



# Modélisation et simulation de la croissance de métastases pulmonaires

Julien Jouganous

## ► To cite this version:

Julien Jouganous. Modélisation et simulation de la croissance de métastases pulmonaires. Modélisation et simulation. Université de Bordeaux, 2015. Français. NNT : 2015BORD0154 . tel-01222935

**HAL Id: tel-01222935**

**<https://theses.hal.science/tel-01222935>**

Submitted on 31 Oct 2015

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

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

# Thèse

pour l'obtention du grade de

**Docteur de l'université de Bordeaux**

École doctorale de mathématiques et d'informatique  
Mathématiques appliquées et calcul scientifique

**Julien JOUGANOUS**

---

MODÉLISATION ET SIMULATION DE LA CROISSANCE  
DE MÉTASTASES PULMONAIRES

---

sous la direction de : **Olivier SAUT**

*Soutenue le 23 Septembre 2015*

*Devant la commission d'examen :*

M. Hervé DELINGETTE  
M. Benoît PERTHAME  
M. Thierry COLIN  
M. Hassan FATHALLAH  
M. Bertrand MAURY  
M. Olivier SAUT

Directeur de Recherche  
Professeur  
Professeur  
Professeur  
Professeur  
Directeur de Recherche

Rapporteur  
Rapporteur  
Examineur  
Examineur  
Examineur  
Directeur





# Modélisation et simulation de la croissance de métastases pulmonaires

**Résumé :** Cette thèse présente des travaux de modélisation mathématique de la croissance tumorale appliqués aux cas de métastases pulmonaires.

La première partie de cette thèse décrit un premier modèle d'équations aux dérivées partielles permettant de simuler la croissance métastatique mais aussi la réponse de la tumeur à certains types de traitements. Une méthode de calibration du modèle à partir de données cliniques issues de l'imagerie médicale est développée et testée sur plusieurs cas cliniques.

La deuxième partie de ces travaux introduit une simplification du modèle et de l'algorithme de calibration. Cette méthode, plus robuste, est testée sur un panel de 36 cas test et les résultats sont présentés dans le troisième chapitre.

La quatrième et dernière partie développe un algorithme d'apprentissage automatisé permettant de tenir compte de données supplémentaires à celles utilisées par le modèle afin d'affiner l'étape de calibration.

**Mots-clés :** modélisation, simulation numérique, cancer, croissance tumorale, métastase pulmonaire, apprentissage automatisé, forêts aléatoires.

---

## Lung metastases growth modeling and simulation

**Abstract :** This thesis deals with mathematical modeling and simulation of lung metastases growth.

We first present a partial differential equations model to simulate the growth and possibly the response to some types of treatments of metastases to the lung. This model must be personalized to be used individually on clinical cases. Consequently, we developed a calibration technic based on medical images of the tumor. Several applications on clinical cases are presented.

Then we introduce a simplification of the first model and the calibration algorithm. This new method, more robust, is tested on 36 clinical cases. The results are presented in the third chapter.

To finish, a machine learning algorithm based on random forests is used to improve the calibration of the model.

**Keywords :** modeling, simulation, cancer, tumor growth, lung metastasis, machine learning, random forests.



## Remerciements

Tout d'abord, je souhaite remercier chaleureusement mon directeur de thèse Olivier Saut. Son aide et son implication ont été déterminantes tout au long de ces trois années, de la recherche de financements jusqu'aux relectures du manuscrit. Ses conseils éclairés dans tous les domaines (avec une mention spéciale pour l'informatique quelque soit le langage) m'ont permis de progresser et d'avancer sur le sujet passionnant qu'il m'a confié. Je tiens aussi à le remercier pour sa gentillesse et la patience sans bornes dont il a su faire preuve et pour tous les moments agréables partagés à l'occasion de conférences ou au sein du labo. Pour tout cela, j'ai du mal à imaginer meilleur directeur de thèse.

Comment ne pas aussi remercier Thierry Colin pour sa bonne humeur contagieuse et son expertise scientifique. Merci pour ces séances (souvent très productives) de calcul au tableau et réunions, toujours ponctuées de quelques blagues. Je pense pouvoir écrire sans trop me tromper qu'il est un pilier fondamental de la bonne ambiance qui règne dans le couloir de l'équipe Monc. . .

Un grand merci aussi à Benoit Perthame et Hervé Delingette qui ont accepté de rapporter mon manuscrit. Leurs nombreux conseils et remarques m'ont permis d'améliorer la présentation de mes travaux. Je tiens aussi à adresser tous mes remerciements à Bertrand Maury ainsi qu'à Hassan Fathallah-Shaykh pour leur participation à mon jury de thèse et pour l'intérêt qu'ils ont montré pour mes travaux.

J'aimerais rendre hommage à nos collaborateurs médicaux, plus particulièrement les Docteurs Jean Palussière (Institut Bergonié) et François Cornélis (CHU Pellegrin). Sans leur grande expertise médicale et leur implication dans le projet, celui-ci n'aurait très certainement pas pu voir le jour.

Je souhaite également remercier les autres membres de l'équipe Monc notamment Clair Poignard et Sébastien Benzekry pour leur gentillesse et leur disponibilité. J'ai aussi une pensée pour l'ensemble des doctorants, post-doctorants et ingénieurs avec qui j'ai eu le plaisir de partager mon bureau : Yannick, Jessica, Hervé, Xin, Thomas, Julie et plus spécifiquement Marie, membre de la bien nommée "team poumon". Je n'oublie pas non plus l'ensemble des membres de l'ancienne équipe MC2 (devenue Monc), permanents et non permanents, et je tiens à saluer l'ambiance exceptionnelle au sein de cette équipe. Je souhaite à tout doctorant d'avoir la chance d'évoluer dans les mêmes conditions de travail.

Je remercie ma famille et plus particulièrement mes parents et mon frère pour leur intérêt et leur soutien indéfectible pendant ces trois années de thèse ainsi que l'ensemble de ma scolarité. Enfin merci à Marine qui elle aussi m'a soutenu et encouragé dans ce projet et pour la joie et la complicité que nous partageons au quotidien.

# Sommaire

<b>1</b>	<b>Introduction</b>	<b>19</b>
1.1	Quelques mots sur le cancer . . . . .	19
1.1.1	L'apparition de la maladie [2] [5] . . . . .	19
1.1.2	Les différentes phases de la maladie . . . . .	24
1.1.3	Quelques données d'épidémiologie . . . . .	26
1.1.4	Les moyens de lutte contre la maladie . . . . .	26
1.1.5	Les spécificités des métastases pulmonaires . . . . .	30
1.2	Imagerie médicale . . . . .	31
1.2.1	Méthodes d'imagerie anatomique . . . . .	31
1.2.2	Méthodes d'imagerie fonctionnelle . . . . .	32
1.2.3	Utilisation de l'imagerie dans le cadre des métastases pulmonaires . . . . .	33
1.3	Problématique et motivations . . . . .	34
1.3.1	Présentation d'un cas clinique . . . . .	34
1.3.2	Orientation des travaux de cette thèse . . . . .	34
1.4	Travaux mathématiques sur le cancer . . . . .	37
1.4.1	Méthodes discrètes . . . . .	38
1.4.2	Modèles continus scalaires . . . . .	38
1.4.3	Méthodes continues spatiales . . . . .	40
1.5	Environnement logiciel . . . . .	42
<b>2</b>	<b>Étude d'un premier modèle de croissance métastatique</b>	<b>45</b>
2.1	Présentation du modèle . . . . .	45
2.1.1	Hierarchie de modèles spatiaux . . . . .	46
2.1.2	Modèle avec angiogenèse simplifié . . . . .	49
2.2	Simulation du modèle et résultats numériques . . . . .	52
2.2.1	Méthodes de simulation numérique . . . . .	52
2.2.2	Utilisation des images médicales dans le code de simulation	56
2.2.3	Premiers résultats de simulations . . . . .	57
2.3	Calibration du modèle : problème inverse . . . . .	59
2.3.1	Objectifs . . . . .	59

2.3.2	Premières hypothèses simplificatrices . . . . .	60
2.3.3	Analyse de sensibilité . . . . .	60
2.3.4	Résolution du problème inverse et optimisation . . . . .	62
2.4	Proper Orthogonal Decomposition (POD) . . . . .	64
2.4.1	Principe de la méthode . . . . .	64
2.4.2	Obtention du la formulation POD . . . . .	65
2.5	Modèle réduit par POD et simulations . . . . .	72
2.5.1	Modèle réduit . . . . .	72
2.5.2	Mise en place du schéma numérique . . . . .	73
2.6	Applications de la méthode . . . . .	76
2.6.1	Données cliniques nécessaires . . . . .	76
2.6.2	Base de données et extraction des modes POD . . . . .	77
2.6.3	Qualité de la base de données et des modes extraits . . . . .	78
2.6.4	Résolution du problème inverse . . . . .	81
2.6.5	Résultats de la méthode sur un premier cas complet . . . . .	81
2.6.6	Étude d'un second cas clinique : aspect spatial . . . . .	89
2.6.7	Étude d'un troisième cas clinique . . . . .	90
2.6.8	Considérations sur la robustesse de la méthode POD . . . . .	90
2.7	Utilisation de données 3D . . . . .	92
2.8	Conclusion . . . . .	94
<b>3</b>	<b>Modèle simplifié de croissance pour les métastases pulmo-</b>	
	<b>naires en vue d'applications cliniques</b>	<b>95</b>
3.1	Limites du modèle complet . . . . .	95
3.2	Simplifications du modèle de croissance tumorale . . . . .	96
3.2.1	Premiers essais numériques . . . . .	96
3.2.2	Transport de la tumeur par méthode de Front Tracking . . . . .	98
3.2.3	Analyse de sensibilité du modèle . . . . .	107
3.3	Simulations en 3 dimensions . . . . .	109
3.3.1	Formulation du problème . . . . .	109
3.3.2	Reconstruction du champ $T$ à partir des points de l'in-	
	terface . . . . .	110
3.3.3	Simulations . . . . .	119
3.3.4	Application de la méthode de front tracking 3D au cas	
	des méningiomes . . . . .	120
3.4	Réduction 0D du modèle pour les métastases pulmonaires . . . . .	124
3.5	Étude des bornes de l'espace de recherche . . . . .	128
3.6	Résultats cliniques . . . . .	132
3.6.1	Protocole d'étude des cas cliniques . . . . .	132
3.6.2	Premiers résultats sur un cas clinique . . . . .	134
3.7	Méthode de calibration scalaire . . . . .	142

3.8	Étude de la propagation des incertitudes de segmentation . . . .	146
3.8.1	Introduction de perturbations de forme . . . . .	146
3.8.2	Approche statistique . . . . .	147
3.9	Conclusion . . . . .	149
<b>4</b>	<b>Validation clinique</b>	<b>153</b>
4.1	Protocole d'étude . . . . .	153
4.1.1	Données cliniques utilisées . . . . .	153
4.1.2	Méthodologie . . . . .	154
4.2	Résultats . . . . .	154
4.3	Conclusion . . . . .	159
<b>5</b>	<b>Apprentissage automatisé pour la classification de l'agressivité tumorale</b>	<b>161</b>
5.1	Cadre de l'étude et vocabulaire . . . . .	161
5.2	Introduction aux forêts aléatoires . . . . .	163
5.2.1	Construction d'un arbre binaire aléatoire . . . . .	163
5.2.2	Classification du cas et prédiction . . . . .	172
5.2.3	Limites de l'approche de classification par un arbre binaire	173
5.2.4	De l'arbre à la forêt . . . . .	174
5.3	Utilisation des forêts aléatoires sur les cas de métastases pulmo- naires . . . . .	175
5.3.1	Protocole d'étude et base d'apprentissage . . . . .	175
5.3.2	Caractéristiques utilisées . . . . .	176
5.3.3	Utilisation des forêts aléatoires sur la base de données de cas cliniques . . . . .	179
5.4	Conclusion . . . . .	181
<b>6</b>	<b>Conclusion et perspectives</b>	<b>183</b>
<b>A</b>	<b>Implémentation d'un schéma de transport de Lax-Wendroff TVD</b>	<b>187</b>
A.1	Schéma de Lax-Wendroff dans un cas simplifié . . . . .	187
A.2	Schéma de Lax-Wendroff 2D avec vitesses non constantes . . . .	188
A.3	Limiteurs de flux . . . . .	190
A.4	Essais numériques . . . . .	191
<b>B</b>	<b>Suite des résultats de l'étude de cas cliniques</b>	<b>193</b>





# Table des figures

1.1	Présentation du cycle de vie cellulaire. Source : Futura-Sciences.	20
1.2	Étapes de la division cellulaire. Source : RSR Savoirs. . . . .	21
1.3	Utilisation du code génétique par la cellule : phases de transcription et traduction. Source : linternaute.com. . . . .	22
1.4	Schéma représentant le phénomène d'angiogenèse. Source : <a href="http://gfme.free.fr">http://gfme.free.fr</a> .	25
1.5	Invasion métastatique. Source : scientificamerican.com. . . . .	25
1.6	Extraits de biopsies pour un tissu pulmonaire sain (a), une tumeur primitive (b) et une métastase issue d'une tumeur primitive à la thyroïde (c). . . . .	29
1.7	Extrait d'une séquence de CT scans montrant l'évolution d'une métastase pulmonaire issue d'une tumeur primitive à la vessie. .	35
1.8	Critère RECIST : mesure de l'évolution tumorale (croissance ou réponse au traitement). . . . .	36
1.9	Simulation d'un ensemble de cellules présentant une forte adhésion. Source : [14] . . . . .	38
1.10	Déplacement des cellules lié à la prolifération, cas 1D. . . . .	41
2.1	Extrait de biopsie d'une métastase pulmonaire issue d'une tumeur primitive à la thyroïde. . . . .	46
2.2	Tracé des fonctions $\gamma_p$ , $\gamma_d$ ainsi que le taux d'évolution de la densité $\gamma_p - \gamma_d$ avec $\gamma_0 = 1,5$ , $\gamma_1 = 1$ , $M_{th} = 2$ et $K = 1,5$ . . . .	49
2.3	Présentation d'une maille de la grille cartésienne dans le cas 2D. La vitesse de transport $\mathbf{v} = (v^x, v^y)$ est calculée aux bords de la maille ; la composante en $x$ sur les côtés verticaux de la maille et la composante en $y$ sur les côtés horizontaux. Les autres variables ( $T$ , $S$ , $\gamma_p$ , $\gamma_d$ , $M$ et $\pi$ ), matérialisées par la lettre $\rho$ , sont positionnées aux centres des mailles. . . . .	53
2.4	Segmentation de la tumeur : à gauche l'image médicale avant la segmentation, à droite, l'image avec la segmentation de la tumeur.	56

2.5	Traitement des données extraites de la segmentation : à gauche le masque de la tumeur (format vtp) et à droite le champ $T$ reconstruit (format vtk) avec lissage des bords. Pour le champ $T$ l'échelle va de 1 en rouge à 0 en bleu. . . . .	57
2.6	Simulation 1 : évolution de la densité de cellules tumorales, maillage $200 \times 200$ . . . . .	58
2.7	Simulation 2 : évolution de la densité de cellules tumorales, maillage $200 \times 200$ . . . . .	58
2.8	Analyse de sensibilité sur le modèle de croissance tumorale par la méthode de Morris : effet de chaque paramètre sur le volume final de la tumeur. . . . .	62
2.9	Deux cas d'histogramme de répartition des masses tumorales finales dans les bases de données réduites . . . . .	78
2.10	Erreur relative L2 de la projection de la donnée initiale sur la base POD en fonction du nombre de modes. . . . .	80
2.11	Comparaison du modèle réduit POD et du modèle direct. . . . .	82
2.12	Cas 1 : Comparaison de l'évolution de la masse tumorale entre le modèle réduit, le modèle direct et les données cliniques lors de la phase de croissance. . . . .	84
2.13	Comparaison entre la prédiction et les observables : la ligne du haut représente les images médicales et celle du bas les simulations. La colonne de gauche correspond à l'examen du 22/09/2008 (calibration) et celle de droite à l'examen du 10/12/2008 (prédiction). . . . .	86
2.14	Cas 1 : Comparaison de l'évolution de la masse tumorale entre le modèle direct calibré et les données cliniques lors de la phase de décroissance due à la chimiothérapie puis de la rechute. . . . .	87
2.15	Comparaison entre la prédiction et les observables pour la phase de décroissance due au traitement : la ligne du haut représente les images médicales et celle du bas les simulations. De gauche à droite : le premier CT-scan de contrôle du 21/03/2009 (calibration), le second CT-scan de contrôle à la fin du traitement au 27/05/2009 (prédiction) et l'examen fait un mois et demi après la fin de traitement montrant la rechute au 27/07/2009 (prédiction). . . . .	88
2.16	Donnée clinique et résultat de la simulation à l'instant de la calibration $t = 599$ jours. . . . .	89
2.17	Donnée clinique et résultat de la simulation en prédiction à l'instant $t = 1251$ jours. . . . .	90

2.18	Troisième cas test : évolution de la masse et comparaison entre les données médicales et la simulation. . . . .	91
2.19	Évolution de la densité de cellules tumorales : cas tri-dimensionnel d'une métastase sphérique, maillage $100 \times 100 \times 100$ . . . . .	93
3.1	Simulation du modèle simplifié à partir du schéma WENO 5, cas présentant des hétérogénéités dans la distribution spatiale des cellules cancéreuses. Paramètres : $\eta = 3$ , $M_0 = 3$ , $M_1 = 0$ et $t = 3$ . . . . .	97
3.2	Extrait de CT-scan montrant une métastase pulmonaire issue d'une tumeur primitive de la vessie. . . . .	98
3.3	Représentation de l'interface $\Phi$ et des points de discrétisation notés $\times$ . . . . .	99
3.4	Calcul du winding number, notations. . . . .	100
3.5	Représentation de la frontière $\Phi$ dans le domaine de calcul $\Omega$ . La zone du domaine à tester pour la reconstruction du champ $T$ correspond au rectangle hachuré. . . . .	101
3.6	Méthode des courbes de Jordan : la demi-droite d'origine p coupe 3 fois le polygone (nombre impair), p est donc à l'intérieur du polygone ; la demi-droite d'origine q coupe 2 fois le polygone, q est donc à l'extérieur du polygone. . . . .	102
3.7	Interpolation bilinéaire des vitesses. Les noeuds du maillage $x$ correspondent avec le centre des pixels. . . . .	103
3.8	Exemple de croissance en couronne, en haut la densité de cellules tumorales T, en bas les zones proliférantes (produit $T \times M$ ) ; de gauche à droite : $t = 0$ , $t = 1,5$ et $t = 3$ . Paramètres : $M_0 = 2,5$ , $M_1 = 0,2$ , $\eta = 2$ . . . . .	104
3.9	Exemple de croissance exponentielle, en haut la densité de cellules tumorales T, en bas les zones proliférantes (produit $T \times M$ ) ; de gauche à droite : $t = 0$ , $t = 1,5$ et $t = 3$ . Paramètres : $M_0 = 1,1$ , $M_1 = 1,1$ , $\eta = 1$ . . . . .	105
3.10	Densités de cellules tumorales issues d'une simulation du modèle sur un cas test de polygone présentant une excroissance, à gauche la donnée initiale et à gauche le résultat à $t = 3$ ( $\eta = 0,12$ , $M_0 = 0,65$ et $M_1 = 0$ ). . . . .	106
3.11	Comparaison des formes finales à masses égales pour une condition initiale elliptique. À droite, croissance de type exponentielle ( $\eta = 2,5$ , $M_0 = 1$ et $M_1 = 0,8$ ) ; à gauche, croissance présentant une couche extérieure proliférante ( $\eta = 2,67$ , $M_0 = 3$ et $M_1 = 0$ ). . . . .	107

3.12	Analyse de sensibilité sur le modèle simplifié par la méthode de Morris : effet de chaque paramètre sur le volume final de la tumeur. . . . .	108
3.13	Présentation en vue étalée des points de deux couches consécutives de la segmentation d'une métastase fictive. . . . .	111
3.14	Illustration de la méthode de triangulation des faces latérales sur un cas fictif : première étape. . . . .	111
3.15	Deuxième étape de triangulation des faces latérales : la zone hachurée en rouge correspond à la surface restant à triangulariser.	112
3.16	Deuxième étape de triangulation des faces latérales : formation des deux triangles de la couche inférieure en gris foncé. . . . .	112
3.17	Illustration de la méthode de triangulation des faces latérales sur un cas fictif : résultat final obtenu. . . . .	112
3.18	Cas fictif de discrétisation de la couche supérieure de la tumeur.	113
3.19	Triangulation de la couche supérieure de la tumeur sur le cas fictif. . . . .	113
3.20	Schéma du triangle ABC. . . . .	114
3.21	Représentation de la projection du point $P(0,1;0,1;0,1)$ sur le plan $(A,B,C)$ suivant l'axe des ordonnées. . . . .	116
3.22	Coordonnées barycentriques $(r,s)$ du point $P'$ dans la base $(A, \overrightarrow{AB}, \overrightarrow{AC})$ .	116
3.23	Reconstruction d'une tumeur en 3 dimensions à partir d'un échantillonnage de la surface : en haut la triangulation de la surface, en bas le champ $T$ reconstruit. . . . .	118
3.24	Cas de raffinement de la triangulation par ajout d'un point. Le triangle $ABC$ est remplacé dans la liste des triangles par les triangles $ABG$ , $ACG$ et $BCG$ . . . . .	119
3.25	Simulation 3D basée sur la méthode de front tracking repositionnée dans la cage thoracique du patient. Les paramètres de la simulation sont les suivants : $M_0 = 0,6$ , $M_1 = 0$ et $\eta = 0,177$ .	121
3.26	Localisation des cellules proliférantes sur le cas 3D avec les paramètres $M_0 = 0,6$ ; $M_1 = 0$ et $\eta = 0,177$ . La zone rouge représente la zone la plus dense en cellules proliférantes. . . . .	122
3.27	Schéma présentant la disposition des tissus enveloppant le cerveau. Source : teachmeanatomy.info . . . . .	122
3.28	Extrait d'une séquence d'IRM sous injection de gadolinium montrant la présence d'un méningiome. . . . .	123

3.29	Simulation 3D de l'évolution d'un méningiome basée sur la méthode de front tracking. La tumeur est représentée en jaune et le crâne en gris clair. Les paramètres de la simulation sont les suivants : $M_0 = 0,6$ et $\eta = 0,177$ et les temps (adimensionnés) correspondant aux snapshots sont de haut en bas : $t = 0$ , $t = 2$ et $t = 4$ . . . . .	125
3.30	Évolution de $M$ en fonction du temps. Les paramètres sont, pour la courbe rouge : $M_0 = 1$ , $M_1 = 0$ et $\eta = 1$ (cas décroissant), pour la courbe bleue : $M_0 = 1$ , $M_1 = 2$ et $\eta = 1$ (cas croissant).	130
3.31	Diagramme synthétique décrivant le processus de calibration. . .	133
3.32	Scéance de CT-scans montrant l'évolution d'une métastase. . .	136
3.33	Premier cas test : Faisceau de courbes obtenu avec deux données.	137
3.34	Répartition des prédictions sur le premier cas test en utilisant deux points. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite noire correspond au maximum de la densité de répartition. . . . .	138
3.35	Premier cas test : Faisceau de courbes obtenu avec trois données.	139
3.36	Répartition des prédictions sur le premier cas test en utilisant trois examens. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite bleue correspond au maximum de la densité de répartition. . . .	140
3.37	Premier cas test : Faisceau de courbes obtenu avec quatre données.	141
3.38	Répartition des prédictions sur le premier cas test en utilisant quatre examens. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite bleue correspond au maximum de la densité de répartition. . . .	142
3.39	Utilisation de la méthode de calibration basée sur la formule scalaire du volume tumoral sur le cas de la Figure 3.33 : en rouge les 5 <sup>ème</sup> et 95 <sup>ème</sup> centiles (90% des simulations se situent entre ces deux courbes), en bleu la simulation moyenne et en noir la médiane. . . . .	144
3.40	Rendu spatial de la méthode de calibration scalaire : superposition des simulations sur l'image initiale de la tumeur. On retrouve à gauche l'image du poumon et à droite un zoom sur la tumeur. Le contour en rouge représente la simulation médiane et les contours jaunes délimitent la zone comprenant 90% des simulations. . . . .	145

3.41	Comparaison des simulations du modèle avec forme perturbée (à droite) et non perturbée (à gauche). La ligne du haut correspond aux conditions initiales ( $t=0$ ) et celle du bas aux résultats de simulation à $t=2$ . Le rayon initial est de 0,1cm et l'écart-type est $\sigma = 0,007\text{cm}$ . . . . .	148
3.42	Répartition des réalisations de la variable aléatoire $\Xi_1^0$ . Les estimateurs de la moyenne et de l'écart-type sont : $\overline{m} = -1,5 \times 10^{-3}$ , $\overline{\sigma} = 6,9 \times 10^{-2}$ . . . . .	149
3.43	Répartition des réalisations de la variable aléatoire $\Xi_1^2$ . Les estimateurs de la moyenne et de l'écart-type sont : $\overline{m} = -9,2 \times 10^{-3}$ , $\overline{\sigma} = 1,5 \times 10^{-2}$ . . . . .	150
3.44	Propagation de l'incertitude de contourage sur l'ensemble des points : à l'instant initial, puis aux instants $t = 1$ , $t = 1,5$ et $t = 2$ . . . . .	151
4.1	Résultat de la méthode sur les cas 1, 2, 3 et 4 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . . . .	155
4.2	Cas 5, 6, 7 et 8 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . . . .	156
5.1	Représentation de la base d'apprentissage test, les cas d'apprentissage dont la variable cible vaut 1 sont représentés par une croix, ceux dont la variable cible vaut 0 sont marqués d'un rond. Les deux caractéristiques utilisées sont les coordonnées $x$ et $y$ . . . . .	165
5.2	Découpe optimale en $x$ de l'espace des caractéristiques pour la base d'apprentissage du cas test selon le critère de Gini, $d = 0,04$ . . . . .	167
5.3	Découpe optimale en $x$ de l'espace des caractéristiques pour la base d'apprentissage du cas test selon le critère $\Gamma$ , $d = 0,425$ . La droite rouge représente la découpe optimale selon le critère, la courbe bleue représente l'évolution en fonction de $x$ du critère $\Gamma$ . . . . .	169
5.4	Première découpe : formation des deux premières branches de l'arbre. Critère de séparation $y = 0,385$ . . . . .	170
5.5	Découpe de l'espace des caractéristiques du cas test correspondant à l'arbre binaire construit. . . . .	172
5.6	Positionnement du cas B dans l'espace des caractéristiques dans le cas de la découpe correspondant à l'arbre initial, puis de la découpe correspondant à un arbre perturbé. . . . .	174
5.7	Position dans le plan $(a,b)$ de jeux de paramètres du modèle de Gompertz pour une population de métastases pulmonaires. . . . .	177

A.1	Grille cartésienne bidimensionnelle. Les noeuds du maillage correspondent avec le centre des pixels. . . . .	189
A.2	Simulation numérique de transport d'une ellipse à vitesse constante utilisant le schéma de Lax-Wendroff 2D non TVD. À gauche la condition initiale, à droite le champ transporté. . . . .	191
A.3	Simulation numérique de transport d'une ellipse à vitesse constante utilisant le schéma de Lax-Wendroff 2D TVD. À gauche la condition initiale, à droite le champs transporté. . . . .	192
B.1	Cas 9, 10, 11 et 12 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	194
B.2	Cas 13, 14, 15 et 16 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	195
B.3	Cas 17, 18, 19 et 20 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	196
B.4	Cas 21, 22, 23 et 24 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	197
B.5	Cas 25, 26, 27 et 28 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	198
B.6	Cas 29, 30, 31 et 32 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	199
B.7	Cas 33, 34, 35 et 36 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . .	200
B.8	Cas 37, à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ». . . . .	201





# Liste des tableaux

2.1	Paramètres utilisés pour les simulations présentées en Figures 2.6 et 2.7. . . . .	59
2.2	Espace de paramètres utilisé pour l'analyse de sensibilité. . . . .	61
2.3	Jeu de paramètres obtenu par le problème inverse. . . . .	83
2.4	Indicateurs spatiaux pour la phase de croissance du premier cas clinique : DICE et concordance en volume. . . . .	85
2.5	Indicateurs spatiaux pour la phase de décroissance du premier cas clinique : DICE et concordance en volume. . . . .	88
2.6	Jeu de paramètres utilisé pour la simulation 3D. . . . .	93
3.1	Espace de paramètres utilisé pour l'analyse de sensibilité du modèle simplifié. . . . .	108
3.2	Données statistiques brutes de la série de prédictions en mm <sup>2</sup> . La surface réelle au troisième examen est de 31,06 mm <sup>2</sup> . . . . .	135
3.3	Données statistiques brutes de la série de prédictions en mm <sup>2</sup> . La surface réelle au quatrième examen est de 50,82 mm <sup>2</sup> . . . . .	140
3.4	Données statistiques brutes de la série de prédictions en mm <sup>2</sup> . La surface réelle au cinquième examen est de 70,36 mm <sup>2</sup> . . . . .	141
3.5	Jeu de paramètres utilisé pour l'étude de propagation des incertitudes de segmentation. . . . .	147
3.6	Moyenne et écart-type de l'écart à la position non perturbée sur l'ensemble des points à l'instant initial et aux instants $t = 1$ , $t = 1,5$ et $t = 2$ . . . . .	152
4.1	Résultats synthétiques de l'étude des cas cliniques (les valeurs sont données en mm <sup>2</sup> sauf pour l'erreur relative en %). . . . .	158
5.1	Résultat du test de la méthode des forêts aléatoires sur la base de données de métastases pulmonaires. De la gauche vers la droite : identifiant de la tumeur, résultat de l'algorithme, prédiction (arrondi du résultat), valeur cible. . . . .	180



# Chapitre 1

## Introduction

### 1.1 Quelques mots sur le cancer

#### 1.1.1 L'apparition de la maladie [2] [5]

Le cancer est un des grands enjeux de santé publique du XXI<sup>ème</sup> siècle, amplifié par l'allongement de l'espérance de vie mais aussi les modifications de notre environnement. Quoiqu'étant l'une des maladies les plus étudiées à l'heure actuelle, de nombreux mécanismes impliqués dans l'apparition et le développement du cancer restent à éclaircir.

Son origine est due aux effets de mutations génétiques provoquant des dysfonctionnements au niveau cellulaire. Les principales caractéristiques des cellules cancéreuses par rapport aux cellules saines sont d'une part une vitesse de division accrue, et d'autre part, la disparition de l'apoptose ou mort cellulaire programmée. Ces deux caractéristiques semblent liées à l'apparition d'une insensibilité de la cellule aux différents signaux et mécanismes régissant la vitesse de multiplication et bloquant la prolifération dans l'organisme. Une tumeur n'est alors rien d'autre qu'un amas de cellules cancéreuses présentant ces comportements ou phénotypes anormaux. Une théorie majoritairement admise dans le monde scientifique est l'origine commune ou monoclonale des cellules malades formant un cancer (voir [70]). Autrement dit, l'ensemble des cellules d'une même tumeur serait, à l'origine, issu d'une même cellule mère à force de divisions. De plus, les cellules cancéreuses sont capables de se déplacer dans l'organisme, via notamment les systèmes sanguin et lymphatique, et coloniser des tissus distants de la tumeur initiale. Cependant, l'origine monoclonale a été récemment remise en question.

Afin de comprendre comment une simple mutation génétique peut engendrer de telles conséquences au niveau macroscopique, il est important de s'intéresser au cycle de vie de la cellule (voir Figure 1.1). Ce cycle est divisé en quatre

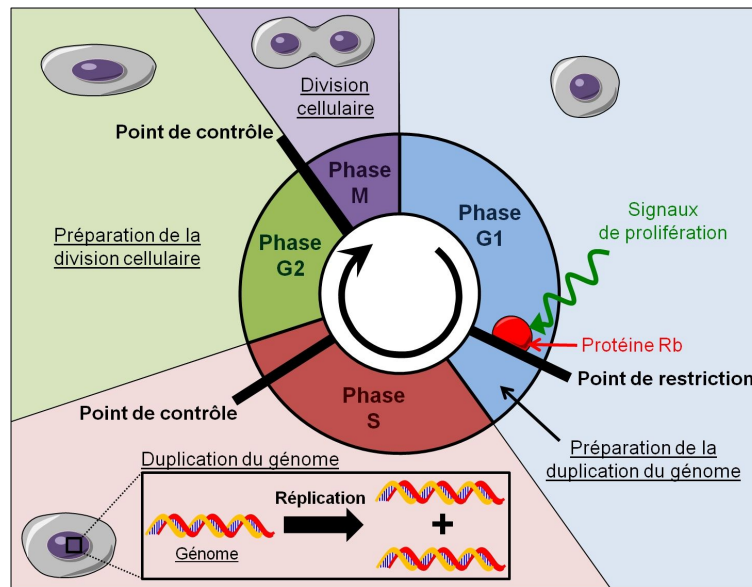


FIGURE 1.1 – Présentation du cycle de vie cellulaire. Source : Futura-Sciences.

phases :  $G_1$ ,  $S$ ,  $G_2$  et  $M$ .

- La phase  $G_1$  est une phase préparatoire à la phase  $S$  de duplication de l'ADN (acide désoxyribonucléique). Durant cette phase, la cellule fonctionne normalement et grossit jusqu'à atteindre une taille critique déclenchant la phase suivante. Afin de se préparer à dupliquer son matériel génétique, la cellule synthétise des protéines notamment la protéine Rb. Une fois le **point de restriction** de la phase  $G_1$  atteint, la cellule n'a plus besoin des facteurs de croissances extérieurs et peut proliférer indifféremment de ceux-ci. Avant la transition vers la phase  $S$ , un point de contrôle permet à la cellule de vérifier si son ADN ne présente pas d'altérations et peut poursuivre son cycle.
- La phase  $S$  correspond à la synthèse d'une copie de l'ADN de la cellule. Une enzyme, l'ADN polymérase duplique ainsi le patrimoine génétique de la cellule. Les deux brins complémentaires de la molécule d'ADN sont séparés et chacun d'entre eux sert de modèle pour la formation d'un nouveau brin. L'ADN polymérase « lit » le brin d'ADN et en reconstitue la séquence complémentaire. En effet les nucléotides, molécules faisant partie de l'ADN et dont la séquence définit le code génétique, s'associent par paires : l'adénine (A) avec la thymine (T) d'une part, et la guanine (G) avec la cytosine (C). Schématiquement, l'enzyme ajoute un A en face d'un T ou inversement, et un G en face d'un C.
- La phase  $G_2$ , tout comme la phase  $G_1$  d'ailleurs, matérialise un in-

tervalle (gap). Pendant cette phase, la cellule se prépare à la division cellulaire ou mitose. La cellule ne grossit plus mais poursuit son fonctionnement métabolique. Un nouveau point de contrôle jalonne la fin de cette phase avant le début de la mitose.

- La **mitose** (représentée plus en détails à la Figure 1.2), ou division cellulaire, est représentée par la phase *M*. Il s'agit d'une duplication non sexuée (contrairement à la méiose) d'une cellule mère en deux cellules filles. La mitose se divise en plusieurs phases. Durant la **prophase** et la **prométaphase** les différents chromosomes (chacun formé de deux chromatides) s'individualisent et la membrane nucléaire disparaît. Durant la **métaphase**, les chromosomes s'alignent dans un plan partageant la cellule en deux (on parle d'alignement à l'équateur). Chaque chromosome se divise en deux chromatides qui vont chacune migrer séparément vers les pôles opposés : il s'agit de l'**anaphase**. Pour finir, la **télophase** met fin au partage de la cellule avec la reconstitution de la membrane autour de chaque lot de chromatides.

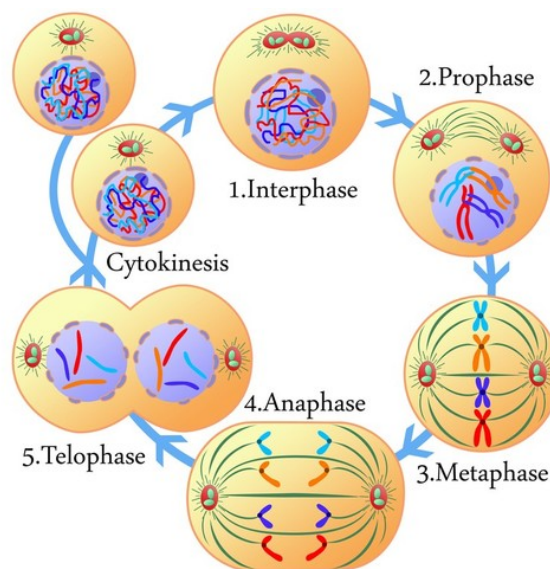


FIGURE 1.2 – Étapes de la division cellulaire. Source : RSR Savoirs.

La fidélité de réplication de l'ADN lors de la phase *S* est très grande grâce notamment à la relecture de la chaîne par les ADN polymérases leur permettant de vérifier que le dernier nucléotide inséré est bien le bon. Cette relecture ainsi que les différents points de contrôle à franchir au cours du cycle font chuter le taux d'erreur d'environ  $10^{-5}$ , c'est à dire une erreur sur cent mille nucléotides assemblés, à environ  $10^{-7}$ . Cela dit, le génome humain étant composé d'environ 3,2 milliards de paires de nucléotides, l'apparition de ce genre

d'erreurs de réplication est très probable durant un cycle cellulaire et courante tout au long de la vie d'un être humain.

Afin de comprendre les implications de tels changements dans la molécule d'ADN, il faut comprendre comment est utilisé le code génétique dans le fonctionnement de la cellule (voir Figure 1.3).

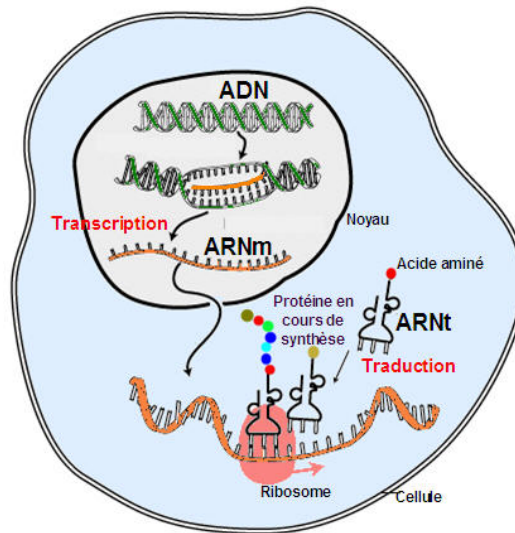


FIGURE 1.3 – Utilisation du code génétique par la cellule : phases de transcription et traduction. Source : linternaute.com.

L'ADN n'est pas directement lu par la cellule dans le noyau. Il est tout d'abord transcrit en une molécule intermédiaire, l'ARNm (acide ribonucléique) qui sert de messenger entre l'ADN confiné au noyau, et le reste de la cellule. Cette **transcription** se fait au sein du noyau par complétion du brin d'ADN, de façon un peu similaire à ce qu'il se passe pendant la phase *S* du cycle cellulaire. Durant cette transcription, il peut, là encore, apparaître des erreurs dans le code. Une fois la molécule d'ARNm synthétisée, celle-ci migre au travers de pores au niveau de la membrane nucléaire pour atteindre le cytoplasme extra-nucléaire. L'ARNm est alors traduit par des organites de la cellule : les ribosomes. Ces derniers parcourent la molécule d'ARNm et lisent les nucléotides trois par trois (on parle aussi de codons). Chaque codon correspond à un des 22 acides aminés ou à un codon-stop indiquant la fin de la chaîne moléculaire. Le ribosome assemble les acides aminés les uns à la suite des autres dans l'ordre défini par l'ARN et synthétise ainsi une protéine qui sera utilisée dans le métabolisme de la cellule. Un gène est une portion d'ADN, c'est à dire un ensemble de nucléotides, codant pour une même protéine.

La plupart des mutations issues des erreurs de réplication n'ont aucune répercussion sur le phénotype de la cellule fille car on considère qu'une majorité du génome humain est composée de séquences ne codant pas pour des gènes. Certaines mutations vont cependant pouvoir toucher un gène. La protéine codée par ce gène peut alors être modifiée dans sa configuration spatiale et ne plus remplir sa fonction initiale. Cela peut avoir pour conséquence la mort de la cellule ou encore une modification dans son fonctionnement. Ces modifications ne sont pas nécessairement négatives pour l'organisme et peuvent même être à l'origine d'évolutions avantageuses (principe même de l'évolution des espèces). Cependant l'apparition d'erreurs touchant les gènes qui régulent la prolifération peut influencer sur les premières phases du cycle cellulaire (*G1* et *S*). Cela peut avoir pour conséquence d'augmenter la capacité de prolifération de la cellule. De plus, la croissance et la prolifération des cellules sont régulées par diverses molécules lues par des récepteurs membranaires. Les mutations impliquées dans l'apparition de cancers ont souvent des répercussions sur la chaîne permettant à ce signal régulateur d'atteindre la cellule.

En plus de cette prolifération accrue, il est important de mentionner l'altération de l'apoptose, c'est à dire de la mort cellulaire programmée. Ce phénomène intervient pour réguler la population de cellules de l'organisme et surtout pour limiter l'instabilité génétique. Ainsi, en fonctionnement normal, lorsqu'une altération de l'ADN se produit dans la cellule, celle-ci meurt automatiquement. Si ce phénomène est enrayé par une mutation, la protection qu'il assure disparaît et le terrain devient propice au développement du cancer.

Une fois que la cellule ne répond plus à l'apoptose et présente une vitesse de duplication accrue, elle se multiplie de nombreuses fois en propageant son matériel génétique altéré aux générations suivantes. Il en résulte l'apparition d'une tumeur, amas de cellules issues de ce même clone initial.

Il existe une très grande variété de cancers. Pour les plus courants, on peut distinguer les carcinomes qui touchent un épithélium (surface uniquement composée de cellules), les sarcomes qui sont des cancers proliférant dans des tissus conjonctifs (comme les os), ou encore les cancers hématopoïétiques qui concernent les cellules sanguines (leucémies, lymphomes...). On classe aussi les cancers en fonction de l'organe touché : sein, côlon, poumon, pancréas, moelle osseuse (myélome), peau (mélanome)... Il existe en réalité presque autant de sources de cancer que de variétés de tissus dans l'organisme, cependant certains sont plus courants que d'autres tout comme certains sont plus malins que d'autres.



### 1.1.2 Les différentes phases de la maladie

À partir d'une seule cellule, la maladie peut, si elle n'est pas prise en charge suffisamment tôt, s'étendre à d'autres parties du corps du patient. On parle alors de cancer généralisé. Cela dit, avant d'en arriver à ce stade, le cancer passe par plusieurs phases de développement bien distinctes.

**La phase avasculaire.** Il s'agit de la première phase de la maladie. Pour croître, une tumeur a besoin de place mais aussi d'un apport suffisant en nutriments et dioxygène. Durant la phase avasculaire, la tumeur est de petite taille, de quelques cellules à un amas d'environ  $1\text{mm}^3$  et se contente donc des nutriments qui lui parviennent par diffusion via les capillaires les plus proches. Ces nutriments ne peuvent pas pénétrer au cœur de la tumeur et sont généralement consommés par les cellules cancéreuses périphériques. Il peut donc arriver que les cellules du centre meurent faute d'approvisionnement et on voit apparaître un cœur nécrosé.

**La phase vasculaire et l'angiogenèse.** Lorsque la tumeur a atteint sa limite de croissance avasculaire, elle développe un réseau sanguin propre afin d'être davantage alimentée et de poursuivre sa croissance. Ce phénomène de néovascularisation clé dans la croissance tumorale est appelé angiogenèse.

L'angiogenèse (voir [15] ou [29]), dont un schéma est présenté en Figure 1.4, est un mécanisme normal dans le développement d'un organisme et nécessaire notamment durant la croissance embryonnaire. Cependant certaines cellules cancéreuses présentent une capacité anormalement prononcée à susciter l'angiogenèse et il s'agit alors d'un processus pathologique. En pratique, les cellules tumorales sécrètent une variété d'agents chimiques qui vont provoquer la fabrication des nouveaux vaisseaux sanguins émergeant du système vasculaire pré-existant en direction de la tumeur. Un des agents pro-angiogéniques les plus connus est le VEGF (Vascular Endothelial Growth Factor) mais il s'agit d'un phénomène complexe dont les voies moléculaires sont nombreuses et encore un thème de recherche majeur.

