based non-negative matrix factorization for bioluminescence imaging with warping variability.

Further validation of these results represents a valuable perspective of our work, for example by testing the robustness of our model to the magnitude of the perturbations involved in synthetic data generation (see Chapter II) or by comparing its performance with some of the methods used in Hyperspectral Imaging Unmixing [154]. It could also be interesting to precisely model the kinetics $\mathbf{S}$ in the optimization procedure with the pharmacokinetics model described in Chapter II.

The performances of our source separation algorithms could benefit from a deeper understanding of kinetics variability, by studying the influence of physiological constraints such as respiratory movements [116] or tumor vascularization, and technical constraints such as the impact of pre-processing or the response of the imaging device.

In this chapter, we detailed the use of NMF to separate the photon flux coming from the different tumors within the same BLI. These kinetics have to be quantified to monitor tumor status during *in vivo* experiments (see Chapter I). In the next chapter, we will then describe how we used a pharmacokinetics model to improve the quantification of BLI.

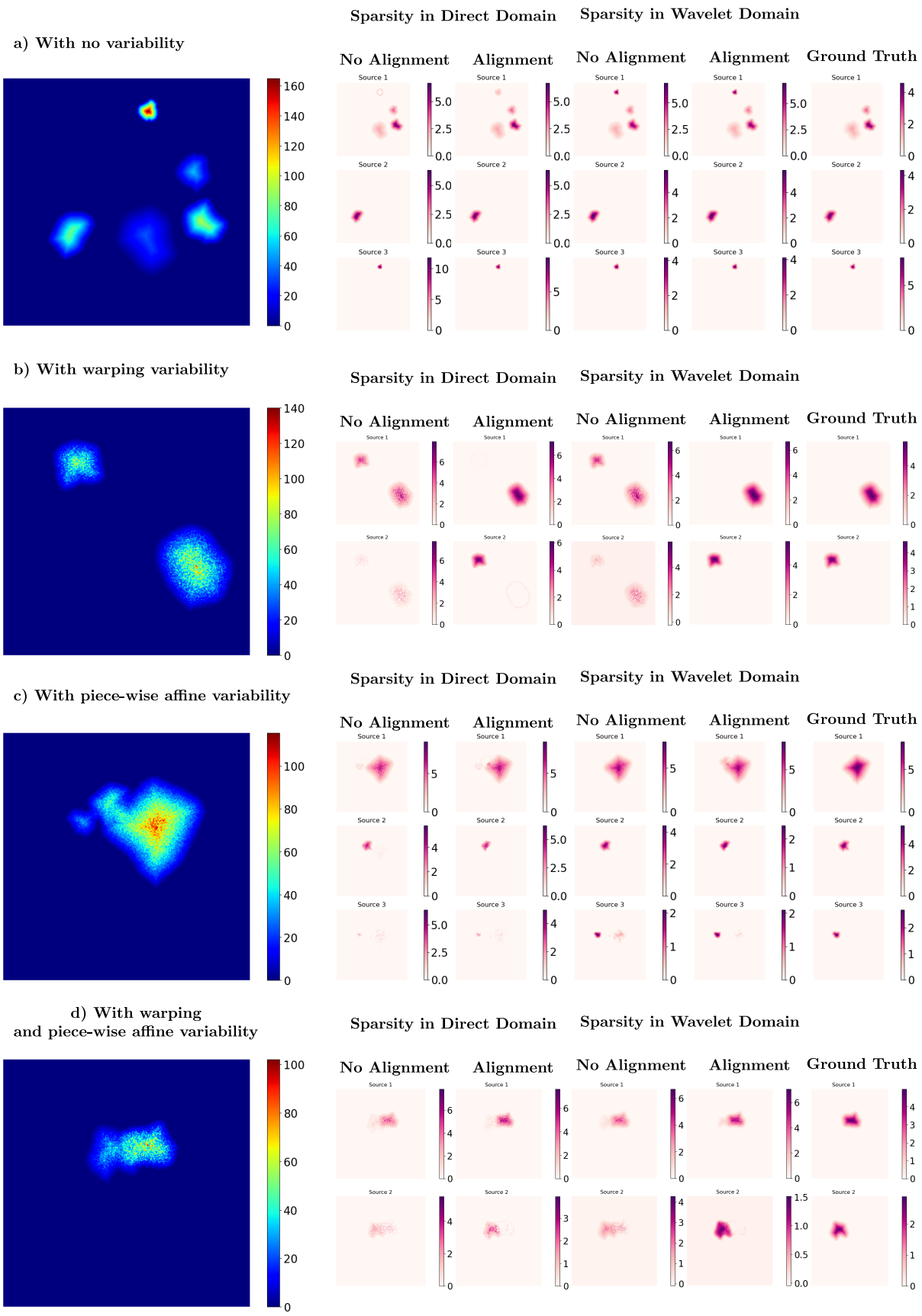

Figure III.25: Results of the different PALM algorithms, with and without the alignment pre-processing, illustrating their behavior for the different scenarios of synthetic data in terms of variability: None, time warping (subsubsection II.3.2.2), piece-wise affine (subsubsection II.3.2.1) and time warping + piece-wise affine.

Chapter IV

Quantitative analysis of BLI with pharmacokinetics model

IV.1 Challenges in Bioluminescence Imaging quantification

IV.1.1 Introduction on quantification challenges

Quantifying the state of the tumor using BLI remains an active field of research [57,72]. Indeed, the current quantification methods do not take into account hypoxia, necrosis or photon scattering through tissues when estimating the tumor volume. This results in a poor correlation between the quantification of the BL signal and the real volume measured with other modalities [106], especially in the last few days of experiments before the sacrifice of the animals as illustrated in Figure I.21.

On top of that, we also showed a lack of repeatability in the BLI experiments, causing a loss of confidence in the interpretation of the current methods of BLI quantification. This is especially the case for the methods using the Average Photon Flux, as illustrated in Figure I.26 where we can see a strong variability in the BL signals at different time stamps of the same day for 2 tumors. The work of [100] seemed promising to quantify the kinetics in a more precise way and could help to better understand the kinetics of the different tumors. They used a multi-compartment pharmacokinetic model in which they studied the flow of luciferin in the tissues and quantified the tumor with exchange rates of luciferin between the tumor and the blood. This work focused on modelling the bioluminescence reaction in the context of an IP injection of luciferin by making the assumptions that the rate of photon emitted from the tumor is proportional to the amount of substrate in the tumor area and that the reaction follows a Michaelis-Menten kinetics. Their work has been deeper studied by Dai et al. in [212] and [84]. In [84] the authors used a 3 compartments model to take into account the metabolism of the luciferin but [100] argued that the results are nearly identical with a 2 compartments model. The work in [100] seems to be a good trade-off complexity-wise, it is therefore the model we considered in this work. We describe this model and the associated challenges in the following part, in particular how we propose to enhance the relevance of the estimated exchange rates.

IV.1.2 Presentation of pharmacokinetics model and associated challenges

Sim et al. [100] described the bioluminescent reaction in the context of an IP administration of luciferin in a mouse by the model illustrated in Figure IV.1.

The equations describing the balance of luciferin between the different compartments are

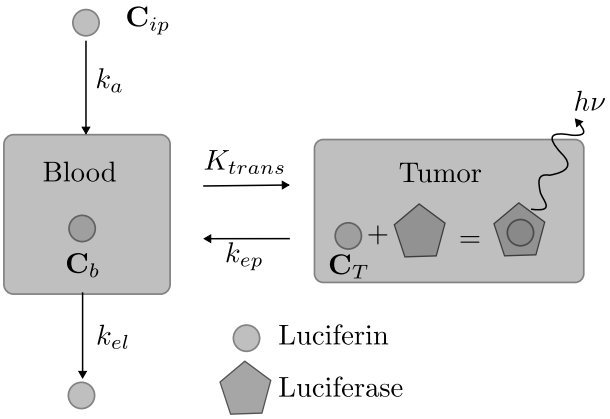

Figure IV.1: BLI Pharmacokinetics model described in [100] that models the emission of photons by the tumor from the interactions luciferin-luciferase.

the following:

$$\begin{aligned}
\frac{d\mathbf{C}_{ip}}{dt} &= -k_a\mathbf{C}_{ip} \\
\frac{d\mathbf{C}_b}{dt} &= -(K_{trans} + k_{el})\mathbf{C}_b + k_{ep}\mathbf{C}_T + k_a\mathbf{C}_{ip} \\
\frac{d\mathbf{C}_T}{dt} &= K_{trans}\mathbf{C}_b - k_{ep}\mathbf{C}_T,
\end{aligned} \tag{IV.1}$$

where $\mathbf{C}_{ip}, \mathbf{C}_b$ and $\mathbf{C}_T$ are the luciferin concentration respectively in peritoneum, blood and tumor. k_a is the rate constant describing the absorption of luciferin from the peritoneum into the blood and k_{el} is the elimination rate constant from the blood. K_{trans} and k_{ep} are the rate constant describing respectively the transfer of luciferin from the blood to the tumor and from the tumor to the blood.

For the sake of simplicity, in that part, we note $\mathbf{Y} \in \mathbb{R}^{M\times N\times T}$ the BLI acquired by the device described earlier in a hand-drawn ROI around a specific isolated tumor, where $\mathbf{Y}(x, y, t)$ is the number of photons acquired by the pixel (x, y) at the frame t. Let's note $\mathbf{Y}_T(t)$ the total photon flux of the tumor, then:

$$\mathbf{Y}_T(t) = \sum_{x,y} \mathbf{Y}(x, y, t) \tag{IV.2}$$

If we are in the context of a mouse bearing several tumors, $\mathbf{Y}_T(t)$ can be obtained thanks to a BSS algorithm.

This model makes the assumption that, as each cell expresses luciferase, the concentration of luciferase in a tumor remains constant over the experiment. It also assumes that the level of other reagents, such as oxygen or ATP, is sufficient to enable the enzymatic reaction of luciferase with luciferin in the tumor which should, according to [100], follow the Michaelis-Menten kinetics. The authors of [100] state that since they are in regime where the substrate concentration is much less than the Michaelis constant K_m (which depends on the reaction) then:

$$\nu \mathbf{Y}_t(t) = \frac{d\mathbf{P}}{dt} = \mu \mathbf{C}_t, \tag{IV.3}$$

where μ is the conversion factor that relates the emitted photon flux $\mathbf{P}$ to the concentration of luciferin and ν is the parameter encoding the signal loss through tissues due to photon scattering and absorption, described more precisely in Chapter I, under a linear hypothesis.

We first determined in what extent we can apply this model to our data. Even though it is a strong assumption, we also assumed that all the reagents of the reaction but the luciferin are present. We know this may not always be true since we have sometimes necrotic or hypoxic tumors but we did not fine-tune the model to take into account the level of others reagents.

First, since in our case the enzyme concentration in the tumor is much less than the concentration of substrate, we have [213]:

$$\frac{d\mathbf{P}}{dt} = \frac{V_{max}\mathbf{C}_t}{K_m + \mathbf{C}_t} = \frac{k_{cat}\mathbf{C}_e(0)\mathbf{C}_t}{K_m + \mathbf{C}_t}, \tag{IV.4}$$

where V_{max} is the limiting rate of the enzymatic reaction, meaning the rate approached by the system at saturating substrate concentration for a given enzyme concentration, $\mathbf{C}_e(0)$ is the initial concentration of enzyme and k_{cat} is called the catalytic rate constant which is the rate responsible of the formation of the product and the enzyme from the complex enzyme-substrate [214].

Then, according to [215, 216], we have $K_m \approx 1 mmol/L$. Since the molecular weight of luciferin is $280 mg/mmol$ [217], we have $K_m \approx 0.28 mg/mL$.

In the experiments of the present thesis the amount of injected luciferin was 2 mg per mouse via intraperitoneal injection. After injection, the mass of luciferin in blood is about 5% of the injected mass per mL [218], so about $0.1 mg/mL$. Since the tumors are usually badly vascularized tissues according to the estimations of [100], $\mathbf{C}_t << \mathbf{C}_b$ and therefore $\mathbf{C}_t << K_m$, meaning that the substrate concentration in tumors is much less than the Michaelis constant.

We thus have the following:

$$\frac{d\mathbf{P}}{dt} \approx \frac{k_{cat}\mathbf{C}_e(0)\mathbf{C}_t}{K_m}, \tag{IV.5}$$

which justifies the linear model (IV.3).

The work of [100] only relies on the BL signal of the tumors to fit the parameters k_a, k_{el}, K_{trans} and k_{ep}. They use K_{trans} and k_{ep} afterwards to quantify the volume of the tumor thanks to a loglinear regression that yields promising results. However, it is a widely known issue in others modalities described by PK models that it is necessary to have the concentration of agent in the blood, usually called Arterial Input Function, to quantify accurately the parameters K_{trans} and k_{ep} [115]. This issue is not tackled by [84] either, making the estimation of K_{trans} and k_{ep} not robust. Therefore, in this work, we aim at improving the estimation of these PK parameters.

In the literature, there exists different methods to estimate the AIF in other imaging modalities such as DCE-MRI or PET scan. A first approach consist in sampling blood directly from

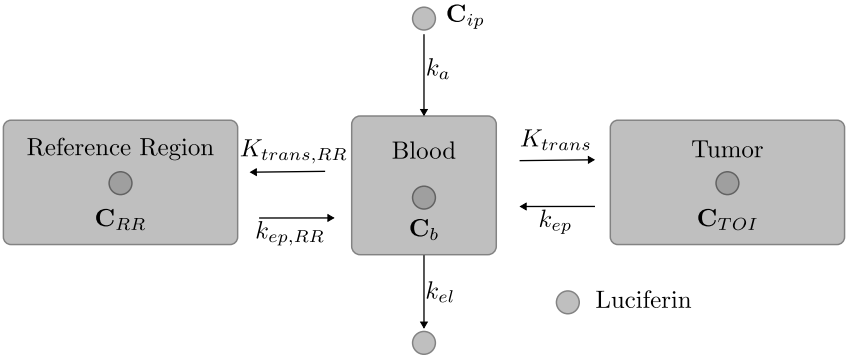

Figure IV.2: Pharmacokinetics model illustrating the exchange of a tracer between the blood and a Tissue of Interest (TOI) and the blood and a Reference Region (RR).

the subject during the imaging process for later analysis [132, 219–221]. This approach is not possible for us because the imaging device is a closed box isolated from the exterior environment. A second type of approaches consists in considering the AIF similar for all subjects, up to a scaling factor [222–225]. Typically, the AIF is measured as a first step via blood samples in a small cohort of subject and the resulting average [226] is used as a template AIF in the estimation of pK parameters [113]. The study of averaging the kinetics of luciferin in blood has also been made in [218] but it is very limited because it does not take into account the possible huge variability between the shape of the AIF of different subjects [222]. A third type of approaches estimates the AIF directly from the imaging data if the blood emits a signal [227,228]. The blood signal can be extracted from the image manually by hand drawing a ROI around an artery [229] or with automatic methods such as artery segmentation [230] or BSS [231]. In the context of BLI, this is not possible since the blood does not contain luciferase therefore it cannot emit any signal. A last type of method frees itself from knowing the AIF by taking the signal coming from another region in the body, called *Reference Region* and estimates PK parameters relative to that Reference Region tissues [115]. In this work, we tried to apply this method to BLI. Hence, in the following part we present how this approach could be adapted to the BLI context, how it works and what information can be learned from it.

IV.2 Proposed approach : Quantification with a Reference Region

IV.2.1 General theory of the model

The Reference Region (RR) model calibrates the signal from the tissue of interest to the signal coming from another region of the subject body, usually a muscle [115, 232]. This method has been vastly used in DCE-MRI [233, 234], PET scan [235, 236] or Dynamic Constrast-Enhanced Ultrasound (DCE-US) [237].

Let's then consider this compartmental model illustrated in Figure IV.2, where we can have access to both the information of $\mathbf{C}_{TOI}$, the concentration of the associated tracer in the Tissue of Interest (TOI) and $\mathbf{C}_{RR}$ the concentration of tracer in the Reference Region.

According to [115], we have:

$$\begin{aligned}\frac{d\mathbf{C}_{TOI}}{dt} &= K_{trans}\mathbf{C}_b - k_{ep}\mathbf{C}_{TOI}\\ \frac{d\mathbf{C}_{RR}}{dt} &= K_{trans,RR}\mathbf{C}_b - k_{ep,RR}\mathbf{C}_{RR}\end{aligned} \tag{IV.6}$$

The objective is to write $\mathbf{C}_{TOI}$ as a function of $\mathbf{C}_{RR}$. Then, we first combine these equations to remove $\mathbf{C}_b$ [115]:

$$\frac{d\mathbf{C}_{TOI}}{dt} + k_{ep}\mathbf{C}_{TOI}(t) = \frac{K_{trans}}{K_{trans,RR}}\frac{d\mathbf{C}_{RR}}{dt} + \frac{K_{trans}k_{ep,RR}}{K_{trans,RR}}\mathbf{C}_{RR}(t) \tag{IV.7}$$

Now, let's consider $\mathbf{I} = e^{k_{ep}t}$ as in [115].

We have:

$$\frac{d(\mathbf{C}_{TOI}\mathbf{I})}{dt} = \frac{d\mathbf{C}_{TOI}}{dt}\mathbf{I} + k_{ep}VC_{TOI}(t)I = \mathbf{I}\left(\frac{d\mathbf{C}_{TOI}}{dt} + k_{ep}\mathbf{C}_{TOI}(t)\right), \tag{IV.8}$$

which is equal to $\mathbf{I}$ multiplied by the left member of (IV.7) [115]. So,

$$\frac{d(\mathbf{C}_{TOI}\mathbf{I})}{dt} = \mathbf{I}\left(\frac{K_{trans}}{K_{trans,RR}}\frac{d\mathbf{C}_{RR}}{dt} + \frac{K_{trans}k_{ep,RR}}{K_{trans,RR}}\mathbf{C}_{RR}(t)\right) \tag{IV.9}$$

By integrating $\mathbf{I}\frac{K_{trans}}{K_{trans,RR}}\frac{d\mathbf{C}_{RR}}{dt}$ by parts with $\mathbf{C}_{RR}(0) = 0$, since we are supposed to inject the tracer at $t = 0$, we obtain [115]:

$$\mathbf{C}_{TOI}\mathbf{I}(t) = \frac{K_{trans}}{K_{trans,RR}}\left(\mathbf{I}(t)\mathbf{C}_{RR}(t) - k_{ep}\int_0^t(\mathbf{I}(\tau))\mathbf{C}_{RR}(\tau)d\tau + k_{ep,RR}\int_0^t(\mathbf{I}(\tau))\mathbf{C}_{RR}(\tau)d\tau\right) \tag{IV.10}$$

By dividing by $\mathbf{I}(t) = e^{k_{ep}t}$, we end up with the Nonlinear Reference Region Model (NRRM) [115, 120]:

$$\mathbf{C}_{TOI}(t) = \frac{K_{trans}}{K_{trans,RR}}\left(\mathbf{C}_{RR}(t) + (k_{ep,RR} - k_{ep})\int_0^t \mathbf{C}_{RR}(\tau)e^{-k_{ep}(t-\tau)}d\tau\right) \tag{IV.11}$$

Alternatively, we can also integrate (IV.7), and then obtain the Linear Reference Region Model (LRRM) [120]:

$$\mathbf{C}_{TOI}(t) = \frac{K_{trans}}{K_{trans,RR}}\left(\mathbf{C}_{RR}(t) + k_{ep,RR}\int_0^t \mathbf{C}_{RR}(\tau)d\tau\right) - k_{ep}\int_0^t \mathbf{C}_{TOI}(\tau)d\tau \tag{IV.12}$$

With both models, we can fit the relative parameters $R_{trans} = \frac{K_{trans}}{K_{trans,RR}}$ and k_{ep} to the available data. The difference between them is that the NRRM uses nonlinear least squares to

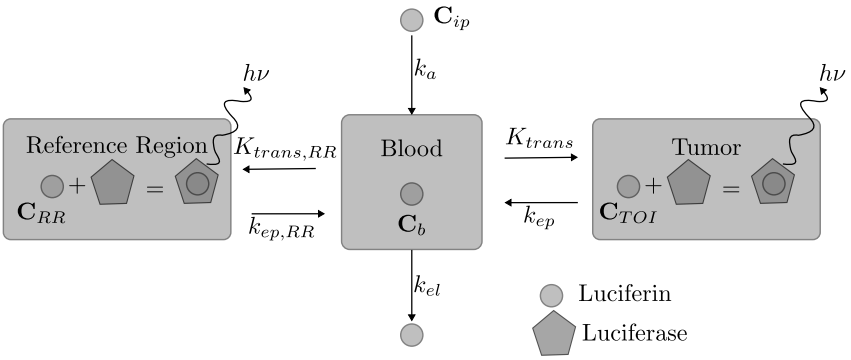

Figure IV.3: Pharmacokinetics model illustrating the exchange of luciferin between the blood and a tumor and the blood and a Reference Region (RR) in the presence of luciferase and the associated emissions of photons.

obtain the parameters while we can use a regular linear least squares with the LRRM [120]. According to [238], the parameters estimated with the LRRM contain more valuable quantitative information than the ones from the NRRM. In the following, we will see how we applied the LRRM to BLI data.

IV.2.2 Application to BLI

Contrary to subsection IV.2.1, in BLI we don't exactly have access to the concentration of luciferin but to the emitted photon flux.

In order to obtain a signal from a reference muscle during a BLI acquisition, an internal control was used. It consists of an injection of $3\mu g$ of recombinant luciferase in $30\mu L$ of PBS into the right cranial tibial muscle of each mouse. For each acquisition, tumor volume was measured with a caliper by using the ellipsoid approximation of tumors shape [85]. Mice were either treated with different of Etoposide formulations, Etoposide Nanocrystals (noted ETO in the following) or with Etoposide/Prednisolone Nanocrystals (noted ETO/PRD in the following) and one group was not treated and used as control (noted CTR in the following). We explained more deeply this protocol to obtain mice emitting signal from tumors and muscles in Chapter II. In total, we have 77 BL images. For each BLI, ROIs were drawn around each tumor and reference muscle. leading up to a total of 147 tumors and their associated 77 muscles. 52 tumors belong to the group CTR, 44 to the group ETO and 51 to the group ETO/PRD.

Since the muscle in a limb of a mouse is about $200mg$ [239], and the density of the muscle is about 1, it means that the concentration of luciferase in the muscle is about $0.015mg/mL$. In that context, the kinetics are described by the Pharmacokinetics Model illustrated Figure IV.3. In that model, as in subsection IV.1.2 and in subsection IV.2.1, $\mathbf{C}_T$ and $\mathbf{C}_{RR}$ are the concentrations of luciferin respectively in the tumor and in the Reference Region. The equations (IV.6) are still describing the exchange in this model by considering $\mathbf{C}_T = \mathbf{C}_{TOI}$.

According to [218], there is almost constantly about 2% of the injected luciferin per gram of muscle after an IP injection. Since the density of the muscle is about 1, there is about $0.04mg/mL$ of luciferin in the muscle. The luciferin is thus in exces compared to the enzyme and we also have $\mathbf{C}_{RR} << K_m$ therefore, as explained in subsection IV.1.2, we can assume that the photon flux in the muscle $\frac{d\mathbf{P}_{RR}}{dt}$ is also proportional to the concentration of luciferin in it $\mathbf{C}_{RR}$:

$$\frac{d\mathbf{P}_{RR}}{dt} \approx \lambda_{RR}\mathbf{C}_{RR}, \tag{IV.13}$$

with λ_{RR} being a constant that depends on K_m, the initial concentration of enzyme in the muscle and k_{cat}.

By noting $\frac{d\mathbf{P}_T}{dt} \approx \lambda_T \mathbf{C}_T$ the photon flux emitted by the tumor, we can then express (IV.12) as:

$$\frac{\mathbf{P}_T(t)}{\lambda_T} = \frac{K_{trans}}{K_{trans,RR}}\left(\frac{\mathbf{P}_{RR}(t)}{\lambda_{RR}} + k_{ep,RR}\int_0^t \frac{\mathbf{P}_{RR}(\tau)}{\lambda_{RR}}d\tau\right) - k_{ep}\int_0^t \frac{\mathbf{P}_T(\tau)}{\lambda_T}d\tau \tag{IV.14}$$

By noting $R_{trans} = \frac{K_{trans}\lambda_T}{K_{trans,RR}\lambda_{RR}}$, we then have:

$$\mathbf{P}_T(t) = R_{trans}\mathbf{P}_{RR}(t) + \frac{K_{trans}\lambda_T k_{ep,RR}}{K_{trans,RR}\lambda_{RR}}\int_0^t \mathbf{P}_{RR}(\tau)d\tau - k_{ep}\int_0^t \mathbf{P}_T(\tau)d\tau \tag{IV.15}$$

As in [120], we can rewrite (IV.15) as $\mathbf{A} = \mathbf{B}\mathbf{x}$ by noting $\mathbf{A} = [\mathbf{P}_T(t_0), \ldots, \mathbf{P}_T(t_f)]$,

$$\mathbf{B} = \begin{bmatrix} \mathbf{P}_{RR}(t_0) & \int_0^{t_0}\mathbf{P}_{RR}(\tau)d\tau & -\int_0^{t_0}\mathbf{P}_{RR}(\tau)d\tau \\ \mathbf{P}_{RR}(t_1) & \int_0^{t_1}\mathbf{P}_{RR}(\tau)d\tau & -\int_0^{t_1}\mathbf{P}_{RR}(\tau)d\tau \\ \vdots & \vdots & \vdots \\ \mathbf{P}_{RR}(t_f) & \int_0^{t_f}\mathbf{P}_{RR}(\tau)d\tau & -\int_0^{t_f}\mathbf{P}_{RR}(\tau)d\tau \end{bmatrix}, \tag{IV.16}$$

and

$$\mathbf{x} = \begin{bmatrix} R_{trans} \\ \frac{K_{trans}\lambda_T k_{ep,RR}}{K_{trans,RR}\lambda_{RR}} \\ k_{ep}, \end{bmatrix} \tag{IV.17}$$

and we have to find $\mathbf{x}$. To do that, [120] used the following formula:

$$\mathbf{x} = \mathbf{B}^{\dagger}\mathbf{A}, \tag{IV.18}$$

where $\mathbf{B}^{\dagger}$ is the pseudo-inverse of $\mathbf{B}$. However, in some hard cases it can output negative parameters, which is a problem given that all these constants are linked to exchange rates and so are positive. Therefore, when the result of (IV.18) yields negative coefficients we in fact use a non-negative linear least-squares (NLLS) algorithm, which is however more costly to compute.

The problem to solve is:

$$\underset{\mathbf{x}}{\operatorname{argmin}}\ \frac{1}{2}||\mathbf{A} - \mathbf{B}\mathbf{x}||_F^2 + i_{\geq\epsilon}(\mathbf{x}), \tag{IV.19}$$

where ϵ is a constant that can be set to 0 or a value strictly greater than 0, depending on the use case. Here, we want the parameters to be exchange rates, therefore to be non-null, so we set $\epsilon > 0$.

The proximal of $i_{\geq\epsilon}(.)$ is :

$$\text{prox}_{i_{\geq\epsilon}(.)}(\mathbf{x}) = [\mathbf{x}]_\epsilon \,, \tag{IV.20}$$

where $[\mathbf{x}]_\epsilon = \max(\mathbf{x}, \epsilon)$.

By noting $f(\mathbf{x} = \frac{1}{2}||\mathbf{A} - \mathbf{B}\mathbf{x}||_F^2)$, we have $\nabla_f(\mathbf{x}) = \mathbf{B}^T(\mathbf{B}\mathbf{x} - \mathbf{A})$, where ∇_f is Lipschitz with a coefficient of Lipschitz $L = ||\mathbf{B}^T\mathbf{B}||_{s,2}$ and $||\cdot||_{s,2}$ is the largest singular value.

We can now solve (IV.19) with the following Forward-Backward Splitting algorithm [162]:

Algorithm 13 $NLLS(\mathbf{A}, \mathbf{B}, \mathbf{x}^{(0)})$

Require: $N_{iter} > 0, \epsilon > 0$
  for $k \leq N_{iter}$ **do**
    $\mathbf{x}^{(k+1)} = \left[\mathbf{x}^{(k)} - \frac{1}{||\mathbf{B}^T\mathbf{B}||_{s,2}}(\mathbf{B}^T(\mathbf{B}\mathbf{x} - \mathbf{A}))\right]_\epsilon$
  end for
  Return $\mathbf{x}$

Estimations of R_{trans} and k_{ep} for each tumor can be computed with these methods. [238] used these relative parameters as features of a classifier to analyze their DCE-MRI images according to the response of tumors to therapy. In our case, we tested their ability to solve 3 tasks that can be relevant for biologists:

- Classify the tumors according to the group of treatment they belong to based on R_{trans} and k_{ep}
- Estimate the tumoral volume based on R_{trans} and k_{ep}
- Estimate the evolution of tumoral volume, meaning the tumoral volume at the next imaging day based on R_{trans}, k_{ep} and the tumoral volume at the imaging day.