**L'invasion métastatique.** Si la vascularisation de la tumeur lui permet de croître, elle a aussi pour effet de permettre la dissémination des cellules malades dans l'ensemble du corps. Si ces dernières arrivent à s'implanter dans de nouveaux tissus, elles vont pouvoir s'y développer. C'est l'invasion métastatique (voir Figure 1.5). Quel que soit l'organe dans lequel elle se développe, la métastase est composée de cellules identiques à celles de la tumeur primaire. Cette phase complique considérablement le traitement puisque les nodules peuvent

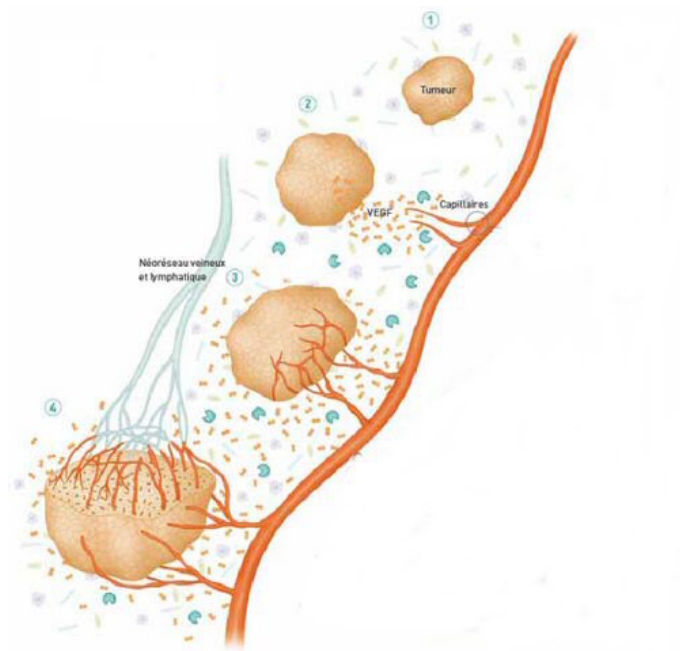


FIGURE 1.4 – Schéma représentant le phénomène d'angiogenèse. Source : <http://gfme.free.fr>.

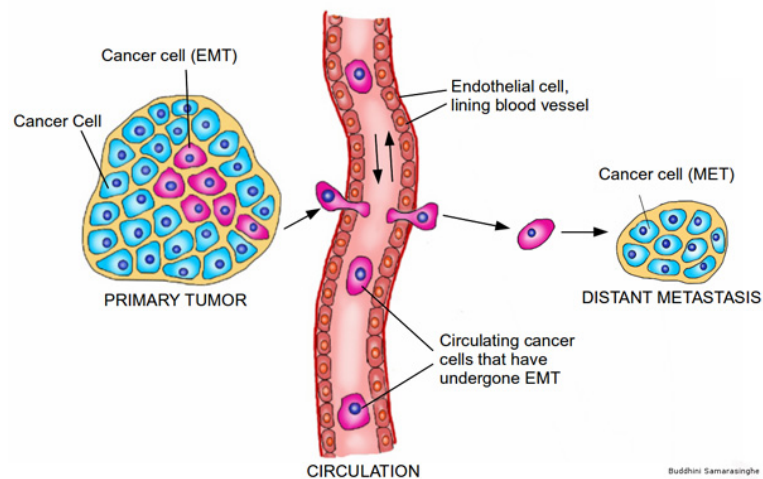


FIGURE 1.5 – Invasion métastatique. Source : [scientificamerican.com](http://scientificamerican.com).

être nombreux dans un organe atteint et les lésions engendrées par des opérations chirurgicales à répétition sont souvent autant de menaces pour la vie du patient.

Il est intéressant de noter que certains organes sont plus souvent touchés par le développement métastatique que d'autres. Ils semblent être un bon terrain d'accueil. De même certains types de tumeurs tendent à métastaser dans des organes particuliers. Ainsi, le cancer du poumon métastase souvent dans le foie. De même, les cancers primitifs du foie et de la vessie provoquent préférentiellement des métastases au poumon.

### 1.1.3 Quelques données d'épidémiologie

De nombreuses études épidémiologiques sont menées sur ce sujet chaque année. Dans le cas de la France les chiffres sont éloquents. Pour l'année 2011, on estime à environ 365000 le nombre de nouveaux cas (soit à peu près 1000 par jour). Les hommes sont les plus touchés avec 207000 cas contre 158000 pour les femmes. Parmi les cancers les plus courants, on retrouve les cancers de la prostate et du poumon chez l'homme et celui du sein chez la femme.

Toujours en 2011 en France, 147500 décès ont pu être attribués au cancer et si l'on considère la période 2004-2008, le cancer est la première cause de mortalité chez l'homme et la seconde chez la femme, le cancer du poumon restant de loin le plus meurtrier (près de 30000 décès contre un peu moins de 20000 pour le cancer côlon-rectal et moins de 12000 pour le cancer du sein).

Si l'incidence (occurrence dans la population) de la maladie ne cesse d'augmenter, peut-être en partie parce que les techniques de dépistage sont en constante progression, on note cependant que la mortalité diminue globalement. Quoique cela soit plus ou moins bien vérifié suivant le type de cancer. En ce qui concerne la survie (à 5 ans) des patients, les données sont très hétérogènes, pouvant varier de 6 à 95% selon le type de cancer mais aussi les conditions de dépistage.

### 1.1.4 Les moyens de lutte contre la maladie

Les méthodes de traitement du cancer sont nombreuses et les progrès sont réels qu'il s'agisse de guérison totale ou de ralentissement de la maladie. Cependant, s'il existe une tendance générale, les chances de guérison peuvent varier du tout au tout selon les différents types de cancer. Il faut aussi avoir à l'esprit que l'efficacité d'un traitement est d'autant plus élevée que le diagnostic est précoce.

**Chirurgie ablative.** Il s'agit de la solution la plus directe puisqu'elle consiste en l'ablation de la tumeur et éventuellement d'une partie des tissus alentours. La chirurgie n'est pas toujours réalisable. Parfois les séquelles engendrées par une intervention seraient trop lourdes (cas de tumeurs intra-crâniennes) ou le patient est trop faible pour se remettre d'une opération lourde. Ce peut être le cas de personnes âgées ou lorsque les métastases sont trop nombreuses. De plus certains cancers comme ceux touchant les cellules sanguines ne sont pas localisés dans un tissu et sont donc inopérables.

**Chimiothérapies cytotoxiques.** La chimiothérapie consiste en un traitement médicamenteux qui fonctionne par arrêt de la mitose et cible donc les cellules ayant tendance à se reproduire rapidement. Ces substances endommagent les cellules, c'est pourquoi elles sont dites cytotoxiques. Les principaux inconvénients de ce type de traitements sont liés à leurs nombreux effets secondaires (fatigues extrêmes, nausées, perte du système pileux...) puisque les cellules cancéreuses ne sont pas les seules à être touchées. Certaines variantes sont à l'étude pour remédier à cet inconvénient comme la photochimiothérapie qui consiste en l'administration de substances chimiques devenant toxiques lorsqu'elles sont exposées à une source lumineuse. De même l'électroporation est une méthode permettant de rendre perméables certaines cellules sous l'effet d'un champ magnétique et d'y faire pénétrer les substances toxiques avec l'avantage de cibler localement les zones à traiter.

**Thérapies ciblées.** Chaque tumeur présente un profil génétique qui lui est propre. Une voie de recherche actuelle consiste à cibler ces mutations spécifiques afin d'accroître l'efficacité du traitement et d'éviter certains effets secondaires de la chimiothérapie notamment dans le cas de cancers du poumon.

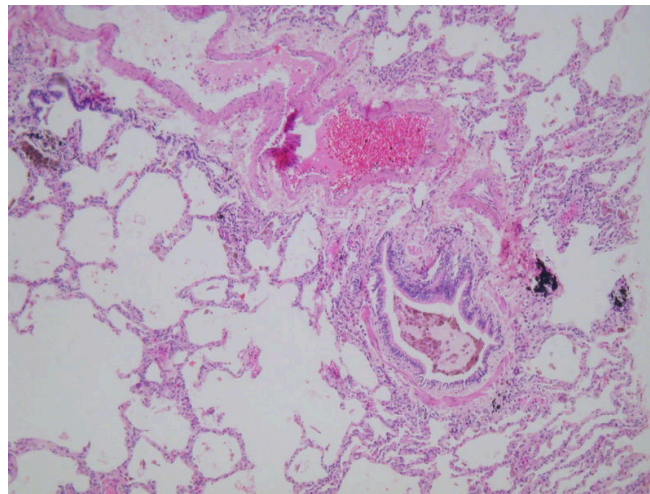
**Traitements anti-angiogéniques.** Il s'agit d'un type de thérapie ciblée. Ces traitements ciblent certains mécanismes propres aux cellules cancéreuses. Comme on l'a vu, l'angiogenèse est un phénomène nécessaire pour la croissance de la tumeur à un certain stade de son développement. Les traitements dits anti-angiogéniques visent à contrarier l'angiogenèse tumorale. Ils peuvent agir sur l'émission des agents angiogéniques par la tumeur ou inhiber les récepteurs tyrosine kinase sensibles au VEGF présents sur les cellules endothéliales. Dans tous les cas les effets de ces traitements sont complexes. On note souvent que le principal effet d'un anti-angiogénique est de normaliser la vascularisation de la tumeur. Ceci est paradoxal puisqu'on pourrait espérer, en inhibant l'angiogenèse, réduire l'approvisionnement de la tumeur alors que dans les fait on l'améliore. Cependant, utilisé avec d'autres traitements (chimiothérapie par

exemple), cela permet d'augmenter les performances de ceux-ci puisqu'ils atteignent ainsi plus efficacement la tumeur. Certains de ces traitements anti-angiogéniques présenteraient, en outre, des effets cytotoxiques.

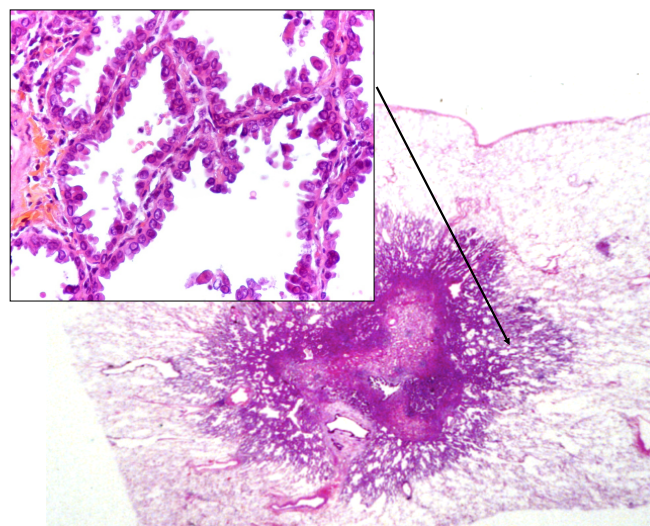
**Cryothérapie.** La cryothérapie utilise un froid intense pour détruire les cellules cancéreuses. Elle est initialement utilisée pour les cancers de la peau facilement accessibles. Plus récemment, on l'utilise sur d'autres types de tumeurs comme le cancer de la prostate, du foie ou du poumon. Dans ces cas, cela nécessite une intervention chirurgicale via une sonde dirigée par imagerie médicale afin de cibler au mieux les cellules malades.

**Hyperthermie.** Au contraire de la cryothérapie, l'hyperthermie consiste en la destruction de cellules par une élévation de la température. Il s'agit bien souvent de protocoles non invasifs utilisant des radiations (radiofréquences). Cela permet d'éviter la chirurgie chez des patients trop faibles. Cependant, ces méthodes posent de nouveaux problèmes technologiques (tracking) puisqu'il faut viser continuellement la tumeur alors même que celle-ci peut être en mouvement. C'est particulièrement vrai avec les tumeurs pulmonaires puisque les mouvements respiratoires obligent à repositionner le faisceau en temps réel. Les difficultés des méthodes par hyperthermie et par cryothérapie relèvent de la difficulté à cibler l'ensemble de la tumeur sans trop détériorer les tissus alentours.

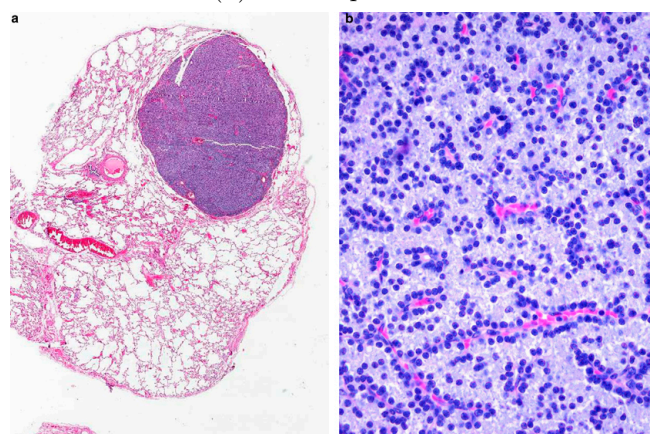
**Thérapies géniques.** Les thérapies géniques sont une nouvelle catégorie de traitement qui représente un réel espoir pour les maladies génétiques ainsi que les cancers. Dans le cas des cancers, les thérapies géniques visent directement la source de la maladie, les gènes ayant muté et à la base de la maladie appelés oncogènes. Plusieurs voies sont envisageables. On peut tenter de remplacer le ou les gènes défectueux par une copie saine. Une autre possibilité est de ré-orienter le système immunitaire contre les cellules cancéreuses afin de les détruire ou encore d'introduire des gènes « suicidaires » dans les cellules malades. Dans ce dernier cas, l'expression de ces gènes, combinée à certains traitements médicamenteux, mènerait à la mort de la cellule. Dans chacun des cas, on modifie le patrimoine génétique même de la tumeur. C'est ce qui fait la force mais aussi la difficulté de l'approche. Ceci se fait généralement grâce à l'utilisation d'un virus capable de pénétrer la membrane cellulaire auquel on administre les gènes souhaités.



(a) histologie normale



(b) tumeur primitive



(c) métastase (cancer de la thyroïde)

FIGURE 1.6 – Extraits de biopsies pour un tissu pulmonaire sain (a), une tumeur primitive (b) et une métastase issue d'une tumeur primitive à la thyroïde (c).

### 1.1.5 Les spécificités des métastases pulmonaires

Les poumons sont une zone de contact privilégiée entre le sang et l'environnement extérieur. La vascularisation y est très dense et ramifiée afin d'optimiser la surface des vaisseaux et ainsi les échanges gazeux y sont facilités. Cette particularité en fait un terrain d'accueil très propice à l'implantation de métastases provenant d'une tumeur distante. De plus, une grande partie du poumon est composée d'air et les métastases y ont suffisamment de place pour se développer.

La plupart des métastases pulmonaires proviennent de cellules cancéreuses pénétrant le poumon via l'artère pulmonaire et se dispersant ensuite dans les capillaires alvéolaires. Plus rarement, l'apparition de métastases peut être liée à une invasion par le système lymphatique ou à un développement de cellules cancéreuses directement dans l'arbre bronchique [59]. Les chances de survie de ces cellules étrangères au poumon sont assez réduites et directement corrélées à la qualité d'accueil de l'environnement [76]. Cependant, celles qui survivent, pénètrent l'endothélium, c'est à dire la couche interne des vaisseaux sanguins, et s'installent dans le parenchyme pulmonaire (tissu fonctionnel du poumon) où elles se développent.

Ainsi, des métastases pulmonaires sont souvent découvertes chez des patients atteints de cancer, quelque soit la localisation de la tumeur primitive et le pronostic lié est souvent assez pessimiste. Cependant, la présence limitée de métastases pulmonaires peut maintenant être traitée avec succès. En contrôlant la tumeur primitive et si bien sûr la maladie ne s'est pas généralisée à l'ensemble des organes du patient, l'ablation des métastases au poumon peut prolonger conséquemment la vie du patient, mais aussi améliorer sa qualité de vie et parfois même permettre de soigner la maladie définitivement [78]. Au cours des dix dernières années, les nombreuses évolutions, notamment dans les domaines de l'imagerie médicale ou des techniques chirurgicales, ont permis d'améliorer la prise en charge de la maladie. On peut ainsi citer des progrès dans la qualité des images et les temps d'acquisition des CT-scans [58] ou encore des avancées en médecine nucléaire et IRM [53] permettant d'obtenir davantage de précisions quant à la localisation ou la géométrie de la tumeur. En particulier, l'utilisation conjointe des modalités PET et CT du scan s'est répandue pour les patients atteints de métastases pulmonaires et permet d'observer de façon précoce une métastase via son activité métabolique [48].

La Figure 2.9 représente des données histologiques issues de biopsies. On peut y voir la différence entre le tissu sain, les tumeurs primitives présentant un caractère infiltrant et les métastases pulmonaires plus compactes (tumeurs solides) et homogènes. Ainsi, dans le cas des métastases, les cellules cancéreuses restent confinées dans la tumeur et ne se diffusent pas à travers l'organe. Cette

caractéristique, qu'il faudra prendre en compte lors de la modélisation, permet d'éradiquer la tumeur avec des taux de réussite satisfaisants, qu'il s'agisse d'ablations chirurgicales ou de méthodes moins intrusives comme l'utilisation des radiofréquences. Les voies de traitement médicamenteuses peuvent aussi donner de bons résultats, qu'il s'agisse de chimiothérapies cytotoxiques, de traitements anti-angiogéniques ou encore de thérapies ciblées (voir [66, 72, 73]).

## 1.2 Imagerie médicale

Les différentes avancées en matière de prise en charge et de compréhension du cancer ne peuvent pas être envisagées sans être reliées aux progrès incessants dans le domaine de l'imagerie médicale. La multiplication des modalités d'examen disponibles et l'amélioration de la qualité des images qui en sont issues jouent un rôle prépondérant dans le dépistage de la maladie, le suivi du patient, mais aussi la compréhension des mécanismes biologiques liés au cancer.

L'imagerie médicale au sens large consiste en l'acquisition de signaux divers (champs magnétiques, rayonnements...) et leur interprétation afin d'obtenir une visualisation d'une information à caractère médical. On peut classer les méthodes d'imagerie en deux grandes familles. L'imagerie dite **structurelle** permet de visualiser des données purement anatomiques comme la taille ou la forme d'un organe, des différences de tissus... Les méthodes d'imagerie dites **fonctionnelles** mettent en évidence le fonctionnement métabolique des tissus. Elles peuvent par exemple distinguer plusieurs zones suivant leur consommation en glucose ce qui sert pour détecter des tumeurs par exemple. On tend de plus en plus à utiliser ces deux types conjointement afin de pouvoir disposer d'une information la plus complète possible.

Dans ce qui suit, on va aborder quelques unes des méthodes les plus utilisées dont celles dont on peut espérer obtenir des données exploitables pour la modélisation mathématique.

### 1.2.1 Méthodes d'imagerie anatomique

**Échographie.** L'échographie consiste en une cartographie des tissus via la mesure de la propagation d'ultrasons. Elle donne généralement une image en deux dimensions. On l'utilise en cancérologie notamment pour diagnostiquer les cancers du sein.

**Radiographie.** Cette fois on utilise des rayons X (et éventuellement des produits de contraste) qui une fois projetés à travers le patient donnent une



image sur un plan bi-dimensionnel. On l'utilise principalement pour observer le système osseux du patient puisque c'est celui qui apparaît le plus contrasté à l'image.

**Tomodensitométrie (ou scanner).** Là encore on utilise des rayons X. Cependant l'acquisition se fait cette fois par coupes millimétriques grâce à un balayage du patient par des faisceaux de rayons. Cela permet, après traitement informatique, d'obtenir des vues 2D ou 3D des organes visés. Le CT-scan (computerized tomography scan) permet ainsi de distinguer les tissus selon leurs densités.

**Imagerie magnétique.** L'imagerie par résonance magnétique repose sur l'alignement des moments magnétiques de spin des atomes (par exemple d'hydrogène) constituant les tissus par un puissant champ magnétique. Ces atomes sont ensuite mis en résonance par des ondes radio. C'est le signal émis lors de la relaxation des atomes qui est enregistré par la machine puis interprété pour reconstruire une image 2D ou 3D.

L'IRM de diffusion est une variante qui permet de mesurer le tenseur de diffusion principalement des molécules d'eau dans le tissu et pour chaque voxel. Cette information peut-être très intéressante puisqu'elle peut nous indiquer si le milieu est isotrope, et dans le cas contraire le sens des fibres.

L'IRM ainsi que le scanner sont les deux seules méthodes d'imagerie anatomique à fournir des images d'une qualité acceptable pour nos applications.

**Angiographie.** L'angiographie permet l'étude du réseau vasculaire du patient. Cet examen consiste en l'injection d'un produit de contraste dans le sang du patient. Ce produit est ensuite observé via une méthode d'imagerie par rayons X.

### 1.2.2 Méthodes d'imagerie fonctionnelle

**TEP scan.** La tomographie par émission de photons, ou TEP scan, repose sur l'injection dans l'organisme d'un traceur marqué par un atome radioactif (carbone, fluor, oxygène...) de durée de demi-vie compatible avec les contraintes de l'examen (entre une et deux heures). Le traceur est choisi en fonction du but de l'examen pour être fixé par les organes cibles. Par exemple, pour détecter des tumeurs, on peut utiliser comme traceur des sucres proches du glucose. En effet, les cellules cancéreuses se multiplient sans cesse et consomment donc beaucoup de glucose. Cette consommation excessive fait que le traceur se fixera préférentiellement sur les cellules cancéreuses. Cependant les traceurs utilisés,

s'ils ressemblent à du glucose, ne sont pas utilisables directement par les cellules et vont donc se retrouver concentrés dans la zone cancéreuse sans être dégradés. La désintégration de l'atome radioactif du traceur produit un positon qui lui même va s'annihiler avec un électron du milieu en émettant deux photons dans la même direction mais dans un sens opposé. Les capteurs de la machine, situés de part et d'autre du patient détectent alors ces photons en coïncidence (c'est à dire émis en même temps) ce qui permet de déterminer la ligne sur laquelle se trouve le point d'émission des photons. Ensuite le recouplement de ces informations par un logiciel spécialisé permet d'identifier les zones d'émission des photons. L'une des principales difficultés liées à l'utilisation de ce type de données provient de la complexité à lier les grandeurs mesurées (désintégrations nucléaires) avec des informations utiles pour le mathématicien (densité locale de cellules tumorales, masse réelle de la tumeur...). Des travaux de modélisation existent ayant pour but justement de faire la liaison entre ces différentes grandeurs.

**IRM fonctionnelle.** Cette catégorie d'imagerie est une modalité spécifique de l'IRM principalement utilisée dans l'étude du fonctionnement du cerveau. Elle repose sur la détection de variations locales minimales de flux sanguin. Celles-ci témoignent de l'activation des zones cérébrales concernées. On peut alors obtenir en temps réel une cartographie des zones actives du cerveau et mettre en relation ces zones avec d'éventuels stimuli extérieurs. Cette modalité est utilisable dans l'étude d'autres organes mais l'interprétation des images est généralement plus complexe pour les médecins.

### 1.2.3 Utilisation de l'imagerie dans le cadre des métastases pulmonaires

Compte tenu de la spécificité des métastases au poumon, certaines modalités sont privilégiées par rapport à d'autres. Le poumon est un organe majoritairement composé d'air et donc peu dense contrairement à la métastase qui est un amas compact de cellules. La tomodensitométrie par rayons X, ou CT scan, est donc toute indiquée pour observer ce type de cancers et la différence de densité des tissus assure un contraste suffisant à l'image. Cette modalité structurale est parfois associée à une méthode d'imagerie fonctionnelle comme le TEP scan. Cela permet d'observer l'activité métabolique au sein du poumon et éventuellement de repérer des lésions passées inaperçues sur le CT scan. Dans le cas où un patient présente plusieurs métastases pulmonaires, le TEP scan peut aussi permettre de discriminer les tumeurs suivant l'intensité apparente de leur activité.

## 1.3 Problématique et motivations

### 1.3.1 Présentation d'un cas clinique

La Figure 1.7 regroupe des images issues du suivi longitudinal d'un patient atteint d'une métastase pulmonaire sur plus d'un an. Une tumeur primaire à la vessie s'était déclarée quelques années auparavant et avait été opérée. Plusieurs métastases sont apparues par la suite dont celle qui est présentée. Les images 1.7a, 1.7b et 1.7c correspondent à l'histoire naturelle, c'est à dire sans traitement, de la lésion. À la vue des deux premiers examens, le médecin oncologue a planifié une ablation par radiofréquences pour la date du 10/12/2008. Cependant, lors du scan pré-opératoire, il s'est rendu compte que la lésion avait grossi bien plus vite que prévu comme on peut le voir sur l'image 1.7c. L'ablation par radiofréquences présente l'avantage d'être semi-invasive puisqu'elle ne nécessite pas d'ouvrir la cage thoracique du patient. Elle est donc indiquée pour des patients âgés ou affaiblis qui ne supporteraient peut-être pas une intervention chirurgicale, comme c'est le cas pour le patient concerné. Une limite de cette approche thérapeutique est qu'elle ne donne des résultats satisfaisants que pour des tumeurs de petite taille (diamètre inférieur à 3cm). Dans notre cas, la tumeur avait dépassé cette limite de taille et n'était donc plus opérable par radiofréquences.

Le praticien a donc dû mettre en place un protocole de traitement par chimiothérapie cytotoxique afin d'enrayer la croissance menaçante de cette tumeur et éventuellement de la faire réduire en taille. Ce protocole standard s'étale sur 6 mois durant lesquels une dose de médicaments est administrée au patient de façon régulière. Deux examens de contrôle jalonnent ce protocole afin de s'assurer des effets du traitement. Le premier intervient au milieu du traitement, et le second en fin de protocole. Ces deux examens correspondent aux figures 1.7d et 1.7e et on peut constater que la tumeur répond assez bien à la chimiothérapie administrée puisqu'elle diminue en taille tout au long du protocole. Au bout des 6 mois, le traitement est arrêté et un examen est programmé un mois plus tard qui est représenté en figure 1.7f. On constate sur cette dernière image que l'effet du traitement a pris fin et que les cellules n'ayant pas été tuées par la chimiothérapie ont continué à proliférer et donc que la tumeur s'est remise à grossir. L'utilisation d'un traitement par chimiothérapie n'a donc pas permis de soigner le patient mais a ralenti la progression de la maladie.

### 1.3.2 Orientation des travaux de cette thèse

Le cas clinique précédent nous amène à nous poser les questions suivantes :

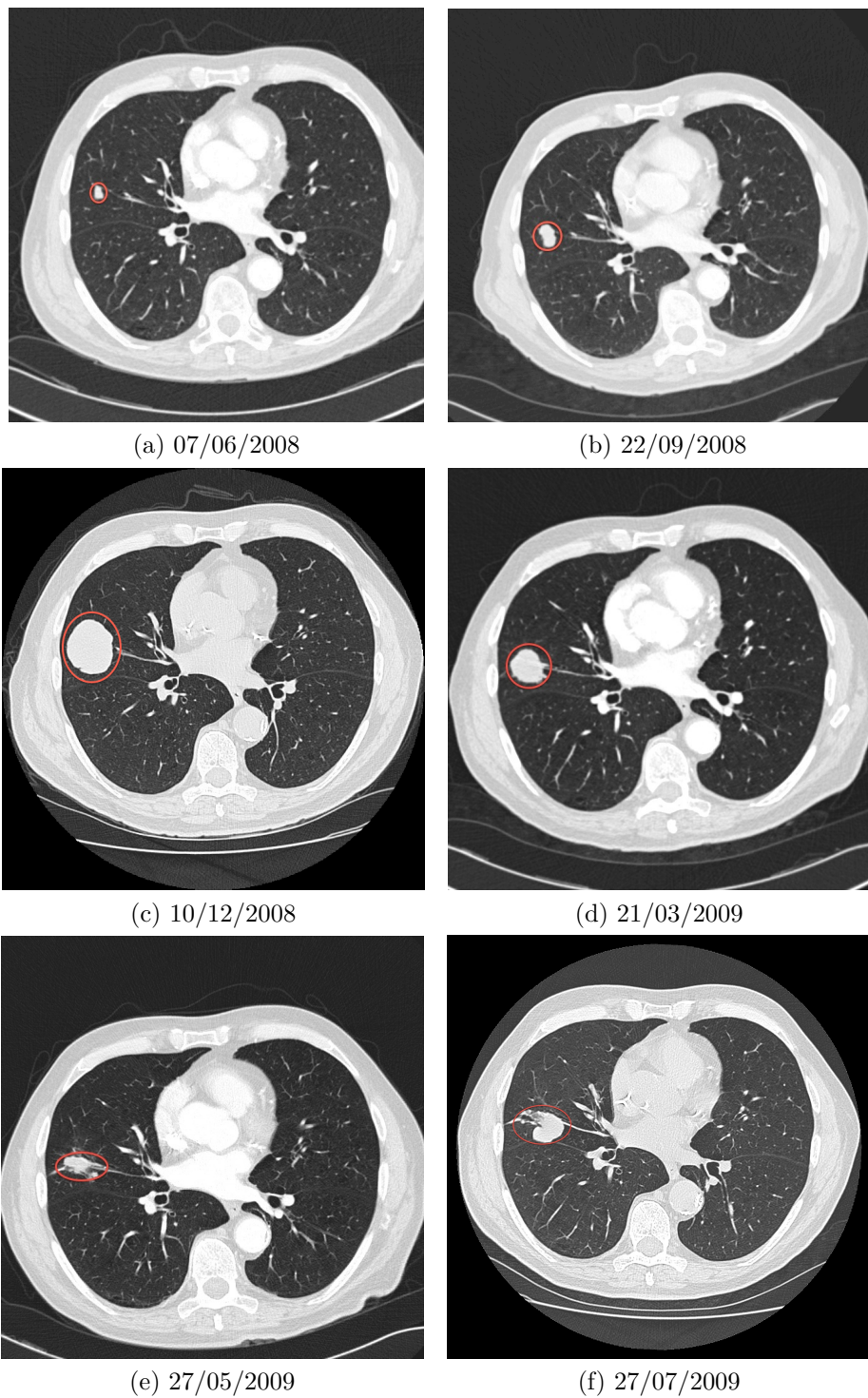


FIGURE 1.7 – Extrait d'une séquence de CT scans montrant l'évolution d'une métastase pulmonaire issue d'une tumeur primitive à la vessie.

- À partir des deux premiers examens (1.7a et 1.7b) serait-il possible d'évaluer l'agressivité de la tumeur et donc de quantifier l'évolution à venir de la tumeur et sa taille à la date du troisième examen (1.7c) ? On comprend que, dans le cas présent, une meilleure prédiction de la croissance de la tumeur que celle, intuitive, établie par le médecin aurait pu permettre d'adapter la planification de la thérapie par radiofréquence et donc d'éviter les désagréments de la chimiothérapie.
- De la même manière, serait-il possible, à partir des précédents examens ainsi que du premier scanner de contrôle durant la chimiothérapie (1.7d) d'avoir une idée de la réponse finale au traitement (1.7e) et éventuellement de l'ampleur de la rechute suite à l'arrêt des traitements (1.7f) ?

Les travaux réalisés dans le cadre de cette thèse ont pour but d'étudier ces questions et de tenter d'apporter des éléments de réponse. Ce cas clinique est révélateur de la réelle difficulté qu'ont les médecins à évaluer la croissance d'une tumeur. À l'heure actuelle, le critère le plus utilisé par les radiologues dans un cadre clinique pour évaluer l'agressivité d'une métastase pulmonaire est le RECIST (Response Evaluation Criteria In Solid Tumors). Ce critère se base sur la mesure de diamètres d'une ou plusieurs métastases directement sur l'examen médical et la comparaison de ces diamètres au cours du temps.

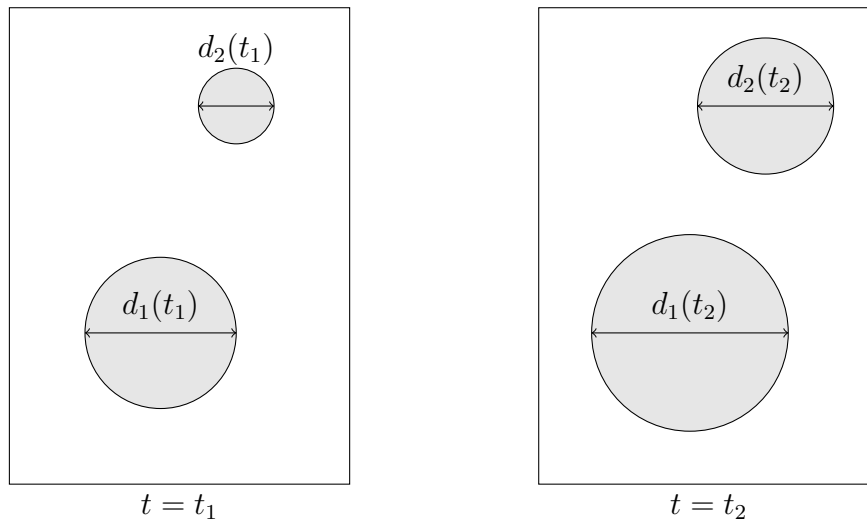


FIGURE 1.8 – Critère RECIST : mesure de l'évolution tumorale (croissance ou réponse au traitement).

Cet indicateur est utilisé afin de quantifier l'évolution de la tumeur ou sa réponse au traitement dans un cadre clinique. On calcule simplement la somme  $S$  des diamètres de lésions sélectionnées (au maximum 5) pour chaque examen. Dans notre cas (Fig. 1.8),  $S(t) = d_1(t) + d_2(t)$ . Ensuite on détermine la progression ou la réponse au traitement de la maladie en fonction de l'évolution de cette somme. Si l'ensemble des lésions a disparu, la réponse au traitement est complète. Dans le cas où  $S$  a diminué de plus de 30%, on a une réponse partielle. Si au contraire,  $S$  augmente de plus de 20%, la maladie a progressé. Enfin, si on est dans une gamme de critères située entre la réponse partielle (−30%) et la progression tumorale (+20%), on considère que l'évolution n'est pas significative et que la maladie est stable.

Ce critère montre assez vite certaines limites. Tout d'abord, il ne permet d'évaluer que l'évolution passée de la tumeur mais n'est pas prédictif de son évolution future sauf dans de rares cas. De plus il est **fortement dépendant de l'utilisateur** puisqu'il s'appuie sur la mesure du diamètre d'un objet non sphérique et varie donc suivant la coupe choisie de la tumeur ainsi que le rayon établi par l'utilisateur. Enfin, le choix des lésions utilisées pour le calcul influence aussi grandement le résultat.

Étant donnée la qualité des images médicales actuelles, il semble pertinent de tenter d'utiliser au mieux les quantités d'informations qu'elles renferment. Pour cela, l'ensemble de notre approche est orienté sur les données médicales disponibles. De même les différentes étapes des travaux, notamment la phase de modélisation de la maladie ou d'analyse des résultats, ont nécessité la participation et l'expertise de médecins oncologues. Le but de cette approche orientée clinique est multiple. Cela nous permet tout d'abord d'améliorer notre compréhension, mais aussi de hiérarchiser les mécanismes régissant l'évolution de la maladie. Il est ainsi primordial de saisir les spécificités des métastases pulmonaires par rapport aux autres types de cancer et d'adapter les méthodes employées en fonction de ces particularités. Enfin, et c'est peut-être là le point le plus important, cette collaboration permet de garder en tête les préoccupations et les attentes des médecins et de tenter ainsi d'apporter des outils permettant de simplifier leurs pratiques.

## 1.4 Travaux mathématiques sur le cancer

Plusieurs types d'approches peuvent être utilisés pour modéliser mathématiquement le développement de tumeurs cancéreuses. Ces méthodes peuvent être classifiées suivant différents types de critères comme l'échelle de l'étude (macroscopique ou microscopique), les données utilisées ou encore les applications possibles. On distinguera trois principales grandes familles de méthodes

de modélisation appliquées au cancer.

### 1.4.1 Méthodes discrètes

Ces méthodes consistent à considérer indépendamment chaque cellule (Fig.1.9), on parle alors d'automates cellulaires ou de modèles d'agent (voir [55]). Les différents phénomènes tels que la prolifération ou la mort des cellules sont décrits par l'utilisation de lois de probabilités. On peut aussi tenir compte du déplacement des cellules et des contraintes mécaniques dans le tissu et de bien d'autres paramètres (effet de traitements, différentes populations de cellules...). Ils sont adaptés à l'échelle microscopique et à privilégier pour étudier notamment les stades très précoces de la maladie. Au contraire, dès que le nombre de cellules devient grand, la méthode devient très coûteuse en calculs et n'est donc pas adaptée aux échelles macroscopiques. De même, étant données les échelles considérées, il est très difficile de confronter ces méthodes à des données cliniques ou issues d'imagerie médicale classique.

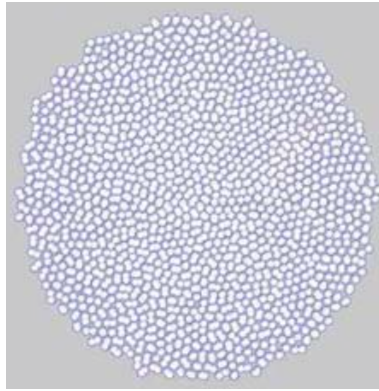


FIGURE 1.9 – Simulation d'un ensemble de cellules présentant une forte adhésion. Source : [14]

### 1.4.2 Modèles continus scalaires

Cette catégorie de modèles décrit l'évolution continue de grandeurs scalaires de la tumeur comme le volume ou la masse dans le temps (voir [60]). Ces modèles sont bien souvent basés sur l'utilisation d'équations différentielles ordinaires (EDO). L'un des modèles EDO les plus utilisés pour la croissance tumorale est le modèle de Gompertz (voir [32] ou [61]). Cette équation a été écrite au XIX<sup>ième</sup> siècle pour décrire l'évolution d'une population humaine en fonction de la capacité d'accueil du milieu (quantité de nourriture par exemple).

D'après la théorie de Gompertz retranscrite dans ce modèle mathématique, la population croît (de façon exponentielle) tant que son environnement le permet. Dès que ce n'est plus le cas, la croissance ralentit. Ce modèle décrit bien le comportement d'une population lorsque la quantité de nourriture est limitée. L'équation décrivant ce processus est la suivante :

$$\frac{dN}{dt}(t) = \alpha \ln\left(\frac{K}{N(t)}\right)N(t), \quad (1.1)$$

où  $N(t)$  est la taille de la population à l'instant  $t$ ,  $\alpha$  un paramètre influant sur la vitesse de prolifération et  $K$  la capacité d'accueil du milieu, c'est à dire la taille asymptotique de la population correspondant à une quantité de nourriture donnée.

Ce modèle peut être utilisé pour l'étude de la croissance de tumeurs cancéreuses. En effet la tumeur peut être vue comme une population de cellules évoluant dans un milieu dont la capacité d'accueil est limitée. Plusieurs travaux ont permis de démontrer la pertinence de l'application de ce modèle à différents types de tumeur, chez l'animal comme chez l'homme [52].

Le modèle de Gompertz constitue bien souvent un point de départ pour l'étude de la croissance tumorale et de nombreux modèles ont été élaborés à partir de celui-ci. On peut par exemple citer le modèle utilisé dans les travaux de Hahnfeldt (voir [34]). Ce modèle est donné par les équations :

$$\frac{dV}{dt} = \alpha \ln\left(\frac{K}{V}\right)V, \quad (1.2)$$

$$\frac{dK}{dt} = -\lambda K + \beta V - \delta K V^{\frac{2}{3}} - \eta K f(t). \quad (1.3)$$

L'équation (1.2) est l'équation de Gompertz sur le volume  $V$  de la tumeur. Cependant ici la capacité d'accueil  $K$  n'est plus une fonction mais elle varie au cours du temps suivant l'équation (1.3). Cette variable  $K$  est d'ailleurs ici vue comme la vascularisation des tissus environnant la tumeur. Dans cette équation, on tient compte de différents phénomènes pouvant influencer sur cette vascularisation.

Le terme  $-\lambda K$  est un terme de dégradation naturelle alors que le terme  $\beta V$  correspond à une synthèse de vascularisation provoquée par la présence de la tumeur et peut donc être assimilé à un terme d'angiogenèse.

Les termes  $-\delta K V^{\frac{2}{3}}$  et  $-\eta K f(t)$  représentent quant à eux la consommation de nutriments par la tumeur à sa surface et la destruction du réseau vasculaire par l'effet d'un éventuel traitement anti-angiogénique  $f(t)$ . Ce modèle est donc plus complet que l'équation de Gompertz et tient compte de différents phénomènes impliqués dans l'évolution tumorale ainsi que des effets de traitements.



Cependant, il comporte beaucoup plus de paramètres, au minimum six contre deux pour le Gompertz. Ce grand nombre de degrés de liberté permet d'obtenir des comportements variés mais rend aussi plus compliquée la calibration du modèle sur des données.

De nombreux modèles scalaires génériques pour la croissance tumorale ont été ainsi développés, une comparaison de certains d'entre eux est établie dans [7]. S'ils permettent d'étudier l'évolution de variables macroscopiques comme la taille de la tumeur, ces modèles ne peuvent pas tenir compte de données spatiales, comme la forme des tissus, qui pourtant sont très importantes pour décrire la croissance tumorale.

### 1.4.3 Méthodes continues spatiales

Afin d'intégrer à la modélisation des aspects spatiaux et des grandeurs locales tels que l'anisotropie des tissus, la densité locale de cellules cancéreuses ou la distribution spatiale des nutriments, il faut s'orienter vers des modèles spatiaux continus bien adaptés à l'échelle macroscopique. Ceux-ci sont souvent composés d'équations aux dérivées partielles décrivant l'évolution dans le temps et dans l'espace de diverses variables. Ce type d'approche présente l'avantage de permettre l'utilisation des données spatiales directement issues de l'imagerie médicale et ne nécessite plus de se restreindre à des variables intégrales comme dans le cas des modèles scalaires. Les informations qu'on peut espérer tirer de cette catégorie de modèles sont donc bien plus riches qu'une simple courbe d'évolution de la taille de la tumeur. De même, les particularités de l'environnement dans lequel évolue la lésion peuvent être prises en compte plus finement. On peut par exemple tenter d'expliquer les effets de la présence d'un vaisseau sanguin sur l'évolution de la forme d'une tumeur, ou encore prendre en compte les effets mécaniques liés à la présence du crâne dans la croissance de tumeurs intracrâniennes.

La plupart des travaux de modélisation de la croissance tumorale utilisant des systèmes d'équations aux dérivées partielles sont basés sur la description de l'évolution de densités locales de différentes populations de cellules ([56, 13]). Le point de départ de ces équations est un principe physique : la conservation de la masse. Ensuite, en fonction des hypothèses de modélisation mais aussi des types de cancers considérés, les phénomènes de prolifération et d'invasion par les cellules cancéreuses peuvent être traduits de différentes manières. On retrouve des modèles de type réaction-diffusion, principalement pour les tumeurs infiltrantes présentant un caractère diffusif (voir [74] sur les tumeurs du rein ou [69, 35, 17, 18] sur les tumeurs cérébrales). Ceux-ci sont basés sur une

équation de réaction-diffusion telle que celle présentée ci-dessous (1.4) :

$$\frac{\partial T}{\partial t} = \nabla \cdot (D(x)\nabla T) + \rho T. \quad (1.4)$$

Dans cette équation, le champ  $T$  représente la densité de cellules tumorales, dépendant du temps et de l'espace. Le terme  $\nabla \cdot (D(x)\nabla T)$  traduit la diffusion des cellules tumorales dans les tissus environnants alors que le terme  $\rho T$  correspond à la prolifération. D'autres équations peuvent intervenir pour réguler le taux de prolifération  $\rho$  ou encore le tenseur de diffusion  $D(x)$  et les relier à des facteurs comme la présence et la concentration de nutriments dans le tissu. Ces modèles sont donc construits sur l'hypothèse que le mouvement des cellules cancéreuses est assimilable à un phénomène de diffusion. Si cela semble réaliste dans le cas de certains cancers, cette hypothèse est nettement moins vérifiée dans le cas des métastases pulmonaires. Ces lésions sont en effet compactes et l'interface entre le tissu sain et la tumeur est nette et bien définie, comme on peut le voir sur les images de la Figure 1.7. La croissance de la tumeur est à relier au mouvement passif des cellules sous l'effet de la prolifération : l'apparition de nouvelles cellules au sein d'un milieu déjà saturé pousse vers l'extérieur les cellules environnantes et la tumeur grossit. La Figure 1.10 illustre ce phénomène dans un cas unidimensionnel.

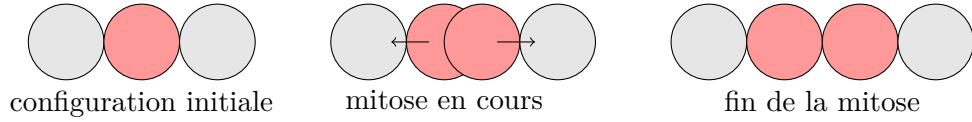


FIGURE 1.10 – Déplacement des cellules lié à la prolifération, cas 1D.

Dans ce cas, il semble plus pertinent d'utiliser un modèle basé sur une équation d'advection ou transport conservative (voir [22]) telle que l'équation suivante (1.5) :

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = \rho T. \quad (1.5)$$

Le terme  $\nabla \cdot (\mathbf{v}T)$  traduit le transport des cellules tumorales à la vitesse  $\mathbf{v}$ . Cette vitesse est déterminée par ailleurs, en fonction des hypothèses de modélisation, et peut par exemple incorporer le comportement mécanique des tissus. De même que précédemment, on retrouve un terme de prolifération  $\rho T$ . Dans les travaux effectués durant cette thèse, on se limite à l'étude de modèles de réaction-advection qui nous semblent plus pertinents pour décrire le comportement des métastases pulmonaires.

On vient de voir le principe de base de deux grandes classes de modèles EDP pour la description de la croissance tumorale. De nombreux mécanismes peuvent être intégrés dans ces modèles : angiogenèse, comportement mécanique des tissus, distribution des nutriments, effet de thérapies... On peut alors imaginer des modèles de plus en plus complexes, comportant un grand nombre d'équations et de variables mais aussi de paramètres. Si cette complexité doit pouvoir permettre de raffiner la modélisation et donc potentiellement d'obtenir un modèle très précis, il faut garder à l'esprit que plus le nombre de paramètres sera important, plus il sera difficile de calibrer le modèle et de le personnaliser. Il est donc important de garder à l'esprit la finalité du travail de modélisation pour adapter la complexité du modèle. Dans notre cas, le but est réellement d'apporter un outil utilisable cliniquement et donc à la fois précis et personnalisable à partir d'un nombre restreint de données. L'équilibre à trouver entre la pertinence biologique du modèle et la simplicité de celui-ci sera donc une des principales problématiques de ces travaux de recherche.

## 1.5 Environnement logiciel

L'ensemble des travaux présentés dans cette thèse a été mené à partir de code développé au sein de l'équipe projet INRIA Monc. Les codes de simulation numérique utilisés pour la résolution des systèmes d'équations aux dérivées partielles sont basés sur les outils Elyse puis Cadmos mis au point par Olivier Saut. Plus qu'une simple librairie, Cadmos fournit un environnement complet destiné principalement à la mise en oeuvre de codes de simulation. Il rassemble notamment une collection de solveurs 2D et 3D pour différents types d'équations. On retrouve par exemple des solveurs de transport (WENO 5, schéma amont, Lax Wendroff) ou pour l'équation de Poisson qui sont directement utilisés pour la simulation des systèmes d'équations aux dérivées partielles modélisant la croissance des métastases pulmonaires, de même que les méthodes de suivi d'interface. En plus de cela, Cadmos permet, entre autres, la gestion des paramètres du modèle, du maillage ou encore des outils de lecture des fichiers sources et d'écriture des résultats. L'ensemble est codé en C++ et l'utilisation de l'approche orientée objet permet une modularité de l'outil et simplifie les ajouts de fonctionnalités mais aussi l'utilisation.

La plupart des algorithmes utilisés ont été codés en C++ sauf certaines méthodes d'optimisation, comme l'algorithme des essaims particuliers utilisé pour calibrer le modèle, qui ont été programmés en langage ruby. Ce langage est plus simple et rapide à mettre en oeuvre bien que moins performant d'un point de vue temps de calcul. Il a donc souvent été utilisé afin d'établir des preuves de concept ou de tester certaines méthodes. Les courbes ont prin-

cipalement été tracées à l'aide des logiciels Scilab et R.

Le logiciel Osirix a été utilisé afin de visualiser les examens médicaux mais aussi de segmenter les tumeurs. L'export des données depuis Osirix se fait via des fichiers au format xml. Ces fichiers sont ensuite convertis au format vtk afin d'être lus par les codes de simulation. L'algorithme des forêts aléatoires a aussi été entièrement codé en C++. La plupart des codes, notamment les codes d'optimisation qui impliquent de nombreux calculs, sont compatibles avec une utilisation parallèle en multi-thread.

Les calculs les plus lourds ont été menés sur des clusters de calculs disponibles depuis l'Université de Bordeaux. Il s'agit de la plateforme PlaFRIM et des moyens de calcul mis à disposition par le Mésocentre de Calcul Intensif Aquitain via le cluster Avakas.



## Chapitre 2

# Étude d'un premier modèle de croissance métastatique

Les travaux présentés dans ce chapitre ont fait l'objet de la publication des articles [20] et [21].

### 2.1 Présentation du modèle

Notre objectif de base dans cette étude est de mettre au point un modèle mathématique permettant de décrire la croissance de métastases au poumon. La finalité est de pouvoir étudier des cas cliniques et donc de prendre en compte certains types de traitement. Pour cela le modèle adopté devra tenir compte du phénomène d'angiogenèse induit par la croissance tumorale dans son stade de développement vasculaire. Nous souhaitons ainsi décrire, d'une part le mécanisme par lequel la tumeur va se vasculariser, et d'autre part les effets de cette néovascularisation sur la croissance de la métastase (apport de nutriments et de dioxygène...). De plus, ce modèle devra être le plus simple possible pour permettre, lors de la phase des études cliniques, un calibrage simplifié sur les données médicales.

Pour la conception de ce modèle, les hypothèses biologiques prises en compte sont les suivantes :

- Un seul type de nutriments ou facteur de croissance est considéré. Sa concentration contrôle la prolifération mais aussi la mort des cellules.
- La quantité totale de nutriments disponible est proportionnelle à la concentration de vaisseaux sanguins dans le tissu. On ne tient pas compte de la diffusion des nutriments dans le tissu car elle est négligeable à l'échelle où on se situe.
- Le seul mouvement pris en compte est le transport passif des cellules

résultant des variations du volume dues à la prolifération (ou à la mort) cellulaire.

- Les cellules cancéreuses passent continuellement du phénotype proliférant à quiescent et on ne considère qu’une seule population de cellules cancéreuses (voir les images de biopsie présentées en Figure 2.1).

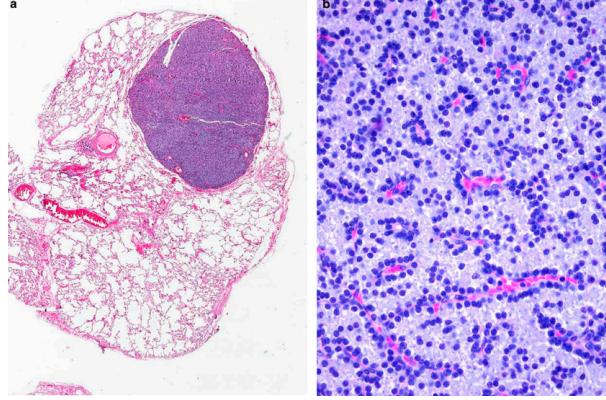


FIGURE 2.1 – Extrait de biopsie d’une métastase pulmonaire issue d’une tumeur primitive à la thyroïde.

### 2.1.1 Hiérarchie de modèles spatiaux

Cette partie décrit les différents modèles qui ont été étudiés, du plus simpliste au plus complet, et qui ont mené au modèle finalement adopté.

**Premier modèle de croissance à une équation :** Dans ce modèle on ne considère qu’une population de cellules, les cellules tumorales dont la densité est notée  $T$ . Ces cellules se déplacent par diffusion dans le tissu sain et se divisent. On obtient l’équation suivante :

$$\frac{\partial T}{\partial t} = \underbrace{\nabla \cdot (D \nabla T)}_{\text{diffusion}} + \underbrace{\rho T}_{\text{prolifération}} .$$

Si ce modèle a le mérite d’être simple, il comporte aussi de nombreux défauts. Le premier est que rien n’assure que la densité de cellules soit bornée. Pour palier cela, on introduit le modèle suivant.

**Modèle à une équation avec terme logistique :** On reprend le modèle précédent mais on sature la croissance de  $T$  avec un terme logistique :

$$\frac{\partial T}{\partial t} = \nabla \cdot (D \nabla T) + \underbrace{\rho T(1 - T)}_{\text{prolifération logistique}}.$$

Le terme de prolifération logistique permet de maintenir une densité de cellules tumorales entre 0 et 1, la valeur 1 correspondant à un tissu saturé en cellules cancéreuses.

**Modèle avec loi de Darcy :** Le modèle précédent reste trop grossier. En particulier, la partie diffusive, si elle permet le déplacement des cellules, semble inadapté. En effet, et comme on l'a vu en introduction, un modèle basé sur une équation de diffusion semble peu réaliste pour décrire des métastases pulmonaires. Les cellules tumorales se déplacent sous l'effet de l'augmentation du volume de la lésion poussant les tissus alentour. C'est pourquoi le modèle suivant met en œuvre un terme de transport.

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v} T) = \rho T(1 - T),$$

$$\mathbf{v} = -\nabla \pi,$$

$$\nabla \cdot \mathbf{v} = \rho T(1 - T).$$

La vitesse  $\mathbf{v}$  est calculée en utilisant une loi de Darcy, c'est à dire qu'elle dérive d'une pression  $\pi$ . La dernière équation portant sur la divergence de la vitesse permet de fermer le système et de calculer  $\mathbf{v}$ . Elle est obtenue en utilisant une équation de transport sur le tissu sain  $S = 1 - T$  sans terme de prolifération.

**Modèle avec angiogenèse :** À la fin de la phase avasculaire de son développement, la tumeur atteint une telle taille que son environnement immédiat n'est plus capable de la fournir efficacement en nutriments pour lui permettre de poursuivre sa croissance. Les cellules cancéreuses se mettent alors à émettre des signaux chimiques tels que le VEGF (Vascular Endothelial Growth Factor) qui déclenchent le phénomène d'angiogenèse c'est à dire l'émergence d'une néo-vascularisation irriguant la tumeur (voir [15] et [49]). La recherche est très active sur ce phénomène et ce particulièrement pour mettre au point des traitements, dits anti-angiogéniques, agissant sur ce levier. Comme la plupart des métastases visibles sur les images médicales ont déjà atteint le stade vascularisé de leur développement, il est important de tenir compte de l'angiogenèse pour tenter de modéliser leur croissance. L'étape suivante est donc de coupler le modèle de croissance tumorale à un modèle d'angiogenèse. Pour cela, on doit



ajouter de nouvelles variables comme la vascularisation et l'agent angiogénique (de type VEGF). Il est aussi nécessaire de décrire les nouvelles dépendances entre ces variables et la densité de cellules tumorales. On reprend le principe du modèle avec loi de Darcy mais cette fois, on considère que les cellules tumorales peuvent se multiplier si les conditions sont favorables et mourir dans le cas contraire :

$$\begin{aligned} \frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) &= \underbrace{\gamma_p T}_{\text{prolifération}} - \underbrace{\gamma_d T}_{\text{apoptose}}, \\ \mathbf{v} &= -\nabla \pi, \\ \nabla \cdot \mathbf{v} &= (\gamma_p - \gamma_d)T. \end{aligned}$$

On note  $M$  la densité de vascularisation. On peut donc considérer que cette densité influe sur les valeurs de  $\gamma_p$  et  $\gamma_d$ . En effet, si  $M$  est insuffisant (c'est à dire que la tumeur n'est pas correctement approvisionnée), l'environnement ne sera pas forcément favorable à la multiplication des cellules cancéreuses (et  $\gamma_d$  prendra le dessus sur  $\gamma_p$ ). On peut quantifier cette notion d'apport suffisant par la prise en compte d'un seuil d'hypoxie  $M_{th}$  et modéliser les coefficients  $\gamma_p$  et  $\gamma_d$  par des fonctions de type échelon (on préférera des tangentes hyperboliques à des heavy sides classiques pour une meilleure régularité). La Figure 2.2 illustre la transition entre le cas hypoxique ( $\gamma_d > \gamma_p$ ) et le cas d'approvisionnement suffisant autour de la valeur de  $M_{th}$  (2 dans le cas de la figure). Si  $M$  est au dessus du seuil  $M_{th}$  la tumeur va donc proliférer, sinon les cellules cancéreuses auront tendance à mourir. Le paramètre  $K$  est une constante de raideur qui permet d'ajuster la pente de la phase de transition entre les deux états.

$$\begin{aligned} \gamma_p(M) &= \gamma_0 \frac{1 + \tanh(K(M - M_{th}))}{2}, \\ \gamma_d(M) &= \gamma_1 \frac{1 - \tanh(K(M - M_{th}))}{2}. \end{aligned}$$

L'agent angiogénique, dont la densité est notée  $\xi$ , est *a priori* sécrété par les cellules hypoxiques dont la densité est donnée par l'expression  $(1 - \frac{\gamma_p}{\gamma_0})T$ . Ces molécules sont consommées par les cellules endothéliales qu'elles attirent. De plus, on peut considérer que l'agent angiogénique se répand dans les tissus par diffusion. On obtient l'équation suivante :

$$\frac{\partial \xi}{\partial t} = \underbrace{\alpha \left(1 - \frac{\gamma_p}{\gamma_0}\right) T (1 - \xi)}_{\text{production}} - \underbrace{\lambda \xi M}_{\text{consommation}} + \underbrace{\nabla \cdot (\beta \nabla \xi)}_{\text{diffusion}}.$$

La sécrétion est saturée par le terme  $(1 - \xi)$  pour éviter une explosion de la densité d'agent angiogénique ou prendre en compte la saturation du signal.

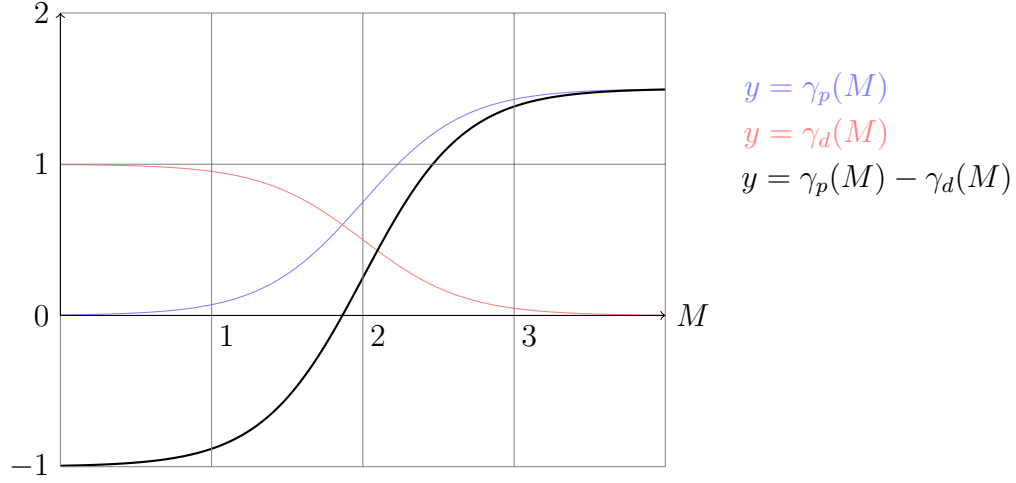


FIGURE 2.2 – Tracé des fonctions  $\gamma_p$ ,  $\gamma_d$  ainsi que le taux d'évolution de la densité  $\gamma_p - \gamma_d$  avec  $\gamma_0 = 1,5$ ,  $\gamma_1 = 1$ ,  $M_{th} = 2$  et  $K = 1,5$ .

Pour décrire l'évolution de la vascularisation autour de la tumeur, on considère tout d'abord que la néovascularisation provient du tissu sain et suit le gradient d'agent angiogénique (transport). D'autre part la tumeur dégrade une partie de cette vascularisation.

$$\frac{\partial M}{\partial t} = \underbrace{-\eta T M}_{\text{destruction par la tumeur}} + \underbrace{C_0 S(1 - M)}_{\text{production par le tissu sain}} + \underbrace{\nabla \cdot (M \nabla \xi)}_{\text{transport}}.$$

Pour une modélisation plus poussée du phénomène d'angiogenèse, on pourra se référer à l'article [23].

### 2.1.2 Modèle avec angiogenèse simplifié

Au paragraphe précédent, nous avons vu une hiérarchie de quelques modèles spatiaux allant du plus simple à un modèle assez complexe. Étant donnée notre problématique, le modèle retenu doit être pertinent biologiquement et tenir compte des phénomènes clés de la croissance tumorale tout en restant suffisamment simple pour être calibré (peu de paramètres, faible complexité en terme de calcul). Pour cela, nous sommes partis du dernier modèle présenté, et avons simplifié légèrement le modèle d'angiogenèse. Le modèle obtenu est présenté ci-dessous.

**Dynamique cellulaire et conservation de la masse :** On utilise les mêmes notations que précédemment. L'équation sur la densité de cellules tu-

morales  $T$  est donnée par (2.1).

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = (\gamma_p - \gamma_d)T, \quad (2.1)$$

où  $\mathbf{v}$  est la vitesse de transport. Les coefficients pour la prolifération et la mort par hypoxie (manque de nutriments),  $\gamma_p$  et  $\gamma_d$ , sont donnés par les mêmes équations (2.2) et (2.3) que dans le modèle précédent.

$$\gamma_p(M) = \gamma_0 \frac{1 + \tanh(K(M - M_{th}))}{2}, \quad (2.2)$$

$$\gamma_d(M) = \gamma_1 \frac{1 - \tanh(K(M - M_{th}))}{2}. \quad (2.3)$$

En plus des cellules cancéreuses, on tient aussi compte du tissu sain dont la densité est notée  $S$ . On considère donc que les cellules saines sont transportées à la même vitesse que  $T$ . Cependant, elles ne prolifèrent globalement pas ni ne disparaissent. Plus précisément, à l'échelle du tissu et des durées que l'on considère, la mitose et l'apoptose se compensent l'une l'autre et on a sur  $S$  une équation de transport sans terme source (2.4).

$$\frac{\partial S}{\partial t} + \nabla \cdot (\mathbf{v}S) = 0. \quad (2.4)$$

Comme on suppose que le tissu est saturé (cellules de volumes constants), la somme des deux densités  $T$  et  $S$  vaut 1 (2.5) (voir [12]).

$$S = 1 - T, \quad (2.5)$$

En sommant (2.1) et (2.4) on obtient une équation sur  $\mathbf{v}$  (2.6).

$$\nabla \cdot \mathbf{v} = (\gamma_p - \gamma_d)T. \quad (2.6)$$

Cela n'est pas suffisant pour calculer la vitesse de transport. Pour fermer le système d'équations (voir [3]), on utilise une loi de Darcy dans l'équation (2.7) :  $\mathbf{v}$  dérive d'une pression (ou d'un potentiel)  $\pi$  dans le tissu.

$$\mathbf{v} = -\nabla \pi. \quad (2.7)$$

Cette formulation revient à dire que les cellules tumorales sont poussées vers l'extérieur alors que si elles meurent, la tumeur se rétracte sur elle même. On pourrait aussi utiliser une loi de comportement de type équation de Stokes pour décrire la vitesse (voir [12]) mais cela compliquerait le modèle sans améliorer grandement sa précision ou sa pertinence du point de vue mécanique.

**Angiogenèse :** Le micro-modèle que nous avons utilisé est contenu dans les équations (2.8) et (2.9). La variable scalaire  $\xi$  représente la quantité totale d'agent pro-angiogénique dans le tissu. Ce facteur de croissance est sécrété par les cellules quiescentes (ou non proliférantes) dont la densité est donnée par  $\int_{\Omega}(1 - \frac{\gamma_p}{\gamma_0})T d\omega$ , où  $\Omega$  est le domaine de calcul, et évacué par l'organisme. Contrairement au modèle précédent,  $\xi$  est ici une variable scalaire représentant une quantité intégrale. Cela permet de simplifier le modèle tout en conservant une description suffisante du phénomène d'angiogenèse. On perd tout de même la répartition spatiale des facteurs angiogéniques.

$$\frac{\partial \xi}{\partial t} = \alpha \int_{\Omega} (1 - \frac{\gamma_p}{\gamma_0}) T d\omega - \lambda \xi. \quad (2.8)$$

Comme on fait l'hypothèse que la quantité de nutriments disponible est proportionnelle à la densité vasculaire, on amalgame ces deux notions en un champ  $M$  que l'on appellera pour simplifier « vascularisation ». On peut aussi voir  $M$  comme la capacité de l'environnement de la cellule à satisfaire ses besoins en terme de nutriments. On peut rapprocher cette idée de celle de « carrying capacity » introduite dans le modèle EDO de Hahnfeldt [54]. La densité vasculaire  $M$  est endommagée par les cellules tumorales est produite là où sont présentes les cellules quiescentes proportionnellement à  $\xi$  par le terme  $\beta \xi (1 - \frac{\gamma_p}{\gamma_0}) T$ .

$$\frac{\partial M}{\partial t} = -\eta T M + \beta \xi (1 - \frac{\gamma_p}{\gamma_0}) T. \quad (2.9)$$

Ainsi, ce modèle prend en compte d'importants mécanismes impliqués dans la croissance tumorale tels que la prolifération, l'hypoxie ou encore l'angiogenèse. De plus, il est relativement simple (on peut par exemple le comparer à celui utilisé dans [22]) puisqu'on ne considère qu'un type de cellules cancéreuses et qu'on ne tient pas compte du transport ou de la diffusion de la vascularisation. Ce compromis nous permet de garder un modèle pertinent biologiquement et suffisamment simple pour qu'on puisse espérer le calibrer efficacement.

**Prise en compte des effets thérapeutiques :** L'architecture du modèle rend la prise en compte de différents types de traitements aisée. Les effets de thérapies cytotoxiques et donc notamment des chimiothérapies peuvent être simulés par l'ajout d'un terme de destruction dans (2.1) tel que  $-\delta T$  pour le plus simple ou éventuellement  $-\delta M T$  si on veut prendre en compte le fait que le traitement atteint la tumeur via la vascularisation. Dans ce cas (2.1) devient :

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v} T) = (\gamma_p - \gamma_d) T - \delta T. \quad (2.10)$$

Pour satisfaire à l'hypothèse de saturation des tissus, l'équation (2.6) est modifiée comme suit :

$$\nabla \cdot \mathbf{v} = (\gamma_p - \gamma_d - \delta)T. \quad (2.11)$$

On peut aussi prendre en compte un traitement de type anti-angiogénique qui joue sur la production d'agent pro-angiogénique. On ajoute alors une fonction d'inhibition dans (2.8) :

$$\frac{\partial \xi}{\partial t} = \alpha f(t) \int_{\Omega} (1 - \frac{\gamma_p}{\gamma_0}) T d\omega - \lambda \xi.$$

Si le traitement inhibe aussi les récepteurs de VEGF positionnés sur la membrane des cellules endothéliales, on peut ajouter un terme correspondant dans l'équation sur la vascularisation (2.9) :

$$\frac{\partial M}{\partial t} = -\eta T M + \beta \xi f(t) (1 - \frac{\gamma_p}{\gamma_0}) T.$$

Dans les deux cas, on définit un indicateur thérapeutique à travers la fonction  $f$  définie ci-dessous :

$$f(t) = \begin{cases} 1 & \text{sans traitement,} \\ < 1 & \text{avec traitement.} \end{cases}$$

## 2.2 Simulation du modèle et résultats numériques

Cette partie présente l'implémentation numérique du modèle mathématique mis au point pour ce travail. C'est une étape primordiale puisqu'elle doit nous permettre de valider le comportement du modèle. D'un point de vue programmation, des tests ont été effectués sous Scilab en amont, puis le modèle a été implémenté en C++ via la plateforme de calcul scientifique eLYSe dans un premier temps, puis la librairie Cadmos. Celle-ci contient une collection de solveurs dédiés aux différents types d'équations que l'on peut rencontrer (solveur de transport, solveur de diffusion...). Cela permet de simuler une équation de façon simple et modulaire.

### 2.2.1 Méthodes de simulation numérique

La première étape ici est le choix du maillage. On adopte une grille cartésienne basique qui correspond à la distribution des intensités sur une image

médicale pour la discrétisation en espace. Le nombre de mailles choisi dépend en partie de la qualité et de la définition des images médicales dont nous disposons, mais aussi de la précision numérique et des temps de calcul induits. Compte tenu de ces différentes considérations, la plupart des simulations présentées dans ce qui suit ont été réalisées sur des maillages de taille  $200 \times 200$  en 2D et  $100 \times 100 \times 100$  en 3D. Pour des raisons de simplicité par rapport aux schémas numériques utilisés, les grandeurs telles que la pression  $\pi$  et les densités cellulaires ( $T$  et  $S$ ) ou de vascularisation ( $M$ ) sont données aux centres des mailles, alors que les vitesses sont calculées sur la frontière des mailles (voir Figure 2.3).

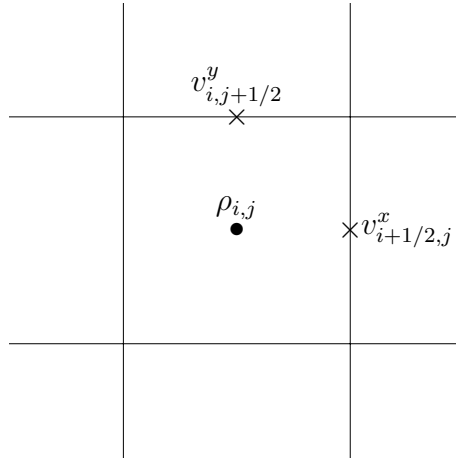


FIGURE 2.3 – Présentation d’une maille de la grille cartésienne dans le cas 2D. La vitesse de transport  $\mathbf{v} = (v^x, v^y)$  est calculée aux bords de la maille ; la composante en  $x$  sur les côtés verticaux de la maille et la composante en  $y$  sur les côtés horizontaux. Les autres variables ( $T$ ,  $S$ ,  $\gamma_p$ ,  $\gamma_d$ ,  $M$  et  $\pi$ ), matérialisées par la lettre  $\rho$ , sont positionnées aux centres des mailles.

Les différentes équations du système EDP étudié sont traitées comme suit :  
**Initialisation :**

- On connaît la répartition initiale des cellules tumorales  $T(t = 0)$  (extraite du premier examen, voir paragraphe 2.2.2), et on connaît donc aussi  $S(t = 0) = 1 - T(t = 0)$ .
- On initialise la vascularisation en se basant sur l’hypothèse simplificatrice développée au paragraphe 2.3.2

$$M(t = 0) = S(t = 0) + 0,8 \times T(t = 0),$$

de même, on prend arbitrairement  $\xi(t = 0) = 0,1$ .

- À partir de cela, on calcule les valeurs initiales des champs  $\gamma_p$ ,  $\gamma_d$  ainsi que la divergence de  $\mathbf{v}$ .

**Itérations : passage du temps  $t_n$  à  $t_{n+1}$**

- On estime la pression et la vitesse en résolvant le système composé des équations (2.6) et (2.7).
- On calcule une valeur approchée de  $T(t_{n+1})$  (puis de  $S(t_{n+1})$ ) à partir de l'équation (2.1) en utilisant les valeurs  $\mathbf{v}(t_{n+1})$ ,  $\gamma_p(t_n)$  et  $\gamma_d(t_n)$ .
- On met ensuite à jour les valeurs de  $\gamma_p$  et  $\gamma_d$  puis  $\xi$  et  $M$  (en utilisant les équations (2.8) et (2.9)).

Dans la thèse de Jean-Baptiste Lagaert ([44]), une discrétisation plus fine est utilisée, elle ne semble pas nécessaire dans notre cas (et elle est plus coûteuse).

Afin de suivre la trame présentée précédemment, il faut mettre en oeuvre différents schémas numériques permettant une résolution approchée des équations du système EDP. Les deux principales difficultés d'un point de vue numérique, qui seront davantage détaillées par la suite, sont la résolution de l'équation d'advection (2.1) d'une part, et la résolution de l'équation de Poisson découlant du système sur la vitesse (2.6) et (2.7) d'autre part. Les équations (2.8) et (2.9) ne présentent pas de difficultés particulières et ne seront donc pas détaillées. Pour les équations du type  $\frac{\partial X}{\partial t} = aX$ , nous utilisons un schéma exponentiel classique. De façon générale, les différents termes des équations sont traités séparément (méthodes de splitting de Strang, voir [44]).

### Schéma de transport pour la densité de cellules tumorales

L'équation d'advection (2.1) est l'élément central du système d'équations aux dérivées partielles puisqu'elle régit l'évolution des cellules tumorales. Ici on s'intéresse uniquement à l'équation d'advection sans second membre :

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = 0,$$

la partie prolifération étant gérée indépendamment dans un second temps. On peut développer le terme de divergence, on obtient alors :

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = -T \nabla \cdot \mathbf{v}.$$

D'après l'équation (2.6), on a l'expression  $\nabla \cdot \mathbf{v} = (\gamma_p - \gamma_d)T$  que l'on peut injecter dans l'expression précédente. On obtient :

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = -(\gamma_p - \gamma_d)T^2.$$

Ici encore le terme de droite peut être vu comme un terme source et traité à part (par splitting) et de façon exacte. Il reste donc à approcher la résolution de l'équation :

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = 0.$$

La méthode utilisée pour la résolution de cette équation est basée sur un schéma WENO (weighted essentially non-oscillatory) d'ordre 5 de type volumes finis avec des conditions aux bords de Dirichlet ( $T = 0$  sur les bords du domaine de calcul). Pour plus de détails, on pourra consulter [37] ou [65]. Ce type de schéma permet de minimiser les phénomènes de perte de masse ou encore de diffusion numérique.

Le passage du temps  $n$  au temps  $n+1$  pour la densité de cellule (2.1) est donc finalement géré selon le splitting suivant :

- Calcul d'un champ intermédiaire  $T^{temp}$  en appliquant la résolution de l'équation de transport sans second membre à  $T^n$  pendant l'intervalle de temps  $dt$ ,
- Résolution de la partie réaction (schéma exponentiel) à partir de  $T^{temp}$  pendant  $dt$  pour obtenir  $T^{n+1}$  (sans traitement) :

$$T^{n+1} = T^{temp} \times e^{(\gamma_p - \gamma_d)dt}.$$

### Schéma numérique pour l'équation de Poisson

À partir des équations (2.6) et (2.7), on obtient une équation de Poisson sur la pression  $\pi$  :

$$\Delta \pi = -\Gamma T, \quad (2.12)$$

où  $\Gamma = \gamma_p - \gamma_d$ . Dans le cas 2D, le Laplacien peut s'écrire comme suit :  $\Delta \pi = \frac{\partial^2 \pi}{\partial x^2} + \frac{\partial^2 \pi}{\partial y^2}$ . On discrétise alors les dérivées spatiales en utilisant des différences finies centrées. Il vient, en notant  $\delta x$  et  $\delta y$  les pas d'espace,

$$\frac{\pi_{i+1,j} - 2\pi_{i,j} + \pi_{i-1,j}}{\delta x^2} + \frac{\pi_{i,j+1} - 2\pi_{i,j} + \pi_{i,j-1}}{\delta y^2} = -\Gamma_{i,j} T_{i,j}. \quad (2.13)$$

Il s'agit là du schéma 5 points classique pour le Laplacien. La pression est déterminée à une constante près, on considère donc qu'il s'agit d'une pression relative et on la fixe à 0 loin de la tumeur, ce qui revient à imposer une condition aux limites de Dirichlet (pression nulle aux bords). On peut ainsi calculer le champ de pression  $\pi$  et en déduire la vitesse  $\mathbf{v}$ .



### 2.2.2 Utilisation des images médicales dans le code de simulation

Afin d'utiliser notre modèle ainsi que le code de simulation évoqué précédemment, il faut extraire les données utiles des images médicales dont nous disposons. Ici nous utilisons des extraits de CT-scans (CT : computerized tomography), on y observe directement la densité des tissus. La procédure explicitée dans le paragraphe qui suit correspond au cas de simulations 2D mais le principe est le même pour les cas 3D (sauf qu'il n'est plus nécessaire de choisir une coupe).

Dans un premier temps, il faut segmenter la tumeur, c'est à dire dessiner sur l'image médicale la frontière entre la tumeur et les tissus sains dans une coupe pertinente préalablement choisie par le médecin. Cette phase de segmentation est primordiale et particulièrement sensible puisqu'on estime que l'erreur relative inter-opérateurs sur les surfaces peut atteindre 12%. Ce chiffre est issu d'une étude comparant les segmentations de plusieurs lésions par 20 opérateurs (médecins ou membres de l'équipe de recherche). On utilise le logiciel Osirix pour la visualisation et la segmentation de la tumeur (voir Figure 2.4).

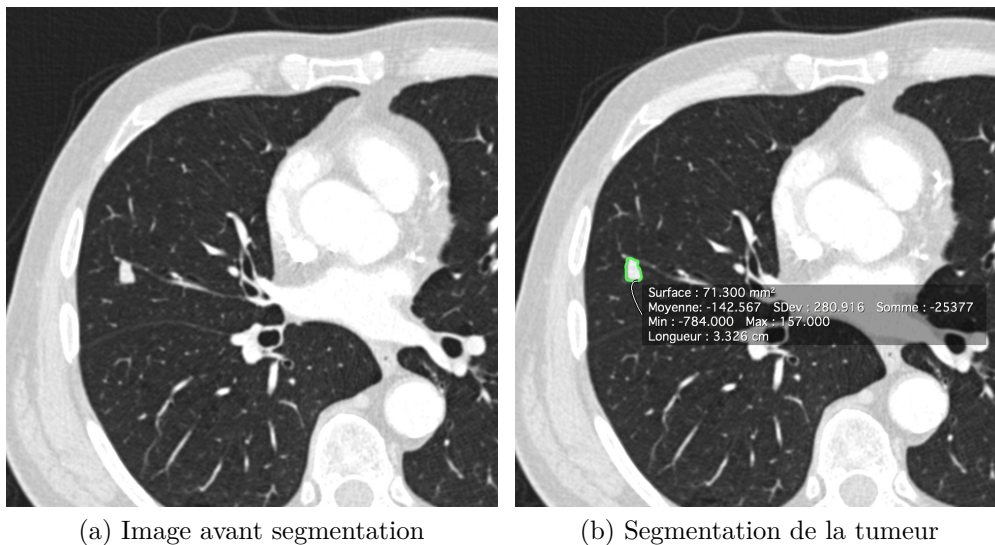


FIGURE 2.4 – Segmentation de la tumeur : à gauche l'image médicale avant la segmentation, à droite, l'image avec la segmentation de la tumeur.

La segmentation obtenue suite à l'étape présentée en Figure 2.4b se présente comme un polygone. On peut exporter la liste des points qui le composent sous la forme d'un fichier xml. On reconstruit le masque de la tumeur (voir Figure 2.5a). Ensuite on récupère les pixels sous la forme de niveaux de gris contenus à

l'intérieur du masque. L'échelle de gris est re-normalisée entre 0 pour les zones les moins denses (intérieur des bronches) et 1 pour les zones les plus denses et on obtient ainsi la densité de cellules tumorales  $T$  (Figure 2.5b).

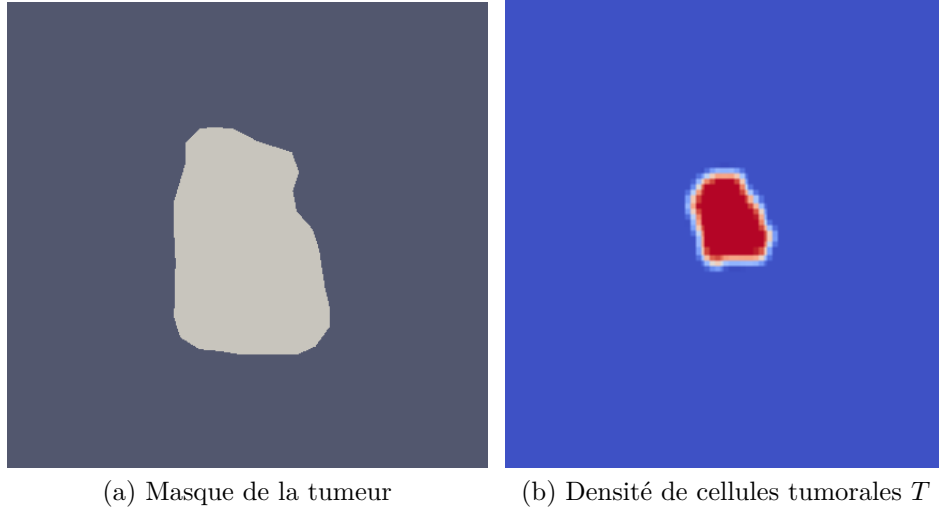


FIGURE 2.5 – Traitement des données extraites de la segmentation : à gauche le masque de la tumeur (format vtp) et à droite le champ  $T$  reconstruit (format vtk) avec lissage des bords. Pour le champ  $T$  l'échelle va de 1 en rouge à 0 en bleu.

Il peut être judicieux de garder une certaine marge entre les bords de la tumeur et ceux du domaine de calcul. Cette marge est évaluée empiriquement en fonction de la croissance maximale attendue pour la tumeur. De même, le bord de la tumeur peut être régularisé tout en conservant la masse (c'est le cas sur la Figure 2.5b) afin d'éviter des problèmes numériques liés à une variation spatiale trop rapide (gradient élevé) de  $T$ .

### 2.2.3 Premiers résultats de simulations

Voici deux résultats de simulation (Fig.2.6 et Fig.2.7) qui représentent l'évolution de la densité de cellules tumorales à partir d'une donnée initiale issue d'un scanner de patient.

Les paramètres ayant servi pour ces simulations sont présentés dans le Tableau 2.1.

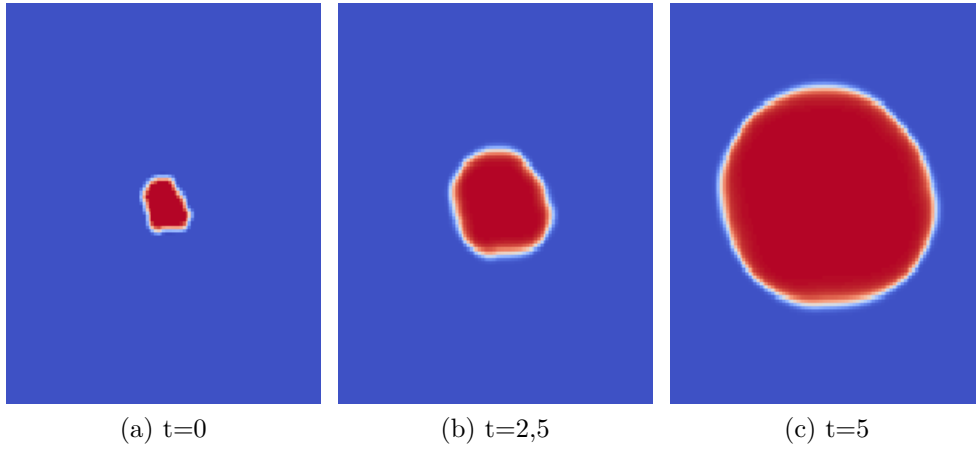


FIGURE 2.6 – Simulation 1 : évolution de la densité de cellules tumorales, maillage  $200 \times 200$

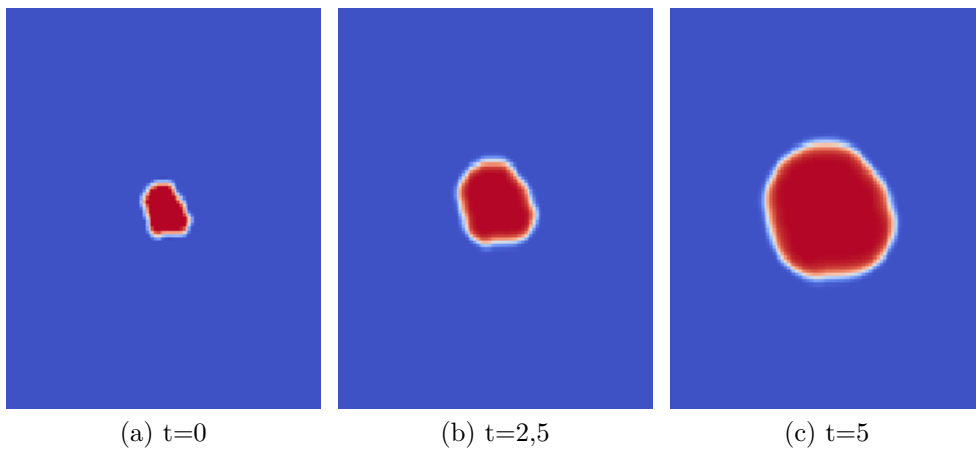


FIGURE 2.7 – Simulation 2 : évolution de la densité de cellules tumorales, maillage  $200 \times 200$

paramètre	simu. 1	simu. 2
$\alpha$ (en $\text{mm}^{-2}$ )	7,5	5
$\beta$	10	5
$\eta$	0,5	0,5
$\gamma_0$	0,6	0,4
$\gamma_1$	0,6	0,6
$\lambda$	0,3	0,3
Seuil d'hypoxie	0,85	0,8

TABLE 2.1 – Paramètres utilisés pour les simulations présentées en Figures 2.6 et 2.7.

## 2.3 Calibration du modèle : problème inverse

### 2.3.1 Objectifs

L'objectif principal de ces travaux est, rappelons-le, de pouvoir être applicables à des cas cliniques pour, à terme, constituer une aide aux médecins. Nous avons un modèle qui semble pertinent et reproduit des dynamiques d'évolution comparables à celles observées cliniquement. Cependant ce modèle comporte un nombre important de paramètres qu'il convient de déterminer pour chaque tumeur étudiée. C'est pourquoi il est crucial de savoir calibrer le modèle étant donné un nombre très limité d'informations sur la métastase. En pratique, les médecins disposent d'un certain nombre d'images médicales (scanner par exemple) pour chaque métastase plus ou moins espacées dans le temps. À partir de ces données cliniques, il faut résoudre un problème inverse qui va avoir pour but de déterminer quel(s) jeu(x) de paramètres nous permet(tent) d'approcher au mieux le comportement de la métastase avec notre modèle.

Ce problème inverse est généralement mal posé, ce qui fait qu'il peut n'y avoir aucune solution satisfaisante comme il peut y en avoir plusieurs, voir une infinité. Toujours est-il que plus on a d'informations à notre disposition, meilleures sont nos chances d'aboutir à une solution correcte. Une condition nécessaire pour pouvoir tenter de résoudre le problème inverse est de disposer de deux images (ou snapshots) de la tumeur en deux instants distincts. La première servira de condition initiale pour la simulation du modèle alors que la deuxième (et éventuellement les suivantes) sera utilisée pour la calibration à proprement parler.

### 2.3.2 Premières hypothèses simplificatrices

Les inconnues et paramètres du problème sont nombreux. Ainsi, on compte 9 paramètres scalaires :  $\alpha$ ,  $\beta$ ,  $\gamma_0$ ,  $\gamma_1$ ,  $\eta$ ,  $\lambda$ ,  $M_{th}$ ,  $K$  et la valeur initiale de la quantité d'agent pro-angiogénique  $\xi_0$ . Le paramètre  $K$  n'est rien d'autre qu'une constante de raideur utilisée pour décrire la transition hypoxique, on fixe arbitrairement sa valeur à 5. On fixe de même la valeur initiale de  $\xi$ , par exemple à 0,1 considérant qu'à l'instant initial, le processus d'angiogenèse est déjà engagé. De plus, l'effet de la variable  $\xi$  est contrôlé par le paramètre  $\beta$  dans l'équation sur la vascularisation. On peut ainsi considérer que  $\beta$  et  $\xi_0$  sont liés et ainsi fixer l'un des deux.