To follow the work of [238], for the first task we used Gaussian SVM [240]. For the other 2, we used SVR (Gaussian and Linear), as well as basic linear and loglinear regressions. In the next part, we will test our algorithms.

IV.3 Experiments and Discussion

There are 2 parts in our work : the estimation of relative Pharmacokinetics Parameters and the quantification of tumors thanks to these parameters.

IV.3.1 Experiments with synthetic data to test the fitting algorithms

In a first step, we evaluated how well the parameters are estimated. To do that, we used the 81 kinetics generated in Chapter II, with the kinetic parameters summarized in Table II.1 and with the sinusoidal perturbations, illustrated in Figure II.11(d). For the reference region, we

	R_{trans}	k_{ep}
NMSE	0.17 ± 0.30	0.39 ± 0.40

Table IV.1: Results of the estimation of the kinetics parameters

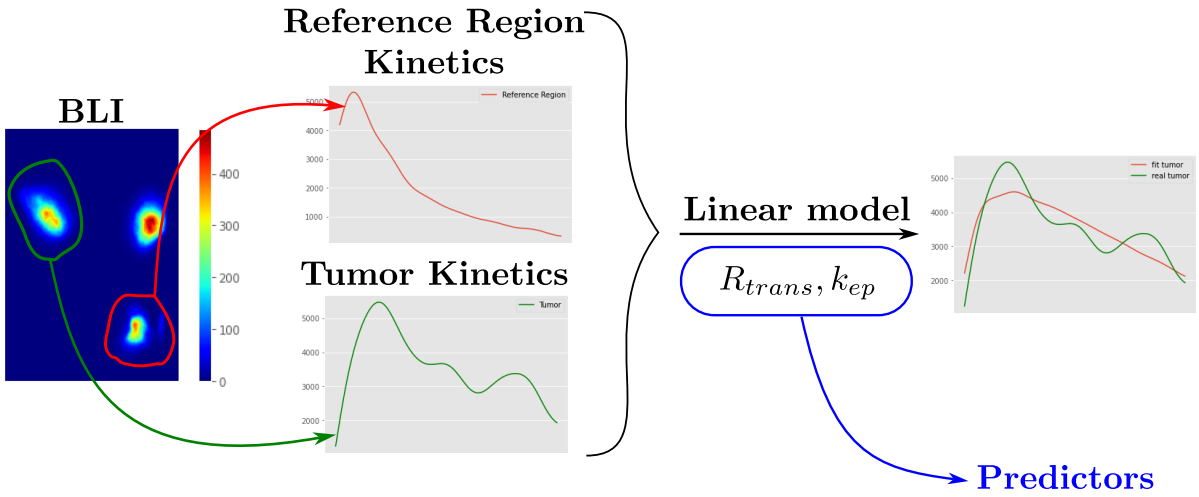

Figure IV.4: Extraction of Pharmacokinetics parameters from a BLI image using a Reference Region. The estimated PK coefficients are used as features for classifiers.

used $K_{trans,RR} = 0.1$ and $k_{ep,RR} = 1$ as in [120]. We also used $\lambda_{RR} = \lambda_T = 1$. For the NLLS algorithm, we used $N_{iter} = 10000$ and $\epsilon = 10^{-10}$. The initialization of the parameters is $\mathbf{x}^{(0)} = [10, 10, 10]$.

The metric we used is the $NMSE$ between the estimated parameters and the real ones. We get the results reported in Table IV.1.

These results show that the algorithm is able to estimate with accuracy the relative kinetic parameters, especially R_{trans}. The estimation of k_{ep} is not as good, making a point of possible improvement. In addition, we did not consider multiple possible values for $K_{trans,RR}$ and $k_{ep,RR}$, and set arbitrarily $\lambda_T = \lambda_{RR} = 1$. Estimating the robustness of the algorithm with respect to these parameters represents an promising perspective.

In this work, we considered those estimated parameters and we now will see how we can use them to quantify tumors in BLI.

IV.3.2 Experiments with real data : quantification of tumors

In this part we tested the ability of the estimated relative PK parameters to represent relevant features to quantify tumors as illustrated Figure IV.4. Here, we used empircally $\mathbf{x}^{(0)} = [10, 10, 10]$ to take into account the possible higher variability inter-tumors than in our synthetic dataset. To compare them with the more classical methods of quantifying BLI described in Chapter I, we tested several other features for our predictors : the average photon flux over the whole experiment (noted AUC), the average photon flux for 10 minutes, 8 minutes after the beginning of the experiment (noted AUC_{10}) and the Maximal Photon Emission (noted PE_{max}). We did not use the Luminoscore because we don't have access to the BLI from the back of the animal and we argued that our pre-processing makes the NBLI irrelevant because the background noise is null.

To be fair, we also wanted to find a way to test the PK parameters K_{trans} and k_{ep} to be consistant with the work of [100]. As we explained in subsection IV.1.2, it is not possible to estimate these parameters without the AIF, which we can not get during a BLI experiment. However, we also mentioned that some works in DCE-MRI or PET measured the AIF in a cohort of subjects and used the average as an approximation of the AIF to estimate PK parameters. The work of [218] provides the average on 8 mice of the evolution of luciferin in the blood after an IP injection. We then also used that theoretical average AIF to estimate K_{trans} and k_{ep}. Nevertheless, we are aware that there can be a huge variability in the AIF of the different subjects which would lead to biased parameters. This represents an interesting point of discussion and perspectives.

In order to estimate them, the procedure is quite the same as for the Reference Region parameters. Indeed, if we integrate the first equation of (IV.6), we have the following Linear Model:

$$\mathbf{C}_T(t) = K_{trans} \int_0^t \mathbf{C}_b(\tau)d\tau - k_{ep} \int_0^t \mathbf{C}_T(\tau)d\tau \tag{IV.21}$$

Therefore, since the link between the emitted photon flux and the concentration of luciferin in the tumor is linear, we derive:

$$\mathbf{P}_T(t) = K_{trans}\lambda_T \int_0^t \mathbf{C}_b(\tau)d\tau - k_{ep} \int_0^t \mathbf{P}_T(\tau)d\tau \tag{IV.22}$$

Hence, by noting $\mathbf{A} = [\mathbf{P}_T(t_0), \dots, \mathbf{P}_T(t_f)]$,

$$\mathbf{B} = \begin{bmatrix} \int_0^{t_0} \mathbf{C}_b(\tau)d\tau & -\int_0^{t_0} \mathbf{P}_{RR}(\tau)d\tau \\ \int_0^{t_1} \mathbf{C}_b(\tau)d\tau & -\int_0^{t_1} \mathbf{P}_{RR}(\tau)d\tau \\ \vdots & \vdots \\ \int_0^{t_f} \mathbf{C}_b(\tau)d\tau & -\int_0^{t_f} \mathbf{P}_{RR}(\tau)d\tau \end{bmatrix}, \tag{IV.23}$$

and

$$\mathbf{x} = \begin{bmatrix} R_{trans}^{AIF} \\ k_{ep} \end{bmatrix}, \tag{IV.24}$$

where $R_{trans}^{AIF} = K_{trans}\lambda_T$ we have $\mathbf{A} = \mathbf{B}\mathbf{x}$. We can then use the same algorithm as described in subsection IV.2.2 to estimate $\mathbf{x}$. The only difference is that we used empirically $\mathbf{x}^{(0)} = [10^6, 10]$ to simplify the NLLS algorithm because the linear factor λ_T is extremely high [100]. In the following, we will note k_{ep}^{AIF} the exchange rate from the tumor to the blood estimated with this method to avoid confusion with the one estimated from the Reference Region.

Based on the work of [218] in which they measured the concentration of luciferin in the blood after an IP injection on a cohort a mice and averaged it, the theoretical AIF we used is the one illustrated in Figure IV.5.

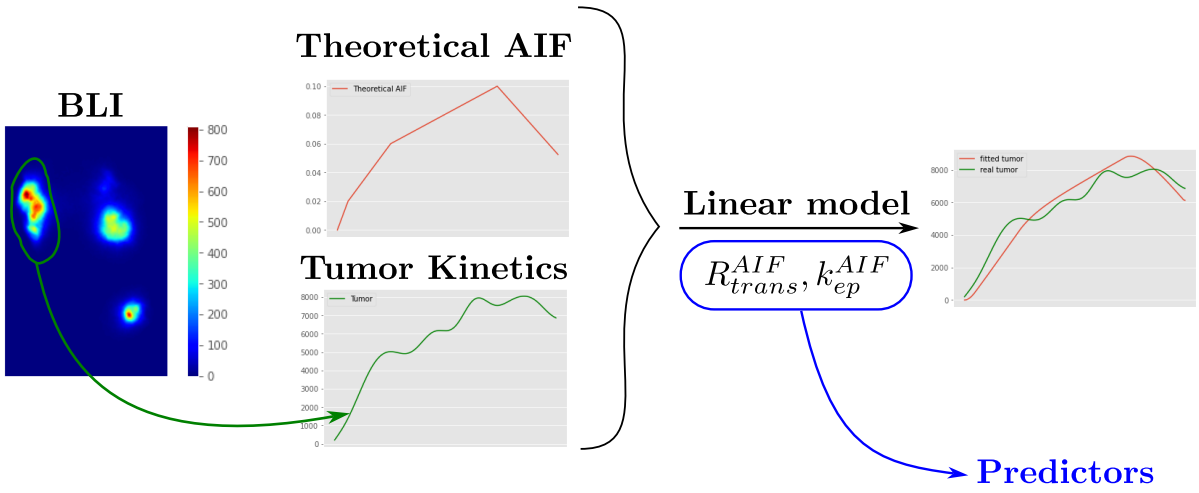

Figure IV.5: Extraction of Pharmacokinetics parameters from a BLI image using a theoretical AIF. The estimated PK coefficients are used as features for classifiers.

Predictor	Balanced Accuracy	Cohen's Kappa
R_{trans}	0.39 ± 0.08	0.08 ± 0.11
k_{ep}	$\mathbf{0.41 \pm 0.05}$	0.12 ± 0.08
AUC	0.30 ± 0.08	-0.05 ± 0.12
AUC_{10}	0.28 ± 0.07	-0.08 ± 0.10
PE_{max}	0.31 ± 0.05	-0.03 ± 0.08
$R_{trans} + k_{ep}$	$\mathbf{0.41 \pm 0.05}$	$\mathbf{0.12 \pm 0.07}$
R_{trans}^{AIF}	0.29 ± 0.03	-0.06 ± 0.05
k_{ep}^{AIF}	0.32 ± 0.02	-0.01 ± 0.03
$R_{trans}^{AIF} + k_{ep}^{AIF}$	0.30 ± 0.03	-0.05 ± 0.05

Table IV.2: Performance metrics for the predictive models generated with the PK parameters and the classical features extracted from BLI.

IV.3.2.1 Classification of tumors

First, we checked if the estimated relative pharmacokinetics parameters can be relevant features to classify tumors according to their treatment. This could be useful for biologists, as it would mean that we can extract physiological information from the tumor, which is different from the usual tumor volume, directly from the BLI. This information takes into account tumor hemodynamics and the specific exchanges between the tumor and its environment.

We used Gaussian SVMs, as in [238], with an empirical kernel coefficient $\gamma = 100$ and a one-vs-rest decision function, with a 5-fold cross validation scheme. As metrics, we used as in [238] the Balanced Accuracy and the Cohen's Kappa metric (that quantifies the level of agreement during the cross validation). We reported the results on the dataset presented in Chapter II in Table IV.2.

These results show that these estimated relative parameters are more relevant features for classification of tumors according to their associated treatment than the other features usually derived from BLI. Indeed, all the other predictors have a balanced accuracy of about $\frac{1}{3}$, which would be the score of a random classifier and the Cohen's Kappa metric show that their associated cross validated classifiers strongly disagree. In particular, we note that the PK parameters estimated with the theorical AIF did not perform better than the other predictors. The variability inter-subjects of the AIF can explain why the algorithm did not always output the same estimation for k_{ep} with the Reference Region and with the theoretical AIF as illus-

Kernel	Predictor	$NMSE$	R^2
Gaussian	R_{trans}	2.4 ± 0.58	-0.11 ± 0.15
	k_{ep}	2.9 ± 0.78	-0.14 ± 0.14
	AUC	2.2 ± 0.77	-0.09 ± 0.15
	AUC_{10}	2.1 ± 0.73	-0.09 ± 0.16
	PE_{max}	2.2 ± 0.77	-0.10 ± 0.15
	$R_{trans} + k_{ep}$	2.6 ± 0.65	-0.12 ± 0.15
	R_{trans}^{AIF}	2.8 ± 0.71	-0.17 ± 0.14
	k_{ep}^{AIF}	2.0 ± 0.62	-0.07 ± 0.17
	$R_{trans}^{AIF} + k_{ep}^{AIF}$	2.5 ± 0.72	-0.13 ± 0.16
Linear	R_{trans}	2.4 ± 0.59	-0.12 ± 0.13
	k_{ep}	3.4 ± 1.2	-0.12 ± 0.24
	AUC	2.4 ± 0.67	0.09 ± 0.16
	AUC_{10}	2.4 ± 0.62	$\mathbf{0.10 \pm 0.15}$
	PE_{max}	2.4 ± 0.69	0.05 ± 0.15
	$R_{trans} + k_{ep}$	2.8 ± 0.92	-0.10 ± 0.13
	R_{trans}^{AIF}	2.5 ± 0.66	-0.16 ± 0.15
	k_{ep}^{AIF}	1.5 ± 0.48	0.03 ± 0.18
	$R_{trans}^{AIF} + k_{ep}^{AIF}$	$\mathbf{1.3 \pm 0.35}$	0.06 ± 0.19

Table IV.3: Performance metrics for the regression models generated with the PK parameters and the classical features extracted from BLI.

trated in Figure IV.7, even though they should be the same. The predictor using R_{trans} and k_{ep} performs the best, which means that physiological information about the tumor exchange rates with the blood seems to be obtainable with the BLI using a Reference Region and that it is in a way representative of the treatment used on the tumor. This is illustrated in Figure IV.6 that displays the evolution of the average relative parameters for each group. In that figure, we can see that the R_{trans} in the group treated with ETO seems to have a distinct strong growth whereas the k_{ep} of the control group looks higher for the treated groups. This would mean that the tumors treated with ETO take more luciferin from the blood day after day and that the control tumors return more luciferin to the blood than treated tumors. Overall, there seems to be distinct patterns in the evolution of these relative PK parameters, even if this has to be further tested and validated with other experiment. However, the metrics are far from perfect, meaning that there is still a lot of improvement possible in quantifying physiological information of the tumors with BLI to classify them according to their treatment.

IV.3.2.2 Estimation of tumoral volume

a) Estimation of the tumoral volume based only on the PK parameters

We tested if we would be able to evaluate the tumoral volume thanks to the PK parameters we estimated, since it is the quantitative information usually estimated with BLI.

To do that, we used Gaussian SVRs with a kernel coefficient $\gamma = 100$ with a 5-fold cross validation scheme to be consistent with the classification task. However, since the relationship is supposed to be linear between the photon flux and the tumoral volume, we also tested Linear SVR with the same hyperparameters. As metrics, we used the $NMSE$ and the coefficient of determination R^2. We also tested and evaluated the Pearson correlation coefficient between the estimated parameters and the tumoral volume (noted $pearson$), and between the logarithm of the parameters and the logarithm of the volume to be consistant with [100] (noted $pearson_{log}$). We reported the results on our dataset in Table IV.3 and Table IV.4.

Table IV.3 shows that it is difficult to estimate precisely the tumor volume with the features

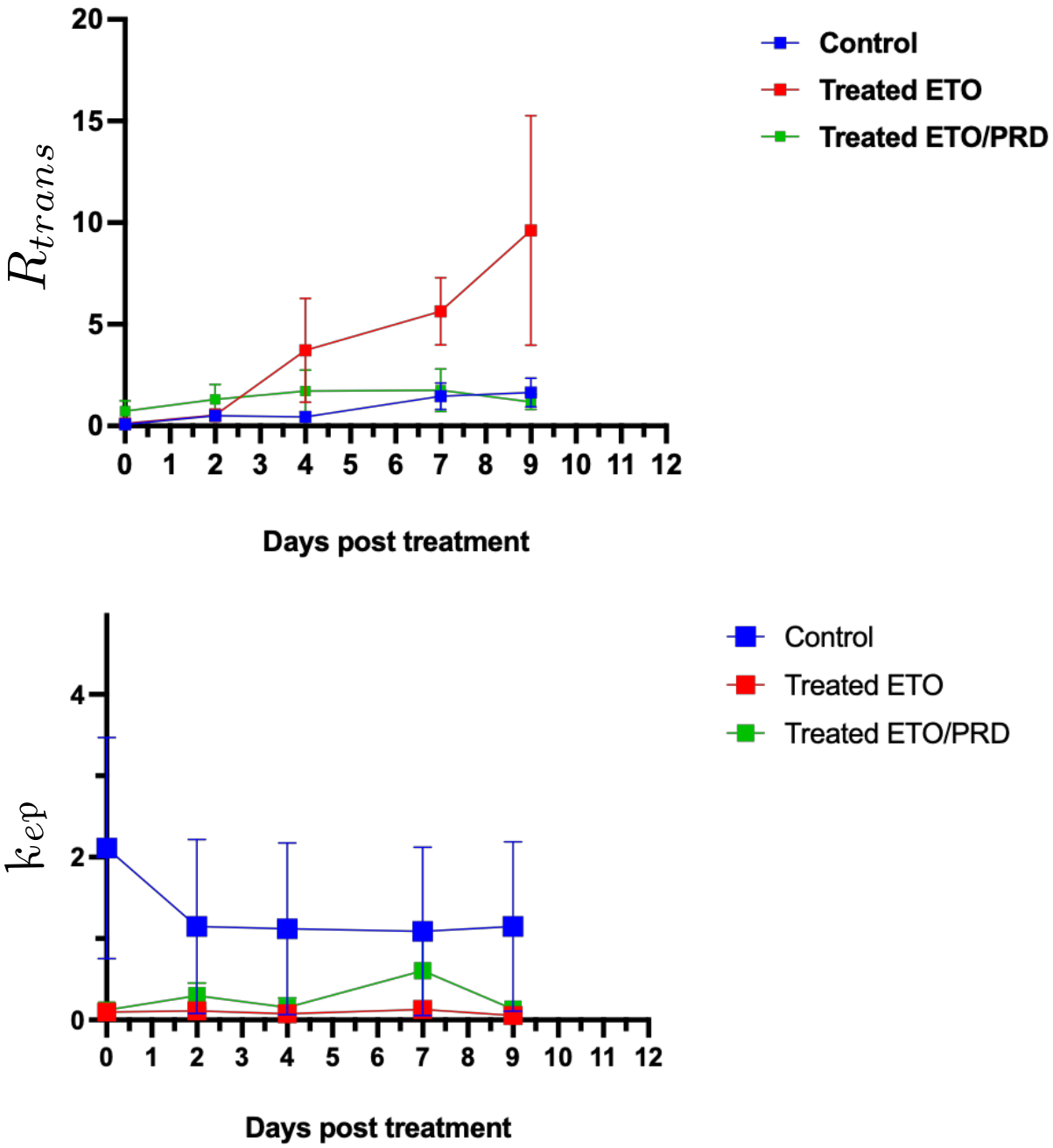

Figure IV.6: Evolution of relative parameters per group of treatment. For each time day are displayed the average relative parameters for each group and its standard deviation.

extracted by BLI. More precisely, the relative parameters R_{trans} and k_{ep} do not perform better as predictors for this task than any of the previous way of quantifying BLI. It is however interesting to note that the Linear Predictor using both R_{trans}^{AIF} and k_{ep}^{AIF} is the one that performs better. An interpretation of why it works better than the relative parameters despite the inter subjects variability is that k_{ep}^{AIF} may encode different information than k_{ep}, which can be explained by the simplicity of our model that uses only 2 compartments for describing the interaction of luciferin and luciferase in the tissues and that can induce biases. Table IV.4 shows that k_{ep}^{AIF} seems to indeed encode a feature negatively correlated with the tumoral volume, which could explain why the Linear SVR using it performs somehow better than the others. Table IV.4

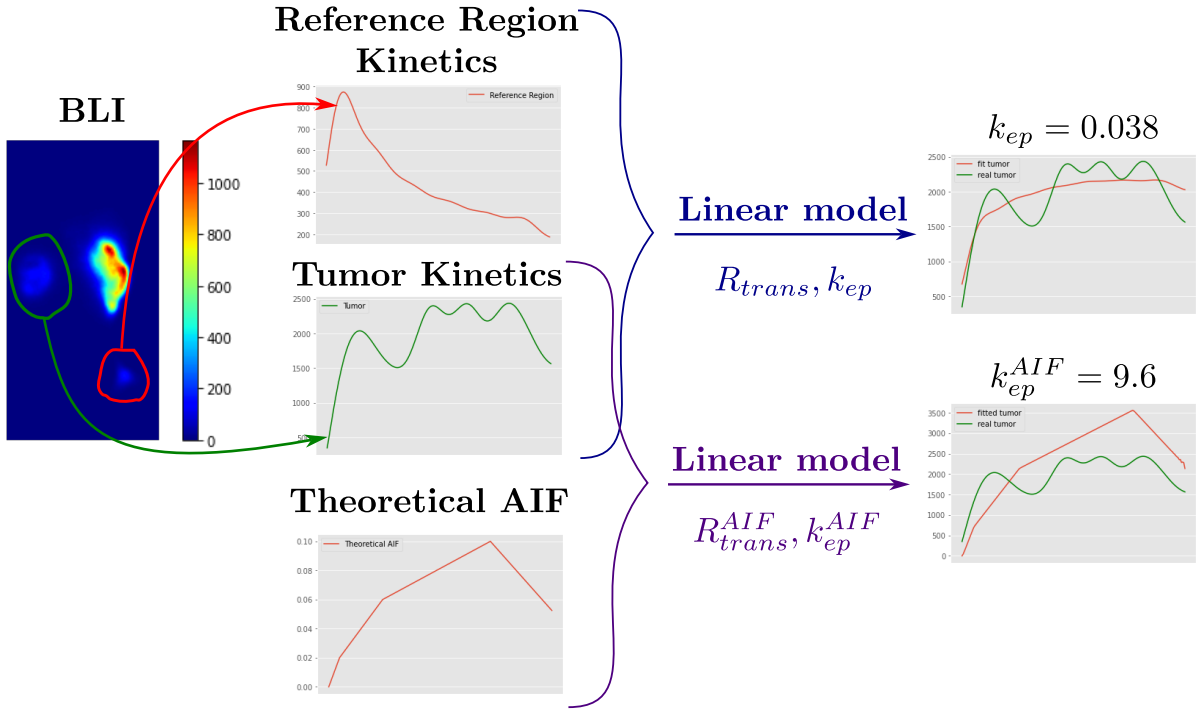

Figure IV.7: Estimation of Pharmacokinetics parameters with the Reference Region and with the Theoretical AIF yields different results for k_{ep}.

Predictor	$pearson(p_{value})$	$pearson_{log}(p_{value})$
R_{trans}	$0.35(1 \times 10^{-4})$	$0.20(0.03)$
k_{ep}	$-0.15(0.11)$	$-0.22(0.02)$
AUC	$\mathbf{0.56(3 \times 10^{-10})}$	$0.52(6 \times 10^{-9})$
AUC_{10}	$0.54(1 \times 10^{-9})$	$\mathbf{0.54(1 \times 10^{-9})}$
PE_{max}	$0.50(2 \times 10^{-8})$	$0.49(3 \times 10^{8})$
R_{trans}^{AIF}	$0.13(0.16)$	$0.14(0.14)$
k_{ep}^{AIF}	$-0.46(3 \times 10^{-7})$	$-0.34(3 \times 10^{4})$

Table IV.4: Performance metrics for the predictive models generated with the PK parameters and the classical features extracted from BLI.

also shows that even the standard quantification methods only have moderate correlation with the tumoral volume although the highest one is obtained with AUC, which seems natural. This reflects the widely discussed BLI quantification issues [57, 72, 96, 98, 105–108]. Table IV.4 however does not on show any particular polynomial relationship between k_{ep}^{AIF}, R_{trans}^{AIF} or k_{ep} and the tumoral volume as was suggested in [100].

b) Estimation of the evolution of the tumoral volume based on the PK parameters and the current volume

We finally tested if the PK parameters we estimated and the measured tumoral volume at a day D were able to predict the tumoral volume at day $D + 2$ (next imaging day). We then removed from the dataset presented in Chapter II the images of the last day of experiments, and were left with 111 tumors. We used Gaussian and Linear SVRs with the same hyperparameters as previously (kernel coefficient $\gamma = 100$ with a 5-fold cross validation scheme) and the same metrics $NMSE$ and R^2. We reported the results in Table IV.5.

Kernel	Predictor	$NMSE$	R^2
Gaussian	R_{trans}	1.3 ± 0.49	-0.12 ± 0.14
	k_{ep}	1.4 ± 0.52	-0.14 ± 0.14
	AUC	1.3 ± 0.47	-0.13 ± 0.14
	AUC_{10}	1.3 ± 0.51	-0.13 ± 0.14
	PE_{max}	1.3 ± 0.48	-0.14 ± 0.14
	$R_{trans} + k_{ep}$	1.4 ± 0.51	-0.14 ± 0.14
	R_{trans}^{AIF}	1.4 ± 0.53	-0.15 ± 0.14
	k_{ep}^{AIF}	1.3 ± 0.46	-0.12 ± 0.14
	$R_{trans}^{AIF} + k_{ep}^{AIF}$	1.4 ± 0.53	-0.15 ± 0.13
Linear	R_{trans}	0.83 ± 0.31	-0.12 ± 0.11
	k_{ep}	0.95 ± 0.28	-0.11 ± 0.13
	AUC	0.75 ± 0.27	$\mathbf{0.10 \pm 0.09}$
	AUC_{10}	0.76 ± 0.25	0.09 ± 0.07
	PE_{max}	0.61 ± 0.14	0.07 ± 0.11
	$R_{trans} + k_{ep}$	0.89 ± 0.29	-0.08 ± 0.13
	R_{trans}^{AIF}	0.70 ± 0.28	-0.12 ± 0.16
	k_{ep}^{AIF}	0.55 ± 0.13	0.03 ± 0.13
	$R_{trans}^{AIF} + k_{ep}^{AIF}$	$\mathbf{0.47 \pm 0.10}$	0.09 ± 0.15

Table IV.5: Performance metrics for the regression models generated with the PK parameters and the classical features extracted from BLI to estimate the tumoral volume at the next imaging day.

Table IV.5 shows that is seems easier to estimate the evolution of tumor volume than the current volume, knowing the current pharmacokinetics parameters. As before, the model that performs the best is the one using both parameters R_{trans}^{AIF} and k_{ep}^{AIF}. The interpretation here too is that maybe the simplicity of the model induces biases that make the parameters extracted with the reference region and the theoretical AIF encoding different informations. Overall, these results seem to show that it should be possible to estimate the tumoral volume at the next imaging day with the current volume and the PK parameters estimated from a theoretical AIF. It would be useful for biologists to better anticipate the growth of the animal's tumors and thus the times of sacrifice since the animals are sacrificed when the tumors become too large, which was not possible with the previous ways of quantifying BLI.

To conclude, this method provided encouraging results showing that the pharmacokinetics model could improve the quantification of BLI. The relative parameters extracted from the Reference Region seem to be representative of the tumor treatment and the parameters extracted from a theoretical AIF seem to be linked to the evolution of the tumoral volume. Although we tried to explain it by some biases induced by the simplicity of our model, it is not yet clear why the parameters generated with the reference region are not always similar to those generated with the theoretical AIF. More experiments have to be conducted to better validate these results and make them more robust. New experiments could include a comparison of reconstruction performances between the Reference Region based approach and the theoretical AIF based approach, as well as an exhaustive display of the evolution of the different parameters involved as in Figure IV.6.

Better modeling the relationship between photon flux and luciferin concentration could also help to enhance the model. Some prospects to achieve that are: adding compartements or taking into account others reagents necessary to the reaction (such as oxygen) with for instance dedicaded probes, accurately estimating the parameters related to the Michaelis-Menten kinetics (see (IV.13)), modeling the interactions between the photons and their environment [100], and differentiating quantity and concentration of substrate. This study has to be done jointly with a deeper analysis of the kinetics and their variability, since their oscillating pattern, pos-

sibly induced by physiological (such as respiratory movements [116]) or technical constraints (such as the pre-processing, see Chapter II), is a key factor that complicates the quantification process. One interesting observation is that, as displayed in Figure IV.4 and Figure IV.6, the Reference Regions kinetics seems less oscillating than the tumors kinetics. The robustness of the quantification with regard to the choice of Reference Region has also to be measured.

In the objective of a robust quantification of the tumor state, we could introduce another physiological indicator with the notion of distribution volume [132].

From a more methodological point of view, the fitting algorithm can be improved by integrating prior knowledge or adapted constraints. Finally, this quantification process could be applied on the results of the Blind Source Separation algorithms described in Chapter III to have a complete pipeline of automatic tumor state quantification for BLI.