De plus, il faut initialiser les champs  $T$  et  $M$ . La valeur initiale de la densité de cellules tumorales  $T_0$  provient directement de la première image médicale dont nous disposons. On fait l'hypothèse que la densité de vascularisation est constante dans le tissu sain d'une part, et dans la tumeur d'autre part, on a donc  $M_0 = a \times T_0 + b \times S_0$  (sachant que  $S_0 = 1 - T_0$ ). On remplace donc l'indétermination sur un champ bi-dimensionnel par deux inconnues scalaires. On considère que le tissu sain n'est pas hypoxique, soit  $b > M_{th}$ . De plus, la néo-vascularisation tumorale provient du tissu sain et on suppose qu'à l'instant initial la tumeur n'est pas complètement vascularisée (début de la phase vasculaire). On fait par conséquent l'hypothèse que la tumeur est moins vascularisée que le tissu sain soit  $a < b$ . Comme le phénomène d'hypoxie est modélisé à partir de fonctions seuil, seule importe réellement la position relative de  $a$ ,  $b$  et  $M_{th}$ . On fixe arbitrairement  $a = 0,8$  et  $b = 1$  et on garde  $M_{th}$  libre que l'on fera varier autour de la valeur de  $a$  pour décrire à la fois les comportements hypoxiques et non-hypoxiques de la tumeur.

Après ces premières considérations, il reste 7 paramètres scalaires à déterminer pour calibrer le modèle :  $\alpha$ ,  $\beta$ ,  $\gamma_0$ ,  $\gamma_1$ ,  $\eta$ ,  $\lambda$  et  $M_{th}$ .

### 2.3.3 Analyse de sensibilité

Pour tenter de simplifier davantage le problème inverse, on a procédé à une analyse de sensibilité sur le modèle dans le but de déterminer si certains paramètres sont moins influents que les autres sur les résultats des simulations. Dans notre cas, cela pourrait signifier que les intervalles de variation pour les paramètres les moins sensibles ont été mal choisis ou encore qu'on peut tout simplement fixer ces paramètres à des valeurs nominales. Pour cette analyse, nous avons utilisé la méthode de Morris (voir [51]). Pour chaque paramètre  $p_i$ , on considère l'effet élémentaire de sa variation sur le résultat de la simulation comme une variable aléatoire notée  $V_{p_i}$ . Plus précisément, ici on regarde les variations de la masse finale de la tumeur induites par des perturbations des

paramètres. Ainsi l'expression

$$\frac{\int_{\Omega} T_{model}(p_1, \dots, p_i + \epsilon, \dots, p_k, t_f) d\omega - \int_{\Omega} T_{model}(p_1, \dots, p_i, \dots, p_k, t_f) d\omega}{\epsilon}$$

est une réalisation de  $V_{p_i}$ , où  $(p_1, \dots, p_k)$  est le jeu de paramètres et  $t_f$  le temps final de simulation (dans notre cas, on a utilisé  $t_f = 4$ ). On calcule un estimateur de la moyenne  $m_{p_i}$  et de l'écart-type  $S_{p_i}$  de chaque variable  $V_{p_i}$ . Les points correspondant sont placés dans un plan  $(m, S)$ .

Cette méthode qualitative est bien adaptée à notre problématique puisqu'elle nous fournit un moyen visuel de comparer entre elles les influences des différents paramètres sur les simulations du modèle. De plus, il s'agit d'une méthode globale d'analyse de sensibilité puisqu'elle ne dépend pas d'un choix de jeu de paramètres initiaux (voir Table 2.2). Les intervalles de valeurs utilisés pour les paramètres ont été déterminés empiriquement pour que le domaine ainsi défini contienne les comportements intéressants, c'est à dire proches de ce que l'on observe cliniquement. Les résultats de l'analyse présentés Figure 2.8 sont basés sur environ 200 simulations du modèle.

paramètre	intervalle de variation
$\alpha$ (en $\text{mm}^{-2}$ )	[5; 25]
$\beta$	[5; 25]
$\eta$	[0,1; 1]
$\gamma_0$	[0,2; 1,4]
$\gamma_1$	[0; 0,3]
$\lambda$	[0,2; 1,2]
$M_{th}$	[0,7; 1,1]

TABLE 2.2 – Espace de paramètres utilisé pour l'analyse de sensibilité.

Plus le point est loin de l'origine du repère, plus le modèle est sensible au paramètre correspondant. De plus, si l'écart-type est grand, cela signifie que la dépendance du modèle au paramètre est non-linéaire ou qu'il y a des interactions croisées avec d'autres paramètres. Dans notre cas, le paramètre le plus influent est  $\gamma_0$  ce qui n'est pas étonnant vu qu'il contrôle la croissance exponentielle de la tumeur. Au contraire, le modèle n'est quasiment pas sensible aux variations de  $\alpha$  et  $\beta$  dans les intervalles de variation que l'on considère. On peut donc envisager de les fixer pour simplifier le problème inverse.

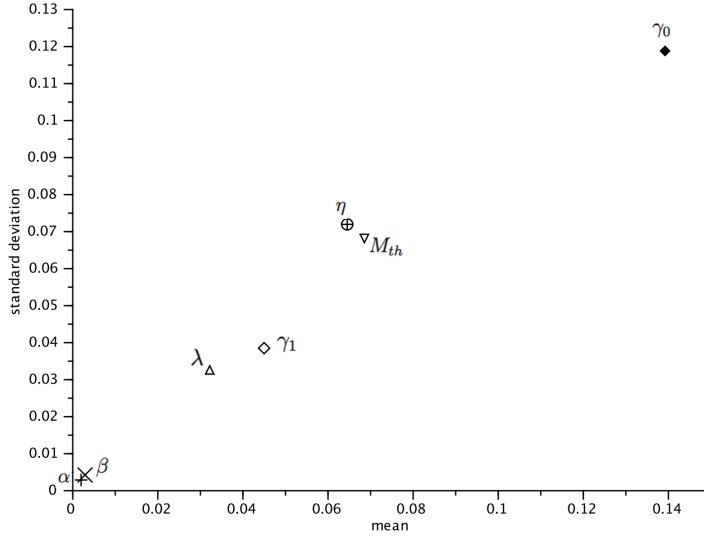


FIGURE 2.8 – Analyse de sensibilité sur le modèle de croissance tumorale par la méthode de Morris : effet de chaque paramètre sur le volume final de la tumeur.

### 2.3.4 Résolution du problème inverse et optimisation

On se donne une fonction coût des paramètres qui représente l'erreur L2 entre l'observable (c'est à dire les images médicales ou snapshots de la métastase en différents instants) et le comportement du modèle pour les paramètres testés  $(p_i)_{i=1}^k$ .

$$f_{obj}(p_1, \dots, p_k) = \sum_{i=1}^{ns} \int_{\Omega} (T_{snap}(t_i) - T_{modele}(p_1, \dots, p_k, t_i))^2 d\omega \quad (2.14)$$

Ensuite, on utilise une méthode d'optimisation classique de type gradient (voire une métaheuristique comme par exemple un algorithme génétique) pour trouver le minimum de  $f_{obj}$ . On contraint ce problème de minimisation en imposant des bornes à l'espace des paramètres parcouru lors de l'optimisation. On peut en effet cibler à partir d'observations simples (volume final de la tumeur...) certaines plages de variation crédibles pour les différents paramètres. Plus cet espace sera restreint et plus on peut espérer converger rapidement vers une bonne solution. La fonction objectif peut avoir plusieurs minima dans l'espace des paramètres étudié et un algorithme de descente risque de rester bloqué sur un minimum local. Si elles sont plus complexes à implémenter et parfois plus coûteuses en calcul, les métaheuristiques peuvent permettre d'éviter ce genre

de soucis.

On s'aperçoit assez vite que l'utilisation d'une telle fonction objectif a ses limites. En effet, si la tumeur est positionnée différemment entre deux images médicales (rotation, problèmes de recalage de l'image...), l'erreur L2 va être grande même si les tumeurs sont semblables (même forme générale, même masse...). Pour cela, on ajoute à la fonction objectif, une comparaison entre les masses tumorales de la donnée clinique et celle issue du modèle :

$$f_{obj}(p_1, \dots, p_k) = \sum_{i=1}^{ns} \left[ a \int_{\Omega} (T_{snap}(t_i) - T_{modele}(p_1, \dots, p_k, t_i))^2 d\omega + b \left| \int_{\Omega} T_{snap}(t_i) d\omega - \int_{\Omega} T_{modele}(p_1, \dots, p_k, t_i) d\omega \right| \right]. \quad (2.15)$$

Les paramètres  $a$  et  $b$  servent à pondérer la prise en compte de la norme L2 et des masses. On pourrait aussi ajouter un terme de régularisation sur les paramètres, comme par exemple la somme de leurs carrés.

**Méthode des sensibilités :** On se donne un jeu de paramètres de départ, et on se déplace de manière itérative dans l'espace des paramètres admissibles à partir de cette condition initiale. Le déplacement entre l'itéré  $n$  et l'itéré  $n+1$  se fait en suivant la direction opposée au gradient de la fonction objectif comme suit :

$$\forall i \in \{1, \dots, k\}, p_i^{n+1} = p_i^n - \alpha \frac{\partial f_{obj}}{\partial p_i}(p_1, \dots, p_k).$$

Où  $\frac{\partial f_{obj}}{\partial p_i}(p_1, \dots, p_k)$  est approximé par :

$$\frac{\partial f_{obj}}{\partial p_i}(p_1, \dots, p_k) = \frac{f_{obj}(p_1, \dots, p_i + \epsilon_i, \dots, p_k) - f_{obj}(p_1, \dots, p_i, \dots, p_k)}{\epsilon_i}.$$

Le pas  $\alpha$  est ici constant mais il peut être optimisé (variante du gradient à pas optimal).

**Algorithme génétique :** Cette méthode métaheuristique bio-inspirée se base sur l'adaptation algorithmique d'une méthode d'optimisation naturelle : la Théorie de Darwin. Ici un individu représente un jeu de l'espace des paramètres admissibles (un  $k$ -uplet). On les compare selon la même fonction objectif  $f_{obj}$  donnée en (2.15). À chaque itération (ou génération) les individus sont sélectionnés en fonction de leur valeur de  $f_{obj}$  (équivalent de la sélection naturelle) puis croisés entre eux (reproduction). C'est à dire que les jeux de paramètres sont mélangés, les meilleurs individus ayant plus de chance de transmettre leur patrimoine mais sans écarter nécessairement les moins performants. Lors de cette étape de mise à jour de la population, un paramètre peut



être modifié sans raison précise, il s'agit d'une « mutation spontanée » qui est un élément clé de la Théorie de l'Évolution et de l'efficacité de l'algorithme.

**Algorithme des essaims particuliers :** Ici encore, il s'agit d'une méta-heuristique inspirée par des phénomènes naturels notamment le déplacement des vols d'oiseaux. Chaque « particule » est un point de l'espace des paramètres initialement tiré aléatoirement. Au fil des itérations, les particules se déplacent dans l'espace pour minimiser leur valeur de fonction coût. Ces déplacements se font en fonction de trois critères :

- la meilleure position passée de la particule (position pour laquelle la particule avait la plus petite valeur de fonction coût),
- la meilleure position connue pour l'ensemble des particules,
- l'inertie de la particule, c'est à dire sa résistance au mouvement.

Une fois ces différentes composantes paramétrées, cette méthode de minimisation permet une fouille efficace de l'espace des paramètres du modèle puisqu'elle permet d'éviter de tomber dans des minima locaux.

En pratique, on utilise souvent une stratégie mixte. On lance tout d'abord un algorithme génétique ou plus simplement une méthode de Monte Carlo (tirages aléatoires dans l'espace des paramètres). Le résultat de cette première étape est affiné par la suite à l'aide d'un algorithme de type gradient.

Quelle que soit la méthode d'optimisation adoptée, la résolution du problème inverse nécessite un grand nombre d'évaluations de la fonction objectif, soit autant de simulations du modèle ( $k+1$  par itération pour l'algorithme du gradient et autant que d'individus par itération pour l'algorithme génétique). Ainsi on peut vite atteindre des temps de calcul très importants, surtout si on souhaite utiliser un maillage relativement fin et se laisser la possibilité de traiter des cas en trois dimensions. C'est pour répondre à cette problématique et alléger les calculs que nous avons mis en place un modèle réduit par POD.

## 2.4 Proper Orthogonal Decomposition (POD)

### 2.4.1 Principe de la méthode

La Décomposition Orthogonale aux valeurs Propres ou POD ([38, 39]) est une méthode permettant de simplifier la résolution d'un système d'équations aux dérivées partielles en l'approximant par un système de dimension inférieure. Pour cela, on cherche à séparer les variables en temps et en espace, c'est à dire que si l'on considère une variable  $u(X,t)$ , on souhaite déterminer

les fonctions du temps  $a_i(t)$  et de l'espace  $\Phi_i(X)$  telles que :

$$u(X,t) = \sum_{i=1}^d a_i(t) \Phi_i(X) + \epsilon \quad (2.16)$$

Où  $\epsilon$  est le plus petit possible.

La méthode POD, telle que nous l'utilisons, s'appuie sur l'exploitation d'une base de données qui contient des snapshots  $(S_i)_i$  représentant des captures, à un instant donné, de comportements possibles de la variable que l'on souhaite décomposer. C'est à partir de cette base de données qu'on extrait les fonctions  $\Phi_i$ , ou modes POD. Celles-ci sont prises comme formant une base orthonormée de l'espace vectoriel engendré par les snapshots (espace de dimension  $d$ ). Une caractéristique supplémentaire est que la projection tronquée de  $u$  sur la base des  $\Phi_i$  doit être la plus représentative possible, c'est à dire :

$$\forall h \in \{1, \dots, d\}, \|u(X,t) - \sum_{i=1}^h a_i(t) \Phi_i(X)\| = \min_{(F_i)_i} \|u(X,t) - \sum_{i=1}^h a_i(t) F_i(X)\|$$

Cette propriété est très intéressante puisqu'elle nous permet d'utiliser un nombre limité de modes tout en gardant une représentation relativement fidèle de la solution.

### 2.4.2 Obtention du la formulation POD

On considère  $X = (x_1, \dots, x_n)$  la matrice des échantillons ou snapshots de la variable  $x$ . Chaque vecteur colonne de  $X$ , noté  $x_i \in \mathbb{R}^m, \forall i \in (1, 2, \dots, n)$ , correspond à une réalisation de  $x$  de dimension  $m$ . La matrice  $X$  est donc de taille  $m \times n$ .

Dans le cas qui nous intéresse, on a  $m \geq n$ . C'est à dire qu'*a priori* la dimension des  $x_i$  est plus grande que le nombre de snapshots. Concrètement, les  $x_i$  vont être les images de la tumeur mises sous forme de vecteurs et leur taille  $m$  correspond donc au nombre de voxels de l'image.

La matrice  ${}^tXX \in \mathbb{R}^{n \times n}$  correspondant à la matrice d'autocorrélation des snapshots  $x_i$  est semi-définie positive par définition. On note ses valeurs propres (toutes positives ou nulles) de la manière suivante :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0.$$

On note de même  $v_1, v_2, \dots, v_n$  les vecteurs propres de  ${}^tXX$  associés à ces valeurs propres et  $\sigma_i = \sqrt{\lambda_i}, \forall i \in (1, 2, \dots, n)$ , les valeurs singulières de  $X$ .

**Décomposition aux valeurs singulières (SVD) de  $X$  :** Soit  $V = (V_1, V_2)$  où  $V_1 = (v_1, \dots, v_r)$  et  $V_2 = (v_{r+1}, \dots, v_n)$  la matrice ayant pour colonnes les vecteurs propres de  ${}^tXX$ . La matrice  $V$  est une matrice orthonormale de  $\mathbb{R}^{n \times n}$  (les  $v_i$  forment une base orthonormale de  $\mathbb{R}^n$ ).

$${}^tXXV = {}^tXX(v_1, \dots, v_n) = ({}^tXXv_1, \dots, {}^tXXv_n) = (\lambda_1 v_1, \dots, \lambda_n v_n)$$

Car les  $v_i$  sont les vecteurs propres de  ${}^tXX$  associés aux valeurs propres  $\lambda_i$ .

Soit en écriture matricielle :

$${}^tXXV = V \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = V \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}.$$

On multiplie à gauche par  ${}^tV$  :

$$\begin{aligned} {}^tV {}^tXXV &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \text{ car } V \text{ est orthonormale } ({}^tVV = I_n) \\ &= \begin{pmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

où  $\Sigma_r^2 = \text{diag}(\sigma_1^2, \dots, \sigma_r^2)$  est la matrice des valeurs propres strictement positives de la matrice d'autocorrélation  ${}^tXX$  et les blocs 0 sont des matrices rectangulaires aux bonnes dimensions.

Pour la suite, on pose  $U_1 = XV_1 \Sigma_r^{-1}$  avec  $\Sigma_r^{-1} = \text{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1})$ ,  $U_1 \in \mathbb{R}^{m \times r}$ .

$${}^tU_1 U_1 = ({}^t(XV_1 \Sigma_r^{-1}))(XV_1 \Sigma_r^{-1}) = \underbrace{{}^t\Sigma_r^{-1}}_{\Sigma_r^{-1}} \underbrace{{}^tV_1 {}^tXX V_1}_{\Sigma_r^2} \Sigma_r^{-1} = I_r \quad (2.17)$$

De (2.17), on déduit que les colonnes de  $U_1$  sont orthonormales une à une ( $U_1 = (u_1, \dots, u_m)$ ). Par application du théorème de la base incomplète, il existe  $m-r$  vecteurs de  $\mathbb{R}^m$  que l'on notera  $u_{r+1}, \dots, u_m$  tels que la famille de vecteurs  $(u_1, \dots, u_m)$  forme une base orthonormée de  $\mathbb{R}^m$ . On note  $U = (U_1, U_2)$  avec  $U_2 = (u_{r+1}, \dots, u_m)$ . La matrice  $U$  ainsi définie est une matrice orthonormée de taille  $m \times m$ .

Regardons la matrice  ${}^tUXV$  :

$$\begin{aligned} {}^tUV &= {}^t(U_1, U_2)X(V_1, V_2) \\ &= \begin{pmatrix} {}^tU_1XV_1 & {}^tU_1XV_2 \\ {}^tU_2XV_1 & {}^tU_2XV_2 \end{pmatrix}. \end{aligned}$$

Calculons les différents termes de cette matrice :

$${}^tU_1XV_1 = {}^t(XV_1\Sigma_r^{-1})XV_1 = \Sigma_r^{-1} \underbrace{{}^tV_1{}^tXXV_1}_{\Sigma_r^2} = \Sigma_r,$$

$${}^tU_1XV_2 = \Sigma_r^{-1} \underbrace{{}^tV_1{}^tXXV_2}_0 = 0,$$

$${}^tU_2XV_1 = {}^tU_2(XV_1)(\Sigma_r^{-1}\Sigma_r) = {}^tU_2(XV_1\Sigma_r^{-1})\Sigma_r = \underbrace{{}^tU_2U_1}_0 \Sigma_r = 0.$$

0 car U est orthonormale

Pour l'étude du dernier terme, à savoir  ${}^tU_2XV_2$ , repartons du constat que  ${}^tXXV_2 = 0$  (car les colonnes de  $V_2$  sont les vecteurs propres de  ${}^tXX$  associés à la valeur propre 0). On multiplie à gauche par  ${}^tV_2$  :

$${}^tV_2{}^tXXV_2 = {}^t(XV_2)XV_2 = 0,$$

donc

$$\text{tr}({}^t(XV_2)XV_2) = 0,$$

et

$$XV_2 = 0.$$

On conclut donc que  ${}^tU_2XV_2 = 0$ .

$$\text{Ainsi } {}^tUV = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}.$$

On multiplie à gauche par  $U$  et à droite par  $V$  sachant que ces deux matrices sont orthogonales :

$$X = U \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V. \quad (2.18)$$

La formulation ci-dessus (2.18) est la décomposition aux valeurs singulières, ou SVD, de  $X$ . Plus précisément, et ce sera utile par la suite, on note que  $\forall i \in (1, \dots, n), Xv_i = \sigma_i u_i$  et  ${}^tXv_i = \sigma_i u_i$ . Les vecteurs  $u_i$  et  $v_i$  sont appelés vecteurs singuliers respectivement à gauche et à droite de  $X$  et sont vecteurs propres respectivement de  $X^tX$  et de  ${}^tXX$  associés aux valeurs propres  $\lambda_i = \sigma_i^2$ .

**Obtention de la base POD à partir de l'expression SVD de  $X$  :** On cherche à exprimer la base de décomposition POD de  $X$  à partir de la SVD. Pour cela, on part de l'expression (2.18) or on ne sait pas expliciter l'intégralité de la matrice  $U$ . On peut réduire l'expression SVD de  $X$  en remarquant que l'utilisation de  $r$  modes propres suffit. On a donc :

$$X = U_1 \Sigma_r {}^t V_1. \quad (2.19)$$

On a, par définition de  $U_1$  :

$$U_1 = X V_1 \Sigma_r^{-1},$$

or

$$V_1 \Sigma_r^{-1} = \left( \frac{1}{\sigma_1} v_1, \dots, \frac{1}{\sigma_r} v_r \right),$$

donc

$$U_1 = \left( \frac{1}{\sigma_1} X v_1, \dots, \frac{1}{\sigma_r} X v_r \right),$$

soit

$$\forall i \in (1, \dots, r), u_i = \frac{1}{\sigma_i} \begin{pmatrix} \sum_{j=1}^n v_i(j) x_j(1) \\ \vdots \\ \sum_{j=1}^n v_i(j) x_j(m) \end{pmatrix} = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^n v_i(j) x_j,$$

ou encore

$$\forall i \in (1, \dots, r), u_i = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^n v_i(j) x_j. \quad (2.20)$$

La formule (2.20) donne l'expression des modes POD où, rappelons-le,  $x_j$  est le  $j^{\text{ème}}$  snapshot de la variable  $x$ ,  $\lambda_i$  est la  $i^{\text{ème}}$  valeur propre non nulle de la matrice d'autocorrélation  ${}^t X X$  des snapshots et  $v_i(j)$  est la  $j^{\text{ème}}$  composante du  $i^{\text{ème}}$  vecteur propre de  ${}^t X X$ . Les modes POD sont donc les vecteurs singuliers à gauche de  $X$ .

On part maintenant de l'expression (2.19) et on cherche à obtenir la projection sur les modes POD  $u_i$ .

$$\Sigma_r {}^t V_1 = \begin{pmatrix} \sigma_1 v_1(1) & \cdots & \sigma_1 v_1(n) \\ \vdots & \ddots & \vdots \\ \sigma_r v_r(1) & \cdots & \sigma_r v_r(n) \end{pmatrix},$$

$$X = U_1 \Sigma_r {}^t V_1 = U_1 \begin{pmatrix} \sigma_1 v_1(1) & \cdots & \sigma_1 v_1(n) \\ \vdots & \ddots & \vdots \\ \sigma_r v_r(1) & \cdots & \sigma_r v_r(n) \end{pmatrix},$$

ainsi

$$\forall i \in (1, \dots, n), \forall j \in (1, \dots, m), x_i(j) = \sum_{k=1}^r (\sigma_k v_k(i)) u_k(j).$$

On a donc la décomposition de  $x$  sur la base POD :

$$\forall i \in (1, \dots, n), x_i = \sum_{k=1}^r (\sigma_k v_k(i)) u_k = \sum_{k=1}^r a_k^{x_i} u_k, \text{ où } a_k^{x_i} = \sigma_k v_k(i). \quad (2.21)$$

La base des modes POD  $(u_1, \dots, u_r)$  est une base orthonormée, on s'attend donc à pouvoir écrire les coefficients de projection  $a_k^{x_i}$  comme produits scalaires de  $x_i$  avec ces vecteurs de base.

Calculons donc ces produits scalaires (on utilise le produit scalaire usuel sur  $\mathbb{R}^m$ ) :

$$\begin{aligned} \langle x_i, u_k \rangle &= \left\langle x_i, \frac{1}{\sigma_k} \sum_{j=1}^n v_k(j) x_j \right\rangle \\ &= \frac{1}{\sigma_k} \sum_{j=1}^n v_k(j) \langle x_i, x_j \rangle. \end{aligned}$$

Or,  $v_k$  étant vecteur propre de  ${}^tXX$  associé à la valeur propre  $\lambda_k$  ( ${}^tXX v_k = \lambda_k v_k$ ) :

$$\sum_{j=1}^n v_k(j) \langle x_i, x_j \rangle = \sum_{j=1}^n {}^tX X_{ij} v_k(j) = \lambda_k v_k(i) = \sigma_k^2 v_k(i),$$

donc

$$\langle x_i, u_k \rangle = \frac{1}{\sigma_k} \sigma_k^2 v_k(i) = \sigma_k v_k(i).$$

On retrouve bien par projection sur les modes POD, l'expression des coefficients  $a_k^{x_i}$  trouvée en (2.21).

**Optimalité de la base de décomposition :** Soit  $\varphi_1, \varphi_2, \dots, \varphi_m \in \mathbb{R}^m$  une base orthonormée choisie arbitrairement.  $\forall i \in (1, \dots, n), x_i = \sum_{j=1}^m c_j^i \varphi_j$  où  $c_j^i = {}^t\varphi_j x_i$ .

On tronque la projection de  $x_i$  en ne conservant que les  $l \in (1, \dots, m)$  premiers vecteurs de base :

$$x_i(l) = \sum_{j=1}^l c_j^i \varphi_j, \forall i \in (1, \dots, n).$$

Regardons l'erreur de troncature commise sur l'ensemble des snapshots :

$$\begin{aligned}
\epsilon^2(l) &= \sum_{i=1}^n \|x_i - x_i(l)\|^2 = \sum_{i=1}^n \left\| \sum_{j=l+1}^m c_j^i \varphi_j \right\|^2 \\
&= \sum_{i=1}^n \sum_{j=l+1}^m (c_j^i)^2 = \sum_{i=1}^n \sum_{j=l+1}^m {}^t\varphi_j x_i {}^t x_i \varphi_j \\
&= \sum_{j=l+1}^m {}^t\varphi_j X^t X \varphi_j = \|{}^t X \Phi_{m-l}\|_F^2,
\end{aligned}$$

où  $\Phi_{m-l} = (\varphi_{l+1}, \dots, \varphi_m)$  et  $\|\cdot\|_F$  est la norme de Frobenius pouvant être définie de plusieurs façons :  $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2} = \sqrt{\text{tr}(AA^t)}$ .

On cherche les vecteurs de base  $(\varphi_i)_{i=1, \dots, m}$  minimisant cette erreur et tels que la base  $\Phi = (\varphi_1, \dots, \varphi_m)$  soit orthonormée.

$$\begin{cases} \min_{\varphi_j} \epsilon^2(l) = \sum_{j=l+1}^n {}^t\varphi_j X^t X \varphi_j, \\ {}^t\varphi_i \varphi_j = \delta_{ij}. \end{cases} \quad (2.22)$$

Pour résoudre ce problème de minimisation sous contraintes, on introduit les multiplicateurs de Lagrange  $w_{i,j}$  et on écrit le lagrangien :

$$L = \sum_{j=l+1}^m {}^t\varphi_j X^t X \varphi_j - \sum_{i=l+1}^m \sum_{j=l+1}^m w_{i,j} ({}^t\varphi_i \varphi_j - \delta_{ij}).$$

On dérive ensuite par rapport à  $\varphi_j$  :

$$\begin{aligned}
\forall j \in (l+1, \dots, m), \frac{\partial L}{\partial \varphi_j} &= 2(X^t X \varphi_j - w_{i,j} \sum_{i=l+1}^m \varphi_i) \\
&= 2X^t X \varphi_j - 2\Phi_{m-l} w_j,
\end{aligned}$$

où  $w_j = (w_{l+1,j}, \dots, w_{m,j})$ . En reprenant cette expression sous forme matricielle, avec  $W_{m-l} = (w_{l+1}, \dots, w_m)$ , on obtient :

$$\frac{\partial L}{\partial \Phi_{m-l}} = 2X^t X \Phi_{m-l} - 2\Phi_{m-l} W_{m-l},$$

puis en annulant la dérivée, il reste :

$$X^t X \Phi_{m-l} = \Phi_{m-l} W_{m-l}. \quad (2.23)$$

On multiplie (2.23) à gauche par la matrice  ${}^t\Phi_{m-l}$  :

$$W_{m-l} = {}^t\Phi_{m-l} X^t X \Phi_{m-l}.$$

On note donc que  $W_{m-l}$  est une matrice semi-définie positive de  $\mathbb{R}^{(m-l) \times (m-l)}$ . Il existe donc une matrice orthogonale  $P$  telle que :

$$\Lambda = {}^t P W_{m-l} P = {}^t P {}^t \Phi_{m-l} X^t X \Phi_{m-l} P,$$

où  $\Lambda$  est une matrice diagonale. On multiplie (2.23) à droite par  $P$  :

$$\begin{aligned} X^t X \Phi_{m-l} P &= \Phi_{m-l} P^t P W_{m-l} P \\ &= \Phi_{m-l} P \Lambda. \end{aligned}$$

De l'équation ci-dessus, on déduit que les éléments diagonaux de  $\Lambda$  sont des valeurs propres  $\lambda_i$  de  $X^t X$ , et la matrice  $\Phi_{m-l} P$  a pour colonnes les vecteurs propres de  $X^t X$  correspondant aux  $\lambda_i$ . Il s'agit donc, comme on l'a vu précédemment, des vecteurs singuliers à gauche de  $X$   $(u_i)_{i \in (1, \dots, m)}$ .

Pour la suite on utilise le théorème suivant :

**Théorème .** Soient  $A \in \mathbb{R}^{n \times m}$ ,  $Q \in \mathbb{R}^{m \times m}$  une matrice orthogonale et  $\|\cdot\|_F$  la norme de Frobenius, alors  $\|A\|_F = \|AQ\|_F$ .

*Démonstration.*

$$\begin{aligned} \|AQ\|_F &= \sqrt{\text{tr}({}^t(AQ)AQ)} = \sqrt{\text{tr}({}^t Q {}^t A A Q)} \\ &= \sqrt{\text{tr}(AQ {}^t Q {}^t A)} = \sqrt{\text{tr}(A {}^t A)} \text{ car } Q \text{ est orthogonale} \\ &= \sqrt{\text{tr}({}^t A A)} = \|A\|_F. \end{aligned}$$

□

Reprenons l'expression de l'erreur, par application du précédent théorème,  $P$  étant orthogonale :

$$\begin{aligned} \epsilon^2(l) &= \|{}^t X \Phi_{m-l}\|_F^2 = \|{}^t X \Phi_{m-l} P\|_F^2 \\ &= \text{tr}({}^t({}^t X \Phi_{m-l} P) {}^t X \Phi_{m-l} P) = \text{tr}(\Lambda) \\ &= \sum_{i=l+1}^m \lambda_i = \sum_{i=l+1}^m \sigma_i^2. \end{aligned}$$

On a donc bien montré que l'optimalité de la décomposition au sens du problème de minimisation (2.22) est atteinte pour la base POD de  $X$   $(u_i)_{i=1, \dots, m}$  et on a de plus établi la valeur de l'erreur de troncature sur l'ensemble des snapshots.



## 2.5 Modèle réduit par POD et simulations

Dans cette partie, nous allons expliciter l'utilisation de la POD dans notre cas.

### 2.5.1 Modèle réduit

La Décomposition Orthogonale aux valeurs Propres permet, comme on l'a vu, de réduire la dimension d'un problème ou plutôt d'approximer un système par un système réduit de dimension moindre. Ici, nous souhaitons utiliser cette méthode afin de simplifier le modèle EDP (2.1) à (2.9), la visée finale étant de réduire au maximum les temps de calcul tout en gardant une précision suffisante pour l'utilisation du modèle sur des cas concrets.

Lorsqu'on résout le système direct, on se rend vite compte que les équations les plus coûteuses du point de vue de la résolution sont l'équation sur la densité de cellules métastatiques (2.1) et le système de Darcy (2.7) et (2.6). C'est pourquoi on choisit de décomposer les variables  $T$  et  $\pi$ .

Pour cela, on a vu la nécessité de disposer d'une base de snapshots représentant les comportements admissibles des variables. Cette base est donc générée une fois pour toutes. On discrétise tout d'abord l'espace des paramètres  $\Pi$  : si l'on considère que notre modèle comporte, en absence de traitement, six paramètres indépendants ( $\alpha$ ,  $\beta$ ,  $\gamma_0$ ,  $\gamma_1$ ,  $\eta$  et  $\lambda$ ), alors  $\Pi$  est inclus dans  $\mathbb{R}^6$ . Les plages de variations de chaque paramètre sont établies de façon empirique et en fonction du type de comportement que l'on veut pouvoir reproduire. Une fois la base de snapshots obtenue, on calcule les modes d'après la procédure explicitée à la partie précédente.

Soient  $\Phi_i^T$  et  $\Phi_i^\pi$  les modes propres respectifs de  $T$  et  $\pi$ . On notera de même  $n_T$  et  $n_\pi$  le nombre de modes utilisés pour ces deux décompositions. On fait donc l'approximation  $T = \sum_{i=1}^{n_T} a_i^T(t) \Phi_i^T(x, y)$  et  $\pi = \sum_{i=1}^{n_\pi} a_i^\pi(t) \Phi_i^\pi(x, y)$ .

#### POD sur la densité de cellules tumorales $T$ :

On injecte l'expression POD de  $T$  dans l'équation (2.1) :

$$\sum_i \frac{\partial a_i^T}{\partial t} \Phi_i^T + \sum_i a_i^T \nabla \cdot (\mathbf{v} \Phi_i^T) = \sum_i a_i^T \Gamma \Phi_i^T.$$

Puis on projette cette expression sur les modes propres de  $T$ , ceux-ci formant une base orthonormée de l'espace vectoriel engendré par les snapshots. On obtient ainsi un système d'équations différentielles sur les coefficients  $a_i^T$  :

$$\forall j \in \{1, \dots, n_T\}, \frac{\partial a_j^T}{\partial t} + \sum_i a_i^T \langle \nabla \cdot (\mathbf{v} \Phi_i^T), \Phi_j^T \rangle = \sum_i a_i^T \langle \Gamma \Phi_i^T, \Phi_j^T \rangle,$$

ou encore :

$$\forall j \in \{1, \dots, n_T\}, \frac{\partial a_j^T}{\partial t} = \sum_i a_i^T \langle \Gamma \Phi_i^T - \nabla \cdot (\mathbf{v} \Phi_i^T), \Phi_j^T \rangle. \quad (2.24)$$

### POD sur la pression $\pi$ :

En composant les équations 2.7 et 2.6 on obtient :

$$-\Delta \pi = \Gamma T.$$

Puis on utilise la décomposition POD de  $\pi$  :

$$-\sum_j a_j^\pi \Delta \Phi_j^\pi = \Gamma T.$$

Enfin on projette cette équation sur les modes de  $\pi$  (base orthonormée) afin d'obtenir un système d'équations linéaires sur les coefficients  $a_i^\pi$  :

$$\forall i \in \{1, \dots, n_\pi\}, -\sum_j a_j^\pi \langle \Delta \Phi_j^\pi, \Phi_i^\pi \rangle = \langle \Gamma T, \Phi_i^\pi \rangle. \quad (2.25)$$

### 2.5.2 Mise en place du schéma numérique

Le but de cette partie est de décrire le schéma numérique utilisé pour approcher les solutions du système réduit décrit précédemment.

#### Conditions initiales :

Pour chaque variable du problème, on se fixe des conditions initiales. Pour trouver les valeurs à  $t=0$  des coefficients POD de  $T$ , on projette  $T(t=0)$  (donnée du problème) sur la base  $(\Phi_i^T)_i$ . En ce qui concerne les coefficients  $a_i^\pi$ , leur calcul étant issu d'une simple résolution de système linéaire, leur connaissance à l'instant initial est inutile.

Ensuite, on prend :

- $S(t=0) = 1 - T(t=0)$  ;
- $M(t=0) = aS + bT$  (avec  $a \geq M_{th}$ ) ;
- $\xi(t=0) = 0,1$  ;
- Les variables  $\gamma_p$  et  $\gamma_d$  sont initialisées à partir des expressions (2.2) et (2.3).
- La pression et la vitesse initiales ( $\pi(t=0)$  et  $\mathbf{v}(t=0)$ ) sont calculées à partir de  $T(t=0)$  et des équations (2.7) et (2.6).

### Schéma numérique :

On décrit ici, pour chaque variable du problème, le passage du temps  $t$  au temps  $t + dt$ .

**calcul des  $a_i^T$  :** Repartons de l'expression (2.24) :

$$\forall j \in \{1, \dots, n_T\}, \frac{\partial a_j^T}{\partial t} = \sum_i a_i^T \langle \Gamma \Phi_i^T - \nabla \cdot (\mathbf{v} \Phi_i^T), \Phi_j^T \rangle.$$

On peut noter que,  $\nabla \cdot (\mathbf{v} \Phi_i^T) = (\nabla \cdot \mathbf{v}) \Phi_i^T + \mathbf{v} \cdot \nabla \Phi_i^T$ , or d'après l'équation (2.6),  $\nabla \cdot \mathbf{v} = \Gamma T$ , ce qui donne :

$$\forall j \in \{1, \dots, n_T\}, \frac{\partial a_j^T}{\partial t} = \sum_i a_i^T \langle \Gamma \Phi_i^T (1 - T) - \mathbf{v} \cdot \nabla \Phi_i^T, \Phi_j^T \rangle.$$

Notons  $(a_i^T)^n$  la valeur de  $a_i^T$  à l'instant  $t_n = t_0 + n \cdot dt$  (dans le cas où le pas de temps est constant) et  $M_n$  la matrice de composantes  $M_n(i, j) = \langle \Gamma^n \Phi_j^T (1 - T^n) - \mathbf{v}^n \cdot \nabla \Phi_j^T, \Phi_i^T \rangle$ . On utilise un schéma de Crank-Nicholson pour le passage de  $t_n$  à  $t_{n+1}$ . En écriture matricielle et en notant  $A_n^T$  le vecteur colonne des  $(a_i^T)^n$ , on a :

$$(I_{n_p} - \frac{dt}{2} M_n) A_{n+1}^T = (I_{n_p} + \frac{dt}{2} M_n) A_n^T. \quad (2.26)$$

Les valeurs des coefficients au temps  $t_{n+1}$  sont donc calculées par résolution du système linéaire (2.26).

La matrice  $M_n$  doit être recalculée à chaque pas de temps, cependant, certaines parties du calcul concernant les modes propres peuvent être effectuées un fois pour toutes (notamment les  $\nabla \Phi_i$ ).

Une fois que les coefficients  $(a_i^T)^{n+1}$  sont identifiés, on peut calculer  $T^{n+1}$  (qui sera utile par la suite) avec la formule de décomposition sur les modes propres :  $T^{n+1} = \sum_{i=1}^{n_T} (a_i^T)^{n+1} \Phi_i^T$ .

**mise à jour des variables non réduites :** Maintenant qu'on dispose de la valeur à jour de la densité de cellules tumorales, on peut recalculer les divers champs et variables du problème (les calculs des coefficients  $a_i^T$  et de la vitesse seront cependant détaillés dans le paragraphe suivant).

**tissu sain :**  $S^{n+1} = 1 - T^{n+1}$ .

**quantité d'agent angiogénique :** On part de l'expression (2.8) et on splitte la résolution en deux :

1. première partie explicite :

$$\xi^{temp} = \xi^n + dt[\alpha \int_{\Omega} (1 - \frac{\gamma_p^n}{\gamma_0}) T^{n+1} d\omega],$$

2. partie exponentielle :

$$\xi^{n+1} = \xi^{temp} \exp(-\lambda dt).$$

**vascularisation :** Là encore on part de l'équation (2.9) et on effectue le calcul en deux temps :

1. une partie explicite :

$$M^{temp} = M^n + dt[\beta \xi^{n+1} (1 - \frac{\gamma_p^n}{\gamma_0}) T^{n+1}],$$

2. et une partie exponentielle :

$$M^{n+1} = M^{temp} \exp(-\eta T^{n+1} dt).$$

**taux de mort cellulaire et de réplication :** À partir des équations (2.2) et (2.3), on obtient :

$$\begin{aligned} \gamma_p^{n+1} &= \gamma_0 \frac{1 + \tanh(K(M^{n+1} - M_{th}))}{2}, \\ \gamma_d^{n+1} &= \gamma_1 \frac{1 + \tanh(K(M^{n+1} - M_{th}))}{2}. \end{aligned}$$

Et on a donc finalement :

$$\Gamma^{n+1} = \gamma_p^{n+1} - \gamma_d^{n+1} - \delta \frac{\gamma_p^{n+1}}{\gamma_0} M^{n+1}.$$

**calcul des  $a_i^\pi$  :** On utilise cette fois l'expression (2.25) :

$$\forall i \in \{1, \dots, n_\pi\}, \sum_j a_j^\pi \langle \Delta \Phi_j^\pi, \Phi_i^\pi \rangle = - \langle \Gamma T, \Phi_i^\pi \rangle.$$

Ce qui donne en écriture matricielle au temps  $t_{n+1}$  :

$$N A_{n+1}^\pi = -B_{n+1}. \quad (2.27)$$

Où  $N$  est la matrice de taille  $n_\pi \times n_\pi$  telle que  $N(i, j) = \langle \Delta \Phi_i^\pi, \Phi_j^\pi \rangle$ , et  $B_{n+1}$  le vecteur colonne de composante  $B_{n+1}(i) = \langle \Gamma^{n+1} T^{n+1}, \Phi_i^\pi \rangle$ .

Ici, la matrice  $N$  n'est calculée qu'une fois et seul le vecteur  $B_{n+1}$  doit être mis à jour à chaque pas de temps. Il est aussi intéressant de noter qu'il ne s'agit plus d'équations aux dérivées partielles ou ordinaires mais d'un simple système linéaire.

À partir des valeurs des  $(a_i^\pi)^{n+1}$ , on peut ainsi calculer  $\mathbf{v}^{n+1}$  (en remarquant par exemple que  $\mathbf{v}^{n+1} = - \sum_i (a_i^\pi)^{n+1} \nabla \Phi_i^\pi$ ).

## 2.6 Applications de la méthode

Dans cette partie, nous allons reprendre l'ensemble de la méthode théorique décrite précédemment et expliciter sa mise en oeuvre sur des cas pratiques. On mettra ainsi en relief les problèmes rencontrés et les réponses apportées à ces derniers et on finira sur l'étude d'exemples cliniques.

### 2.6.1 Données cliniques nécessaires

Lorsque l'on veut prédire l'évolution d'une métastase chez un patient, on a comme on l'a vu, besoin d'un certain nombre de données. Il faut une donnée initiale qui va servir pour les simulations du modèle ainsi qu'une ou plusieurs données « intermédiaires » qui vont servir à calibrer le modèle. La POD est une méthode de réduction d'ordre qui supporte assez mal les irrégularités spatiales. Plus les formes de la tumeur seront irrégulières, et plus le nombre de modes à utiliser pour obtenir une approximation acceptable de la solution sera grand. On prend donc soin de régulariser les bords de la tumeur. Enfin, il est important de s'assurer que les espacements temporels entre les snapshots utilisés sont d'un même ordre de grandeur.

### 2.6.2 Base de données et extraction des modes POD

La première difficulté lors de la génération de la base de données qui servira pour mettre en place la POD consiste en l'adimensionnement en temps du système. On met donc en place une échelle de temps qui tient compte de l'espacement en jours entre les différentes données dont nous disposons. La première donnée correspond à la tumeur à  $t=0$  puisqu'il s'agit de la condition initiale. Le temps final est celui du dernier snapshot que l'on souhaite utiliser pour le problème inverse.

Par exemple, si on dispose de trois images d'une tumeur régulièrement espacées, on place la première à  $t=0$  et la troisième à un temps arbitraire, par exemple  $t=5$ . La deuxième image sera ensuite positionnée dans l'échelle temporelle adimensionnée en fonction des intervalles de temps réels entre les images. Lorsqu'on va contrôler la pertinence du calibrage du modèle sur une quatrième donnée, il suffira de calculer l'instant correspondant à cette image dans notre échelle temporelle adimensionnée. *Il est important de noter que les jeux de paramètres pour une tumeur donnée, et donc la résolution du problème inverse, sont conditionnés par ce choix d'échelle temporelle.*

Ensuite vient le choix de l'espace des paramètres à parcourir. En effet il faut déterminer les plages de variation de chacun des paramètres que l'on souhaite calculer via le problème inverse. Cela se fait de façon empirique en tenant compte de l'échelle temporelle déterminée au préalable, de la taille de la tumeur initiale et de l'évolution de la tumeur dans l'intervalle de temps considéré.