Chapter V

Conclusion and Perspectives

V.1 Thesis summary

Bioluminescence Imaging is an imaging modality based on the emission of photons by a living tissue during a biological reaction called bioluminescence, that does not require any exciting light. This reaction is usually induced by the presence in a tissue of an enzyme, the luciferase, that oxidizes a substrate, called luciferin, in the presence of other reagents like ATP, oxygen or magnesium, resulting in the emission of photons from within the tissue. It has many applications in biomedical resarch, like in preclinical oncology where it is used to estimate non-invasively tumoral volumes.

In this work, we provided valuable insights for improving the robustness of analysis of BLI in this context. We indeed showed that BLI has many biases, starting with the input of the user who has to manually draw a Region of Interest around each tumor, which can be troublesome and lead to tumors not being correctly quantified. We then proposed a model to automatically separate the photon flux emitted by different tumors present within the same BLI thanks to Blind Source Separation algorithms. To our knowledge, this is the first work that applies Blind Source Separation methods to BLI. We first described the data as a Linear Mixing Model and applied non-negativity constraint to be consistent with photon counting, formulating a Non-Negative Matrix Factorization (NMF). The first results we obtained showed strong spatial mixtures of the tumors. Therefore, we added in our model several prior informations to better separate the tumors. In a first contribution, we showed that adding a sparsity constraint on the sources was able to improve the spatial separation by enforcing as many pixels as possible to be null. This constraint was weighted by hyperparameters set automatically with a fixed point argument. This sparsity constraint led the cost function to optimize to be non-smooth and non-convex, and required to use dedicated optimization algorithms.

In a second contribution, we showed that our sources are in fact not really sparse in the direct domain leading to some artifact pixels in the outputs of the source separation algorithms. We therefore adapted our previous algorithm to enforce them to be sparse in the wavelet domain. This improved the spatial consistence of the sources estimated with the algorithm because it encouraged the neighboring pixels to share the same value, at different scales.

We next tackled the fact that this algorithm sometimes create halos around isolated pixels, because of intra-tumoral pixel kinetics variability. A last contribution to our model is thus to take that variability into account. We have shown that pixels belonging to the same tumor essentially share the same general temporal patterns up to a temporal warping defined as a diffeomorphism. In our model, as a preprocessing step before our NMF algorithms, pixels

belonging to a single tumor are thus grouped together using a watershed algorithm, then synchronized with an algorithm for aligning their kinetics. These algorithms have been tested and validated on two real BLI datasets and on one synthetic dataset generated with our proposed framework. To simulate BLIs, we generated parametric maps and kinetics using a pharmacokinetics model, which were degraded by adding variability to make the resulting images more realistic. To our knowledge, this is the first BLI simulator able to generate spatio-temporal images of tumors. Another limitating factor of BLI is its lack of reproductibility and robustness in the quantification of tumor state, which is the last challenge we tackled in this work.

In preclinical BLI, the tumoral state is usually quantified with the average photon flux emitted by the tumor, which is supposed to be correlated to the tumoral volume. Poor reproductibility results in a lack of confidence in the experiments involving tumoral quantification with BLI. In a last contribution, we showed how a pharmacokinetics model can help to improve the quantification of BLI. The biologists we are working with injected luciferase, the enzyme responsible for the bioluminescence reaction, in a muscle that helped to calibrate the signal coming from the tumor. We then proposed a linear reference region model of BLI and extracted relative pharmacokinetics parameters from the bioluminescent signals. We tested the ability of these parameters to quantify tumor state and compared them with other more usual ways of BLI quantification. These experiments provided interesting results showing that these relative parameters, and thus the exchange rates of luciferin between the tumor and the blood, seemed to represent significant features of the treatments received by the mice. However, the pharmacokinetics parameters obtained with a theoretical knowledge of the concentration of luciferin in the blood (AIF) yielded better results for the estimation of the tumoral volume. Overall, the results obtained by our pharmacokinetic model have shown encouraging results that could improve the quantification of BLI.

V.2 Perpectives

Although they give promising results as shown, our blind sources separation algorithms still present several limitations. They are iterative optimization algorithms and, as such, they do not need any training and are unsupervised. However, the computing time necessary for processing a new BLI is quite high compared to newer algorithms using Deep Learning, which after training have a low inference time and can in addition extract more complex features from the images to separate the photon flux. It would be possible to formulate our iterative algorithm as an unrolling problem [153] or to use a neural network that would provide an optimized constraint [154]. Also, the ℓ_2 loss is not robust enough to intensity imbalance, causing the algorithm to yield poor results in the case of strongly imbalanced tumors within the same BLI. We can use a normalization, such as a log transform, to compensate for this imbalance but this could lead to pixels with very low SNR to have an excessive importance in the optimization scheme, leading to badly estimated source kinetics. Hence, in this setting, it would then be of a great importance to take into account the remaining variability in our model, for instance by using a perturbated linear mixing model [122]. In any case, the robustness of our algorithm to the magnitude of the variability should be challenged, thanks to our data simulator.

In addition, kinetics estimation has been somewhat impaired, probably because not all kinetics were pre-aligned during the pre-processing step, even though the strong improvement in spatial separation compensates for this loss overall. Computing it directly in the optimization pipeline could lead to better results as it would also be applied to the pixels belonging to several tumors, which is not possible currently. The principal limitation to integrate it in the optimization pipeline is that it is not straightforward to obtain an explicit proximal of the indicator

function of the space of diffeomorphisms. To overcome that hurdle, we propose for future work to restrict the optimized time-warping functions to a specific family of diffeomorphisms defined by a few coefficients, as the polynomial functions with strictly positive coefficients. This family seems generalizable enough to cover the space of diffeomorphisms we need and we can optimize the warping functions with only a few coefficients by enforcing a non-negativity constraint on them, which would limit the complexity of the algorithm.

In this work we also chose to work with the ℓ_2 distance in the data fidelity terms for our optimization settings, which can be applied to our data only after having pre-processed them to remove the Poisson noise and most of the additive noise. Another perspective for our work would be to use a loss adapted to our raw data, such as the Kullback-Leibler divergence that acts as a maximum likelihood when the sources are corrupted with a Poisson noise [159]. A problem would be to tackle the non-differentiability of the Kullback-Leibler divergence at 0.

A possible enhancement to our work would be to automatically determine the number of sources in the image. This is a vast issue in most of blind sources separation algorithms and we could use criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [138]. It could also maybe enable us to identify heterogeneous areas in the tumors.

An interesting applicative perspective of this work on blind source separation would be to apply our algorithms on FLI data to monitor longitudinally the activity on tumors during several days. This could be useful, as the excitation light in the FLI causes a strong superposition of signals, thus making the quantification more difficult. This idea can lead to several interesting challenges, such as image registration issues to compensate the evolution of position due to the longitudinal nature of the data.

The pharmacokinetics model we propose showed promising results but it needs to be further validated with other experiments, that must also test the quality of the reconstruction, the evolution of all the parameters involved and the robustness of the quantification with respect to the choice of Reference Region. New data could confirm the significance of the extracted features. They would potentially help to explain the differences seen between the parameters calibrated with a reference region and the ones calibrated with the theoretical AIF. This could be carried out in conjunction with an enhancement of the model used to more precisely explain the interaction with luciferin and the other reagents necessary to the bioluminesence reaction. Oxygen is for instance a crucial part of the reaction since the necrotic or hypoxic parts of the tumors do not emit any light [57, 72]. We can maybe use specific probes to quantify the concentration of oxygen and extend the model to take it into account, for instance by leveraging the work of [110] that models its influence on bioluminescence. We could also follow the idea of [84] and add a compartment to simulate the metabolism of the luciferin. This new compartement is supposed to model the potential loss of luciferin in the tumor due possibly to high extracellular pressure and degenerated vascularity in tumors. All these ideas, as well as an accurate estimation of the parameters related to the Michaelis-Menten kinetics and of the interactions between the photons and their environment [100], would improve our model of the relationship between the photon flux and luciferin quantity, which in future works will have to be differentiated from the concentration currently considered. In the scope of enhancing the quantification of tumor state, another relevant perspective is the introduction of the notion of distribution volume [132].

This pharmacokinetics model could also be used directly in the optimization process to have a joint estimation of both the source separation and the kinetics parameters. However, a more

direct way to have a complete automatic tumor state quantification pipeline for BLI would be to apply our quantification method to the blind source separation results.

A general perspective among of our work, shared by the different projects, is that our analysis would benefit from a deeper understanding of the kinetics and their variability: it could lead to better source separation and more robust tumor quantification, and it would enable us to improve the synthetic BLI generation framework we proposed to simulate more realistic images. We identified physiological constraints, such as tumor vascularization, and technical constraints, such as the applied pre-processing, that should be studied in that goal.

We could also extract more precise PK parameters by improving the non-negative least squares algorithm. Adding new information to this model can help to achieve this objective, since it removes degrees of freedom. The fitting algorithm could also benefit from the knowledge of the physiological amplitudes of the exchange rates to constrain them in a specific range as done in [120].

Finally, it would be desirable for biologists if the tools we have developed could be integrated into a GUI. The most direct way seems to be to wrap the `python` code we developed and its associated dependencies into a module. This would enable to manually or automatically draw ROIs around mice (for the blind source separation algorithms) or around tumors (for the quantification algorithms) and launch the processes. This could be integrated in frameworks able to read BLI as M3Vision (the one of BiospaceLab, our industrial collaborator in this work) or Icy [241].

Bibliography

[1] J.-P. Dillenseger, *Imagerie préclinique multimodale chez le petit animal: qualification des instruments et des méthodes (IRM, μTDM et μTEMP)*, Ph.D. thesis, Université de Strasbourg, 2017. (See page 1)

[2] E. L. C. Roncali, *Développements en imagerie de luminescence in vivo chez le petit animal: système d'acquisition sur animal vigile et tomographie en temps réel*, Ph.D. thesis, Châtenay-Malabry, Ecole centrale de Paris, 2008. (See pages 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11)

[3] L. Cunha, I. Horvath, S. Ferreira, et al., "Preclinical imaging: an essential ally in modern biosciences," *Molecular Diagnosis & Therapy*, vol. 18, pp. 153–173, 2014. (See pages 1, 2, 3, 5, 6, and 8)

[4] A. Dubois, J. Dauguet, A.-S. Herard, et al., "Automated three-dimensional analysis of histological and autoradiographic rat brain sections: application to an activation study," *Journal of Cerebral Blood Flow & Metabolism*, vol. 27, no. 10, pp. 1742–1755, 2007. (See page 1)

[5] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, no. 4, pp. 128–138, 1988. (See page 1)

[6] G. Vereb, J. Szöllősi, J. Matko, et al., "Dynamic, yet structured: the cell membrane three decades after the singer–nicolson model," *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8053–8058, 2003. (See page 1)

[7] P. Bandi, O. Geessink, Q. Manson, et al., "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 550–560, 2018. (See page 2)

[8] A. Barua, P. Bitterman, J. S. Abramowicz, et al., "Histopathology of ovarian tumors in laying hens: a preclinical model of human ovarian cancer," *International Journal of Gynecologic Cancer*, vol. 19, no. 4, 2009. (See page 1)

[9] P. K. Thanos, N. B. Taintor, D. Alexoff, et al., "In vivo comparative imaging of dopamine D2 knockout and wild-type mice with ^{11}C-raclopride and microPET," *Journal of Nuclear Medicine*, vol. 43, no. 11, pp. 1570–1577, 2002. (See page 2)

[10] H. Bridge and S. Clare, "High-resolution MRI: in vivo histology?," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1465, pp. 137–146, 2006. (See page 2)

[11] K. Cui, X. Xu, H. Zhao, and S. T. Wong, "A quantitative study of factors affecting in vivo bioluminescence imaging," *Luminescence*, vol. 23, no. 5, pp. 292–295, 2008. (See pages 2 and 8)

[12] M. Savinaud, *Recalage de flux de données cinématiques pour l'application à l'imagerie optique*, Ph.D. thesis, Ecole Centrale Paris, 2010. (See pages 2, 4, 5, 6, 7, 8, 9, 11, 12, and 19)

[13] R. K. Jain, L. L. Munn, and D. Fukumura, "Dissecting tumour pathophysiology using intravital microscopy," *Nature Reviews Cancer*, vol. 2, no. 4, pp. 266–276, 2002. (See page 2)

[14] M. J. Cobb, Y. Chen, R. A. Underwood, et al., "Noninvasive assessment of cutaneous wound healing using ultrahigh-resolution optical coherence tomography," *Journal of Biomedical Optics*, vol. 11, no. 6, pp. 064002–064002, 2006. (See page 2)

[15] M. Ruggeri, H. Wehbe, S. Jiao, et al., "In vivo three-dimensional high-resolution imaging of rodent retina with spectral-domain optical coherence tomography," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 4, pp. 1808–1814, 2007. (See page 2)

[16] F. Helmchen and W. Denk, "New developments in multiphoton microscopy," *Current Opinion in Neurobiology*, vol. 12, no. 5, pp. 593–601, 2002. (See page 2)

[17] R. W. Cootney, "Ultrasound imaging: principles and applications in rodent research," *ILAR Journal*, vol. 42, no. 3, pp. 233–247, 2001. (See page 2)

[18] N. Deshpande, A. Needles, and J. K. Willmann, "Molecular ultrasound imaging: current status and future directions," *Clinical Radiology*, vol. 65, no. 7, pp. 567–581, 2010. (See page 2)

[19] G. C. Kagadis, G. Loudos, K. Katsanos, et al., "In vivo small animal imaging: current status and future prospects," *Medical Physics*, vol. 37, no. 12, pp. 6421–6442, 2010. (See page 2)

[20] E. Roncali, M. Savinaud, O. Levrey, et al., "New device for real-time bioluminescence imaging in moving rodents," *Journal of Biomedical Optics*, vol. 13, no. 5, pp. 054035–054035, 2008. (See pages 3 and 10)

[21] T. F. Massoud and S. S. Gambhir, "Molecular imaging in living subjects: seeing fundamental biological processes in a new light," *Genes & Development*, vol. 17, no. 5, pp. 545–580, 2003. (See pages 3, 4, and 8)

[22] E. L. Lundin, M. Stauber, P. Papageorgiou, et al., "Automatic registration of 2D histological sections to 3D microCT volumes: trabecular bone," *Bone*, vol. 105, pp. 173–183, 2017. (See page 3)

[23] C. T. Badea, M. Drangova, D. W. Holdsworth, and G. Johnson, "In vivo small-animal imaging using micro-CT and digital subtraction angiography," *Physics in Medicine & Biology*, vol. 53, no. 19, pp. R319, 2008. (See page 3)

[24] N. Lee, S. H. Choi, and T. Hyeon, "Nano-sized CT contrast agents," *Advanced Materials*, vol. 25, no. 19, pp. 2641–2660, 2013. (See page 3)