Cette phase est primordiale. Si on a, dans la base de données, des comportements aberrants, cela va compromettre l'extraction des modes. Un des cas classiques et à éviter est celui d'une tumeur qui croît jusqu'à sortir du domaine de calcul. Comme il est difficile de s'assurer au cas par cas que chacune des simulations de la base (qui peut en contenir plus de 15000) est admissible, on utilise une procédure de tri automatisée qui sélectionne les simulations sur un critère prédéfini. Le critère le plus utilisé est le rapport  $\frac{\text{masse finale de la tumeur}}{\text{masse initiale de la tumeur}}$ , qui permet déjà d'écarter un certain nombre de simulations pouvant fausser l'extraction des modes.

La base de données est générée en parallèle. On conserve, pour chaque simulation, un certain nombre de snapshots répartis dans l'intervalle de temps. Dans la plupart des cas on a ainsi cinq images de la tumeur (y compris la condition initiale) par simulation. Ainsi, pour 5000 simulations et comme on ne tient compte qu'une fois de la condition initiale, cela ferait pas moins de 20001 images à utiliser dans l'extraction des modes. Il est donc souvent impossible de procéder à partir de l'ensemble de la base de données, pour des raisons de lourdeur des calculs, tant du point de vue de la durée que de la mémoire né-

cessaire. Pour cela, on sélectionne des simulation de la base de façon aléatoire afin d'obtenir une base réduite. Dans les cas présentés à la suite, on sélectionne environ 300 simulations.

### 2.6.3 Qualité de la base de données et des modes extraits

On vient de voir qu'une part d'aléa a été introduite dans la sélection des simulations utilisées pour la POD. Il est donc légitime de se demander comment influe cette sélection sur la qualité des modes POD et donc sur la fiabilité du modèle réduit. De même, il est important de se donner des critères qui vont permettre de quantifier cette fiabilité. Pour s'éviter des calculs inutiles, il serait intéressant de pouvoir évaluer le processus le plus en amont possible, c'est à dire dès l'extraction aléatoire de la base réduite. Pour cela, nous avons regardé la distribution des masses tumorales finales dans la base de données réduite et généré des histogrammes sur plusieurs cas. La Figure 2.9 reprend

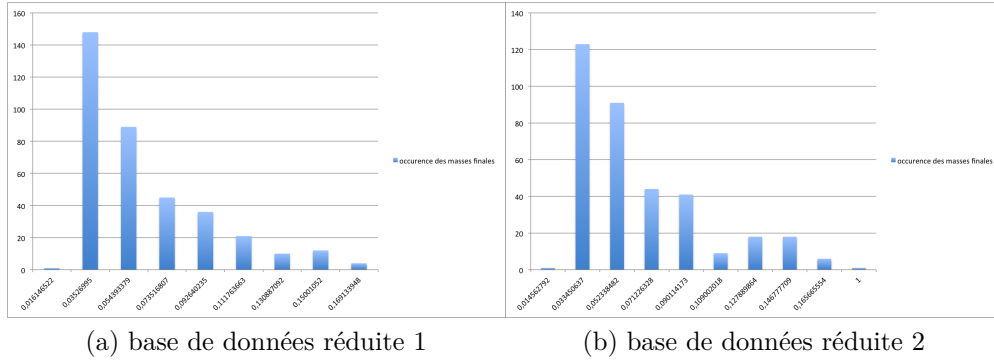


FIGURE 2.9 – Deux cas d'histogramme de répartition des masses tumorales finales dans les bases de données réduites

la répartition des masses tumorales pour deux bases réduites extraites de la même base de données. Dans chaque cas, les modes POD ont été extraits à partir des bases réduites et on a procédé à une test simple : on projette la donnée initiale sur les modes POD et on calcule l'erreur commise. Cela permet de voir si les modes reproduisent fidèlement cette donnée, et le cas échéant, de quantifier le nombre de modes minimum pour reproduire cette donnée avec une erreur ne dépassant pas un seuil fixé. L'histogramme 2.9a correspond à une base réduite qui permet d'obtenir une bonne base de modes POD, c'est à dire qui reproduit fidèlement la donnée initiale avec peu de modes. Au contraire, la seconde base réduite dont est extrait l'histogramme 2.9b, produit des modes

POD peu fiables puisqu'en projetant la condition initiale sur l'ensemble des modes ainsi extraits, on obtient une erreur relative L2 de plus de 7% (contre une erreur relative L2 inférieure à  $10^{-4}$  dans le premier cas). En regardant ces histogrammes on voit bien que la répartition est plus régulière dans le premier cas que dans le second.

En reproduisant l'expérience, on obtient des résultats similaires avec d'autres bases réduites ce qui nous permet de dégager une tendance générale et d'écarter ainsi *a priori* certaines bases réduites qui restent cependant assez rares. Il ne s'agit là que d'un test empirique et aucune analyse théorique n'a été faite concernant la relation entre la distribution des simulations de la base réduite et la qualité de la base POD extraite. De même, ce test n'est effectué qu'à partir de la répartition des masses tumorales à un instant donné et ne tient compte ni de la dynamique de croissance ni des géométries des simulations de la base de données. Cependant on utilise ce critère afin d'écarter des bases réduites potentiellement mauvaises.

Si la base réduite passe ce premier test, on extrait les modes puis on projette comme explicité auparavant la condition initiale sur l'ensemble des modes en relevant l'erreur relative commise en fonction du nombre de modes utilisés pour la projection. La Figure 2.10 correspond au tracé de ces données dans le cas de la base réduite dont est issu l'histogramme 2.9a. On voit bien ici qu'on arrive à reproduire fidèlement la condition initiale avec relativement peu de modes. Dans le cas présent (voir Figure 2.10), avec seulement 41 modes, on reproduit la donnée initiale avec moins de 0,5% d'erreur relative L2. Ce tracé peut donc nous permettre de choisir le nombre de modes qui sera utilisé par la suite dans le modèle réduit (ici 40).

On utilise donc la projection de la donnée initiale comme critère pour établir la qualité d'une base et des modes POD extraits. Cela se justifie de plusieurs façons. Tout d'abord, parmi tous les tests de projection faits, les erreurs relatives L2 les plus grandes sont rencontrées avec la donnée initiale. Cela provient probablement du fait que la donnée initiale est la plus petite tumeur de la base de données puisqu'on considère que les tumeurs grossissent. C'est donc un cas extrême, pour cela plus dur à représenter par la base POD. De plus, c'est une démarche simple puisque la condition initiale est directement accessible dans tous les cas de figure. Enfin, cela semble pertinent de se baser sur la projection de la condition initiale car c'est de cette manière que sont initialisés les coefficients  $a_i^T$  du modèle réduit (cf paragraphe 2.5.2), et on peut raisonnablement penser que si on part d'une erreur conséquente lors de l'initialisation, on n'obtiendra aucun résultat fiable par la suite.

L'ensemble de cette démarche d'évaluation de la base se fait sur la variable T concernant la densité de cellules tumorales. On utilise aussi la POD sur la



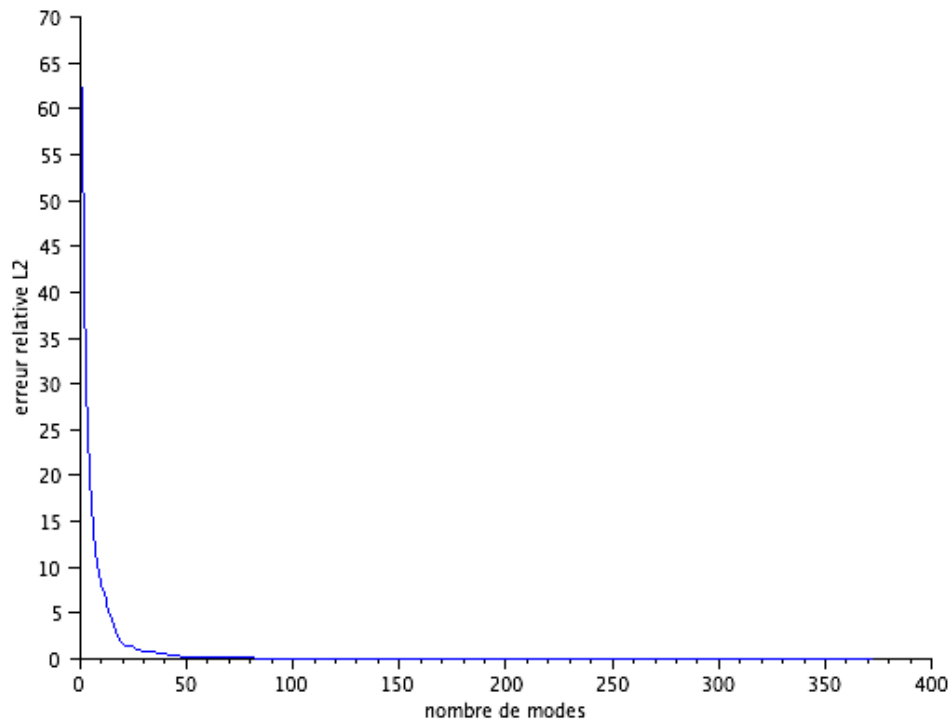


FIGURE 2.10 – Erreur relative L2 de la projection de la donnée initiale sur la base POD en fonction du nombre de modes.

pression, mais on a moins de contrôle sur cette variable d'autant que dans la plupart des cas, elle est bien plus régulière que  $T$ . On extrait donc peu de modes POD sur la pression (tout au plus une trentaine). **Notons que l'utilisation du modèle réduit au lieu du modèle EDP direct peut permettre (selon le nombre de modes utilisés) de diviser par 5 les temps de calcul sur une simulation.**

#### 2.6.4 Résolution du problème inverse

Une fois qu'on a une base de modes POD fiable pour simuler le modèle réduit, on peut s'en servir pour trouver les paramètres les mieux adaptés pour décrire la tumeur que l'on étudie. Dans un premier temps, on calibre le modèle à partir des données dont on dispose à cette fin. On minimise donc la fonction objectif donnée par l'expression (2.15) en utilisant une méthode de sensibilité (algorithme de type gradient). L'algorithme doit être initialisé avec un jeu de paramètres non loin d'une solution du problème d'optimisation. Pour cela, on utilise une méthode de type Monte Carlo : on tire au hasard un certain nombre de jeux de paramètres dans l'espace utilisé pour générer la base de données. Pour chacun de ces jeux, on évalue la valeur de la fonction objectif, et on se sert du jeu correspondant à la plus petite valeur de  $f_{obj}$  comme initialisation. Le modèle réduit est *a priori* valable sur l'intervalle de temps utilisé pour générer la base de données. Si l'on veut prédire l'évolution de la tumeur et donc simuler à des temps supérieurs, le plus prudent est donc d'injecter les paramètres trouvés par le problème inverse dans le système direct. Cela dit, on note que le passage du système direct au système réduit (et inversement) induit un biais. En effet, à jeux de paramètres identiques, les simulations issues du système direct et du système réduit peuvent présenter de légères différences (voir Figure 2.11). Cela provient de l'approximation intrinsèque à la méthode de POD d'autant qu'on utilise relativement peu de modes afin de réduire la complexité des calculs. On peut dans de nombreux cas estimer que l'erreur commise ici est acceptable étant données les nombreuses approximations faites tout au long du processus (segmentation, définition de l'image, zoom...). Cependant, on peut affiner ce premier résultat en utilisant un algorithme de type gradient mais cette fois sur le modèle direct ce qui implique néanmoins des coûts supérieurs en terme de calcul.

#### 2.6.5 Résultats de la méthode sur un premier cas complet

Nous allons ici présenter les résultats de la méthode sur un cas clinique d'une métastase à croissance rapide. Il s'agit du cas présenté en introduction

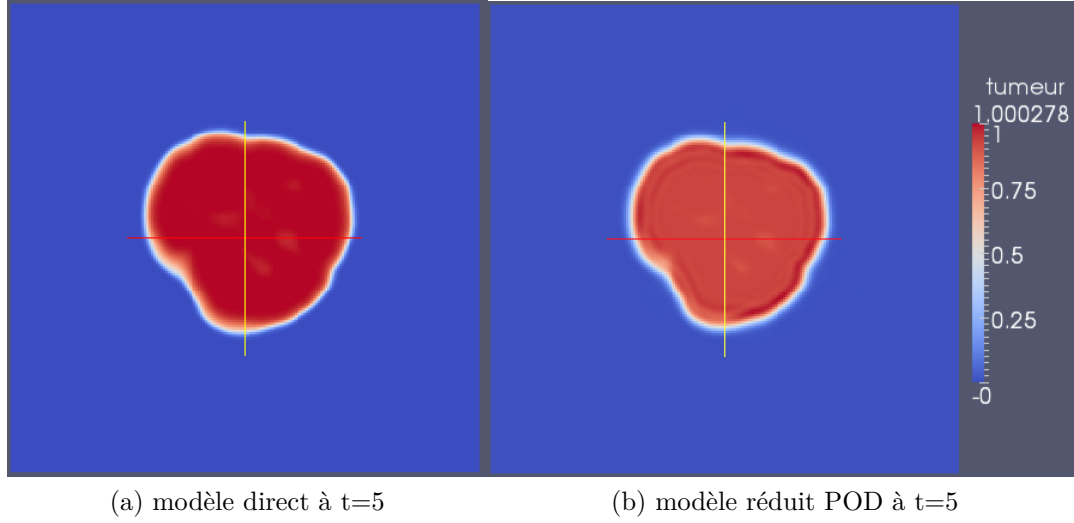


FIGURE 2.11 – Comparaison du modèle réduit POD et du modèle direct.

(1.3). Ce cas est particulièrement intéressant puisqu'on a trois images durant la phase de croissance de la tumeur, puis deux images de suivi lors de la chimiothérapie du patient. On peut donc calibrer le modèle tout d'abord sur la croissance tumorale, puis ajouter et calibrer le terme de chimiothérapie. Pour chacune des phases, nous disposons ainsi de deux données pour calibrer et une donnée pour tester le caractère prédictif de la méthode. La stratégie ici a été de paramétrer le modèle sur la phase de croissance. Dans un second temps, pour la phase de décroissance, on a repris les paramètres ainsi déterminés en ajoutant seulement le paramètre  $\delta$ , correspondant à l'efficacité du traitement chimiothérapeutique.

### Phase de croissance :

Pour cette phase, nous disposons de trois images médicales de l'évolution de la tumeur. La première, à l'instant  $t = 0$  a servi comme donnée initiale du problème. À partir de cette première image de la tumeur, une base de données de plus de 5000 simulations a été générée. Cette base de données couvre un espace de paramètres défini empiriquement. Nous avons, par exemple, décidé de fixer la valeur du paramètre  $\gamma_1$  à 0, pour supprimer un degré de liberté et simplifier la résolution du problème inverse. Les paramètres variables de la base sont donc :  $\alpha$ ,  $\beta$ ,  $\eta$ ,  $\gamma_0$ ,  $\lambda$  et  $M_{th}$  et l'ensemble des paramètres  $\Pi$  est inclus dans  $(\mathbb{R}^+)^6$  (on se limite en effet à des valeurs positives).

La seconde donnée ( $t = 107$  jours) a, quant à elle, servi pour paramétrer le

modèle. On a adimensionné l'échelle en temps en positionnant arbitrairement cette deuxième image de la tumeur à  $t = 4$ . La base de données a donc été générée pour  $t \in [0; 4]$ . Une fois les modes extraits de la base, on peut lancer la résolution du problème inverse. On récupère un jeu de paramètres (voir Table 2.3) que l'on ré-injecte dans le système direct. On peut alors prolonger la simulation jusqu'au temps  $t = 6,95$  (soit 186 jours après le premier examen) qui correspond à la troisième et dernière image de croissance tumorale (juste avant le lancement de la chimiothérapie). On vérifie alors la précision avec laquelle on aurait pu prévoir l'évolution de la tumeur avec notre modèle et disposant seulement des deux premières données.

$\alpha$	facteur de synthèse de l'agent pro-angiogénique	8,109 mm <sup>-2</sup>
$\beta$	facteur de croissance de la vascularisation	7,241
$\eta$	facteur de dégradation de la vascularisation	0,673
$\gamma_0$	taux de prolifération	1,116
$\gamma_1$	taux de mort par hypoxie	0
$\lambda$	taux de destruction de l'agent pro-angiogénique	0,865
$M_{th}$	seuil d'hypoxie	1,045

TABLE 2.3 – Jeu de paramètres obtenu par le problème inverse.

La Figure 2.12 reprend la comparaison du résultat du problème inverse avec les données du patient. On constate ici le biais entre le système réduit et le système direct à jeux de paramètres identiques. La courbe représentant la masse tumorale issue du système réduit, qui est utilisé pour la résolution du problème inverse, est légèrement en dessous de celle obtenue en simulant le système direct avec les mêmes paramètres. Ceci provient de l'approximation inhérente à la méthode de POD. De plus, le modèle inverse a été simulé avec assez peu de modes POD : on en a utilisés 39 soit à peine plus de 10% des modes extraits.

Malgré cette erreur, on obtient un comportement proche de celui du patient. En prédiction, on obtient un retard de moins de 9 jours sur la troisième donnée. Cela peut sembler important. Cependant, certains paramètres sur l'environnement de la tumeur n'ont pas été pris en compte ici, notamment le fait que celle-ci se développe contre une bronche ce qui joue nécessairement sur sa croissance. On peut aussi s'intéresser à l'erreur commise sur la forme de la tumeur. Pour cela on dispose d'indicateurs tels que le DICE ou la concordance

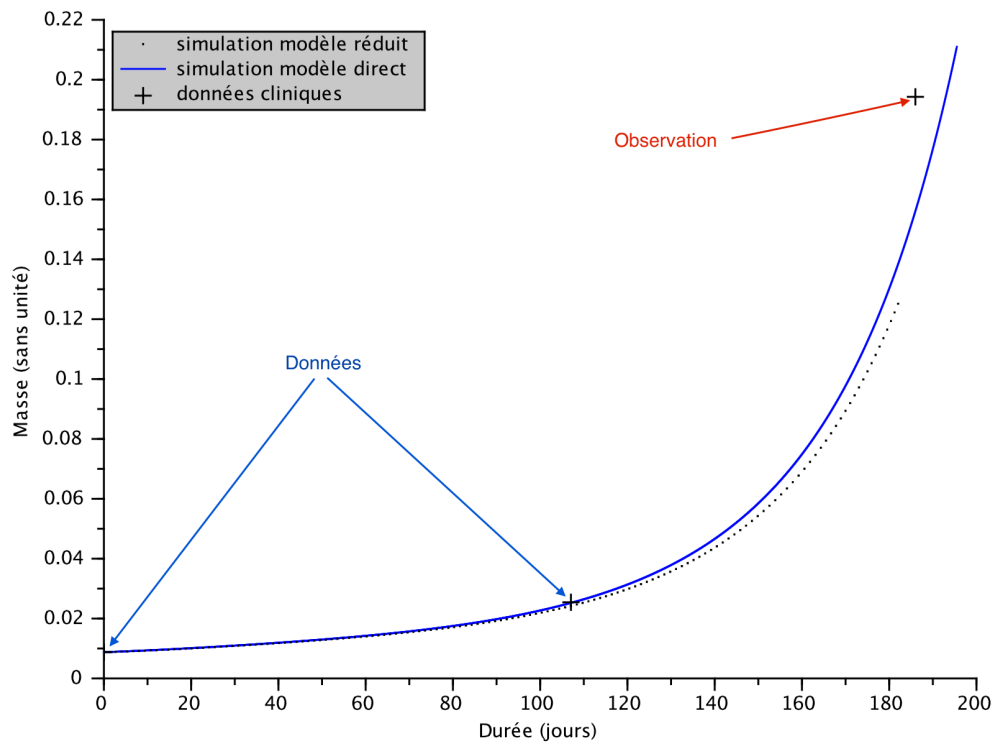


FIGURE 2.12 – Cas 1 : Comparaison de l'évolution de la masse tumorale entre le modèle réduit, le modèle direct et les données cliniques lors de la phase de croissance.

en volume donnés par les formules

$$DICE = 100 \times \left(1 - \frac{2 * |T_{modèle} \cap T_{donnée}|}{|T_{modèle}| + |T_{donnée}|}\right),$$

et

$$VC = 100 \times \left(1 - \frac{|T_{modèle} - T_{donnée}|}{|T_{donnée}|}\right).$$

Leurs valeurs sont données dans la Table 2.4 et attestent d'une reproduction spatiale fidèle de la tumeur. La Figure 2.13 confirme cela en comparant les images médicales de la tumeur et les résultats obtenus par la simulation. Pour finir, avec une erreur inférieure à 9 jours sur une durée totale de 186 jours, on reste sous la marge d'erreur d'environ 10% commise lors de la segmentation de la tumeur.

Date	2008/09/22	2008/12/10
DICE	8,79%	12,51%
Concordance en volume	82,54%	77,76%

TABLE 2.4 – Indicateurs spatiaux pour la phase de croissance du premier cas clinique : DICE et concordance en volume.

### Phase de décroissance : effets de la chimiothérapie

Étant donnée la croissance très rapide de cette métastase, les médecins ont décidé de traiter ce patient par chimiothérapie. Le protocole de traitement commence peu de temps après l'examen du 10/12/2008. Il s'étend sur 6 mois et consiste en des séances de chimiothérapie régulières, la date de la dernière séance étant le 29/06/2009. Pour simplifier la modélisation, on lisse les effets du traitement sur toute sa durée considérant donc que le facteur cytotoxique  $\delta$  est enclenché continuellement entre le 10/12/2008 et le 29/06/2009.

On part donc maintenant de la troisième donnée (à  $t = 186$  jours) correspondant au début du traitement comme donnée initiale et on reproduit la méthode précédente. On calibre ainsi le paramètre restant  $\delta$  grâce à la quatrième image qui correspond à la première image de suivi de la chimiothérapie (au 21/03/2009 soit 101 jours après le début de la chimiothérapie). Pour contrôler, la qualité de la calibration, on prolonge la simulation. Une première vérification se fait à l'instant du scanner de contrôle du 27/05/2009 (soit 166 jours après le début du traitement). Ensuite, on coupe le paramètre de chimiothérapie le 29/06/2009 (on met  $\delta$  à zéro) et on observe une rechute du patient

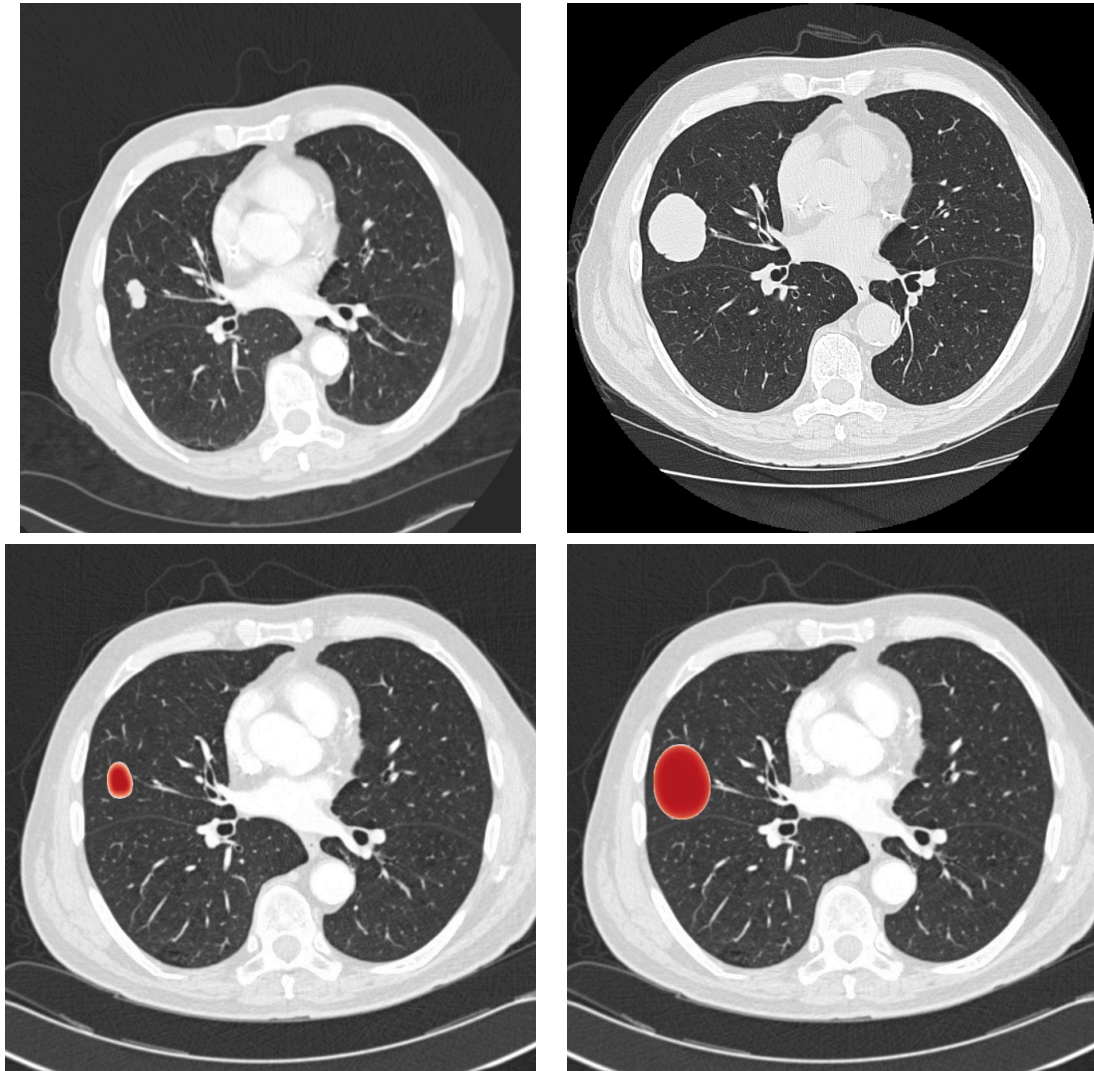


FIGURE 2.13 – Comparaison entre la prédiction et les observables : la ligne du haut représente les images médicales et celle du bas les simulations. La colonne de gauche correspond à l'examen du 22/09/2008 (calibration) et celle de droite à l'examen du 10/12/2008 (prédiction).

puisque la tumeur reprend sa croissance. On dispose justement d'un examen de contrôle post-chimiothérapie le 27/07/2009, c'est à dire un mois après la fin du traitement, que l'on va pouvoir comparer à la prédiction du modèle. Le Figure 2.14 donne l'évolution de la masse dans le temps pour le modèle ainsi calibré et pour le patient.

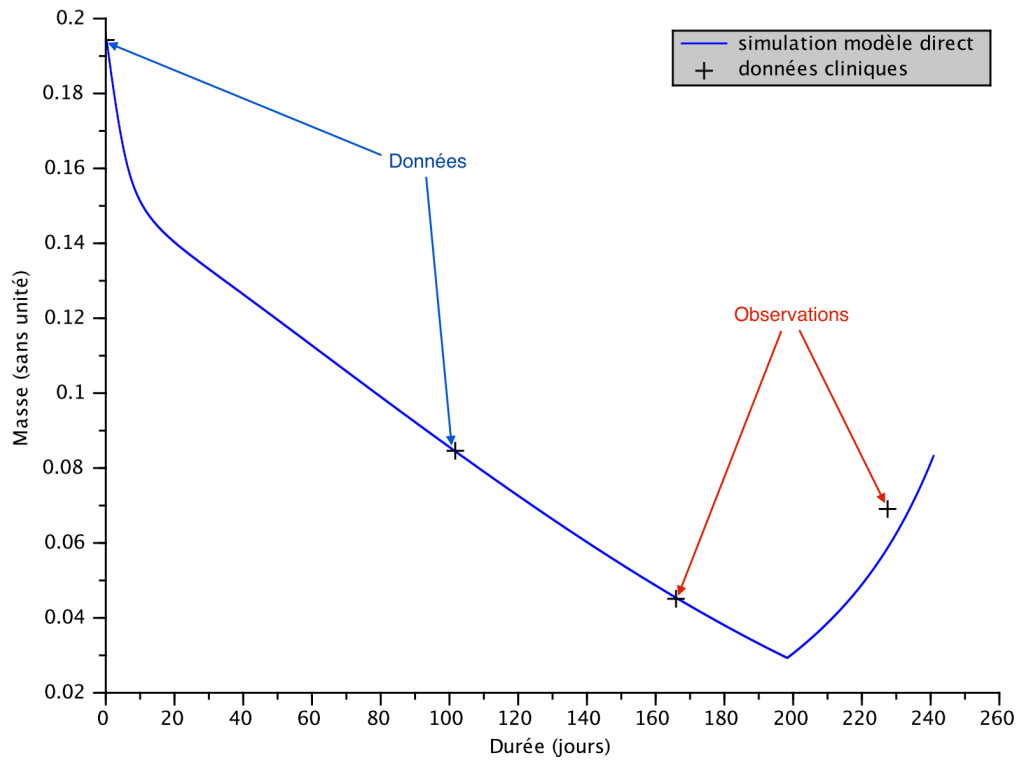


FIGURE 2.14 – Cas 1 : Comparaison de l'évolution de la masse tumorale entre le modèle direct calibré et les données cliniques lors de la phase de décroissance due à la chimiothérapie puis de la rechute.

On voit là encore que la prédiction sur le dernier point (à  $t = 166$  jours) est très bonne. Les indicateurs spatiaux pour la phase de traitement sont donnés Table 2.5 et les comparaisons avec les images médicales Figure 2.15. La rechute est quantitativement bien décrite par le modèle ainsi calibré puisqu'on commet là encore une erreur de moins de 10 jours. Pour ce dernier point cependant, la comparaison des formes est moins pertinente puisque la tumeur s'est déplacée dans le poumon. En fait, la rechute provient de la croissance d'une partie excentrée de la tumeur initiale ce qui explique ce déplacement. De plus, notre modèle ne tient pas compte de l'hétérogénéité ou de la distribution spatiale de



la chimiothérapie donc on ne peut pas espérer obtenir une prédiction en forme fiable.

Date	2009/03/21	2009/05/27
DICE	7,63%	12,36%
Concordance en volume	85,07%	77,22%

TABLE 2.5 – Indicateurs spatiaux pour la phase de décroissance du premier cas clinique : DICE et concordance en volume.

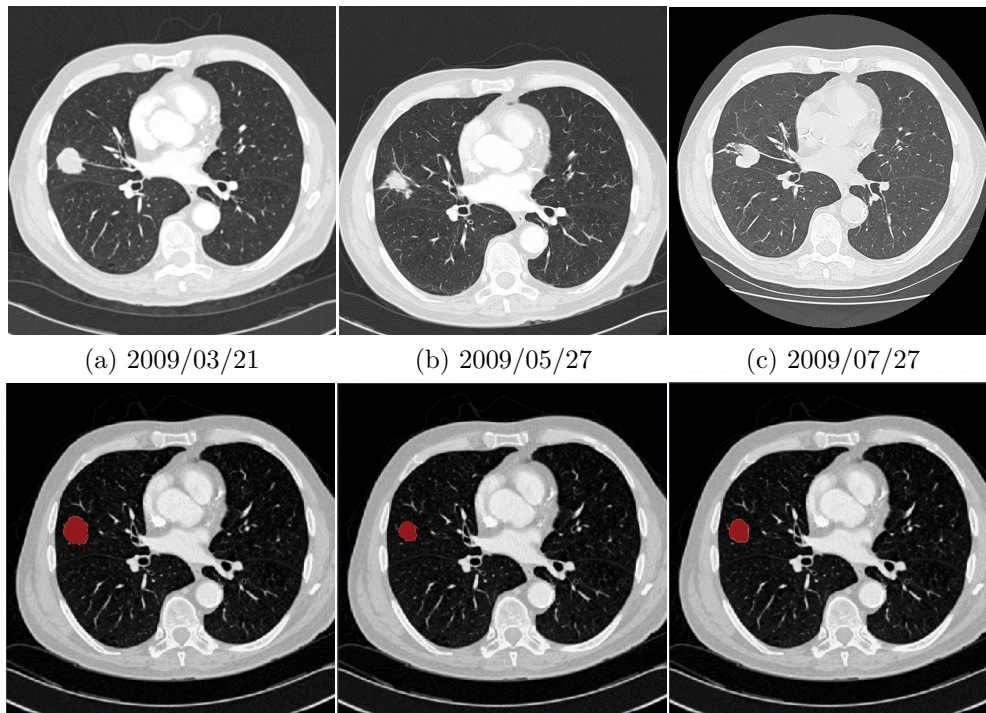


FIGURE 2.15 – Comparaison entre la prédiction et les observables pour la phase de décroissance due au traitement : la ligne du haut représente les images médicales et celle du bas les simulations. De gauche à droite : le premier CT-scan de contrôle du 21/03/2009 (calibration), le second CT-scan de contrôle à la fin du traitement au 27/05/2009 (prédiction) et l'examen fait un mois et demi après la fin de traitement montrant la rechute au 27/07/2009 (prédiction).

### 2.6.6 Étude d'un second cas clinique : aspect spatial

Nous avons vu précédemment que l'on pouvait obtenir de bonnes prédictions sur l'évolution de la masse tumorale grâce à notre méthode. Cependant, nous utilisons un modèle spatial et il serait bien dommage de limiter nos résultats à la masse tumorale. Dans ce second exemple, nous reprenons la même méthodologie mais nous allons nous intéresser à la géométrie de la tumeur. Il s'agit d'un cas de métastase pulmonaire provenant cette fois d'une tumeur de la thyroïde. Nous ne disposons pour ce cas que de données sans traitement et nous ne nous intéressons donc qu'à la phase de croissance. Nous utilisons là encore trois images de la tumeur issues de scanners à trois instants différents : la première comme condition initiale, la deuxième pour calibrer le modèle et la troisième pour vérifier la capacité prédictive du modèle calibré. Les résultats de simulation à l'instant de la calibration et en prédiction sont donnés Figures 2.16 et 2.17. Notons que la croissance tumorale est bien plus lente que pour le cas précédent.

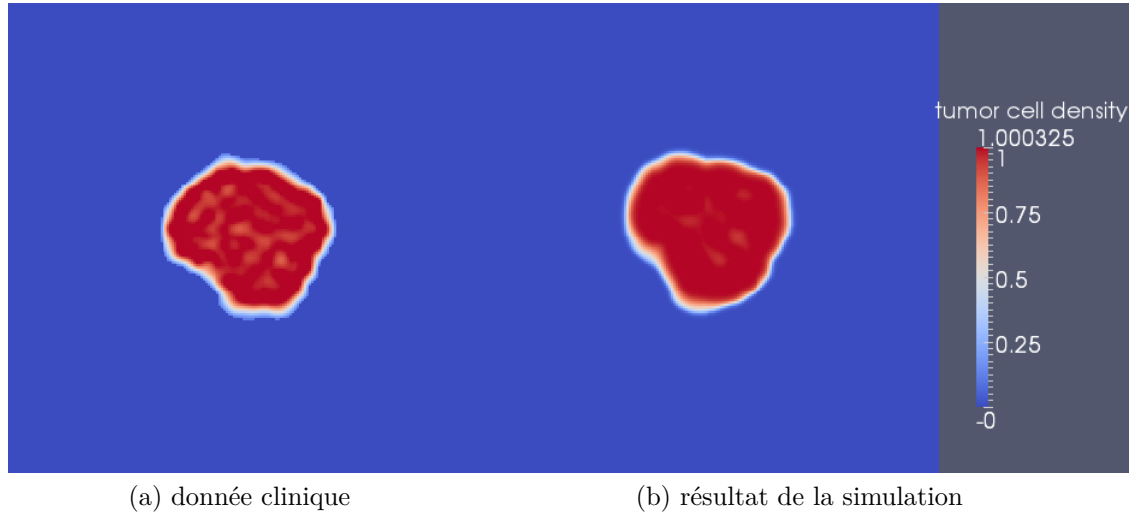


FIGURE 2.16 – Donnée clinique et résultat de la simulation à l'instant de la calibration  $t = 599$  jours.

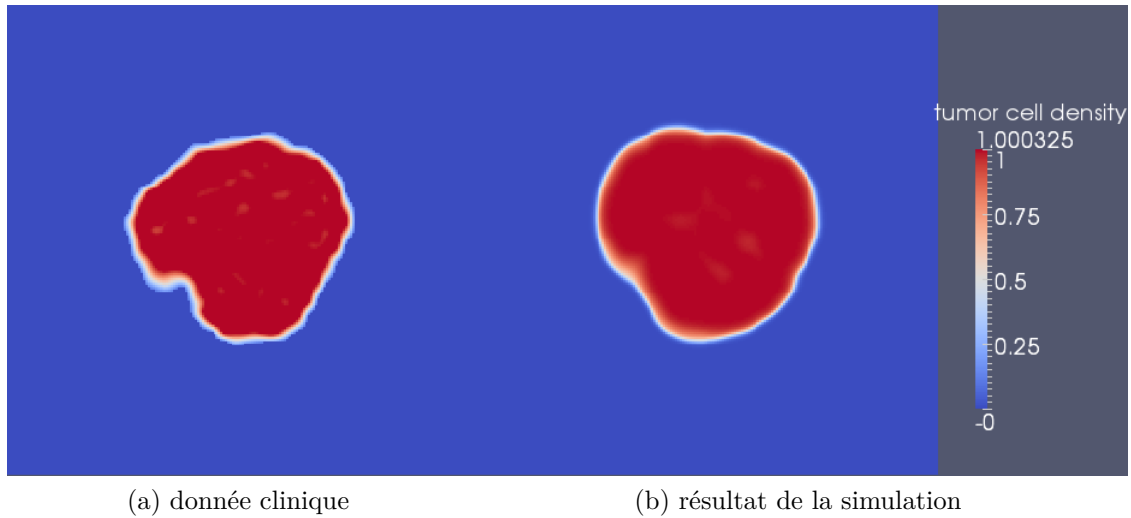


FIGURE 2.17 – Donnée clinique et résultat de la simulation en prédiction à l'instant  $t = 1251$  jours.

### 2.6.7 Étude d'un troisième cas clinique

La même méthodologie est utilisée sur un troisième cas clinique d'une tumeur là encore moins agressive que celle du premier cas. Il s'agit d'une métastase pulmonaire issue d'un cancer du colon. On dispose à nouveau de trois images sans traitement, la qualité en prédiction s'évaluant sur la troisième. La dynamique de croissance est présentée Figure 2.18.

### 2.6.8 Considérations sur la robustesse de la méthode POD

On a vu que, malgré un léger biais entre le système réduit et le système direct, la méthode de personnalisation basée sur la POD est assez robuste pour donner des résultats fiables et permet d'accélérer considérablement les calculs. Cela est vrai tant qu'on simule le système réduit dans les intervalles de paramètres et de temps correspondant à la base de données dont sont issus les modes POD. On peut alors se demander comment se comporte la méthode si l'on sort de ce cadre d'utilisation.

La POD s'appuie sur la synthèse, dans une base aux propriétés intéressantes, des différentes géométries que peut prendre la tumeur. Le cadre de validité du modèle réduit est donc davantage lié à des critères de forme qu'à des valeurs de paramètres ou à des temps de simulations particuliers. Ainsi, si un jeu de paramètres hors de la gamme utilisée pour le calcul des modes donne

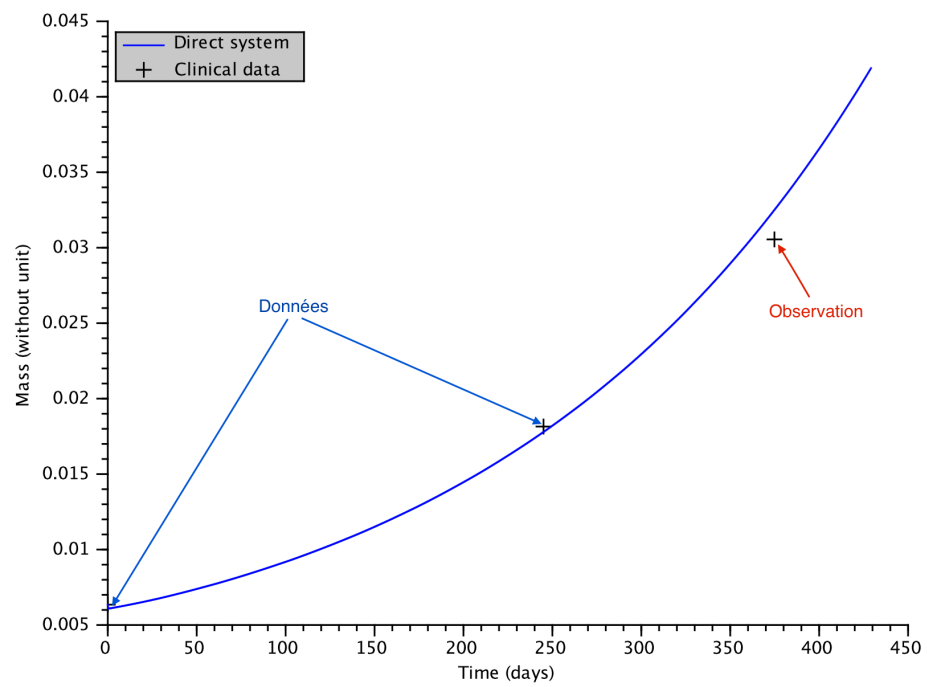


FIGURE 2.18 – Troisième cas test : évolution de la masse et comparaison entre les données médicales et la simulation.

des géométries (formes et tailles des tumeurs simulées) reproductibles à partir des modes, alors la simulation du modèle réduit avec ce jeu de paramètres est possible.

Cela dit, étant donné un jeu de paramètres, il est assez difficile d'avoir une idée *a priori* de la géométrie finale de la tumeur. Pour cela, malgré une certaine robustesse de la méthode basée sur la POD, on préfère se limiter en pratique à la simulation de paramètres et intervalles de temps appartenant à l'espace utilisé pour créer la base de snapshots.

De la même manière, il est possible de simuler la réponse aux traitements à partir du modèle réduit. Il suffit pour cela de générer une base de données de snapshots contenant des cas de tumeurs décroissantes.

## 2.7 Utilisation de données 3D

Jusqu'à présent, on s'est contenté d'utiliser des images de métastases bi-dimensionnelles, celles-ci étant obtenues via des coupes de données 3D. Si l'utilisation de ces données 2D est pratique dans un premier temps pour des raisons de temps de calcul et de simplicité du code, une fois la méthode éprouvée, il peut être intéressant d'utiliser les données 3D. En effet, cela permet de s'affranchir d'un problème de taille lié aux données 2D : le choix de la coupe. De plus on peut ainsi espérer obtenir une prévision de la masse réelle de la tumeur.

Cela engendre certaines modifications au niveau du code notamment en ce qui concerne le système direct. Cependant, la méthode basée sur la POD présente des avantages certains lors du passage au 3D. En effet, la seule différence réside dans le fait que les modes POD deviennent des fonctions spatiales à trois dimensions (au lieu de deux). Les surcoûts engendrés par le passage au 3D dans cette méthode sont essentiellement liés aux phases de « pré-traitement », c'est à dire la génération d'une base de données 3D à partir de la donnée initiale du patient, l'extraction des modes 3D à partir de cette base et le calcul des dérivées spatiales des modes. Ensuite, la résolution du système réduit est équivalente au cas 2D (seuls les produits scalaires doivent être re-codés en 3D).

Les paramètres ayant servi pour cette simulation sont présentés au tableau suivant :

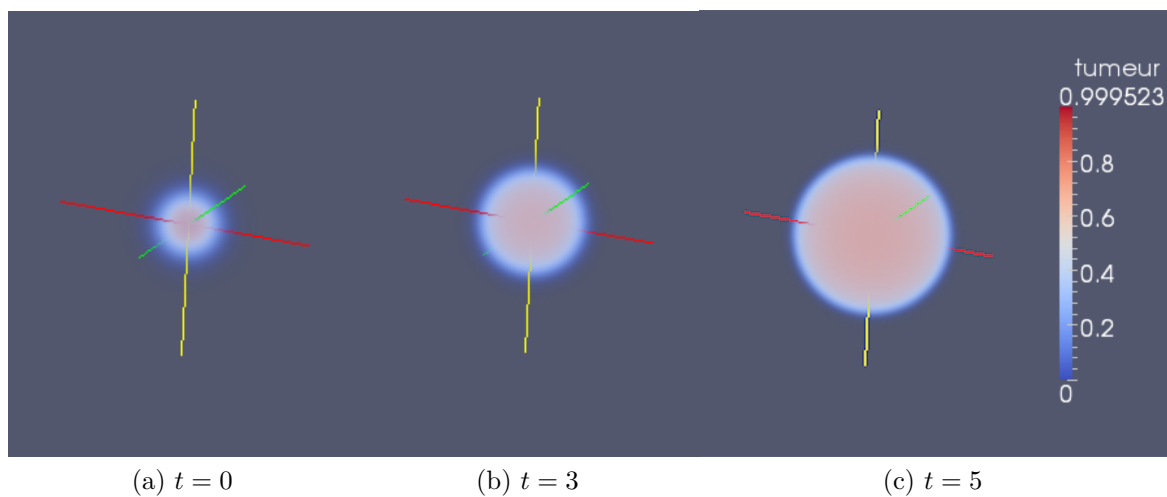


FIGURE 2.19 – Évolution de la densité de cellules tumorales : cas tri-dimensionnel d’une métastase sphérique, maillage  $100 \times 100 \times 100$ .

$\alpha$	$12 \text{ mm}^{-3}$
$\beta$	12
$\eta$	0,2
$\gamma_0$	0,6
$\gamma_1$	0,1
$\lambda$	0,3
Seuil d’hypoxie	0,7

TABLE 2.6 – Jeu de paramètres utilisé pour la simulation 3D.

## 2.8 Conclusion

Dans cette première partie de la thèse, nous avons étudié un premier modèle de croissance de métastases pulmonaires et de réponse à la chimiothérapie basé sur un système d'équations aux dérivées partielles ainsi qu'un algorithme permettant de le calibrer sur des données cliniques. Les premiers résultats de calibration et de prédiction de la croissance tumorale, voir de la réponse au traitement, sont satisfaisants. Cependant cette approche présente certains défauts. Le modèle comporte de nombreux paramètres ce qui complique l'étape de calibration surtout dans le cadre d'une utilisation d'un nombre restreint d'images médicales. De plus la mise en oeuvre de la méthode POD est assez lourde puisqu'elle nécessite, pour chaque métastase étudiée, de générer une base de données de snapshots. Il semble donc intéressant de tenter de simplifier le modèle sans pour autant nuire aux performances constatées et à la pertinence biologique de celui-ci.

## Chapitre 3

# Modèle simplifié de croissance pour les métastases pulmonaires en vue d'applications cliniques

### 3.1 Limites du modèle complet

Nous avons vu que le modèle présenté lors de la partie précédente donne des résultats assez satisfaisants. Cependant, sa complexité le rend difficile à calibrer. En effet, les données dont nous disposons en pratique ne sont pas suffisantes pour identifier correctement le modèle, et ce quelque soit l'algorithme de calibration. En plus de cela, l'espace des paramètres utilisé pour la calibration est déterminé de façon totalement empirique. On ne dispose d'aucun moyen analytique pour borner cet espace et, de ce fait, on n'est pas sûr que la meilleure solution se trouve bien là où l'on cherche. Pour finir, on s'aperçoit aussi assez vite que certaines dynamiques de croissance pourtant sensées être reproductibles ne sont jamais obtenues en pratique. Il s'agit notamment des croissances dont le taux décroît dans le temps et dont l'évolution donne une courbe de type sigmoïde présentant une saturation. Ces dynamiques sont pourtant assez courantes dans le cas de métastases pulmonaires. D'un point de vue spatial, le modèle ne permet pas non plus d'observer de façon nette la couronne de cellules proliférantes qui se forme dans certains cas de croissance. Pour toutes ces raisons, le modèle a été remanié et simplifié.



## 3.2 Simplifications du modèle de croissance tumorale

Dans un souci de simplification et d'une meilleure identifiabilité, le modèle a été remanié. La difficulté réside dans notre volonté de ne pas trop sacrifier la pertinence biologique du modèle. Le système d'équations simplifié est le suivant :

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = MT. \quad (3.1)$$

Le taux de prolifération des cellules cancéreuses dans l'équation (3.1) est maintenant directement représenté par  $M$ . La variable  $M$  modélise donc l'approvisionnement des tissus en nutriments et le taux de croissance de la tumeur.

$$\frac{\partial M}{\partial t} = -\eta(M - M_1)T. \quad (3.2)$$

L'équation (3.2) décrit l'évolution de  $M$  que l'on continuera de nommer abusivement vascularisation du tissu. Afin de contrôler le taux de croissance, on impose au champ  $M$  d'être borné, entre sa valeur initiale  $M_0$  et un paramètre  $M_1$ . Pour réduire le nombre de paramètres, la variable scalaire  $\xi$  a été supprimée. Le phénomène d'angiogénèse reste cependant implicitement compris dans (3.2). En effet la croissance de  $M$  est contrôlée par la densité de cellules cancéreuses  $T$ . De même que pour le modèle précédent, l'hypothèse de saturation du tissu  $T + S = 1$  (où  $S$  représente le tissu sain) nous donne une équation sur la divergence de la vitesse de déplacement du tissu  $\mathbf{v}$  (3.3).

$$\nabla \cdot \mathbf{v} = MT. \quad (3.3)$$

Pour fermer le système, la vitesse de déplacement du tissu  $\mathbf{v}$  reste modélisée par une loi de Darcy Eq.(3.4).

$$\mathbf{v} = -\nabla \pi. \quad (3.4)$$

Il est toujours possible d'ajouter les effets d'un traitement cytotoxique via l'ajout d'un terme de mort du type  $-\delta T$  dans le membre de droite de l'équation (3.1).

### 3.2.1 Premiers essais numériques

Afin de simuler ce nouveau système, nous avons naturellement utilisé une méthode très similaire à celle adoptée pour la simulation du premier modèle. L'équation d'advection (3.1) est notamment approchée par un schéma WENO

5. Pour la plupart des jeux de paramètres, cette méthode numérique fournit des résultats satisfaisants.

Lorsque  $M_1$  est proche de 0, on peut observer la formation d'une couche extérieure de cellules proliférantes alors que le coeur de la tumeur est quiescent. Cette couronne proliférante est alors la seule zone de la tumeur où se multiplient les cellules cancéreuses. Elle se comporte comme un front se propageant à la vitesse  $\mathbf{v}$ . Si l'épaisseur est faible par rapport aux dimensions de la tumeur (de l'ordre du voxel) et la vitesse suffisamment élevée, des hétérogénéités apparaissent lors des simulations (voir Figure 3.1). On voit là apparaître la limite

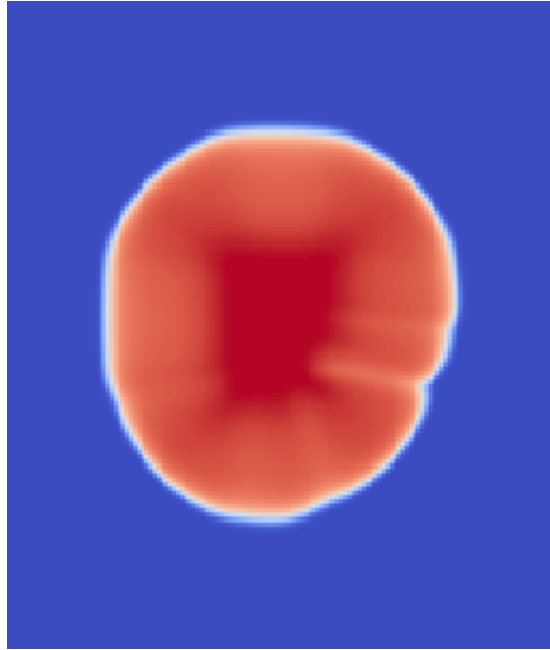


FIGURE 3.1 – Simulation du modèle simplifié à partir du schéma WENO 5, cas présentant des hétérogénéités dans la distribution spatiale des cellules cancéreuses. Paramètres :  $\eta = 3$ ,  $M_0 = 3$ ,  $M_1 = 0$  et  $t = 3$ .

de l'utilisation du schéma de transport WENO 5 qui ne parvient pas à capturer précisément le comportement de cette couronne de cellules proliférantes. D'autres schémas de transport ont été testés dont un schéma de Lax-Wendroff d'ordre 2 (voir en Annexe A) sans davantage de succès. Pour résoudre ce problème numérique, nous avons décidé de mettre au point une méthode de suivi d'interface nécessitant une nouvelle formulation du problème.

### 3.2.2 Transport de la tumeur par méthode de Front Tracking

#### Principe de la méthode

Afin d'éviter les hétérogénéités de la tumeur lors des simulations, nous avons décidé d'adopter une méthode de suivi d'interface (Front Tracking) pour évaluer l'évolution de la tumeur. Au lieu de transporter la densité de cellules tumorales, la tumeur est matérialisée par son interface avec le tissu sain (que l'on appellera aussi frontière) notée  $\Phi$ . Dans notre cas, ce choix est influencé par plusieurs critères. Tout d'abord, les métastases qui se développent dans le poumon sont des tumeurs solides, globalement hétérogènes et non infiltrantes comme on le voit sur l'extrait de scan présenté en Figure 3.2.

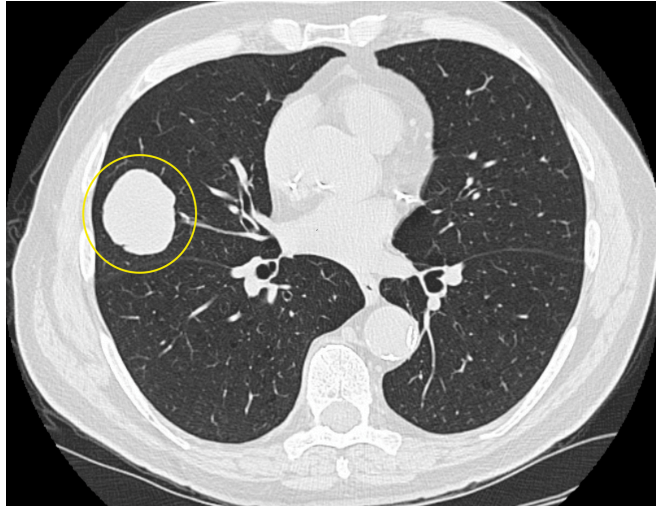


FIGURE 3.2 – Extrait de CT-scan montrant une métastase pulmonaire issue d'une tumeur primitive de la vessie.

La densité de cellules tumorales vaut donc 1 dans la tumeur et 0 à l'extérieur et il semble légitime de considérer une frontière tumeur/tissu sain. De plus, lorsqu'on segmente la tumeur sur la coupe de l'examen depuis le logiciel OsiriX, on exporte un masque polygone sous la forme d'une liste ordonnée de coordonnées de points ou marqueurs d'interface (voir Figure 3.3). Ces coordonnées sont données en millimètres dans le repère propre à la machine ayant servi à faire l'examen. On a donc immédiatement une discrétisation de la frontière  $\Phi$  et on gagne une étape de traitement des données.

L'évolution de la tumeur peut donc se résumer au déplacement de son interface avec le tissu sain. L'équation de transport sur la densité de cellules

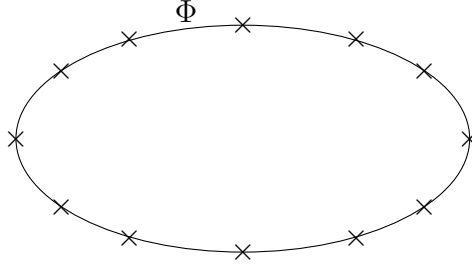


FIGURE 3.3 – Représentation de l'interface  $\Phi$  et des points de discrétisation notés  $\times$ .

tumorales avec termes sources est ainsi remplacée par une équation de transport sans second membre sur la frontière  $\Phi$  (Eq.(3.5)).

$$\frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \nabla \Phi = 0. \quad (3.5)$$

Lorsqu'on connaît la position de la frontière, on obtient la densité de cellules tumorales de la façon suivante :

$$T = \begin{cases} 1 & \text{à l'intérieur de } \Phi, \\ 0 & \text{à l'extérieur de } \Phi. \end{cases}$$

En plus de l'équation sur la frontière, nous devons résoudre le système d'équations permettant d'obtenir la vitesse. Celle-ci est finalement calculée comme auparavant. Les effets de prolifération et d'un éventuel traitement sont contenus dans l'équation sur la divergence de la vitesse.

$$\frac{\partial M}{\partial t} = -\eta(M - M_1)T,$$

$$\nabla \cdot \mathbf{v} = MT,$$

$$\mathbf{v} = -\nabla \pi.$$

### Reconstruction de la tumeur à partir de l'interface

Connaissant la position des points de discrétisation de l'interface, il est important de savoir reconstruire le champ  $T$ . En effet, la densité de cellules tumorales doit être connue afin de déterminer l'évolution de la vascularisation et de la vitesse de transport. Les marqueurs de l'interface n'ont aucune raison de correspondre aux noeuds de la grille cartésienne utilisée. Pour simplifier, on considère qu'un pixel donné est à l'intérieur de la tumeur si son centre

y est. On passe donc en revue chaque pixel  $(i; j)$ . Si le centre de celui-ci est dans la tumeur  $T_{i,j} = 1$ , sinon  $T_{i,j} = 0$ . Reste à trouver un moyen de tester l'appartenance d'un point à la tumeur, ce qui revient à tester si un point se trouve à l'intérieur d'un polygone, ce dernier n'étant pas forcément convexe mais simplement connexe.

Pour cela on calcule le nombre de tours ou winding number (voir [57]) du polygone par rapport au point courant, c'est à dire le nombre de fois que l'on tourne autour du point considéré en parcourant le polygone. Notons  $P = \{p_1, \dots, p_{np}\}$  le polygone et  $c_{i,j}$  le point courant qui est, en pratique, le centre de la maille  $(i, j)$ . Les coordonnées du point  $p_i$  sont données par  $(x_i, y_i)$  et celles de  $c_{i,j}$  par  $(x_{c_{i,j}}, y_{c_{i,j}}) = ((i + \frac{1}{2})dx, (j + \frac{1}{2})dy)$ ,  $dx$  et  $dy$  étant les pas d'espace de la grille. Le nombre de tours de  $\phi$  par rapport à  $c_{i,j}$  est donné par :

$$wn_{i,j} = \frac{1}{2\pi} \sum_{k=1}^{np} \angle(p_k c_{i,j} p_{k+1}),$$

où l'on ferme le polygone en imposant  $p_{np+1} = p_1$ . Les angles  $\angle(p_k c_{i,j} p_{k+1})$  sont les angles orientés entre les vecteurs  $\overrightarrow{c_{i,j} p_k}$  et  $\overrightarrow{c_{i,j} p_{k+1}}$ , de valeurs positives s'ils sont orientés dans le sens trigonométrique, négatives sinon (voir Figure 3.4).

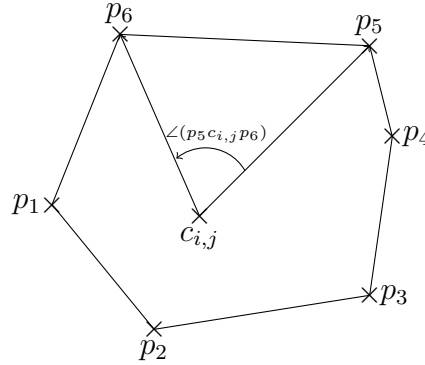


FIGURE 3.4 – Calcul du winding number, notations.

Il reste donc à calculer les angles orientés. Si on considère les deux vecteurs  $\mathbf{U} = (U_0; U_1)$  et  $\mathbf{V} = (V_0; V_1)$ , l'angle orienté de  $\mathbf{U}$  vers  $\mathbf{V}$ , que l'on note  $\theta$ , est donné par l'expression :

$$\theta = \text{atan2}(U_0 V_1 - U_1 V_0, U_0 V_0 + U_1 V_1),$$

où la fonction  $x \rightarrow \text{atan2}(x)$  est une variation de la fonction arctangente à deux arguments permettant d'obtenir des angles orientés. Ensuite et dans le

cas de polygones simples, c'est à dire dont les arrêtes ne se croisent pas, qui nous concerne, si le nombre de tours vaut 1 ou -1, le point  $c_{i,j}$  est à l'intérieur du polygone, s'il vaut 0, le point est à l'extérieur.

Dans la pratique, on n'a pas besoin de tester toutes les mailles de la grille à chaque pas de temps. En effet, en l'absence de traitement, le modèle ne permet pas la décroissance de la tumeur, les points précédemment dans le polygone (pour lesquels  $T_{i,j}$  vaut déjà 1), n'en sortiront pas et il est inutile de les tester. De plus, les mailles extérieures « trop loin » de l'interface resteront, à coup sûr, hors du polygone. On peut donc restreindre l'ensemble de points à tester au plus petit rectangle contenant le polygone  $P$  (voir Figure 3.5).

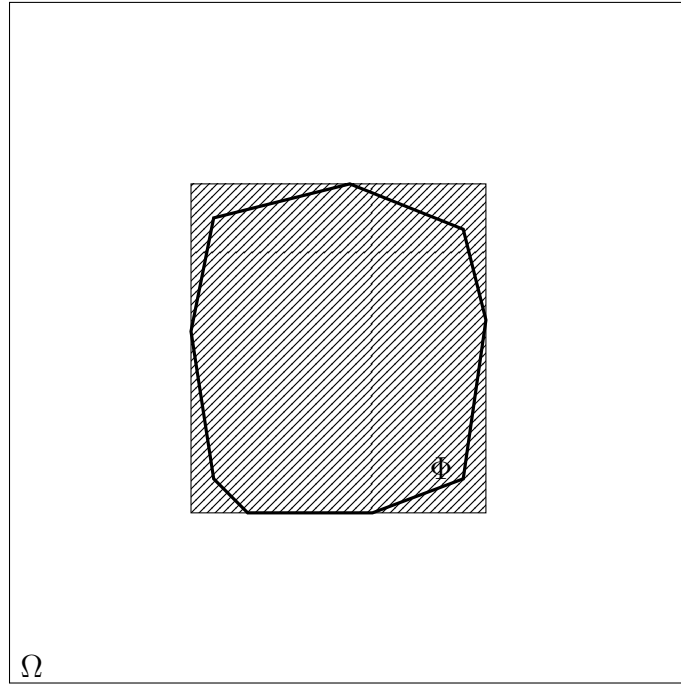


FIGURE 3.5 – Représentation de la frontière  $\Phi$  dans le domaine de calcul  $\Omega$ . La zone du domaine à tester pour la reconstruction du champ  $T$  correspond au rectangle hachuré.

Une amélioration de cette méthode, dite des courbes de Jordan, consiste à tracer une demi-droite à partir du point courant (ici  $c_{i,j}$ ) dans une direction et un sens donnés (par exemple vers la droite) et à compter le nombre de fois que l'on coupe le polygone. Si ce nombre est pair, alors le point testé est à l'extérieur du polygone, sinon, il est à l'intérieur (voir Figure 3.6). Cette variante équivalente à l'utilisation du winding number est plus efficace et est utilisée en

pratique.

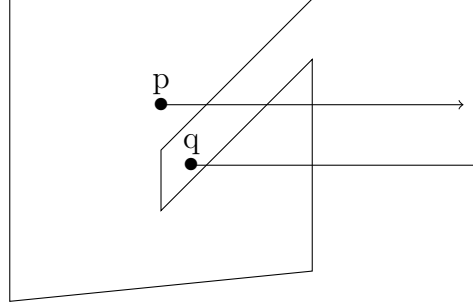


FIGURE 3.6 – Méthode des courbes de Jordan : la demi-droite d'origine p coupe 3 fois le polygone (nombre impair), p est donc à l'intérieur du polygone ; la demi-droite d'origine q coupe 2 fois le polygone, q est donc à l'extérieur du polygone.

### Implémentation numérique de la méthode

D'un point de vue numérique, le schéma utilisé pour le transport de l'interface est assez simple. À partir de l'interface  $\Phi$  on détermine la densité de cellules tumorales  $T$  en utilisant la méthode décrite au paragraphe précédent. On peut alors calculer la vascularisation dans le tissu  $M$  ainsi que la vitesse de transport  $\mathbf{v}$ . Pour cela, on procède de même qu'auparavant. Connaissant le champ de vitesse, on peut mettre à jour la position des points  $p_k$  de discrétisation de celle-ci :

$$p_i^{n+1} = p_i^n + dt \times \mathbf{v}_i^{n+1}.$$

Dans l'expression ci-dessus,  $p_i^n$  et  $p_i^{n+1}$  sont les coordonnées du  $i^{\text{ème}}$  point de discrétisation de l'interface respectivement aux instants  $t_n$  et  $t_{n+1}$  espacés d'un pas de temps :  $t_{n+1} = t_n + dt$ . La vitesse  $\mathbf{v}_i^{n+1}$  de déplacement du point  $p_i$  est calculée par interpolation bilinéaire. Désormais, les vitesses, comme l'ensemble des variables du système, sont calculées aux centres des mailles. On repère les quatre points du maillage les plus proches du point  $p_i^n$  et on calcule ses coordonnées dans le repère défini par ces points (voir Figure 3.7). En reprenant les notations de la Figure 3.7, on fait l'hypothèse que, dans le rectangle défini par  $x_{i,j}, x_{i+1,j}, x_{i+1,j+1}, x_{i,j+1}$ , la vitesse  $\mathbf{v}$  peut être approchée par :

$$\mathbf{v}(u_1, u_2) = a + bu_1 + cu_2 + du_1u_2.$$

Il reste donc à déterminer les coefficients  $a$ ,  $b$ ,  $c$  et  $d$ , et pour cela, on utilise les valeurs connues de la vitesses aux centres des mailles :  $\mathbf{v}_{i,j}$ ,  $\mathbf{v}_{i+1,j}$ ,  $\mathbf{v}_{i+1,j+1}$  et  $\mathbf{v}_{i,j+1}$ . On obtient ainsi :

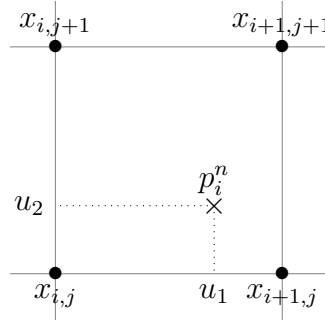


FIGURE 3.7 – Interpolation bilinéaire des vitesses. Les noeuds du maillage  $x$  correspondent avec le centre des pixels.

$$\begin{aligned}
 a &= \mathbf{v}_{i,j}, \\
 b &= \mathbf{v}_{i+1,j} - \mathbf{v}_{i,j}, \\
 c &= \mathbf{v}_{i,j+1} - \mathbf{v}_{i,j}, \\
 d &= \mathbf{v}_{i+1,j+1} + \mathbf{v}_{i,j} - \mathbf{v}_{i+1,j} - \mathbf{v}_{i,j+1}.
 \end{aligned}$$

Dans le cas particulier où l'on se trouve sur le bord du domaine, on utilise la valeur de la vitesse au centre de maille le plus proche. Ce cas est cependant très rare puisque la taille du domaine de calcul est déterminée de manière à ce que la tumeur n'atteigne pas les bords. Nous pouvons ainsi calculer une valeur approchée de la vitesse à la position exacte de chacun des points de discrétisation de l'interface.

L'un des risques inhérents à ce genre d'approche est de voir apparaître des points anguleux prononcés si le nombre de points de discrétisation de l'interface devient trop petit par rapport à la taille de la tumeur. Afin d'éviter cela, on teste à chaque itération du solveur la distance séparant deux points consécutifs. Si celle-ci dépasse un certain seuil, fixé en fonction des paramètres de la géométrie du problème, un point est ajouté entre les deux points préexistants. C'est un moyen simple permettant de conserver une densité de points suffisante.

### Résultats numériques

En jouant avec les paramètres, on peut reproduire plusieurs types de dynamiques de croissance classiques des métastases du poumon. Notamment, il a été fréquemment observé que les cellules les plus actives se concentrent en périphérie de la tumeur. Cela est dû au manque de nutriments au coeur de



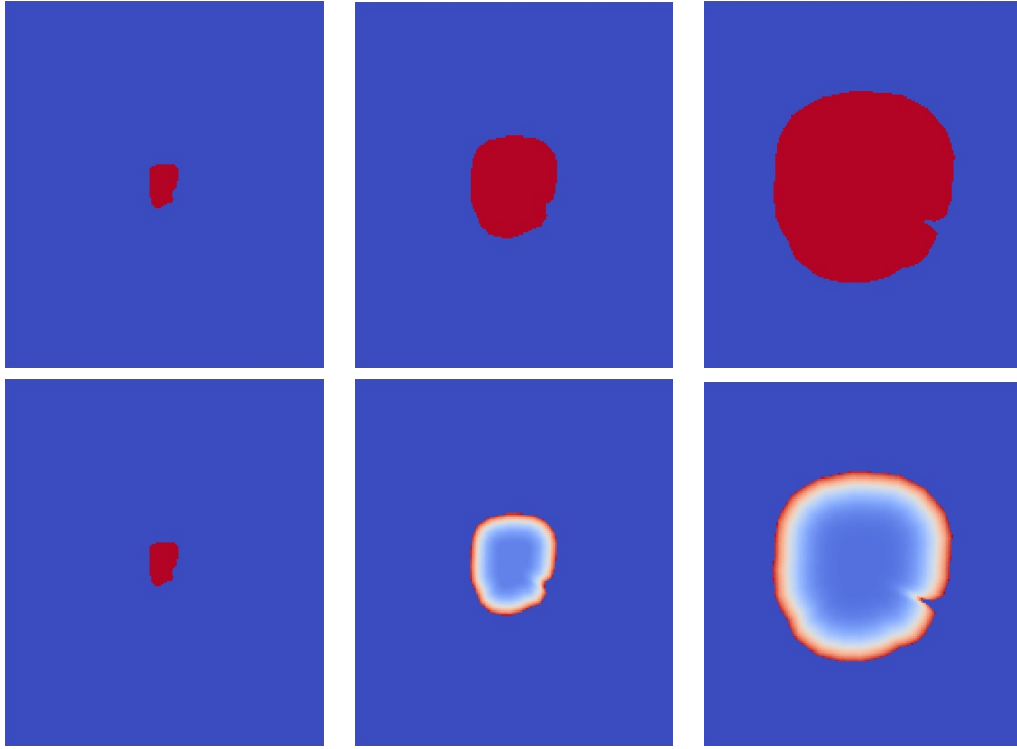


FIGURE 3.8 – Exemple de croissance en couronne, en haut la densité de cellules tumorales  $T$ , en bas les zones proliférantes (produit  $T \times M$ ) ; de gauche à droite :  $t = 0$ ,  $t = 1,5$  et  $t = 3$ . Paramètres :  $M_0 = 2,5$ ,  $M_1 = 0,2$ ,  $\eta = 2$ .

la lésion qui se retrouve alors hypoxique, voire nécrosée (mais on ne tient pas compte de la nécrose dans le modèle). La répartition des cellules proliférantes est donnée par le produit  $M \times T$ . On retrouve aussi plus classiquement les dynamiques que l'on qualifiera d'« exponentielles », où l'ensemble de la tumeur prolifère.

**Couronne proliférante** On voit bien sur la Figure 3.8 que le terme de prolifération  $T \times M$  présente une couronne extérieure, alors que le centre de la tumeur est hypoxique. En poussant ce phénomène, on obtient des évolutions de type « plateau », c'est à dire que la tumeur peut ralentir à l'extrême sa croissance.

**Croissance « exponentielle »** On voit sur la Figure 3.9 que l'ensemble de la tumeur prolifère.

On peut, de même obtenir des croissances que l'on pourrait qualifier de « sur-exponentielles ». Ces cas correspondent à une tumeur entièrement proli-

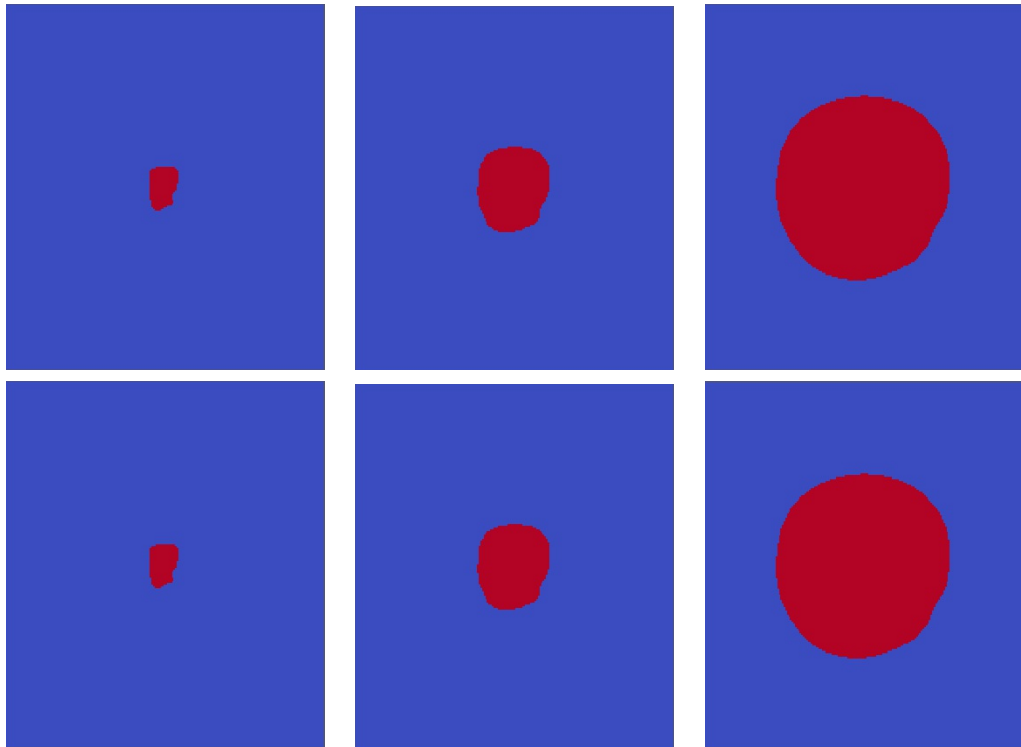


FIGURE 3.9 – Exemple de croissance exponentielle, en haut la densité de cellules tumorales  $T$ , en bas les zones proliférantes (produit  $T \times M$ ) ; de gauche à droite :  $t = 0$ ,  $t = 1,5$  et  $t = 3$ . Paramètres :  $M_0 = 1,1$ ,  $M_1 = 1,1$ ,  $\eta = 1$ .

férente mais donc le taux de prolifération augmente au cours du temps. C'est typiquement le cas lorsque  $M_1 > M_0$ . Dans ce cas la tumeur simulée est particulièrement agressive.

De plus, ces différences de dynamique liées à l'épaisseur de la couche proliférante ont une influence sur la forme finale de la tumeur.

**Cas de tumeurs concaves** La méthode de reconstruction du champ  $T$  à partir du polygone de discrétisation de la tumeur a été conçue pour être utilisable avec des polygones simples quelconques, convexes ou pas. Cela dit, pendant la simulation, les points sont transportés et le polygone est déformé. On peut alors se demander si ces déformations ne peuvent pas être à l'origine de dysfonctionnements de la méthode numérique. Notamment, si le polygone présente une excroissance, plusieurs zones de la frontière peuvent être amenées à se croiser. Certains points de discrétisation se retrouvent alors à l'intérieur du polygone et il convient de tester la méthode sur des cas tests reproduisant ces spécificités. Une simulation à partir d'un cas fictif est présentée en Figure 3.10. Nous pouvons tout d'abord constater sur la Figure 3.10a que la concavité

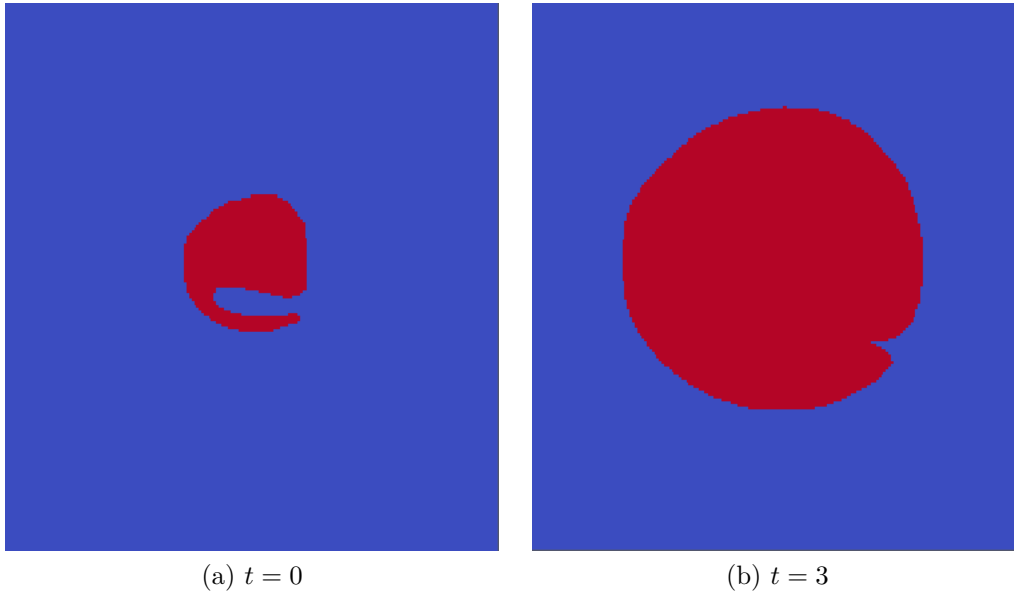


FIGURE 3.10 – Densités de cellules tumorales issues d'une simulation du modèle sur un cas test de polygone présentant une excroissance, à gauche la donnée initiale et à gauche le résultat à  $t = 3$  ( $\eta = 0,12$ ,  $M_0 = 0,65$  et  $M_1 = 0$ ).

du polygone n'est pas un frein à l'algorithme de reconstruction du champ  $T$ . De plus, lors de la simulation, l'excroissance est peu à peu intégrée à la partie

principale de la tumeur et la forme se régularise sans, là encore, que l'on observe d'anomalie.

**Influence de la couronne proliférante sur la forme de la tumeur** Pour tester l'influence de la couronne sur la forme de la tumeur, nous avons utilisé une ellipse comme donnée initiale. Puis on simule le modèle à partir de cette ellipse d'une part pour un jeu de paramètres de type « exponentiel » et d'autre part pour un jeu de paramètres donnant une couronne proliférante fine. Les résultats sont donnés en Figure 3.11.

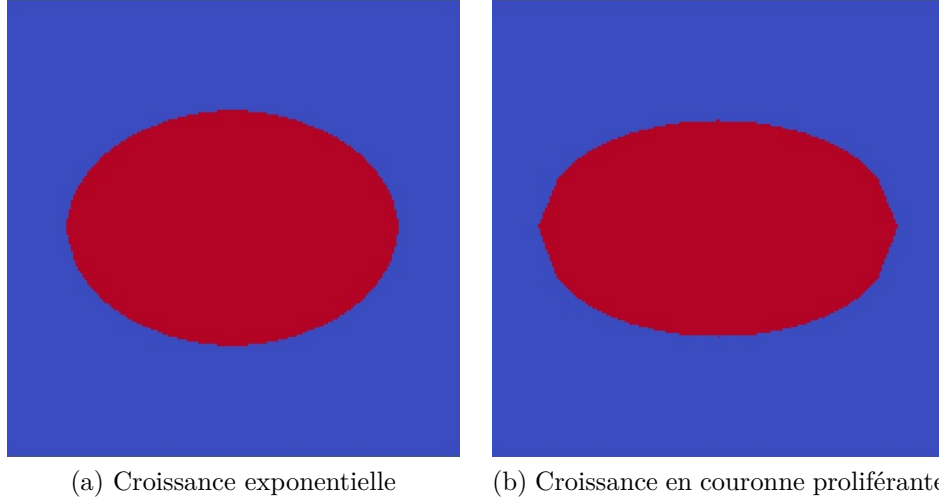


FIGURE 3.11 – Comparaison des formes finales à masses égales pour une condition initiale elliptique. À droite, croissance de type exponentielle ( $\eta = 2,5$ ,  $M_0 = 1$  et  $M_1 = 0,8$ ) ; à gauche, croissance présentant une couche extérieure proliférante ( $\eta = 2,67$ ,  $M_0 = 3$  et  $M_1 = 0$ ).

On constate donc que les paramètres utilisés pour la simulation jouent sur la forme de la tumeur simulée. Cela confirme l'intérêt d'une approche spatiale.

### 3.2.3 Analyse de sensibilité du modèle

Comme pour le modèle précédent, nous procédons à une analyse de sensibilité afin d'étudier l'importance relative des différents paramètres pour le modèle. Les résultats, obtenus via la méthode de Morris [51], sont donnés en Figure 3.12 pour l'espace de paramètres décrit dans la Table 3.1.

Pour l'interprétation de ce genre de diagramme, on peut se reporter à la section 2.3.3. Ici, on peut voir que les valeurs des moyennes et écart-types sont

paramètre	intervalle de variation
$\eta$	$[0,1; 10]$
$M_0$	$[0,2; 1,4]$
$M_1$	$[0; 0,3]$

TABLE 3.1 – Espace de paramètres utilisé pour l'analyse de sensibilité du modèle simplifié.

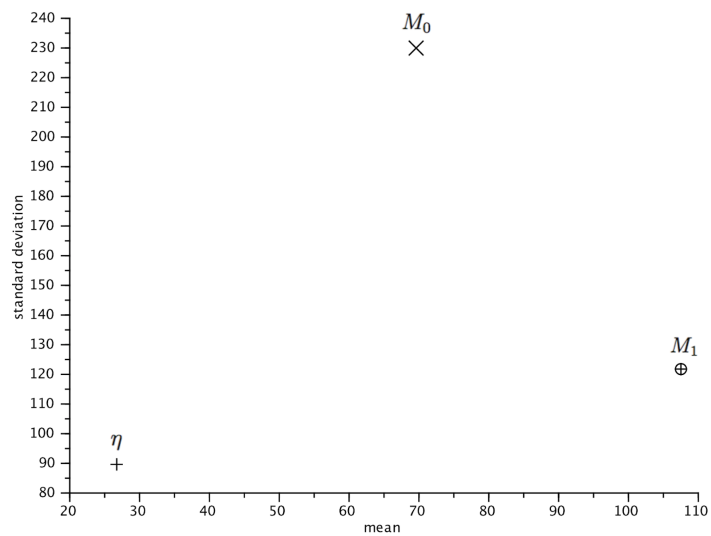


FIGURE 3.12 – Analyse de sensibilité sur le modèle simplifié par la méthode de Morris : effet de chaque paramètre sur le volume final de la tumeur.

du même ordre de grandeur pour les trois paramètres. On conserve donc tous les paramètres.

### 3.3 Simulations en 3 dimensions

La méthode numérique utilisée pour le cas 2D n'est pas directement transposable au cas 3D. En effet, l'algorithme de reconstruction de la tumeur à partir des points d'échantillonnage de l'interface tissu sain / tissu cancéreux, s'appuie sur le fait qu'en sortie de segmentation, on dispose d'une liste ordonnée de points. En trois dimensions, il n'est plus possible d'ordonner de la même manière le nuage de points obtenu lors de la segmentation de la tumeur. De plus, des méthodes classiques existent pour reconstituer l'enveloppe convexe d'un nuage de points, mais ne conviennent pas dans notre cas puisque les métastases n'ont aucune raison d'être convexes. D'autres méthodes fonctionnent sur des formes non convexes mais nécessitent de connaître les directions normales à la surface (voir [33]). Cependant, il est possible d'adapter la méthode décrite pour le cas 2D pour simuler le modèle en trois dimensions en tenant compte des spécificités de notre problème.

#### 3.3.1 Formulation du problème

On repart de la formulation lagrangienne du modèle simplifié de croissance tumorale :

$$\begin{aligned}\frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \nabla \Phi &= 0, \\ T &= \begin{cases} 1 & \text{à l'intérieur de } \Phi, \\ 0 & \text{à l'extérieur de } \Phi, \end{cases} \\ \frac{\partial M}{\partial t} &= -\eta(M - M_1)T, \\ \nabla \cdot \mathbf{v} &= MT, \\ \mathbf{v} &= -\nabla \pi.\end{aligned}$$

Le champ  $T$  intervenant dans les équations sur la vascularisation et sur la divergence de la vitesse, doit, comme pour le cas 2D, être reconstitué à chaque pas de temps. C'est cette reconstruction de  $T$  qui présente la plus grande difficulté, la résolution du reste du système d'équations (vascularisation, équation de Poisson sur la pression...) étant réalisée par des schémas numériques classiques.

### 3.3.2 Reconstruction du champ $T$ à partir des points de l'interface

L'algorithme de reconstruction de la tumeur en 3 dimensions à partir d'une discrétisation de l'interface (nuage de points) s'inspire de celui utilisé pour le cas 2D. Cette fois, au lieu de décomposer l'interface en segments, on la décompose en triangles élémentaires. La première difficulté provient de la triangulation de la surface, *a priori* non convexe, de la tumeur.

#### Triangulation de la surface

Il existe des méthodes classiques pour construire des triangles à partir de nuages de points. La méthode de triangulation de Delaunay (voir [42]) est ainsi régulièrement mise en oeuvre dans des problématiques telles que les maillages utilisés pour les méthodes de type éléments finis. **Dans notre cas, la surface à reconstruire n'est pas convexe et le nuage de points résultant de la segmentation de la tumeur présente des caractéristiques particulières qui peuvent être utilisées pour définir une méthode de triangulation adaptée.** La segmentation d'une métastase pulmonaire en trois dimensions consiste en la segmentation des différentes coupes contenant la tumeur. Il s'agit de coupes transversales de la cage thoracique orthogonales à l'axe  $\vec{n}_z$ . L'ensemble des points issus d'une même coupe se trouve donc dans un même plan d'équation  $z = z_i$  où  $z_i$  est la cote correspondant à la  $i^{\text{ème}}$  coupe portant la tumeur. Le nuage de points échantillonnant la donnée initiale  $\Phi(t = 0)$  se présente comme une succession de listes ordonnées de points de même cote. On procède différemment pour les triangles des faces latérales et pour ceux du « dessus » et du « dessous » du polyèdre, correspondant aux première et dernière coupes contenant la tumeur.

**Triangulation des faces latérales.** La méthode utilisée consiste à relier les points de deux coupes successives de manière à former des triangles élémentaires décrivant la surface. Il est primordial que l'ensemble des triangles forme une partition de la surface, c'est à dire d'éviter toute superposition des triangles mais aussi que certaines portions de la surface ne soient pas incluses dans un triangle. On s'appuie sur un cas simple présenté en Figure 3.13 pour illustrer cette méthode. Pour plus de lisibilité, la figure représente en vue étalée (2D) les points formant les  $i^{\text{ème}}$  et  $i + 1^{\text{ème}}$  couches d'une segmentation fictive. Afin d'assurer la fermeture du polyèdre, le premier point (d'indice 0) de chaque couche est dupliqué en fin de ligne. Pour procéder à la triangulation, deux étapes sont nécessaires.

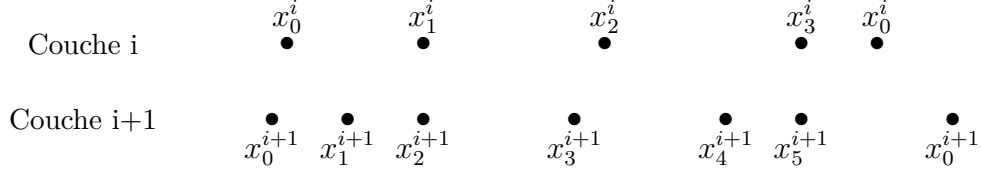


FIGURE 3.13 – Présentation en vue étalée des points de deux couches consécutives de la segmentation d'une métastase fictive.

- Dans un premier temps, on forme une première série de triangles ayant pour base les segments joignant deux points successifs de la couche  $i$ . On connaît ainsi, pour chacun de ces triangles, deux des trois points et on prend un point de la couche inférieure  $i+1$  comme troisième sommet. Ces triangles sont donc de la forme  $(x_j^i, x_{j+1}^i, x_k^{i+1})$  où  $i \in \mathbb{Z}/4\mathbb{Z}$  il reste donc à déterminer quel point de la couche inférieure  $x_k^{i+1}$  choisir. On complète par le point de la couche inférieure minimisant le périmètre du triangle (voir Figure 3.14). Si deux points de la couche inférieure permettent de minimiser le périmètre, ce qui se produit notamment lorsque les points forment un rectangle, on prend celui ayant le plus petit indice (celui de gauche sur l'exemple).