[25] I. C. Simcock, S. C. Shelmerdine, J. C. Hutchinson, et al., "Human fetal whole-body postmortem microfocus computed tomographic imaging," *Nature Protocols*, vol. 16, no. 5, pp. 2594–2614, 2021. (See page 3)

[26] A. Garofalakis, A. Dubois, B. Thézé, et al., "Fusion of [^{18}F] FDG PET with fluorescence diffuse optical tomography to improve validation of probes and tumor imaging," *Molecular Imaging and Biology*, vol. 15, pp. 316–325, 2013. (See page 5)

[27] A. Rahmim and H. Zaidi, "PET versus SPECT: strengths, limitations and challenges," *Nuclear Medicine Communications*, vol. 29, no. 3, pp. 193–207, 2008. (See page 5)

[28] Y. Yang, Y.-C. Tai, S. Siegel, et al., "Optimization and performance evaluation of the microPET II scanner for in vivo small-animal imaging," *Physics in Medicine & Biology*, vol. 49, no. 12, pp. 2527, 2004. (See page 5)

[29] E. J. Sutton, T. D. Henning, B. J. Pichler, et al., "Cell tracking with optical imaging," *European Radiology*, vol. 18, pp. 2021–2032, 2008. (See page 5)

[30] Y. Jung, L. Tong, A. Tanaudommongkon, et al., "In vitro and in vivo nonlinear optical imaging of silicon nanowires," *Nano Letters*, vol. 9, no. 6, pp. 2440–2444, 2009. (See page 5)

[31] R. Weissleder and V. Ntziachristos, "Shedding light onto live molecular targets," *Nature Medicine*, vol. 9, no. 1, pp. 123–128, 2003. (See page 5)

[32] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, no. 2, pp. R41, 1999. (See page 5)

[33] N. Alsawaftah, A. Farooq, S. Dhou, and A. F. Majdalawieh, "Bioluminescence imaging applications in cancer: a comprehensive review," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 307–326, 2020. (See pages 6, 8, 9, 10, 12, and 13)

[34] Q. Le Masne de Chermont, *Nanoparticules à luminescence persistante pour l'imagerie optique in vivo*, Ph.D. thesis, Paris 6, 2007. (See page 6)

[35] X. Chen, P. S. Conti, and R. A. Moats, "In vivo near-infrared fluorescence imaging of integrin $\alpha_V\beta_3$ in brain tumor xenografts," *Cancer Research*, vol. 64, no. 21, pp. 8009–8014, 2004. (See page 7)

[36] K. Liu, H. Shang, X. Kong, and W. Lin, "A novel near-infrared fluorescent probe with a large stokes shift for biothiol detection and application in in vitro and in vivo fluorescence imaging," *Journal of Materials Chemistry B*, vol. 5, no. 21, pp. 3836–3841, 2017. (See page 7)

[37] C. Bremer, V. Ntziachristos, and R. Weissleder, "Optical-based molecular imaging: contrast agents and potential medical applications," *European Radiology*, vol. 13, pp. 231–243, 2003. (See page 7)

[38] W. C. Chan and S. Nie, "Quantum dot bioconjugates for ultrasensitive nonisotopic detection," *Science*, vol. 281, no. 5385, pp. 2016–2018, 1998. (See page 7)

[39] X. Gao, L. Yang, J. A. Petros, et al., "In vivo molecular and cellular imaging with quantum dots," *Current Opinion in Biotechnology*, vol. 16, no. 1, pp. 63–72, 2005. (See page 7)

[40] Q. Le Masne de Chermont, C. Chanéac, J. Seguin, et al., "Nanoprobes with near-infrared persistent luminescence for in vivo imaging," *Proceedings of the National Academy of Sciences*, vol. 104, no. 22, pp. 9266–9271, 2007. (See page 7)

[41] R. M. Hoffman, "Green fluorescent protein imaging of tumour growth, metastasis, and angiogenesis in mouse models," *The Lancet Oncology*, vol. 3, no. 9, pp. 546–556, 2002. (See page 8)

[42] K. Yamauchi, M. Yang, P. Jiang, et al., "Development of real-time subcellular dynamic multicolor imaging of cancer-cell trafficking in live mice with a variable-magnification whole-mouse imaging system," *Cancer Research*, vol. 66, no. 8, pp. 4208–4214, 2006. (See page 8)

[43] O. Shimomura, *Bioluminescence: chemical principles and methods*, World Scientific, 2012. (See page 8)

[44] R. T. Sadikot and T. S. Blackwell, "Bioluminescence imaging," in *Proceedings of the American Thoracic Society*, 2005, vol. 2, pp. 537–540. (See page 8)

[45] C. E. Badr and B. A. Tannous, "Bioluminescence imaging: progress and applications," *Trends in Biotechnology*, vol. 29, no. 12, pp. 624–633, 2011. (See page 8)

[46] J. Myung, S. Hong, F. Hatanaka, et al., "Period coding of Bmal1 oscillators in the suprachiasmatic nucleus," *Journal of Neuroscience*, vol. 32, no. 26, pp. 8900–8918, jun 2012. (See page 8)

[47] J. Cosette, R. B. Abdelwahed, S. Donnou-Triffault, et al., "Bioluminescence-based tumor quantification method for monitoring tumor progression and treatment effects in mouse lymphoma models," *Journal of Visualized Experiments*, vol. 2016, no. 113, jul 2016. (See pages 8, 9, 14, 16, 17, 18, 19, and 27)

[48] O. Shimomura and F. H. Johnson, "Chemical nature of the light emitter in bioluminescence of aequorin," *Tetrahedron Letters*, vol. 14, no. 31, pp. 2963–2966, 1973. (See page 8)

[49] L. F. Greer III and A. A. Szalay, "Imaging of light emission from the expression of luciferases in living cells and organisms: a review," *Luminescence: the Journal of Biological and Chemical Luminescence*, vol. 17, no. 1, pp. 43–74, 2002. (See page 8)

[50] A. Verdes and D. F. Gruber, "Glowing worms: biological, chemical, and functional diversity of bioluminescent annelids," *Integrative and Comparative Biology*, vol. 57, no. 1, pp. 18–32, 2017. (See page 8)

[51] S. H. Haddock, M. A. Moline, and J. F. Case, "Bioluminescence in the sea," *Annual Review of Marine Science*, vol. 2, pp. 443–493, 2010. (See page 9)

[52] E. Dereure, J. Seguin, E. Angelini, et al., "Improved quantification of optical imaging data," in *French Society for Nanomedicine*, dec 2021. (See page 9)

[53] D. E. Jenkins, Y. Oei, Y. S. Hornig, et al., "Bioluminescent imaging (BLI) to improve and refine traditional murine models of tumor growth and metastasis," *Clinical & Experimental Metastasis*, vol. 20, pp. 733–744, 2003. (See pages 9 and 14)

[54] S. Picaud, B. Lambolez, and L. Tricoire, "Bioluminescence imaging of neuronal network dynamics using aequorin-based calcium sensors," in *Live Cell Imaging: Methods and Protocols*, pp. 281–294. Springer, 2021. (See page 9)

[55] S. Li, Z. Ruan, H. Zhang, and H. Xu, "Recent achievements of bioluminescence imaging based on firefly luciferin-luciferase system," *European Journal of Medicinal Chemistry*, vol. 211, pp. 113111, 2021. (See page 9)

[56] R. Weissleder, "A clearer vision for in vivo imaging," *Nature Biotechnology*, vol. 19, no. 4, pp. 316–317, 2001. (See page 9)

[57] M. Keyaerts, J. Verschueren, T. J. Bos, et al., "Dynamic bioluminescence imaging for quantitative tumour burden assessment using IV or IP administration of D-luciferin: effect on intensity, time kinetics and repeatability of photon emission," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 35, no. 5, pp. 999–1007, may 2008. (See pages 9, 14, 15, 17, 18, 19, 88, 101, and 106)

[58] R. Weissleder and U. Mahmood, "Molecular imaging," *Radiology*, vol. 219, no. 2, pp. 316–333, 2001. (See page 9)

[59] A. Rose, "The sensitivity performance of the human eye on an absolute scale," *Journal of the Optical Society of America*, vol. 38, no. 2, pp. 196–208, 1948. (See page 9)

[60] C. Bolin, C. Sutherland, K. Tawara, et al., "Novel mouse mammary cell lines for in vivo bioluminescence imaging (BLI) of bone metastasis," *Biological Procedures Online*, vol. 14, pp. 1–10, 2012. (See page 9)

[61] Y. W. Wang, N. P. Reder, S. Kang, et al., "Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches," *Nanotheranostics*, vol. 1, no. 4, pp. 369, 2017. (See page 9)

[62] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectronics Journal*, vol. 37, no. 5, pp. 433–451, 2006. (See page 9)

[63] S. A. Taylor et al., "CCD and CMOS imaging array technologies: technology review," *UK: Xerox Research Centre Europe*, pp. 1–14, 1998. (See page 9)

[64] S. Liu, Y. Su, M. Z. Lin, and J. A. Ronald, "Brightening up biology: advances in luciferase systems for in vivo imaging," *ACS Chemical Biology*, vol. 16, no. 12, pp. 2707–2718, 2021. (See page 9)

[65] S. J. Kim and H.-Y. Lee, "In vivo molecular imaging in preclinical research," *Laboratory Animal Research*, vol. 38, no. 1, pp. 31, 2022. (See page 9)

[66] L. J. Van Vliet, F. R. Boddeke, D. Sudar, and I. T. Young, "Image detectors for digital image microscopy," *Digital Image Analysis of Microbes: Imaging, Morphometry, Fluorometry, and Motility Techniques and Applications*, pp. 37–64, 1998. (See page 10)

[67] D. J. Denvir and E. Conroy, "Electron-multiplying CCD: the new ICCD," in *Low-Light-Level and Real-Time Imaging Systems, Components, and Applications*. SPIE, 2003, vol. 4796, pp. 164–174. (See page 10)

[68] I. N. Dalinenko, A. V. Malyarov, G. I. Vishnevsky, et al., "Design and fabrication technology of thinned backside-excited CCD imagers and the family of intensified electron-bombarded CCD image tubes," in *Fifth Conference on Charge-Coupled Devices and CCD Systems*. SPIE, 1996, vol. 2790, pp. 67–78. (See page 10)

[69] F. Leblond, S. C. Davis, P. A. Valdés, and B. W. Pogue, "Pre-clinical whole-body fluorescence imaging: review of instruments, methods and applications," *Journal of Photochemistry and Photobiology B: Biology*, vol. 98, no. 1, pp. 77–94, 2010. (See page 10)

[70] M. Suyama, A. Kageyama, I. Mizuno, et al., "Electron bombardment CCD tube," in *Ultrahigh-and High-Speed Photography and Image-based Motion Measurement*. SPIE, 1997, vol. 3173, pp. 422–429. (See page 10)

[71] X. Gu, Q. Zhang, L. Larcom, and H. Jiang, "Three-dimensional bioluminescence tomography with model-based reconstruction," *Optics Express*, vol. 12, no. 17, pp. 3996–4000, 2004. (See page 11)

[72] C. P. Klerk, R. M. Overmeer, T. M. Niers, et al., "Validity of bioluminescence measurements for noninvasive in vivo imaging of tumor load in small animals," *Biotechniques*, vol. 43, no. S1, pp. S7–S13, 2007. (See pages 11, 14, 15, 17, 19, 88, 101, and 106)

[73] J. Mérian, J. Gravier, F. Navarro, and I. Texier, "Fluorescent nanoprobes dedicated to in vivo imaging: from preclinical validations to clinical translation," *Molecules*, vol. 17, no. 5, pp. 5564–5591, 2012. (See page 11)

[74] M. A. Naser and M. S. Patterson, "Improved bioluminescence and fluorescence reconstruction algorithms using diffuse optical tomography, normalized data, and optimized selection of the permissible source region," *Biomedical Optics Express*, vol. 2, no. 1, pp. 169–184, 2011. (See page 11)

[75] Y. Gao, K. Wang, S. Jiang, et al., "Bioluminescence tomography based on gaussian weighted laplace prior regularization for in vivo morphological imaging of glioma," *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2343–2354, 2017. (See page 11)

[76] G. Wang, W. Cong, K. Durairaj, et al., "In vivo mouse studies with bioluminescence tomography," *Optics Express*, vol. 14, no. 17, pp. 7801–7809, 2006. (See page 11)

[77] A. J. Chaudhari, F. Darvas, J. R. Bading, et al., "Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging," *Physics in Medicine & Biology*, vol. 50, no. 23, pp. 5421, 2005. (See pages 11 and 12)

[78] C. Qin, J. Feng, S. Zhu, et al., "Recent advances in bioluminescence tomography: methodology and system as well as application," *Laser & Photonics Reviews*, vol. 8, no. 1, pp. 94–114, 2014. (See pages 11 and 12)

[79] C. Kim, I. H. Kim, S.-i. Kim, et al., "Comparison of the intraperitoneal, retroorbital and per oral routes for F-18 FDG administration as effective alternatives to intravenous administration in mouse tumor models using small animal PET/CT studies," *Nuclear Medicine and Molecular Imaging*, vol. 45, pp. 169–176, 2011. (See page 12)

[80] M. Keyaerts, V. Caveliers, and T. Lahoutte, "Bioluminescence imaging: looking beyond the light," *Trends in Molecular Medicine*, Mar. 2012. (See pages 12 and 19)

[81] S. Iwano, M. Sugiyama, H. Hama, et al., "Single-cell bioluminescence imaging of deep tissue in freely moving animals," *Science*, vol. 359, no. 6378, pp. 935–939, 2018. (See page 13)

[82] Y. Furuhata, A. Sakai, T. Murakami, et al., "Bioluminescent imaging of Arabidopsis thaliana using an enhanced Nano-lantern luminescence reporter system," *PLoS ONE*, vol. 15, no. 1, jan 2020. (See page 13)

[83] A. Mastropietro, A. Tennstaedt, A. Beyrau, et al., "Proof of concept of an automatic tool for bioluminescence imaging data analysis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. nov 2015, vol. 2015-Novem, pp. 6269–6272, Institute of Electrical and Electronics Engineers Inc. (See pages 13, 16, 17, and 27)

[84] Y. Dai, D. Chen, G. Wang, et al., "Kinetic modeling and analysis of dynamic bioluminescence imaging of substrates administered by intraperitoneal injection," *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 2, pp. 389–396, feb 2020. (See pages 14, 15, 16, 88, 90, and 106)

[85] K. M. Parkins, V. P. Dubois, A. M. Hamilton, et al., "Multimodality cellular and molecular imaging of concomitant tumour enhancement in a syngeneic mouse model of breast cancer metastasis," *Scientific Reports*, vol. 8, no. 1, dec 2018. (See pages 14, 17, 26, and 93)

[86] J. D. Rojas, J. B. Joiner, B. Velasco, et al., "Validation of a combined ultrasound and bioluminescence imaging system with magnetic resonance imaging in orthotopic pancreatic murine tumors," *Scientific Reports*, vol. 12, no. 1, pp. 102, 2022. (See page 14)

[87] A.-L. Puaux, L. C. Ong, Y. Jin, et al., "A comparison of imaging techniques to monitor tumor growth and cancer progression in living animals," *International Journal of Molecular Imaging*, vol. 2011, 2011. (See page 14)

[88] A. Thalheimer, D. Korb, L. Bönicke, et al., "Noninvasive visualization of tumor growth in a human colorectal liver metastases xenograft model using bioluminescence in vivo imaging," *Journal of Surgical Research*, vol. 185, no. 1, pp. 143–151, Nov. 2013. (See page 14)

[89] M. Ravoori, O. Margalit, S. Singh, et al., "Magnetic resonance imaging and bioluminescence imaging for evaluating tumor burden in orthotopic colon cancer," *Scientific Reports*, vol. 9, no. 1, pp. 6100, 2019. (See page 14)

[90] K. O'Neill, S. K. Lyons, W. M. Gallagher, et al., "Bioluminescent imaging: a critical tool in pre-clinical oncology research," *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, vol. 220, no. 3, pp. 317–327, 2010. (See page 14)

[91] T. N. Le, H. Lim, A. M. Hamilton, et al., "Characterization of an orthotopic rat model of glioblastoma using multiparametric magnetic resonance imaging and bioluminescence imaging," *Tomography*, vol. 4, no. 2, pp. 55–65, 2018. (See page 14)

[92] S. C. Jost, L. Collins, S. Travers, et al., "Measuring brain tumor growth: combined bioluminescence imaging–magnetic resonance imaging strategy," *Molecular Imaging*, vol. 8, no. 5, pp. 7290–2009, 2009. (See page 14)

[93] P. C. Black, A. Shetty, G. A. Brown, et al., "Validating bladder cancer xenograft bioluminescence with magnetic resonance imaging: the significance of hypoxia and necrosis," *BJU International*, vol. 106, no. 11, pp. 1799–1804, 2010. (See pages 14 and 17)

[94] Z. Paroo, R. A. Bollinger, D. A. Braasch, et al., "Validating bioluminescence imaging as a high-throughput, quantitative modality for assessing tumor burden," *Molecular Imaging*, vol. 3, no. 2, 2004. (See pages 14, 15, and 17)

[95] A. Rehemtulla, L. D. Stegman, S. J. Cardozo, et al., "Rapid and quantitative assessment of cancer treatment response using in vivo bioluminescence imaging," *Neoplasia*, vol. 2, no. 6, pp. 491–495, 2000. (See page 14)

[96] Y. Inoue, S. Kiryu, M. Watanabe, et al., "Timing of imaging after D-luciferin injection affects the longitudinal assessment of tumor growth using in vivo bioluminescence imaging," *International Journal of Biomedical Imaging*, vol. 2010, 2010. (See pages 14, 19, and 101)

[97] J. Burgos, M. Rosol, R. Moats, et al., "Time course of bioluminescent signal in orthotopic and heterotopic brain tumors in nude mice," *Biotechniques*, vol. 34, no. 6, pp. 1184–1188, 2003. (See pages 14 and 19)

[98] K. Cui, X. Xu, H. Zhao, and S. T. Wong, "A quantitative study of factors affecting in vivo bioluminescence imaging," *Luminescence*, vol. 23, no. 5, pp. 292–295, oct 2008. (See pages 14, 18, 19, and 101)

[99] S. Baba, S. Y. Cho, Z. Ye, et al., "How reproducible is bioluminescent imaging of tumor cell growth? Single time point versus the dynamic measurement approach," *Molecular Imaging*, vol. 6, no. 5, pp. 7290–2007, 2007. (See pages 14 and 15)

[100] H. Sim, K. Bibee, S. Wickline, and D. Sept, "Pharmacokinetic modeling of tumor bioluminescence implicates efflux, and not influx, as the bigger hurdle in cancer drug therapy," *Cancer Research*, vol. 71, no. 3, pp. 686–692, feb 2011. (See pages 15, 16, 18, 21, 22, 31, 32, 88, 89, 90, 97, 99, 101, 102, and 106)

[101] Akkoul Smaïl, R. Ledee, R. Leconge, et al., "Validation Expérimentale d'une Chaîne de Traitement d'Images de Bioluminescence chez le Petit Animal," in *GRETSI'09: XXIIème Colloque Francophone de Traitement du Signal et des Images*, 2009, pp. 2–5. (See pages 16, 26, and 27)

[102] S. Pesnel, S. Akkoul, R. Ledée, et al., "Use of an image restoration process to improve spatial resolution in bioluminescence imaging," *Molecular Imaging*, vol. 10, no. 6, pp. 446–452, nov 2011. (See page 16)

[103] K. M. Parkins, A. M. Hamilton, A. V. Makela, et al., "A multimodality imaging model to track viable breast cancer cells from single arrest to metastasis in the mouse brain," *Scientific Reports*, vol. 6, oct 2016. (See page 17)

[104] B. W. Rice, M. D. Cable, and M. B. Nelson, "In vivo imaging of light-emitting probes," *Journal of Biomedical Optics*, vol. 6, no. 4, pp. 432, 2001. (See page 17)

[105] E. H. Moriyama, M. J. Niedre, M. T. Jarvi, et al., "The influence of hypoxia on bioluminescence in luciferase-transfected gliosarcoma tumor cells in vitro," *Photochemical & Photobiological Sciences*, vol. 7, no. 6, pp. 675, 2008. (See pages 17 and 101)

[106] A. A. Khalil, M. J. Jameson, W. C. Broaddus, et al., "The influence of hypoxia and pH on bioluminescence imaging of luciferase-transfected tumor cells and xenografts," *International Journal of Molecular Imaging*, vol. 2013, pp. 1–9, jul 2013. (See pages 17, 18, 88, and 101)

[107] C. D. Scatena, M. A. Hepner, Y. A. Oei, et al., "Imaging of bioluminescent LNCaP-luc-M6 tumors: a new animal model for the study of metastatic human prostate cancer," *The Prostate*, vol. 59, no. 3, pp. 292–303, 2004. (See pages 17 and 101)

[108] S. Sarraf-Yazdi, J. Mi, M. W. Dewhirst, and B. M. Clary, "Use of in vivo bioluminescence imaging to predict hepatic tumor burden in mice," *Journal of Surgical Research*, vol. 120, no. 2, pp. 249–255, 2004. (See pages 17 and 101)

[109] S. Benzekry, C. Lamont, A. Beheshti, et al., "Classical mathematical models for description and prediction of experimental tumor growth," *PLoS Computational Biology*, vol. 10, no. 8, pp. e1003800, 2014. (See page 18)

[110] D. Lambrechts, M. Roeffaers, K. Goossens, et al., "A causal relation between bioluminescence and oxygen to quantify the cell niche," *PLoS ONE*, vol. 9, no. 5, may 2014. (See pages 18 and 106)

[111] J. M. Ignowski and D. V. Schaffer, "Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells," *Biotechnology and Bioengineering*, vol. 86, no. 7, pp. 827–834, 2004. (See page 18)

[112] Y. Zilberman, Y. Gafni, G. Pelled, et al., "Bioluminescent imaging in bone," in *Osteoporosis*, pp. 261–272. Humana Press, 2008. (See page 19)

[113] A. Mikhno, F. Zanderigo, R. T. Ogden, et al., "Toward noninvasive quantification of brain radioligand binding by combining electronic health records and dynamic PET imaging data," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1271–1282, jul 2015. (See pages 21 and 91)

[114] A. Ambikapathi, T. H. Chan, C. H. Lin, et al., "Convex-optimization-based compartmental pharmacokinetic analysis for prostate tumor characterization using DCE-MRI," in *IEEE Transactions on Biomedical Engineering*. apr 2016, vol. 63, pp. 707–720, IEEE Computer Society. (See page 21)

[115] T. E. Yankeelov, J. J. Luci, M. Lepage, et al., "Quantitative pharmacokinetic analysis of DCE-MRI data without an arterial input function: a reference region model," *Magnetic Resonance Imaging*, vol. 23, no. 4, pp. 519–529, 2005. (See pages 22, 26, 90, 91, and 92)

[116] G. Renault, *Etude de la perfusion en échographie de contraste*, Ph.D. thesis. (See pages 22, 86, and 103)

[117] S. Mulé, N. Kachenoura, O. Lucidarme, et al., "An automatic respiratory gating method for the improvement of microcirculation evaluation: application to contrast-enhanced ultrasound studies of focal liver lesions," *Physics in Medicine & Biology*, vol. 56, no. 16, pp. 5153, 2011. (See page 22)

[118] S. Akkoul, R. Ledee, R. Leconge, et al., "Comparison of image restoration methods for bioluminescence imaging," in *Image and Signal Processing: 3rd International Conference, ICISP 2008. Cherbourg-Octeville, France, July 1-3, 2008. Proceedings 3*. Springer, 2008, pp. 163–172. (See pages 27 and 43)

[119] S. Akkoul, R. Lédée, R. Leconge, and R. Harba, "A new adaptive switching median filter," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 587–590, 2010. (See page 28)

[120] Z. Ahmed and I. R. Levesque, "Increased robustness in reference region model analysis of DCE MRI using two-step constrained approaches," *Magnetic Resonance in Medicine*, vol. 78, no. 4, pp. 1547–1557, 2017. (See pages 32, 92, 93, 94, 96, and 107)

[121] W. S. Ouedraogo, B. Nicolas, B. Oudompheng, et al., "A frequency method for blind separation of an anechoic mixture," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 521–525. (See pages 33 and 77)

[122] P. A. Thouvenin, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 525–538, 2016. (See pages 34, 38, 39, 68, 71, 75, 76, 77, 84, and 105)

[123] J. Baranger, B. Arnal, F. Perren, et al., "Adaptive spatiotemporal SVD clutter filtering for ultrafast doppler imaging using similarity of spatial singular vectors," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1574–1586, 2018. (See page 36)

[124] X.-R. Feng, H.-C. Li, R. Wang, et al., "Hyperspectral unmixing based on nonnegative matrix factorization: a comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4414–4436, 2022. (See page 38)

[125] C. Kervazo, J. Bobin, C. Chenot, and F. Sureau, "Use of PALM for ℓ_1 sparse matrix factorization: difficulty and rationalization of a two-step approach," *Digital Signal Processing*, vol. 97, feb 2020. (See pages 38, 39, 47, 58, 62, 63, 64, and 68)

[126] J.-H. Ahn, S. Kim, J.-H. Oh, and S. Choi, "Multiple nonnegative-matrix factorization of dynamic PET images," in *Proceedings of Asian Conference on Computer Vision*. Citeseer, 2004, pp. 1009–1013. (See page 38)

[127] Y. C. Cavalcanti, T. Oberlin, N. Dobigeon, et al., "Factor analysis of dynamic PET images: beyond Gaussian noise," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2231–2241, sep 2019. (See pages 38, 42, 77, and 84)

[128] Y. Xie, J. Zhao, and P. Zhang, "A multicompartment model for intratumor tissue-specific analysis of DCE-MRI using non-negative matrix factorization," *Medical Physics*, vol. 48, no. 5, pp. 2400–2411, may 2021. (See page 38)

[129] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012. (See pages 38, 42, 45, and 61)

[130] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, 2000. (See pages 38, 43, and 44)

[131] Y. C. Cavalcanti, *Factor analysis of dynamic PET images*, Ph.D. thesis, Institut National Polytechnique de Toulouse-INPT, 2018. (See pages 38 and 39)

[132] A. Mikhno, F. Zanderigo, R. T. Ogden, et al., "Toward noninvasive quantification of brain radioligand binding by combining electronic health records and dynamic PET imaging data," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1271–1282, 2015. (See pages 38, 91, 103, and 106)

[133] B. Bodvarsson, L. K. Hansen, C. Svarer, and G. Knudsen, "NMF on positron emission tomography," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. IEEE, 2007, vol. 1, pp. I–309. (See pages 38 and 42)

[134] C. Kervazo, *Optimization framework for large-scale sparse blind source separation*, Ph.D. thesis, Université Paris Saclay (COmUE), 2019. (See pages 39 and 40)

[135] J. M. Nascimento and J. M. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005. (See page 39)

[136] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, et al., "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012. (See page 41)

[137] A. Halimi, P. Honeine, M. Kharouf, et al., "Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3811–3821, jul 2016. (See page 40)

[138] K. Illner, J. Miettinen, C. Fuchs, et al., "Model selection using limiting distributions of second-order blind source separation algorithms," *Signal Processing*, vol. 113, pp. 95–103, 2015. (See pages 40 and 106)

[139] R. Wildeboer, F. Sammali, R. Van Sloun, et al., "Blind source separation for clutter and noise suppression in ultrasound imaging: review for different applications," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 8, pp. 1497–1512, 2020. (See page 40)

[140] A. A. Miranda, Y.-A. Le Borgne, and G. Bontempi, "New routes from minimal approximation error to principal components," *Neural Processing Letters*, vol. 27, pp. 197–207, 2008. (See page 41)

[141] R. Di Paola, C. Penel, J. Bazin, and C. Berche, "Factor analysis and scintigraphy," *International Journal of Nuclear Medicine and Biology*, vol. 3, no. 3-4, pp. 168–169, 1976. (See page 41)

[142] D. Barber, "The use of principal components in the quantitative analysis of gamma camera dynamic studies," *Physics in Medicine & Biology*, vol. 25, no. 2, pp. 283, 1980. (See page 41)

[143] F. Frouin, J.-P. Bazin, and R. Di Paola, "Image sequence processing using factor analysis and compartmental modelling," in *Science and Engineering of Medical Imaging*. oct 1989, vol. 1137, p. 37, SPIE. (See page 41)

[144] A. Sitek, E. V. Di Bella, and G. Gullberg, "Factor analysis with a priori knowledge-application in dynamic cardiac spect," *Physics in Medicine & Biology*, vol. 45, no. 9, pp. 2619, 2000. (See page 41)

[145] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic Press, 2010. (See page 41)

[146] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000. (See page 42)

[147] S.-i. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, 1995. (See page 42)

[148] M. Davies, "Identifiability issues in noisy ICA," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 470–473, 2004. (See page 42)

[149] J. Bobin, F. Sureau, J.-L. Starck, et al., "Joint planck and WMAP CMB map reconstruction," *Astronomy & Astrophysics*, vol. 563, pp. A105, 2014. (See page 42)