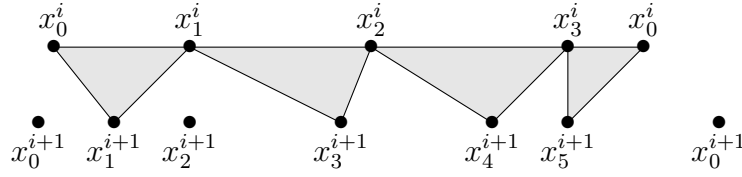


FIGURE 3.14 – Illustration de la méthode de triangulation des faces latérales sur un cas fictif : première étape.

- Une fois la première étape de triangulation effectuée, il reste à combler les zones non couvertes (zones non grisées sur la Figure 3.14). On voit que cela peut se faire par la formation de triangles ayant pour base des segments du type  $[x_j^{i+1}; x_{j+1}^{i+1}]$ ,  $j \in \mathbb{Z}/6\mathbb{Z}$  reliant deux points consécutifs de la couche inférieure. Pour ce faire, on parcourt les couples de triangles successifs précédemment construits. Notre but est de combler les éventuels espaces entre ces triangles. Prenons l'exemple, représenté Figure 3.15, des deux premiers triangles du cas fictif :  $(x_0^i, x_1^i, x_0^{i+1})$  et  $(x_1^i, x_2^i, x_1^{i+1})$ .

La zone restant à trianguler (hachurée en rouge sur la Figure 3.15) peut être facilement découpée en triangles ayant pour base un côté reliant deux points consécutifs de la couche inférieure, c'est à dire du



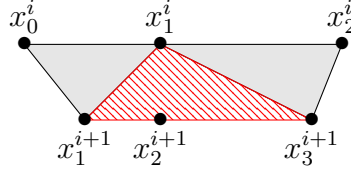


FIGURE 3.15 – Deuxième étape de triangulation des faces latérales : la zone hachurée en rouge correspond à la surface restant à triangulariser.

type  $[x_k^{i+1}, x_{k+1}^{i+1}]$ . Le sommet opposé est alors le point commun des deux triangles de la couche supérieure, c'est à dire  $x_1^i$ . Nous formons ainsi deux triangles supplémentaires  $(x_1^{i+1}, x_2^{i+1}, x_1^i)$  et  $(x_2^{i+1}, x_3^{i+1}, x_1^i)$  comme on peut le voir sur la Figure 3.16.

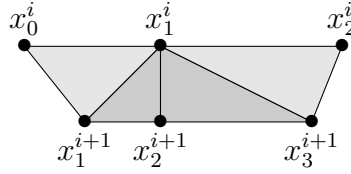


FIGURE 3.16 – Deuxième étape de triangulation des faces latérales : formation des deux triangles de la couche inférieure en gris foncé.

On procède de même pour tous les couples de triangles successifs de la couche supérieure. Le résultat obtenu est représenté en Figure 3.17.

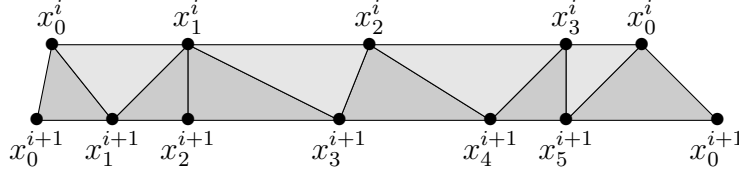


FIGURE 3.17 – Illustration de la méthode de triangulation des faces latérales sur un cas fictif : résultat final obtenu.

On réitère l'opération pour l'ensemble des couples de couches successives. Il reste à trianguler les deux couches aux extrémités.

**Triangulation des extrémités.** Pour les deux couches du « dessus » et du « dessous », la méthode est plus simple puisqu'il s'agit simplement de trianguler un polygone (en deux dimensions). Un cas simple est présenté en vue de dessus en Figure 3.18. Afin de trianguler cette surface, on ajoute le barycentre

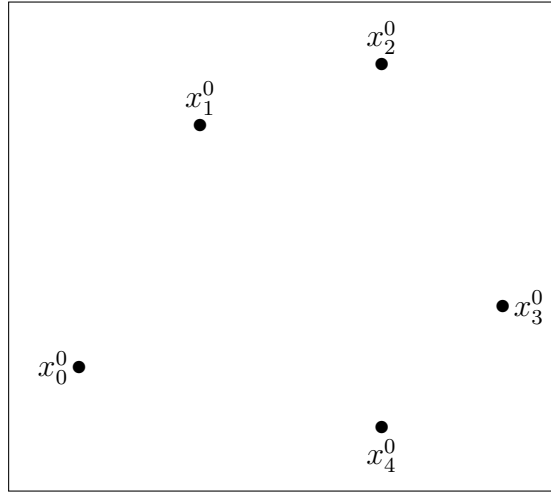


FIGURE 3.18 – Cas fictif de discrétisation de la couche supérieure de la tumeur.

des points du polygone si celui-ci est à l'intérieur du polygone et on génère les triangles formés par les points consécutifs du polygone et le barycentre. Dans le cas particulier où le barycentre du polygone est hors de celui-ci, qui ne s'est jamais produit sur l'ensemble des cas étudiés, on peut utiliser une méthode plus complexe de triangulation comme la méthode dite des oreilles [50]. On obtient donc la triangulation présentée en Figure 3.19.

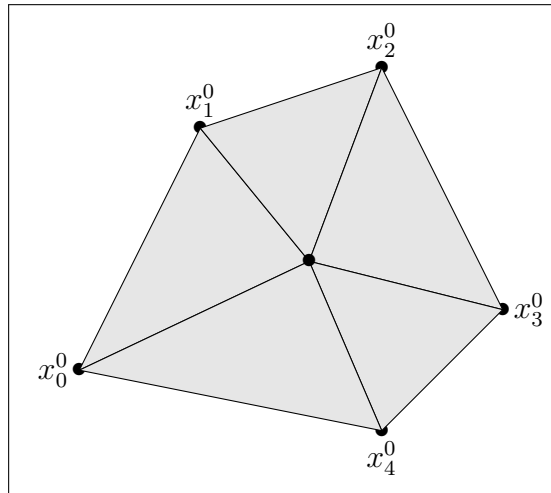


FIGURE 3.19 – Triangulation de la couche supérieure de la tumeur sur le cas fictif.

### Test d'appartenance d'un point à la tumeur

Une fois l'étape de triangulation de la surface achevée, on peut tester l'appartenance de chaque voxel du domaine de calcul à la tumeur. Le critère est le suivant : si le centre du voxel, dont on peut aisément récupérer les coordonnées réelles, est dans la tumeur, alors on considère que le voxel entier est dans la tumeur. Il s'agit bien sûr d'une approximation, mais c'est un critère assez simple à mettre en oeuvre et donnant des résultats suffisamment satisfaisants. Pour un point de coordonnées connues  $P(x_p; y_p; z_p)$ , on cherche donc un moyen de déterminer s'il est à l'intérieur ou à l'extérieur de la tumeur, c'est à dire de la surface délimitée par les triangles générés à l'étape précédente. Pour cela, on s'inspire de la méthode utilisée dans le cas bidimensionnel. Il suffit ainsi de tracer une demi-droite ayant pour origine  $P$  dans une direction définie et de compter le nombre de triangles qu'elle traverse. Comme dans le cas 2D, si ce nombre est impair, le point est à l'intérieur de la tumeur et si ce nombre est pair, il est à l'extérieur. Par défaut, on génère la demi-droite dans la direction des  $y$  croissants, de la forme  $P + u \times \vec{n}_y, u \in \mathbb{R}^+$ . On parcourt donc l'ensemble des triangles formant la surface de la tumeur et on procède en trois temps afin de déterminer si la demi-droite issue du point courant traverse en effet le triangle.

- Tout d'abord, on calcule l'équation cartésienne du plan supportant le triangle. Prenons le cas d'un triangle  $ABC$  tel que  $A(0; 1; 1)$ ,  $B(1; 2; -1)$  et  $C(-1; 2; -1)$  représenté par le schéma donné en Figure 3.20. Ces trois

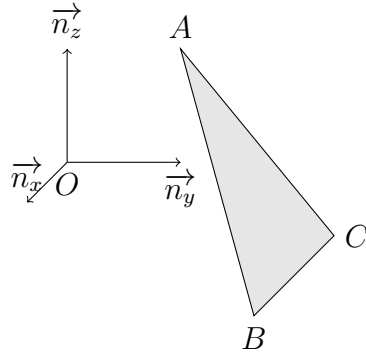


FIGURE 3.20 – Schéma du triangle ABC.

points sont non alignés donc le plan portant le triangle  $ABC$  peut être défini par les vecteurs  $\vec{AB}$  et  $\vec{AC}$  qui forment une base du plan mais aussi par le vecteur normal donné par  $\vec{n} = \vec{AB} \wedge \vec{AC}$ . Dans notre cas,  $\vec{AB}(1; 1; -2)$  et  $\vec{AC}(-1; 1; -2)$ , puis en calculant le produit vectoriel,

on obtient un vecteur normal au plan :  $\vec{n}(0; -4; -2)$ . À partir de ce vecteur on peut déterminer une équation cartésienne du plan, celle-ci se présente sous la forme  $n_x x + n_y y + n_z z + d = 0$  soit dans notre cas et après simplifications,  $-2y - z + d' = 0$ . On détermine la valeur de  $d'$  en appliquant cette équation cartésienne au point  $A$  qui, par construction appartient au plan. Il vient :  $-2 \times 1 - 1 + d' = 0$  soit  $d' = 3$ . Une équation cartésienne du plan est donc :

$$-2y - z + 3 = 0.$$

- L'ensemble des points appartenant à la demi-droite  $P + u \times \vec{n}_y, u \in \mathbb{R}^+$  a pour coordonnées  $(x_p; y_p + u; z_p), u \in \mathbb{R}^+$ .  
Un premier cas doit être traité à part, il s'agit du cas où la demi-droite est parallèle au plan, c'est à dire que son vecteur directeur, ici  $\vec{n}_y$  est orthogonal au vecteur normal au plan  $\vec{n}$ . L'orthogonalité de ces deux vecteurs est équivalente à la nullité de leur produit scalaire. Dans ce cas, on sait d'ores et déjà que la demi-droite n'intersectera pas le plan (elle peut cependant appartenir au plan). Si ce cas particulier est exclu, on calcule la projection  $P'$  du point courant  $P$  sur le plan  $(A, B, C)$ , ce qui revient à résoudre le système d'équations suivant :

$$\begin{cases} -2y_{p'} - z_{p'} + 3 = 0, \\ x_{p'} = x_p, \\ y_{p'} = y_p + u, \\ z_{p'} = z_p. \end{cases}$$

La résolution de ce système équivaut à résoudre l'équation en  $u$  suivante :

$$-2(y_p + u) - z_p + 3 = 0$$

Si on prend, par exemple,  $P$  comme étant le point de coordonnées  $(0,1; 0,1; 0,1)$ , alors on trouve, en résolvant le système précédent,  $u = 1,35$  et  $P'(0,1; 1,45; 0,1)$ . Il est important de noter que la valeur de  $u$  trouvée est bien positive. On vérifie aisément que ce point appartient bien au plan et  $P'$  est bien le projeté de  $P$  sur le plan  $(A, B, C)$  suivant l'axe  $\vec{n}_y$  (voir Figure 3.21).

- Une fois que l'on a les coordonnées du point projeté  $P'$ , il reste à déterminer si ce point est bien à l'intérieur du triangle. Pour cela, on se positionne dans le repère du plan  $(A, B, C)$  déterminé par  $(A, \vec{AB}, \vec{AC})$  et on calcule les coordonnées barycentriques du point  $P'$  dans ce repère. Ces coordonnées  $(r, s)$  sont définies par la relation  $\vec{AP'} = r\vec{AB} + s\vec{AC}$  (voir Figure 3.22).

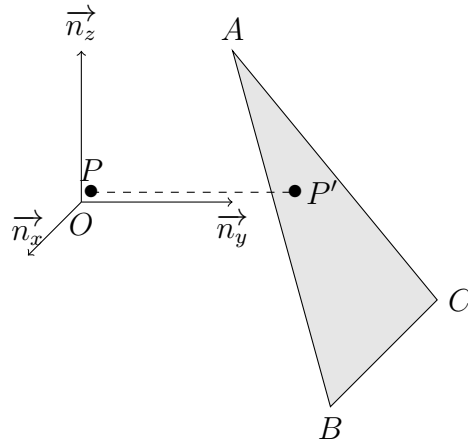


FIGURE 3.21 – Représentation de la projection du point  $P(0,1;0,1;0,1)$  sur le plan  $(A,B,C)$  suivant l'axe des ordonnées.

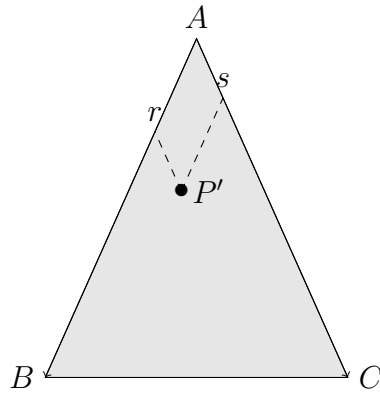


FIGURE 3.22 – Coordonnées barycentriques  $(r,s)$  du point  $P'$  dans la base  $(A, \overrightarrow{AB}, \overrightarrow{AC})$ .

Si  $0 \leq r \leq 1$  et  $0 \leq s \leq 1 - r$  alors, le point  $P'$  est bien dans le triangle et la demi-droite d'origine  $P$  et dirigée par le vecteur  $\vec{n}_y$  traverse le triangle  $ABC$ . On peut alors incrémenter le compteur d'intersections de cette demi-droite avec les triangles formant la surface de 1.

On procède ainsi avec tous les triangles issus de la triangulation de la surface et si le nombre de triangles traversés par la demi-droite est impair, alors le point courant  $P$  est dans la tumeur.

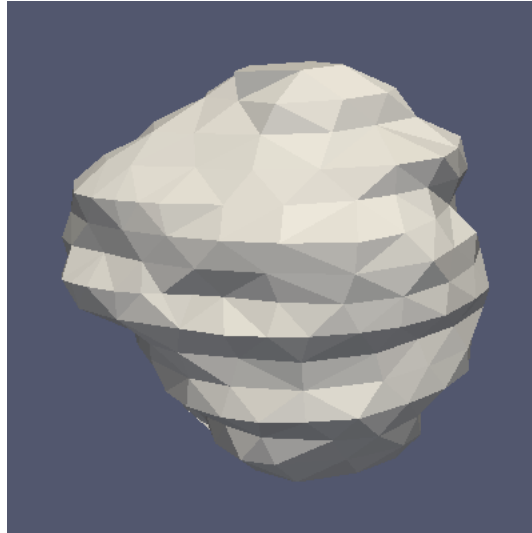
Il existe cependant des cas limites pouvant poser certaines difficultés. Il s'agit des points dont le projeté sur un des triangles appartient à un des côtés du triangle, voire est confondu avec un des sommets. Ces cas correspondent, aux cas d'égalités dans les conditions données sur les coordonnées barycentriques. Si  $P'$  appartient à un côté d'un triangle, il appartient forcément aussi au triangle partageant ce côté avec le premier, il suffit alors d'incrémenter le compteur de  $\frac{1}{2}$  au lieu de 1. De même, si  $P'$  est confondu avec un sommet, il convient d'incrémenter le compteur de l'inverse du nombre de triangles ayant ce sommet en commun.

### Reconstitution du champ $T$ et simulation numérique

Nous avons défini précédemment l'algorithme de test d'appartenance d'un point à la tumeur. On teste donc les centres des voxels du domaine du calcul. Les voxels pour lesquels le centre est dans la tumeur sont considérés intégralement dans la tumeur. Cependant, de même que dans le cas 2D, il n'est en général pas nécessaire de tester l'ensemble des voxels. D'une part, si le patient n'est pas traité, la tumeur ne diminue pas et un voxel atteint par la lésion n'a pas de raison de sortir de la tumeur. On considère donc que les points qui étaient déjà dans la tumeur y restent et on n'a pas besoin de tester ces points. De même, les voxels en dehors du parallélogramme englobant les points de discrétisation de l'interface ne peuvent pas appartenir à la tumeur et ne sont pas testés.

La Figure 3.23 présente la comparaison entre la triangulation du nuage de points et le champ  $T$  reconstruit. L'échantillonnage de la surface de la tumeur comporte 329 points et la surface recomposée compte 834 triangles.

Lors de la simulation, les différents points de discrétisation de la surface  $\Phi$ , qui forment les triangles matérialisant l'interface entre la tumeur et le tissu sain, sont transportés. Il n'y a donc plus aucune raison, pour  $t > 0$ , qu'ils restent organisés en couches. Autrement dit, les ensembles de points qui étaient initialement coplanaires ne le sont *a priori* plus et la méthode de reconstruction du champ  $T$  décrite précédemment ne peut plus être utilisée. Cependant, on fait l'hypothèse que, la vitesse étant assez régulière et centrifuge, les triangles générés sur la segmentation initiale de la tumeur demeurent une bonne des-



(a) triangulation de la surface

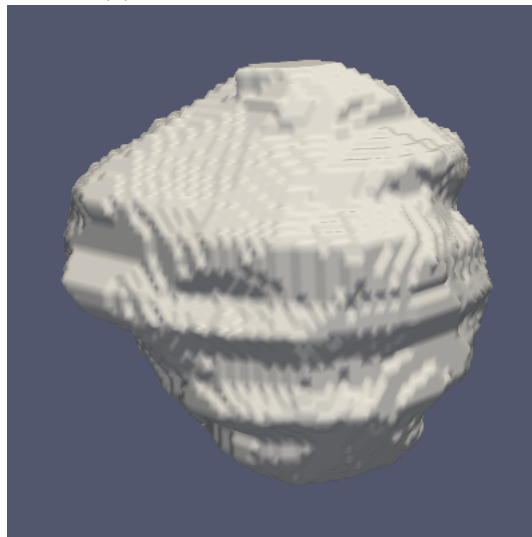
(b) champ  $T$  reconstruit

FIGURE 3.23 – Reconstruction d’une tumeur en 3 dimensions à partir d’un échantillonnage de la surface : en haut la triangulation de la surface, en bas le champ  $T$  reconstruit.

cription de la surface  $\Phi$ . Cela sous-entend, par exemple, que les triangles ne se croisent pas au cours du temps. D'un point de vue informatique, les triangles sont vus comme des listes de références vers les points. Ainsi les coordonnées des triangles évoluent en même temps que celles des points et la phase de triangulation est faite une fois pour toutes (hormis raffinements éventuels). Lorsque la tumeur croît, certaines zones de la tumeurs risquent de ne plus être décrites de façon satisfaisante par les triangles initiaux à cause d'une trop faible définition en points. L'échantillonnage n'est alors plus suffisant. On peut remédier à cela en subdivisant ces triangles. Il suffit d'ajouter à la liste des points le centre de gravité du triangle concerné et de le remplacer dans la triangulation par les trois triangles formés en joignant les côtés du triangle initial au centre de gravité (voir Figure 3.24).

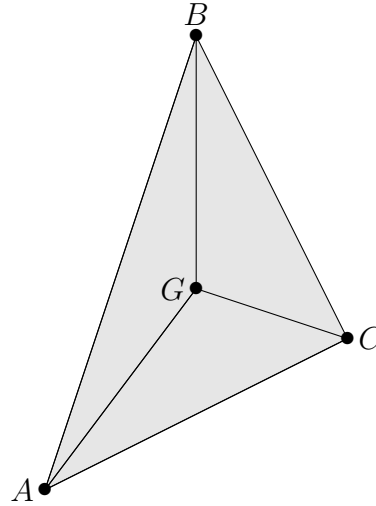


FIGURE 3.24 – Cas de raffinement de la triangulation par ajout d'un point. Le triangle  $ABC$  est remplacé dans la liste des triangles par les triangles  $ABG$ ,  $ACG$  et  $BCG$ .

La méthode employée permet de paralléliser la charge de calcul sur plusieurs threads. On peut notamment répartir les différents voxels du domaine à tester entre les différents threads. Le calcul des équations cartésiennes peut être effectué une fois pour toutes.

### 3.3.3 Simulations

Une fois le problème de reconstruction de la tumeur à partir de l'échantillonnage de l'interface résolu, on peut simuler le modèle à partir de données



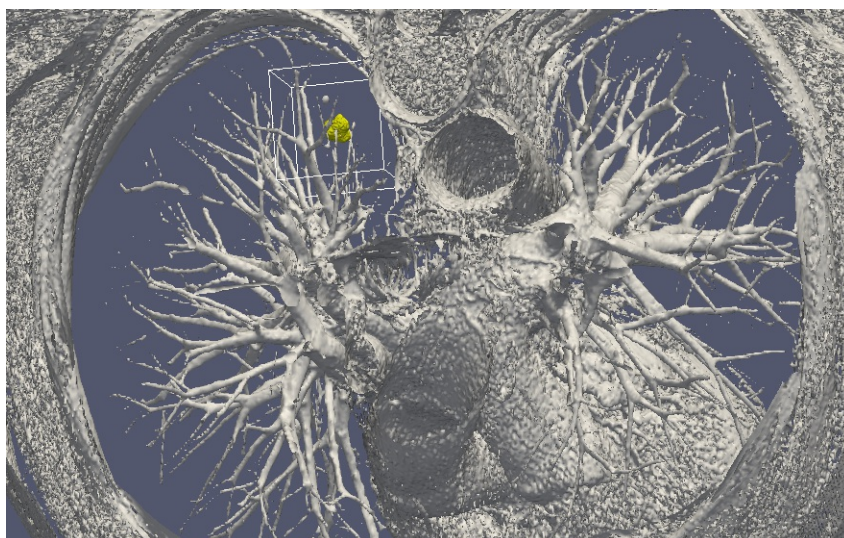
cliniques. On prend comme donnée initiale la tumeur présentée dans la Figure 3.23. On peut convertir une série de coupes de CT-scan du format standard dicom vers le format vtk utilisé pour l'écriture et la visualisation des résultats de simulation. En appliquant un seuillage adéquat des niveaux de gris issus du scan, on peut visualiser la structure de la cage thoracique du patient et repositionner la simulation dans cet environnement. On obtient les résultats présentés en Figure 3.25. On peut visualiser la localisation des cellules proliférantes au sein de la tumeur sur la Figure 3.26. Dans le cas présent, étant donné le jeu de paramètres utilisé, celles-ci forment une couronne périphérique. L'intérêt d'une image comme celle donnée en Figure 3.25 est qu'elle peut permettre au médecin d'identifier des zones sensibles qui pourraient se trouver affectées par la présence de la lésion. Ici, on voit par exemple que la tumeur risque d'atteindre un vaisseau sanguin et donc de nuire à l'irrigation d'une partie du poumon.

### 3.3.4 Application de la méthode de front tracking 3D au cas des méningiomes

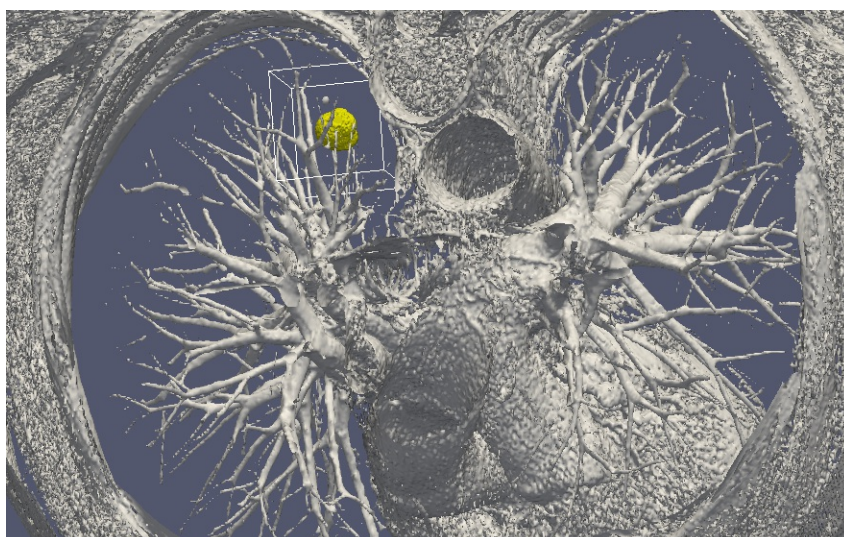
Pour les métastases pulmonaires, dans les cas cliniques, on travaille pour l'instant principalement en 2D, la 3D n'étant généralement pas indispensable et les coûts supplémentaires en temps de calcul et de segmentation pas justifiés. Par contre, pour les méningiomes, l'ensemble de la procédure est en 3D. La méthode de suivi d'interface en 3D développée pour la simulation de la croissance de métastases pulmonaires a aussi été utilisée pour l'étude de l'évolution de méningiomes, en collaboration avec Vivien Pianet dans le cadre de ses travaux de postdoc. Les méningiomes sont des tumeurs cérébrales généralement peu agressives qui se développent à partir des méninges (cellules formant une membrane protectrice autour du cerveau). Dans la plupart des cas (environ 80%), ces tumeurs sont bénignes mais elles se présentent parfois sous des formes plus agressives pouvant engager le pronostic vital du patient. Dans de nombreux cas, le méningiome est implanté sur la dure-mère (voir Figure 3.27), une des membranes constituant les méninges.

La Figure 3.28 présente un cas clinique de méningiome. Il s'agit d'une coupe extraite d'une séquence d'I.R.M sous injection de gadolinium. On voit bien que le méningiome s'est développé à l'extérieur du cerveau à partir de la dure-mère contre l'os crânien.

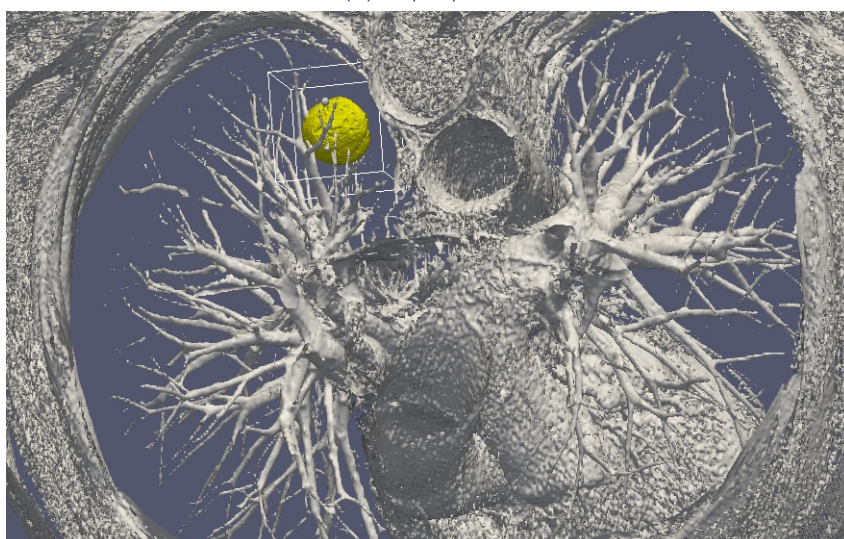
Il s'agit d'une tumeur solide (non infiltrante) tout comme les métastases pulmonaires. Le modèle mathématique utilisé est par conséquent assez proche du modèle simplifié pour les métastases pulmonaires. Il est présenté ci-dessous :



(a) 19/11/2009



(b) 11/03/2010



(c) 16/07/2010

FIGURE 3.25 – Simulation 3D basée sur la méthode de front tracking repositionnée dans la cage thoracique du patient. Les paramètres de la simulation sont les suivants :  $M_0 = 0,6$ ,  $M_1 = 0$  et  $\eta = 0,177$ .

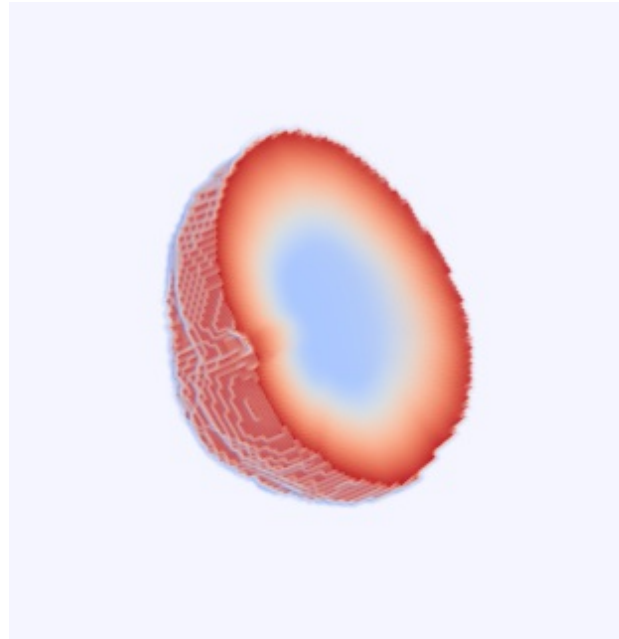


FIGURE 3.26 – Localisation des cellules proliférantes sur le cas 3D avec les paramètres  $M_0 = 0,6$ ;  $M_1 = 0$  et  $\eta = 0,177$ . La zone rouge représente la zone la plus dense en cellules proliférantes.

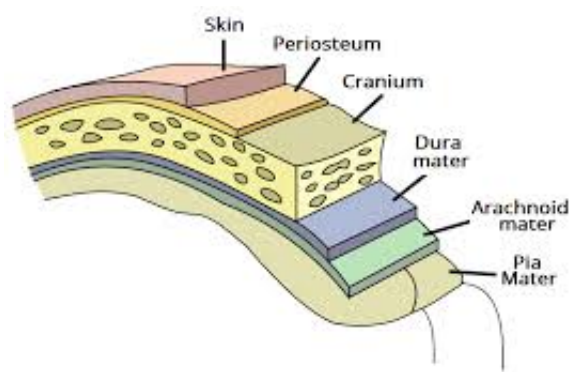


FIGURE 3.27 – Schéma présentant la disposition des tissus enveloppant le cerveau. Source : [teachmeanatomy.info](http://teachmeanatomy.info)

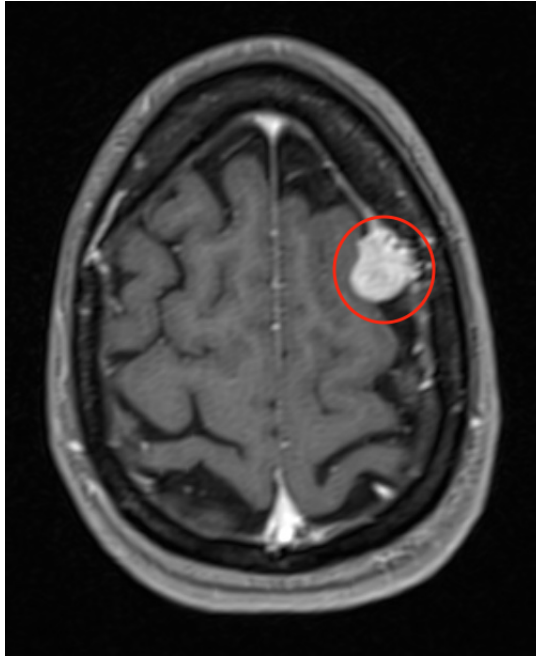


FIGURE 3.28 – Extrait d'une séquence d'IRM sous injection de gadolinium montrant la présence d'un méningiome.

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = MT,$$

$$\frac{\partial M}{\partial t} = -\eta M,$$

$$\nabla \cdot \mathbf{v} = MT,$$

$$\mathbf{v} = -K\nabla\pi,$$

où  $M(t = 0) = M_0$  et  $T(t = 0)$  est donné par l'examen initial. Le champ scalaire  $K$  est la porosité du tissu environnant. La différence majeure du cas du méningiome par rapport aux métastases pulmonaires, est que les méningiomes croissent le long du crâne, en suivant la dure-mère. Il ne se développent pas vers l'extérieur du crâne et on doit retranscrire ce comportement dans le système d'équations. Dans ce but, nous procédons par pénalisation en imposant une très petite valeur de  $K$  à l'extérieur du crâne.

$$K = \begin{cases} 1 & \text{à l'intérieur du crâne,} \\ 0,001 & \text{à l'extérieur du crâne,} \end{cases}$$

ce qui revient à imposer une condition de type Neumann sur la pression. La vitesse normale est donc nulle au bord. De même, l'équation sur la vascularisation  $M$  est plus simple que dans le cas des métastases pulmonaires et ne comporte qu'un paramètre (la donnée initiale  $M_0$ ). Ce système peut tout à fait être simulé via un schéma de transport de type WENO 5. Cependant, des imprécisions numériques apparaissent près du crâne dans le cas 3D. Nous avons donc décidé d'utiliser la méthode de front tracking développée pour les métastases pulmonaires sur le modèle de croissance des méningiomes. Le but est d'une part de résoudre le problème d'imprécisions numériques et d'autre part de tester la méthode de front tracking sur un cas de croissance tumorale légèrement différent.

Pour la mise en oeuvre de la méthode, nous procédons de même que pour les métastases pulmonaires à partir de la formulation lagrangienne du problème. Des résultats de simulation sont présentés Figure 3.29.

### 3.4 Réduction 0D du modèle pour les métastases pulmonaires

Plutôt que de calibrer le modèle EDP sur les données spatiales, ce qui nécessite de nombreuses simulations relativement coûteuses en calcul, on peut tenter d'obtenir une équation sur le volume (surface lorsqu'on est en 2D) de la tumeur. La calibration se fait alors sur une variable scalaire et on perd donc une partie des informations liées à la forme de la tumeur. En contrepartie, on est en mesure de calibrer quasi instantanément le modèle. On repart donc du système d'équations aux dérivées partielles simplifié :

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = MT, \quad (3.6)$$

$$\frac{\partial M}{\partial t} = -\eta(M - M_1)T, \quad (3.7)$$

$$\nabla \cdot \mathbf{v} = MT, \quad (3.8)$$

$$\mathbf{v} = -\nabla \pi. \quad (3.9)$$

On introduit une nouvelle variable scalaire  $\bar{M}$  qui représente la densité moyenne du réseau vasculaire ou de nutriments dans la tumeur :

$$\bar{M}(t) = \frac{1}{V(t)} \int_{\Omega(t)} M(t, \mathbf{x}) d\omega.$$



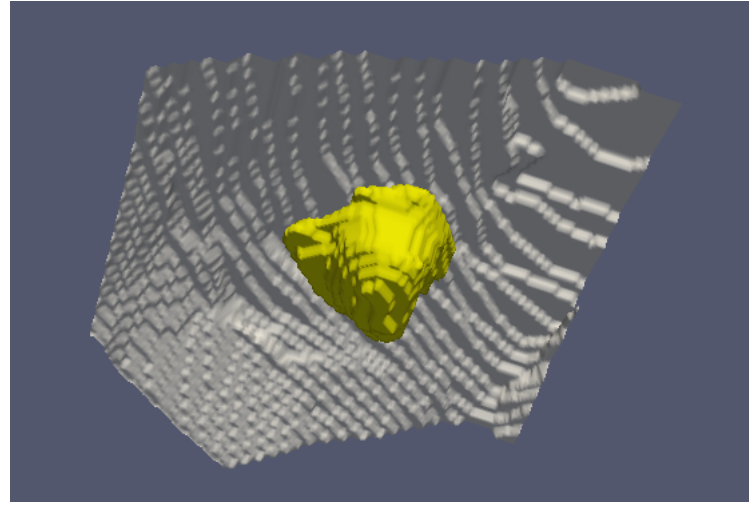
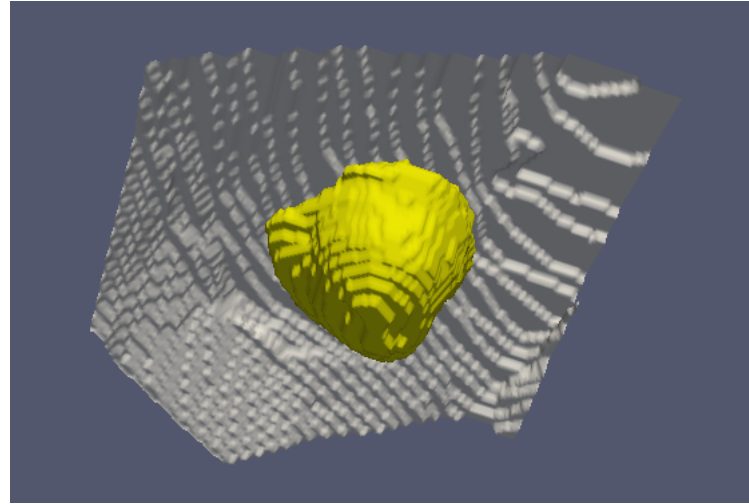
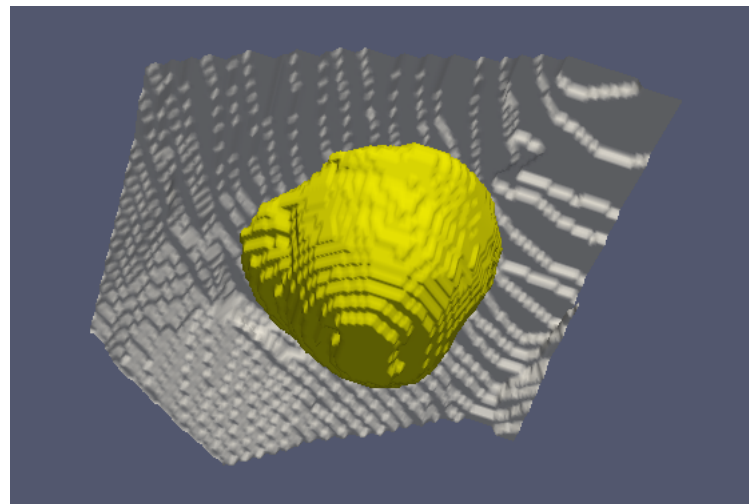

(a)  $t=0$ 

(b)  $t=2$ 

(c)  $t=4$ 

FIGURE 3.29 – Simulation 3D de l'évolution d'un méningiome basée sur la méthode de front tracking. La tumeur est représentée en jaune et le crâne en gris clair. Les paramètres de la simulation sont les suivants :  $M_0 = 0,6$  et  $\eta = 0,177$  et les temps (adimensionnés) correspondant aux snapshots sont de haut en bas :  $t = 0$ ,  $t = 2$  et  $t = 4$ .

Les variables  $V(t)$  et  $\Omega(t)$  représentent respectivement le volume de la tumeur et l'indicatrice du support de la tumeur à l'instant  $t$  ( $\Omega$  vaut 1 dans la tumeur, 0 à l'extérieur de la tumeur). On dérive l'expression du volume tumoral

$$V(t) = \int_{\Omega(t)} T d\omega$$

par

$$\frac{\partial V}{\partial t} = \frac{\partial}{\partial t} \int_{\Omega(t)} T d\omega.$$

En utilisant le théorème de Reynolds, on obtient :

$$\frac{\partial V}{\partial t} = \int_{\Omega(t)} \frac{\partial T}{\partial t} d\omega + \int_{\Omega(t)} \nabla \cdot (\mathbf{v}T) d\omega.$$

Puis on remplace  $\frac{\partial T}{\partial t}$  par son expression donnée par l'équation (3.6) :

$$\frac{\partial V}{\partial t} = \int_{\Omega(t)} MT d\omega - \int_{\Omega(t)} \nabla \cdot (\mathbf{v}T) d\omega + \int_{\Omega(t)} \nabla \cdot (\mathbf{v}T) d\omega.$$

Les deux termes de flux s'annulent et comme, par définition,  $T$  vaut 1 dans le domaine d'intégration  $\Omega(t)$ , on obtient finalement une équation différentielle sur  $V(t)$  :

$$\frac{\partial V}{\partial t} = \int_{\Omega(t)} M d\omega = \bar{M}V. \quad (3.10)$$

Il reste à déterminer une équation sur  $\bar{M}$ . Pour cela, on dérive l'expression  $\bar{M}(t) = \frac{1}{V(t)} \int_{\Omega(t)} M(t, \mathbf{x}) d\omega$  :

$$\begin{aligned} \frac{\partial \bar{M}}{\partial t} &= -\frac{1}{V^2} \frac{\partial V}{\partial t} \int_{\Omega(t)} M d\omega + \frac{1}{V} \frac{\partial}{\partial t} \left( \int_{\Omega(t)} M d\omega \right) \\ &= -\frac{1}{V} \frac{\partial V}{\partial t} \bar{M} + \frac{1}{V} \left[ \int_{\Omega(t)} \left( \frac{\partial M}{\partial t} + M \nabla \cdot \mathbf{v} \right) d\omega \right]. \end{aligned}$$

On utilise ensuite la définition de  $\bar{M}$  et l'équation (3.7) :

$$\frac{\partial \bar{M}}{\partial t} = -\bar{M}^2 + \frac{1}{V} \int_{\Omega(t)} -\eta(M - M_1) d\omega + \frac{1}{V} \int_{\Omega(t)} M \nabla \cdot \mathbf{v} d\omega.$$

Or,

$$\frac{1}{V} \int_{\Omega(t)} -\eta(M - M_1) d\omega = -\eta(\bar{M} - M_1),$$

et

$$\frac{1}{V} \int_{\Omega(t)} M \nabla \cdot \mathbf{v} d\omega = \frac{1}{V} \int_{\partial\Omega(t)} M \mathbf{v} \cdot \mathbf{n} ds,$$

où  $\partial\Omega(t)$  représente la frontière du domaine d'intégration et  $\mathbf{n}$  un vecteur unitaire sortant. Sur le bord de la tumeur  $\partial\Omega(t)$ ,  $M = M_0$  donc

$$\begin{aligned} \frac{1}{V} \int_{\Omega(t)} M \nabla \cdot \mathbf{v} d\omega &= \frac{M_0}{V} \int_{\partial\Omega(t)} \mathbf{v} \cdot \mathbf{n} ds \\ &= \frac{M_0}{V} \int_{\Omega(t)} \nabla \cdot \mathbf{v} d\omega \\ &= M_0 \bar{M}. \end{aligned}$$

On obtient ainsi une équation différentielle sur la variable  $\bar{M}$  :

$$\frac{\partial \bar{M}}{\partial t} = -\bar{M}^2 + (M_0 - \eta)\bar{M} + \eta M_1. \quad (3.11)$$

On intègre l'équation précédente pour obtenir une formulation explicite de  $\bar{M}(t)$ , on factorise le membre de droite :

$$\frac{\partial \bar{M}}{\partial t} = -(\bar{M} - r_1)(\bar{M} - r_2).$$

En notant

$$\begin{aligned} \Delta &= (M_0 - \eta)^2 + 4\eta M_1, \\ r_1 &= \frac{M_0 - \eta + \sqrt{\Delta}}{2}, \\ r_2 &= \frac{M_0 - \eta - \sqrt{\Delta}}{2}. \end{aligned}$$

On a donc :

$$\frac{\partial \bar{M}}{\bar{M} - r_1} - \frac{\partial \bar{M}}{\bar{M} - r_2} = (r_2 - r_1) \partial t.$$

Puis on intègre, il vient :

$$\ln\left(\frac{\bar{M} - r_1}{\bar{M}(t=0) - r_1}\right) - \ln\left(\frac{\bar{M} - r_2}{\bar{M}(t=0) - r_2}\right) = (r_2 - r_1)t + C,$$

où  $C$  est une constante d'intégration. En écrivant cette équation en  $t = 0$  avec  $\bar{M}(t=0) = M_0$ , on obtient  $C = 0$ .