[150] F. Xiong, J. Zhou, S. Tao, et al., "SNMF-Net: Learning a deep alternating neural network for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022. (See page 42)

[151] N. Gillis, D. Kuang, and H. Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2066–2078, 09 2015. (See page 42)

[152] H. Kameoka, L. Li, S. Inoue, and S. Makino, "Supervised determined source separation with multichannel variational autoencoder," *Neural Computation*, vol. 31, no. 9, pp. 1891–1914, 2019. (See page 42)

[153] M. Fahes, C. Kervazo, J. Bobin, and F. Tupin, "Unrolling PALM for sparse semi-blind source separation," in *International Conference on Learning Representations*, 2022. (See pages 42 and 105)

[154] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "UnDIP: Hyperspectral unmixing using deep image prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021. (See pages 42, 86, and 105)

[155] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994. (See pages 43, 44, and 45)

[156] A. Cichocki and R. Zdunek, "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2007, vol. 4493 LNCS, pp. 793–802, Springer Verlag. (See page 43)

[157] P. O. Hoyer, "Non-negative sparse coding," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*. IEEE, 2002, pp. 557–565. (See page 43)

[158] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999. (See page 43)

[159] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007. (See pages 43 and 106)

[160] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: with application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009. (See page 43)

[161] Y. Xu, "Alternating proximal gradient method for nonnegative matrix factorization," *arXiv preprint arXiv:1112.5407*, 2011. (See page 44)

[162] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005. (See pages 45, 56, and 95)

[163] A. Cichocki and A. H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E92-A, no. 3, pp. 708–721, 2009. (See pages 45 and 60)

[164] J. Rapin, J. Bobin, A. Larue, and J. L. Starck, "Sparse and non-negative BSS for noisy data," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5620–5632, aug 2013. (See pages 47, 63, and 68)

[165] C. Boutsidis and E. Gallopoulos, "SVD based initialization: a head start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350–1362, apr 2008. (See pages 48, 63, 64, 71, and 82)

[166] E. Dereure, E. Angelini, A. Garofalakis, et al., "Factorisation non-négative de matrice pour séparation de sources en imagerie par bioluminescence préclinique," in *GRETSI'22: XXVIIIème Colloque Francophone de Traitement du Signal et des Images*, 2022. (See page 48)

[167] E. Dereure, C. Kervazo, J. Seguin, et al., "Sparse non-negative matrix factorization for preclinical bioluminescent imaging," in *International Symposium on Biomedical Imaging*, 2023. (See pages 48, 67, 68, 69, 70, 71, and 84)

[168] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004. (See page 49)

[169] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. (See pages 49 and 56)

[170] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006. (See page 49)

[171] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," *Advances in Neural Information Processing Systems*, vol. 19, 2006. (See page 49)

[172] X. X. Zhu and R. Bamler, "Tomographic SAR inversion by ℓ_1-norm regularization—the compressive sensing approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010. (See page 49)

[173] J. Bobin, J. L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, nov 2007. (See page 49)

[174] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007. (See page 50)

[175] D. L. Donoho, X. Huo, et al., "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001. (See page 50)

[176] R. Bracewell and P. B. Kahn, "The Fourier transform and its applications," *American Journal of Physics*, vol. 34, no. 8, pp. 712–712, 1966. (See page 50)

[177] S. Mallat, *A wavelet tour of signal processing*, Elsevier, 1999. (See pages 50, 68, 69, and 71)

[178] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme*, Georg-August-Universitat, Gottingen., 1909. (See page 50)

[179] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988. (See page 50)

[180] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*, Cambridge University Press, 2010. (See page 50)

[181] I. Daubechies, *Ten lectures on wavelets*, SIAM, 1992. (See pages 51 and 52)

[182] S. Singla and S. Feizi, "Salient ImageNet: how to discover spurious features in deep learning?," in *International Conference on Learning Representations*, 2022. (See pages 51 and 52)

[183] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996. (See page 51)

[184] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, 2013. (See page 53)

[185] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, 2011. (See page 54)

[186] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1199–1226, 2013. (See pages 57 and 69)

[187] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011. (See pages 58 and 69)

[188] J. Douglas and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Transactions of the American Mathematical Society*, vol. 82, no. 2, pp. 421–439, 1956. (See page 58)

[189] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, vol. 109, pp. 475–494, 2001. (See page 58)

[190] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the kurdyka-łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010. (See page 58)

[191] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014. (See pages 58 and 70)

[192] F. Feng and M. Kowalski, "Underdetermined reverberant blind source separation: sparse approaches for multiplicative and convolutive narrowband approximation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 442–456, 2018. (See page 58)

[193] Y. C. Cavalcanti, T. Oberlin, N. Dobigeon, and C. Tauber, "Unmixing dynamic PET images with a PALM algorithm," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 425–429. (See page 58)

[194] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward–backward algorithm," *Journal of Global Optimization*, vol. 66, no. 3, pp. 457–485, 2016. (See page 58)

[195] T. Pock and S. Sabach, "Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1756–1787, 2016. (See page 59)

[196] L. T. K. Hien, N. Gillis, and P. Patrinos, "Inertial block mirror descent method for non-convex non-smooth optimization," *arXiv*, vol. 2019, 2019. (See page 59)

[197] X. Xu, J. Li, S. Li, and A. Plaza, "Curvelet transform domain-based sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4908–4924, 2020. (See pages 63 and 68)

[198] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008. (See page 64)

[199] A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, pp. 20110534, 2013. (See page 64)

[200] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017. (See page 68)

[201] J. Rapin, J. Bobin, A. Larue, and J. L. Starck, "NMF with sparse regularizations in transformed domains," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2020–2047, jul 2014. (See pages 68 and 69)

[202] C. Kervazo, J. Bobin, and C. Chenot, "Blind separation of a large number of sparse sources," *Signal Processing*, vol. 150, pp. 157–165, 2018. (See page 69)

[203] G. Steidl, J. Weickert, T. Brox, et al., "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004. (See page 71)

[204] Y. C. Cavalcanti, T. Oberlin, N. Dobigeon, et al., "Unmixing dynamic PET images with variable specific binding kinetics," *Medical Image Analysis*, vol. 49, pp. 117–127, 2018. (See page 75)

[205] M. A. Veganzones, L. Drumetz, G. Tochon, et al., "A new extended linear mixing model to address spectral variability," *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, 2014. (See page 75)

[206] R. Achanta, A. Shaji, K. Smith, et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012. (See page 77)

[207] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, 2004. (See page 77)

[208] S. Beucher, "Use of watersheds in contour detection," in *Proc. Int. Workshop on Image Processing*, 1979, pp. 17–21. (See page 78)

[209] A. Srivastava, W. Wu, S. Kurtek, and E. Klassen, "Registration of functional data using Fisher-Rao metric," *arXiv preprint arXiv:1103.3817*, 2011. (See pages 79, 80, and 81)

[210] L. M. Sangalli, P. Secchi, S. Vantini, and V. Vitelli, "K-mean alignment for curve clustering," *Computational Statistics & Data Analysis*, vol. 54, no. 5, pp. 1219–1233, 2010. (See page 79)

[211] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, 1945. (See page 80)

[212] Y. Dai, G. Wang, D. Chen, et al., "Intravenous administration-oriented pharmacokinetic model for dynamic bioluminescence imaging," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 3, pp. 843–847, 2018. (See page 88)

[213] E. Seibert and T. S. Tracy, "Fundamentals of enzyme kinetics," *Enzyme Kinetics in Drug Metabolism: Fundamentals and Applications*, pp. 9–22, 2014. (See page 90)

[214] N. C. of the International Union of Biochemistry (NC-IUB), "Symbolism and terminology in enzyme kinetics. Recommendations 1981," *European Journal of Biochemistry*, vol. 128, pp. 281–291, 1982. (See page 90)

[215] J. M. Ignowski and D. V. Schaffer, "Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells," *Biotechnology and Bioengineering*, vol. 86, no. 7, pp. 827–834, 2004. (See page 90)

[216] K. A. Feeney, M. Putker, M. Brancaccio, and J. S. O'Neill, "In-depth characterization of firefly luciferase as a reporter of circadian gene expression in mammalian cells," *Journal of Biological Rhythms*, vol. 31, no. 6, pp. 540–550, 2016. (See page 90)

[217] "MedChemExpress," `https://www.medchemexpress.com/D-Luciferin.html?` (See page 90)

[218] F. Berger, R. Paulmurugan, S. Bhaumik, and S. S. Gambhir, "Uptake kinetics and biodistribution of ^{14}C-D-luciferin—a radiolabeled substrate for the firefly luciferase catalyzed bioluminescence reaction: impact on bioluminescence based reporter gene imaging," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 35, pp. 2275–2285, 2008. (See pages 90, 91, 93, and 97)

[219] T. Fritz-Hansen, E. Rostrup, H. B. Larsson, et al., "Measurement of the arterial concentration of Gd-DTPA using MRI: a step toward quantitative perfusion imaging," *Magnetic Resonance in Medicine*, vol. 36, no. 2, pp. 225–231, 1996. (See page 91)

[220] H. Larsson, M. Stubgaard, J. Frederiksen, et al., "Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors," *Magnetic Resonance in Medicine*, vol. 16, no. 1, pp. 117–131, 1990. (See page 91)

[221] R. V. Parsey, A. Ojha, R. T. Ogden, et al., "Metabolite considerations in the in vivo quantification of serotonin transporters using ^{11}C-DASB and PET in humans," *Journal of Nuclear Medicine*, vol. 47, no. 11, pp. 1796–1802, 2006. (See page 91)

[222] N. E. Simpson, Z. He, and J. L. Evelhoch, "Deuterium NMR tissue perfusion measurements using the tracer uptake approach: I. Optimization of methods," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 1, pp. 42–52, 1999. (See page 91)

[223] H. Degani, V. Gusis, D. Weinstein, et al., "Mapping pathophysiological features of breast tumors by MRI at high spatial resolution," *Nature Medicine*, vol. 3, no. 7, pp. 780–782, 1997. (See page 91)

[224] M. Bentourkia, "Kinetic modeling of PET-FDG in the brain without blood sampling," *Computerized Medical Imaging and Graphics*, vol. 30, no. 8, pp. 447–451, 2006. (See page 91)

[225] T. Tsuchida, N. Sadato, Y. Yonekura, et al., "Noninvasive measurement of cerebral metabolic rate of glucose using standardized input function," *Journal of Nuclear Medicine*, vol. 40, no. 9, pp. 1441–1445, 1999. (See page 91)

[226] H.-J. Weinmann, M. Laniado, and W. Mützel, "Pharmacokinetics of GdDTPA/dimeglumine after intravenous injection into healthy volunteers.," *Physiological Chemistry and Physics and Medical NMR*, vol. 16, no. 2, pp. 167–172, 1984. (See page 91)

[227] R. E. Port, M. V. Knopp, U. Hoffmann, et al., "Multicompartment analysis of gadolinium chelate kinetics: blood-tissue exchange in mammary tumors as monitored by dynamic MR imaging," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 10, no. 3, pp. 233–241, 1999. (See page 91)

[228] M. J. Van Osch, E.-j. P. Vonken, M. A. Viergever, et al., "Measuring the arterial input function with gradient echo sequences," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 49, no. 6, pp. 1067–1076, 2003. (See page 91)

[229] K. Chen, D. Bandy, E. Reiman, et al., "Noninvasive quantification of the cerebral metabolic rate for glucose using positron emission tomography, ^{18}F-fluoro-2-deoxyglucose, the Patlak method, and an image-derived input function," *Journal of Cerebral Blood Flow & Metabolism*, vol. 18, no. 7, pp. 716–723, 1998. (See page 91)

[230] M. Schain, S. Benjaminsson, K. Varnäs, et al., "Arterial input function derived from pairwise correlations between PET-image voxels," *Journal of Cerebral Blood Flow & Metabolism*, vol. 33, no. 7, pp. 1058–1065, 2013. (See page 91)

[231] M. Naganawa, Y. Kimura, K. Ishii, et al., "Extraction of a plasma time-activity curve from dynamic brain PET images based on independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 201–210, 2005. (See page 91)

[232] J. Lee, S. Platt, M. Kent, and Q. Zhao, "An analysis of the pharmacokinetic parameter ratios in DCE-MRI using the reference region model," *Magnetic Resonance Imaging*, vol. 30, no. 1, pp. 26–35, 2012. (See page 91)

[233] C. Yang, G. S. Karczmar, M. Medved, and W. M. Stadler, "Estimating the arterial input function using two reference tissues in dynamic contrast-enhanced MRI studies: fundamental concepts and simulations," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 52, no. 5, pp. 1110–1117, 2004. (See page 91)

[234] A. Faranesh and T. Yankeelov, "Incorporating a vascular term into a reference region model for the analysis of DCE-MRI data: a simulation study," *Physics in Medicine & Biology*, vol. 53, no. 10, pp. 2617, 2008. (See page 91)

[235] R. N. Gunn, A. A. Lammertsma, S. P. Hume, and V. J. Cunningham, "Parametric imaging of ligand-receptor binding in PET using a simplified reference region model," *Neuroimage*, vol. 6, no. 4, pp. 279–287, 1997. (See page 91)

[236] P. D. Acton, S. A. Kushner, M.-P. Kung, et al., "Simplified reference region model for the kinetic analysis of [^{99m}Tc] TRODAT-1 binding to dopamine transporters in nonhuman primates using single-photon emission tomography," *European Journal of Nuclear Medicine*, vol. 26, pp. 518–526, 1999. (See page 91)

[237] M. Doury, A. Dizeux, A. De Cesare, et al., "Quantification of tumor perfusion using dynamic contrast-enhanced ultrasound: impact of mathematical modeling," *Physics in Medicine & Biology*, vol. 62, no. 3, pp. 1113, 2017. (See page 91)

[238] J. B. DeGrandchamp, J. G. Whisenant, L. R. Arlinghaus, et al., "Predicting response before initiation of neoadjuvant chemotherapy in breast cancer using new methods for the analysis of dynamic contrast enhanced MRI (DCE MRI) data," in *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*. SPIE, 2016, vol. 9788, pp. 242–251. (See pages 93, 95, and 98)

[239] V. Visentin, "Mise en place d'un modèle d'étude chez la souris de l'artériopathie oblitérante des membres inférieurs pour l'évaluation de stratégies thérapeutiques," M.S. thesis, Feb. 2012. (See page 93)

[240] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. (See page 95)

[241] F. De Chaumont, S. Dallongeville, N. Chenouard, et al., "Icy: an open bioimage informatics platform for extended reproducible research," *Nature Methods*, vol. 9, no. 7, pp. 690–696, 2012. (See page 107)