Enfin, on obtient une expression pour  $\bar{M}$  :

$$\frac{\bar{M} - r_1}{\bar{M} - r_2} \frac{M_0 - r_2}{M_0 - r_1} = e^{(r_2 - r_1)t},$$



on note  $\alpha = \frac{M_0 - r_1}{M_0 - r_2}$ ,

$$\frac{\bar{M} - r_1}{\bar{M} - r_2} = \alpha e^{(r_2 - r_1)t},$$

et on obtient finalement :

$$\bar{M} = \frac{r_1 - r_2 \alpha e^{(r_2 - r_1)t}}{1 - \alpha e^{(r_2 - r_1)t}} = \frac{r_1 - r_2 \alpha e^{-\sqrt{\Delta}t}}{1 - \alpha e^{-\sqrt{\Delta}t}}. \quad (3.12)$$

On injecte cette expression dans l'équation différentielle sur le volume tumoral Eq. (3.10), puis on intègre :

$$\frac{\partial V}{\partial t} = \frac{r_1 - r_2 \alpha e^{-\sqrt{\Delta}t}}{1 - \alpha e^{-\sqrt{\Delta}t}} V.$$

$$\frac{\partial V}{V} = \frac{r_1 - r_2 \alpha e^{-\sqrt{\Delta}t}}{1 - \alpha e^{-\sqrt{\Delta}t}} \partial t.$$

$$\ln\left(\frac{V}{V_0}\right) = (r_1 - r_2) \left( t - \frac{1}{r_2 - r_1} \ln(\alpha e^{-\sqrt{\Delta}t} - 1) \right) + r_2 t + D,$$

où  $D$  est une constante d'intégration. En écrivant la relation précédente à  $t = 0$ , on trouve  $D = -\ln(\alpha - 1)$ . Après simplification on obtient une formule explicite pour le volume tumoral :

$$V(t) = \frac{V_0 e^{r_1 t}}{\alpha - 1} (\alpha e^{-\sqrt{\Delta}t} - 1). \quad (3.13)$$

### 3.5 Étude des bornes de l'espace de recherche

Une des limites du premier modèle était la difficulté de calibration puisqu'on n'avait aucun moyen analytique de délimiter l'espace des paramètres. *A contrario*, le modèle simplifié nous permet, à partir des données cliniques, de déterminer la zone de l'espace des paramètres à fouiller pour calibrer le modèle. Cela nous permet de limiter l'espace des paramètres à explorer et donc de réduire les coûts en terme de calculs, mais aussi de repenser notre stratégie de calibration et d'exploitation des résultats.

Partons de l'équation (3.7) sur  $M$ . Dans la tumeur,  $T$  vaut 1, on a donc :

$$\frac{\partial M}{\partial t} = -\eta(M - M_1),$$

que l'on peut intégrer en temps :

$$M(t) = (M_0 - M_1)e^{-\eta t} + M_1. \quad (3.14)$$

En injectant cette expression dans (3.6) et en intégrant sur le domaine de calcul, on obtient une équation différentielle sur le volume  $V$  (ou la surface si on est dans les cas 2D) de la tumeur :

$$\frac{\partial V}{\partial t} = [(M_0 - M_1)e^{-\eta t} + M_1]V,$$

qui peut, de même, être intégrée en temps :

$$V(t) = V_0 e^{M_1 t + \frac{M_0 - M_1}{\eta} (1 - e^{-\eta t})}. \quad (3.15)$$

Cette équation donnerait le volume de la tumeur si  $M$  était un scalaire ce qui n'est pas le cas. Cependant, elle va nous permettre de borner l'espace de recherche pour les paramètres  $M_0$  et  $M_1$ . Pour simplifier le raisonnement, on fait ici l'hypothèse que l'on dispose de deux images de la tumeur (scanners dans notre cas). La première donnée, à  $t = t_0$  est la donnée initiale. On prend d'ailleurs  $t_0$  comme origine des temps ( $t_0 = 0$ ). La seconde donnée correspond à une image de la tumeur à l'instant  $t_1 > t_0$ . On calcule facilement le volume (ou la surface si on est dans le cas surfacique) de la tumeur à partir de l'image (intégration spatiale de l'indicatrice de la tumeur) et on dispose ainsi des trois données scalaires suivantes  $t_1$ ,  $V_0$  et  $V_1$ .

Si l'on reprend l'expression locale de  $M$  en chaque point de la tumeur Eq. (3.14), on remarque que  $M$  varie continûment et de façon monotone de  $M_0$  à  $M_1$  comme on le voit sur la Figure 3.30. Ainsi, si  $M_0 < M_1$ ,  $M$  va croître de  $M_0$  vers  $M_1$  et on obtient l'encadrement  $M_0 \leq M(t) \leq M_1$  pour tout  $t$ . Inversement, si  $M_0 > M_1$ ,  $M$  est décroissant et on a  $M_1 \leq M(t) \leq M_0$  pour tout  $t$ .

Les trois paramètres que l'on cherche à calibrer sont  $\eta$ ,  $M_0$  et  $M_1$ . Ce sont donc les trois intervalles de recherche associés que l'on souhaite déterminer au mieux. Le paramètre  $\eta$  conditionne la vitesse de convergence de  $M$  vers la valeur  $M_1$ . Plus  $\eta$  est grand et plus vite  $M$  atteint  $M_1$ . Il semble donc pertinent de s'intéresser au produit  $\eta t_1$  pour adapter la vitesse de convergence à l'échelle des temps que l'on utilise. On n'a pas de moyen simple de déterminer un intervalle de variation pour ce paramètre. **On décide alors de tirer aléatoirement  $\eta t_1$  dans  $[0, 1; 10]$ .** Pour l'étude des paramètres  $M_0$  et  $M_1$ , on doit distinguer deux cas.

**Premier cas,  $M_0 > M_1$  :** Comme on a  $M(t) \leq M_0$ , le volume  $V(t)$  est majoré par le volume qu'on aurait si  $M$  était constante égale à  $M_0$ , on a donc une première inégalité :

$$V_0 e^{M_0 t} \geq V(t). \quad (3.16)$$

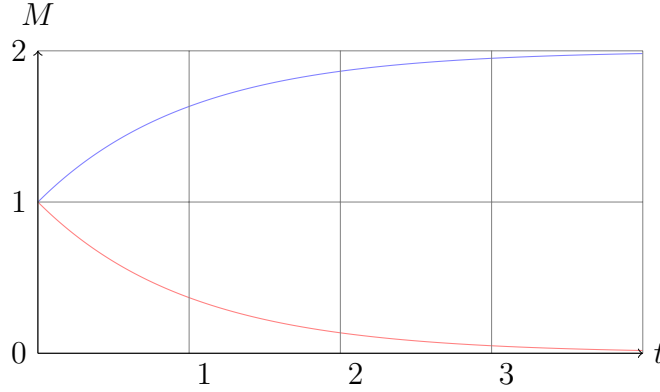


FIGURE 3.30 – Évolution de  $M$  en fonction du temps. Les paramètres sont, pour la courbe rouge :  $M_0 = 1$ ,  $M_1 = 0$  et  $\eta = 1$  (cas décroissant), pour la courbe bleue :  $M_0 = 1$ ,  $M_1 = 2$  et  $\eta = 1$  (cas croissant).

De même, l'expression (3.15) donne la valeur du volume si on fait l'hypothèse d'un champ  $M$  constant (égal à la valeur de  $M$  dans un voxel initialement inclus dans la tumeur). Pour un voxel quelconque d'indice  $(i, j)$  (dans le cas bidimensionnel) on peut écrire :

$$M_{i,j} = M_0 + M_1 \times (1 - e^{-\eta \times \max(t - \tilde{t}_{i,j}, 0)}),$$

où  $\tilde{t}_{i,j}$  correspond à l'instant à partir duquel le voxel  $(i, j)$  est atteint par la tumeur. Cela revient à retarder en temps l'équation (3.14) du délai de propagation des cellules tumorales jusqu'au voxel courant. On peut en déduire l'inégalité :

$$M(x, y, t) \geq (M_0 - M_1)e^{-\eta t} + M_1,$$

Ce qui nous donne, en intégrant en espace la densité de cellules tumorales :

$$V(t) \geq V_0 e^{(M_0 - M_1)e^{-\eta t} + M_1}. \quad (3.17)$$

À l'instant  $t_1$ , on a donc :

$$V_0 e^{M_0 t_1} \geq V_1 \geq V_0 e^{(M_0 - M_1)e^{-\eta t_1} + M_1}. \quad (3.18)$$

De la partie droite de l'inégalité précédente, on déduit :

$$M_1 \leq \frac{\frac{1}{t_1} \ln\left(\frac{V_1}{V_0}\right) - \alpha M_0}{1 - \alpha}, \quad (3.19)$$

avec

$$\alpha = \frac{1 - e^{-\eta t_1}}{\eta t_1}.$$

De plus, par hypothèse de modélisation,  $M$  étant assimilé à une densité de vascularisation des tissus, on a  $M_1 \geq 0$ . Ainsi, d'après l'équation Eq.(3.19),  $\frac{\frac{1}{t_1} \ln(\frac{V_1}{V_0}) - \alpha M_0}{1 - \alpha} \geq 0$ , ce qui nous donne une majoration sur  $M_0$  :

$$M_0 \leq \frac{1}{\alpha t_1} \ln\left(\frac{V_1}{V_0}\right),$$

ou encore une minoration sur  $\frac{1}{M_0}$ ,

$$\frac{1}{M_0} \geq \frac{\alpha t_1}{\ln\left(\frac{V_1}{V_0}\right)}. \quad (3.20)$$

Puis en utilisant l'inégalité de gauche de l'encadrement (3.18), on obtient une majoration sur  $\frac{1}{M_0}$  :

$$\frac{1}{M_0} \leq \frac{t_1}{\ln\left(\frac{V_1}{V_0}\right)}. \quad (3.21)$$

L'espace des paramètres dans lequel il est judicieux de calibrer le modèle est donc le suivant :

$$\begin{aligned} \eta t_1 &\in [0, 1; 10], \\ \frac{1}{M_0} &\in \left[ \frac{\alpha t_1}{\ln\left(\frac{V_1}{V_0}\right)}; \frac{t_1}{\ln\left(\frac{V_1}{V_0}\right)} \right], \\ M_1 &\in \left[ 0; \frac{\frac{1}{t_1} \ln\left(\frac{V_1}{V_0}\right) - \alpha M_0}{1 - \alpha} \right] \end{aligned}$$

**Second cas,  $M_0 \leq M_1$  :** On procède de même que pour le cas  $M_1 < M_0$ , la seule différence est que l'inégalité Eq. (3.18) est inversée. On a donc :

$$V_0 e^{M_0 t_1} \leq V_1 \leq V_0 e^{(M_0 - M_1) e^{-\eta t_1} + M_1}. \quad (3.22)$$

Notre espace de recherche est donc :

$$\eta t_1 \in [0, 1; 10],$$

$$M_0 \in \left[0, 1; \frac{1}{t_1} \ln\left(\frac{V_1}{V_0}\right)\right],$$

$$\frac{1}{M_1} \in \left[0, 1; \frac{1-\alpha}{\frac{1}{t_1} \ln\left(\frac{V_1}{V_0}\right) - \alpha M_0}\right].$$

L'espace dans lequel on calibre le modèle est donc dépendant des données scalaires dont nous disposons *a priori*, à savoir  $t_1$ ,  $V_0$  et  $V_1$ . Ainsi, et contrairement au cas du premier modèle, on détermine automatiquement cet espace pour chaque tumeur étudiée. De plus cet espace ainsi délimité contient à coup sûr les dynamiques de croissance qui nous intéressent et permet donc une fouille bien plus efficace de l'espace des paramètres qu'avec le modèle précédent.

Les calculs présentés au dessus correspondent au cas où l'on ne dispose que de deux images médicales. Si nous avons un nombre plus important  $k$  d'images de la tumeur, il est possible d'utiliser cette information supplémentaire pour affiner la calibration. En pratique, on procède au tirage des paramètres comme précédemment à partir de deux des examens. On prend le premier, qui nous sert de condition initiale et un second qui peut être par exemple l'examen le plus récent ou un des examens suivants tiré au sort. Les  $k - 2$  examens restants sont alors comparés au résultat de la simulation à l'instant correspondant via une fonction distance définie comme le maximum des erreurs relatives entre la simulation et l'observation. On fixe une valeur seuil pour cette distance, par exemple 15% et on élimine les simulations pour lesquelles la distance dépasse le seuil.

## 3.6 Résultats cliniques

Cette partie regroupe certains résultats obtenus sur des cas cliniques.

### 3.6.1 Protocole d'étude des cas cliniques

**Principe général :** La connaissance de bornes précises et adaptées au patient pour l'intervalle de recherche des différents paramètres (voir section précédente) est un atout majeur pour la calibration du modèle sur des données cliniques. Ce point nous a permis de modifier la méthode de prédiction à partir d'images médicales et d'envisager une approche différente.

L'algorithme de prédiction se base sur un algorithme de type Monte Carlo. On tire au sort un jeu de paramètres dans l'espace déterminé à la section précédente. Il s'agit donc d'un tirage non indépendant puisque les bornes des intervalles dépendent des données mais aussi des valeurs d'autres paramètres. On simule le modèle pour le jeu de paramètres ainsi obtenu et on calcule la distance de cette simulation avec les données cliniques au sens de la fonction objectif suivante qui tient compte de l'erreur spatiale L2 ainsi que du volume :

$$f_{obj}(p_1, \dots, p_k) = \sum_{i=1}^{ns} \left[ \sqrt{\frac{\int_{\Omega} (T_{snap}(t_i) - T_{modele}(p_1, \dots, p_k, t_i))^2 d\omega}{\int_{\Omega} (T_{snap}(t_i))^2 d\omega}} \right] \quad (3.23)$$

$$+ a \times \sum_{i=1}^{ns} \left[ \left| \frac{\text{vol}(T_{snap}(t_i)) - \text{vol}(T_{modele}(p_1, \dots, p_k, t_i))}{\text{vol}(T_{snap}(t_i))} \right| \right].$$

En pratique, le coefficient  $a$  est pris suffisamment petit (0,1 par exemple) pour ne pas trop influencer le choix des jeux de paramètres puisqu'on veut tenir compte de la donnée spatiale. On conserve la simulation si la valeur de la fonction objectif est en dessous d'un certain seuil d'acceptation qui sera explicité plus tard. On fait ainsi  $N$  tirages de jeux de paramètres ce qui nous permet de conserver un certain nombre  $n$  de simulations acceptables du point de vue de la fonction objectif que l'on a définie. **On ne cherche donc plus à conserver la simulation qui représente au mieux les données mais un échantillon de simulations plus ou moins proches des données.** Cela se base sur l'idée que, connaissant l'incertitude sur les données elles mêmes (erreur de segmentation par exemple), il faut relativiser l'importance qu'on leur accorde dans le processus de calibration. Ainsi, il se peut qu'en acceptant une erreur un peu plus importante lors de la calibration, on améliore la prédiction de l'évolution de la tumeur. La procédure est résumée dans le diagramme en Figure 3.31.

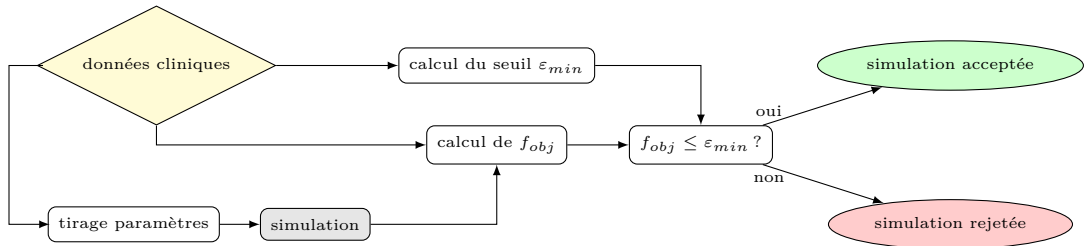


FIGURE 3.31 – Diagramme synthétique décrivant le processus de calibration.

On n'a plus alors (comme c'était le cas au Chapitre 1) une unique courbe

d'évolution, mais un faisceau de courbes qui balaie les comportements possibles de la tumeur selon le modèle.

**Calcul du seuil d'acceptation :** L'ordre de grandeur de l'erreur acceptable pour la simulation par rapport aux données dépend de la qualité des données elles-mêmes. Des hétérogénéités non prises en compte dans le modèle ainsi que les incertitudes de segmentation peuvent faire évoluer la forme de la tumeur d'une façon non exactement prédictible par le modèle. Si cette forme varie beaucoup entre deux examens on va accepter une erreur plus importante que si les formes sont régulières. Le seuil d'acceptation des simulations ne peut donc pas être une valeur fixée une fois pour toutes et doit dépendre de la tumeur étudiée et donc des données dont nous disposons. On le détermine empiriquement pour chaque cas comme étant l'erreur minimale que l'on peut espérer obtenir, à laquelle on ajoute une marge (10% par exemple). Cette erreur minimale est obtenue comme suit, les paramètres  $p_1, \dots, p_k$  étant choisis comme précédemment :

$$\varepsilon_{min} = \sum_{i=1}^{ns} \min_{t \in [0; T_{max}]} \sqrt{\frac{\int_{\Omega} (T_{snap}(t_i) - T_{modele}(p_1, \dots, p_k, t))^2 d\omega}{\int_{\Omega} (T_{snap}(t_i))^2 d\omega}}.$$

**Sorties de la méthode :** Les données brutes recueillies en sortie de l'algorithme sont principalement l'évolution de la surface (ou masse) tumorale ainsi que la répartition spatiale des cellules tumorales à l'instant de prédiction pour chacune des simulations conservées. À partir de ces données, on peut tracer le faisceau de courbes d'évolution de la surface et étudier la répartition des prédictions c'est à dire de la surface obtenue à l'instant de prédiction  $T_{pred}$ . On peut ainsi calculer plusieurs indicateurs statistiques tels que la prévision moyenne (espérance), la prévision médiane, la surface la plus probable (maximum de la fonction de répartition) mais aussi l'écart-type ou encore les intervalles inter-quartiles qui donnent une idée sur la dispersion des résultats. Les données spatiales (images de la tumeur simulée) nous donnent quant à elles la probabilité, selon le modèle, que chaque voxel soit atteint par la tumeur à l'instant de prédiction.

### 3.6.2 Premiers résultats sur un cas clinique

#### Présentation du cas

Le cas clinique étudié dans cette partie est présenté Fig. 3.32. On peut y observer l'évolution temporelle d'une métastase pulmonaire non traitée au travers de cinq extraits de CT-scans régulièrement répartis sur une période

de 21 mois. On observe que la croissance de cette tumeur est assez rapide et régulière. La lésion atteint le bord du poumon, ce qui semble jouer sur la façon dont elle croît. Cependant, on décide de négliger ces effets dans cette étude.

### Application de la méthode de prévision

On dispose donc dans ce cas de cinq examens pour tester la méthode. On va tout d'abord utiliser les deux premiers examens pour établir une prévision à la date correspondant au troisième. On pourra ainsi comparer les résultats obtenus à l'évolution réelle de la métastase, puis on utilisera les trois premiers examens pour tenter de prédire le quatrième et ainsi de suite. Les figures présentées ici ont été obtenues en calibrant le modèle sur la zone de paramètres correspondant à  $M_0 \leq M_1$ .

**Utilisation de deux examens.** Dans ce paragraphe, on utilise les deux premiers examens pour tenter de prédire l'état de la tumeur à la date du troisième. La première image (au 14/09/2011) sert à initialiser le champ de densité de cellules tumorales. La deuxième image (au 18/01/2012) est utilisée pour le calcul de la fonction objectif  $f_{obj}$  et donc pour calibrer le modèle. On obtient le faisceau de courbes suivant (Fig. 3.33). On observe donc que le troisième point, celui que l'on souhaite prédire, se trouve bien dans le faisceau. Cependant, les données semblent assez dispersées. On peut tracer (Fig. 3.34) un histogramme de répartition des volumes (ou surfaces dans le cas 2D) tumoraux issus des simulations à l'instant de prédiction.

moyenne	écart-type	densité maximum	1 <sup>er</sup> quartile	médiane	3 <sup>ème</sup> quartile
39,10	9,57	33,71	31,98	36,14	44,08

TABLE 3.2 – Données statistiques brutes de la série de prédictions en mm<sup>2</sup>. La surface réelle au troisième examen est de 31,06 mm<sup>2</sup>.

On peut normaliser l'écart-type par la moyenne afin d'obtenir un indicateur de dispersion des données relatif et sans unité. On notera cet écart-type normalisé  $\tilde{\sigma} = \frac{\sigma}{m}$ , où  $\sigma$  est l'écart-type classique et  $m$  la moyenne. Dans ce cas, on a  $\tilde{\sigma} = 0,245$ .

La prédiction est donc bonne malgré un écart-type relatif de presque 25%. On voit notamment que la valeur cible correspondant à la valeur réelle appartient à l'intervalle de confiance  $[m - \sigma; m + \sigma]$ . Si l'on regarde les différents indicateurs statistiques, on constate que la moyenne sous-estime la surface de la tumeur d'environ 8mm<sup>2</sup>, le maximum de la densité de répartition d'un peu



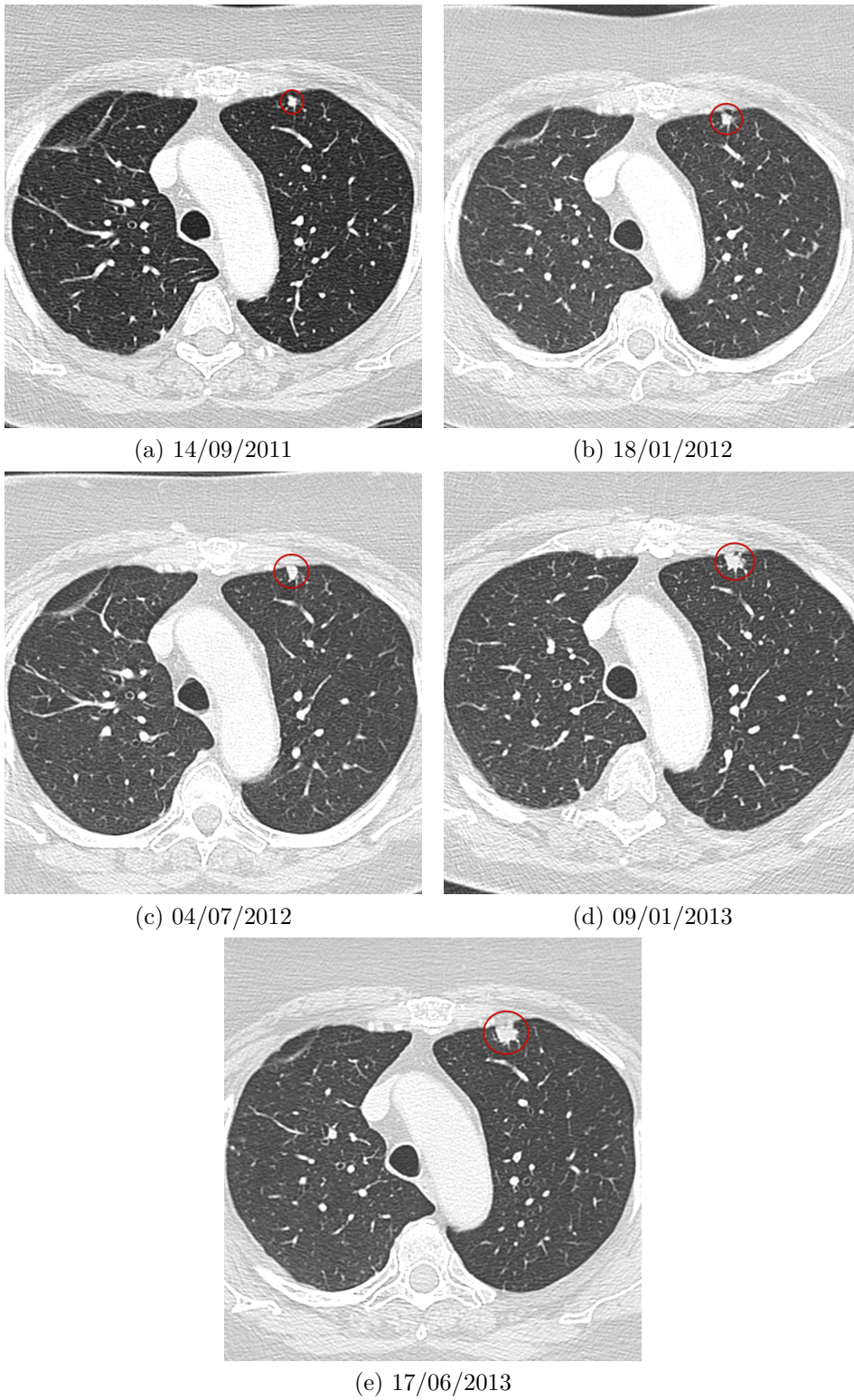


FIGURE 3.32 – Séquence de CT-scans montrant l'évolution d'une métastase.

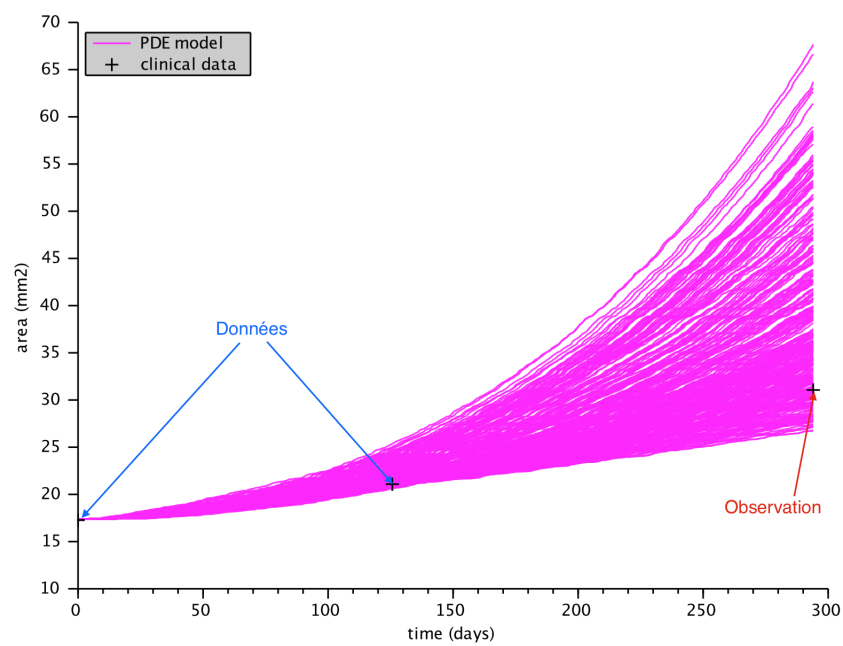


FIGURE 3.33 – Premier cas test : Faisceau de courbes obtenu avec deux données.

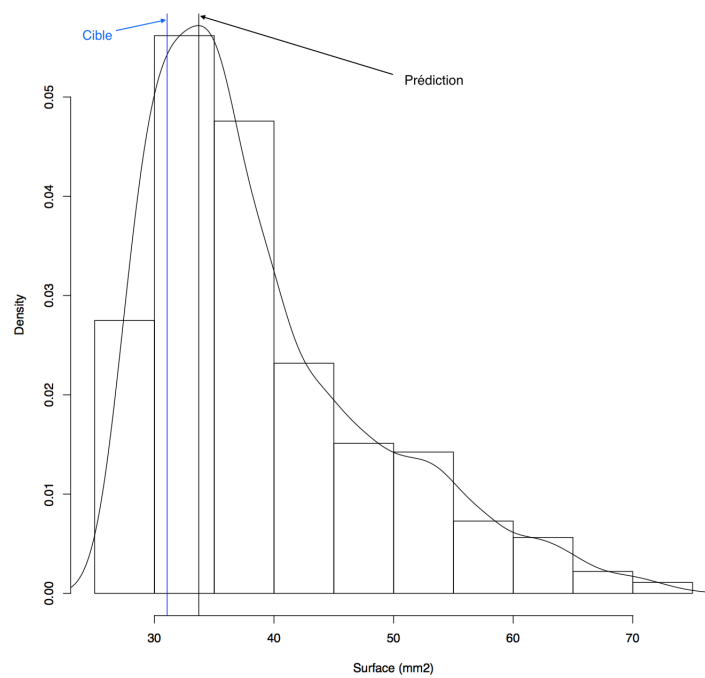


FIGURE 3.34 – Répartition des prédictions sur le premier cas test en utilisant deux points. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite noire correspond au maximum de la densité de répartition.

plus de  $2,5\text{mm}^2$  et la médiane d'environ  $5\text{mm}^2$ . Ces valeurs semblent tout à fait acceptables compte tenu des erreurs de segmentation commises à cette échelle.

**Utilisation de trois examens.** On utilise maintenant les trois premiers examens pour calibrer le modèle, cherchant ainsi à prédire le volume de la tumeur à la date du quatrième examen. Le faisceau de courbes ainsi obtenu est présenté Figure 3.35.

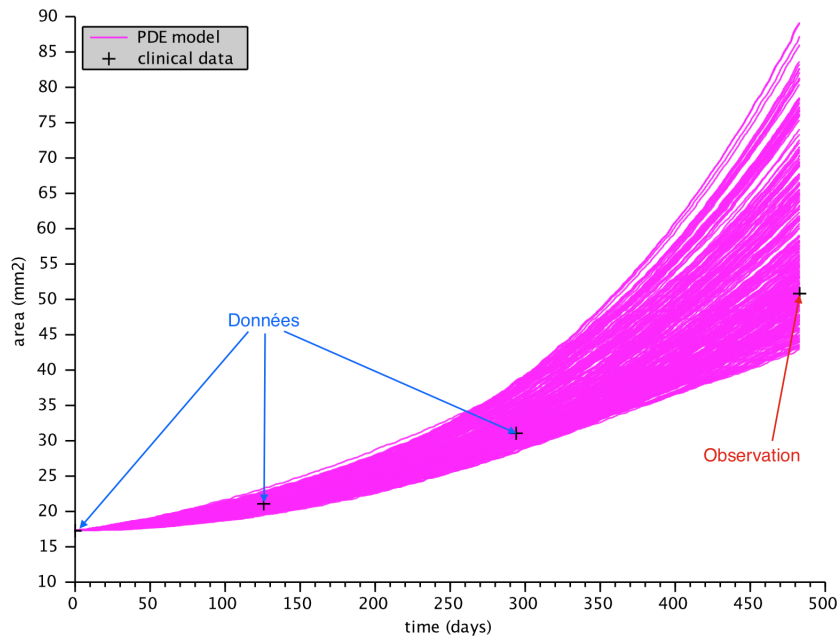


FIGURE 3.35 – Premier cas test : Faisceau de courbes obtenu avec trois données.

Ici encore, la valeur cible est dans le faisceau. On remarque de même que le faisceau semble plus étroit que dans le cas précédent. Cela vient du fait qu'en ajoutant un examen, et donc de l'information, dans le processus de calibration, on discrimine davantage lors du tri des simulations. La répartition des prédictions ainsi que les données statistiques correspondantes sont données dans les tableaux 3.36 et 3.3.

Ici encore, les différents indicateurs donnent une bonne prédiction de la surface tumorale. De plus, l'écart-type normalisé vaut dans ce cas 0,184 et est donc inférieur au cas précédent. Cela confirme que la prédiction s'est affinée avec l'ajout d'informations.

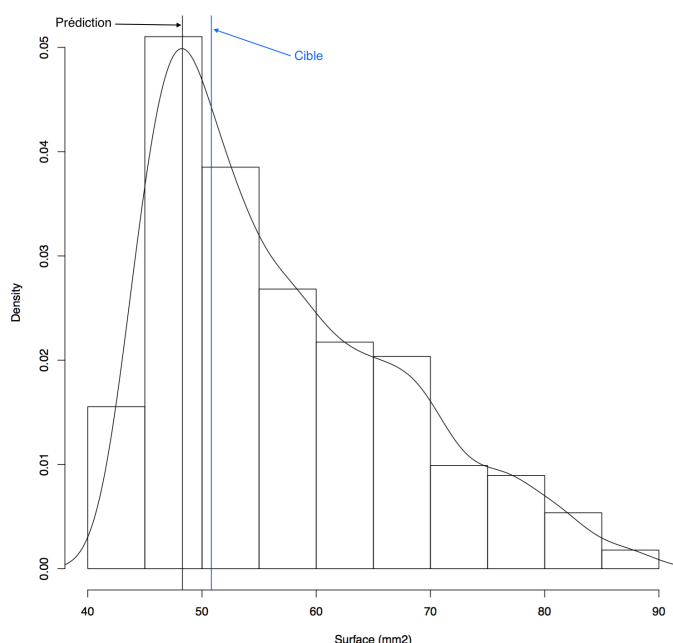


FIGURE 3.36 – Répartition des prédictions sur le premier cas test en utilisant trois examens. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite bleue correspond au maximum de la densité de répartition.

**Utilisation de quatre examens.** Pour finir, on utilise les quatre premiers examens pour prédire la taille de la tumeur au cinquième. Le faisceau de courbes correspondant est donné en Figure 3.37, la répartition des prédictions en Figure 3.38 et les indicateurs statistiques dans la Table 3.4.

Les prédictions semblent qualitativement correctes puisque, une fois encore, la valeur cible se trouve dans le faisceau qui semble s'être encore affiné par rapport au cas précédent.

Les indicateurs statistiques donnent là encore une bonne prédiction et l'écart-type normalisé vaut 0,138 et a donc encore diminué.

On voit ainsi que, comme on aurait pu s'y attendre, plus on utilise de

moyenne	écart-type	densité maximum	1 <sup>er</sup> quartile	médiane	3 <sup>ème</sup> quartile
57,04	10,50	48,28	48,44	54,14	64,15

TABLE 3.3 – Données statistiques brutes de la série de prédictions en mm<sup>2</sup>. La surface réelle au quatrième examen est de 50,82 mm<sup>2</sup>.

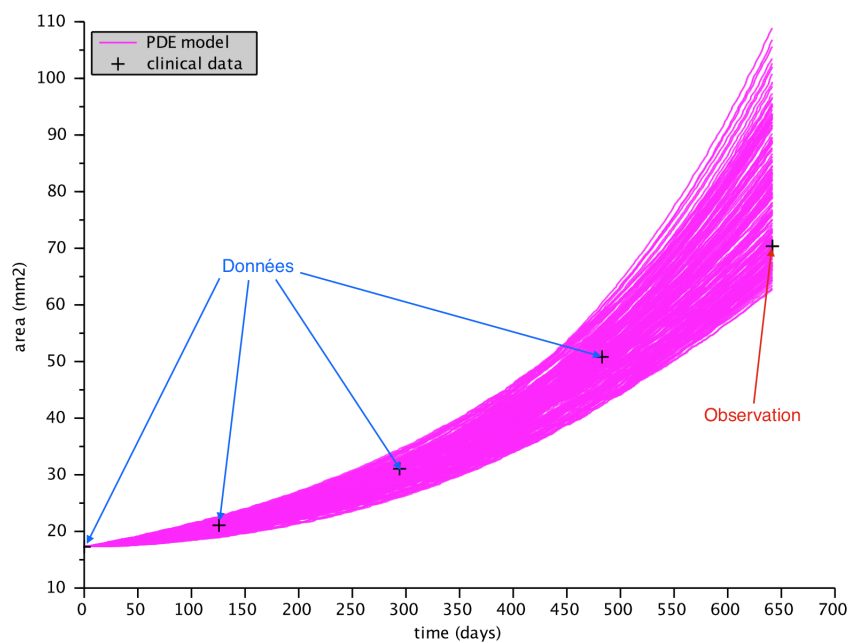


FIGURE 3.37 – Premier cas test : Faisceau de courbes obtenu avec quatre données.

moyenne	écart-type	densité maximum	1 <sup>er</sup> quartile	médiane	3 <sup>ème</sup> quartile
78,24	10,83	68,80	69,01	75,84	85,86

TABLE 3.4 – Données statistiques brutes de la série de prédictions en mm<sup>2</sup>. La surface réelle au cinquième examen est de 70,36 mm<sup>2</sup>.

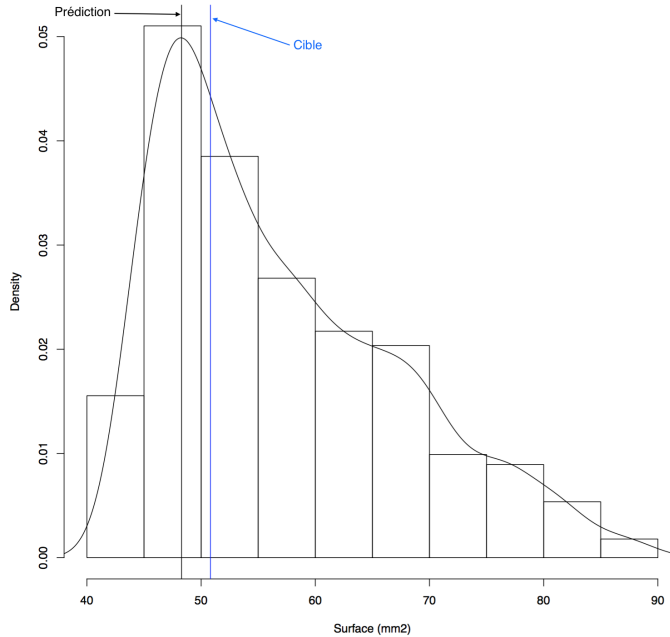


FIGURE 3.38 – Répartition des prédictions sur le premier cas test en utilisant quatre examens. La droite bleue indique la valeur réelle de la surface au troisième examen (valeur cible à prédire) et la droite bleue correspond au maximum de la densité de répartition.

données, plus la prédiction est précise au sens de la diminution de la taille de l'intervalle de confiance. De plus, la méthode donne sur ce cas des résultats qualitativement et quantitativement satisfaisants.

Pour une analyse à plus grande échelle de l'utilisation de la méthode sur des cas cliniques, on peut se reporter au chapitre 4.

### 3.7 Méthode de calibration scalaire

Nous avons vu précédemment une méthode de calibration basée sur la minimisation d'une fonction distance permettant de comparer la forme et la taille des tumeurs simulées aux données cliniques. Cette méthode nécessite comme on l'a vu de simuler un grand nombre de fois le modèle EDP. Cela permet d'axer la personnalisation du modèle sur la forme (erreur  $L^2$ ). Cependant, la méthode ne fournit pas de résultats en temps réel puisque, même en répartissant la charge de calcul sur huit coeurs, il faut compter environ une heure et demi de temps de calcul pour simuler le modèle sur 5 000 jeux de paramètres. Cette considération technique peut sembler peu importante dans un premier

temps mais, dans l'optique de l'utilisation par des médecins d'un outil basé sur cette méthode, cela pourrait être un inconvénient majeur. Afin de pallier cette difficulté et de diminuer les temps de calcul, nous avons mis au point une méthode de calibration « légère » basée sur l'utilisation de la formule donnant le volume tumoral (Eq. (3.13)) obtenue à la section 3.4. Cette formule est rappelée ci-dessous :

$$V(t) = \frac{V_0 e^{r_1 t}}{\alpha - 1} (\alpha e^{-\sqrt{\Delta} t} - 1).$$

L'avantage de cette formule est qu'elle nous permet de connaître le volume tumoral en tout temps sans passer par le système EDP et donc de réduire considérablement le nombre de calculs. Pour cette variante de la méthode de calibration, on conserve la procédure de génération des jeux de paramètres utilisée précédemment et décrite à la section 3.5. Cependant la fonction distance utilisée ne compare plus que les volumes aux instants des examens entre les observations et le résultat fourni par le modèle au travers de la formule scalaire précédente. Ainsi, il n'est plus nécessaire de simuler le système EDP pour la phase de calibration et les gains en terme de temps de calcul sont conséquents. En pratique, on teste 100 000 jeux de paramètres en moins d'une seconde. La fonction distance correspond à un critère des moindres carrés relatif sur les volumes aux instants des examens (voir Eq. (3.24)).

$$f_d = \sum_{i=1}^{n_s} \frac{(V_{obs}(t_i) - V_{simu}(t_i))^2}{V_{obs}^2(t_i)}, \quad (3.24)$$

où  $V_{obs}(t_i)$  est le volume observé de la tumeur réelle à l'instant  $t_i$ , c'est à dire mesurée sur le  $i^{ème}$  examen et  $V_{simu}(t_i)$  est le volume calculé avec la formule (3.13) au même instant. On se fixe toujours un seuil d'acceptation sur la valeur de cette fonction distance, par exemple 15%, et on ne conserve que les jeux de paramètres avec lesquels on se situe en dessous de ce seuil.

On obtient donc, en sortie de l'algorithme, une liste de jeux de paramètres correspondant à différentes dynamiques de croissance tumorale admissibles. On classe ces différents jeux de paramètres en fonction du volume final de la tumeur obtenu. En pratique on récupère une liste ordonnée de la forme  $(\pi_1; V_1(t_f), \dots, (\pi_i; V_i(t_f)), \dots, (\pi_N; V_N(t_f)))$  où les  $\pi_i$  sont les jeux de paramètres sélectionnés, les  $V_i(t_f)$  les volumes finaux correspondant et  $N$  le nombre de jeux de paramètres ayant satisfait le critère de sélection. On peut ordonner la liste précédente dans l'ordre croissant :  $V_1(t_f) \leq \dots \leq V_i(t_f) \leq \dots \leq V_N(t_f)$ . On identifie ensuite certains jeux de paramètres particuliers :

- **le comportement médian**, c'est à dire le jeu  $\pi_m$  tel que le volume final correspondant  $V_m(t_f)$  représente la médiane de l'ensemble des volumes  $(V_i(t_f))_{1 \leq i \leq N}$ ,



- les deux jeux  $\pi_{q5}$  et  $\pi_{q95}$  associés aux volumes  $V_{q5}(t_f)$  et  $V_{q95}(t_f)$  correspondant aux **premier et dernier 20-quantiles**, c'est à dire que 90% des volumes finaux sont compris entre ces deux valeurs,
- **le comportement le plus proche de la dynamique moyenne**, c'est à dire le jeu  $\pi_M$  tel que  $V_M(t_f)$  est, parmi l'ensemble des  $(V_i(t_f))_{1 \leq i \leq N}$  celui le plus proche de la moyenne  $\bar{V}_f = \frac{1}{N} \sum_{i=1}^N V_i(t_f)$ .

Ces jeux de paramètres particuliers sont alors utilisés dans le modèle EDP pour effectuer des simulations. On obtient ainsi la répartition spatiale des cellules cancéreuses.

Une illustration du résultat de la méthode de calibration scalaire est donnée en Figure 3.39. Il s'agit du même cas que celui traité dans la Figure 3.33 et on constate que les prédictions fournies sont très proches avec les deux méthodes ce qui confirme l'intérêt de la méthode scalaire. On peut aussi visualiser la

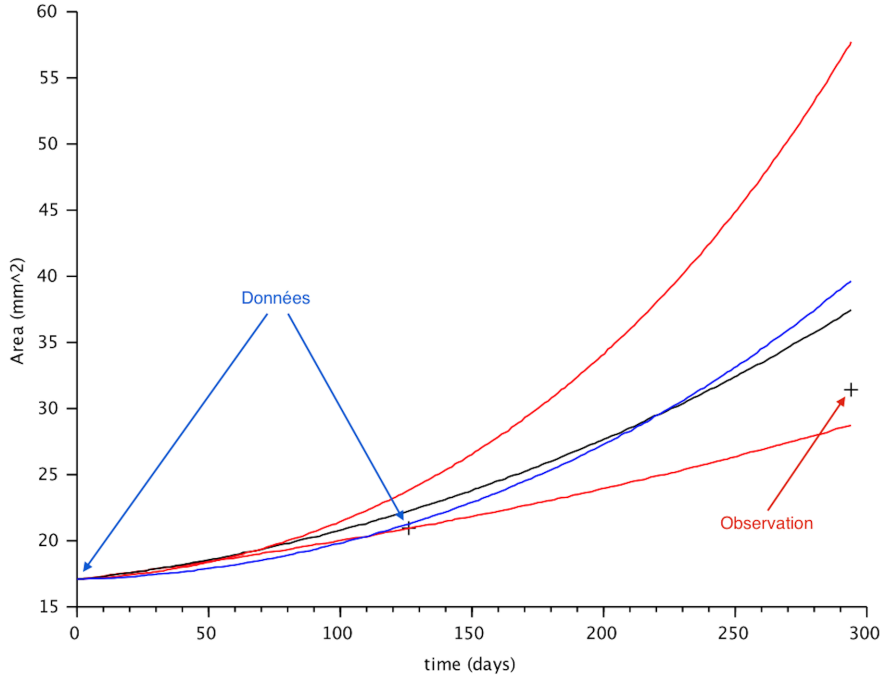


FIGURE 3.39 – Utilisation de la méthode de calibration basée sur la formule scalaire du volume tumoral sur le cas de la Figure 3.33 : en rouge les 5<sup>ème</sup> et 95<sup>ème</sup> centiles (90% des simulations se situent entre ces deux courbes), en bleu la simulation moyenne et en noir la médiane.

prédiction de croissance tumorale en superposant les simulations aux images tirées des examens (voir Figure 3.40).

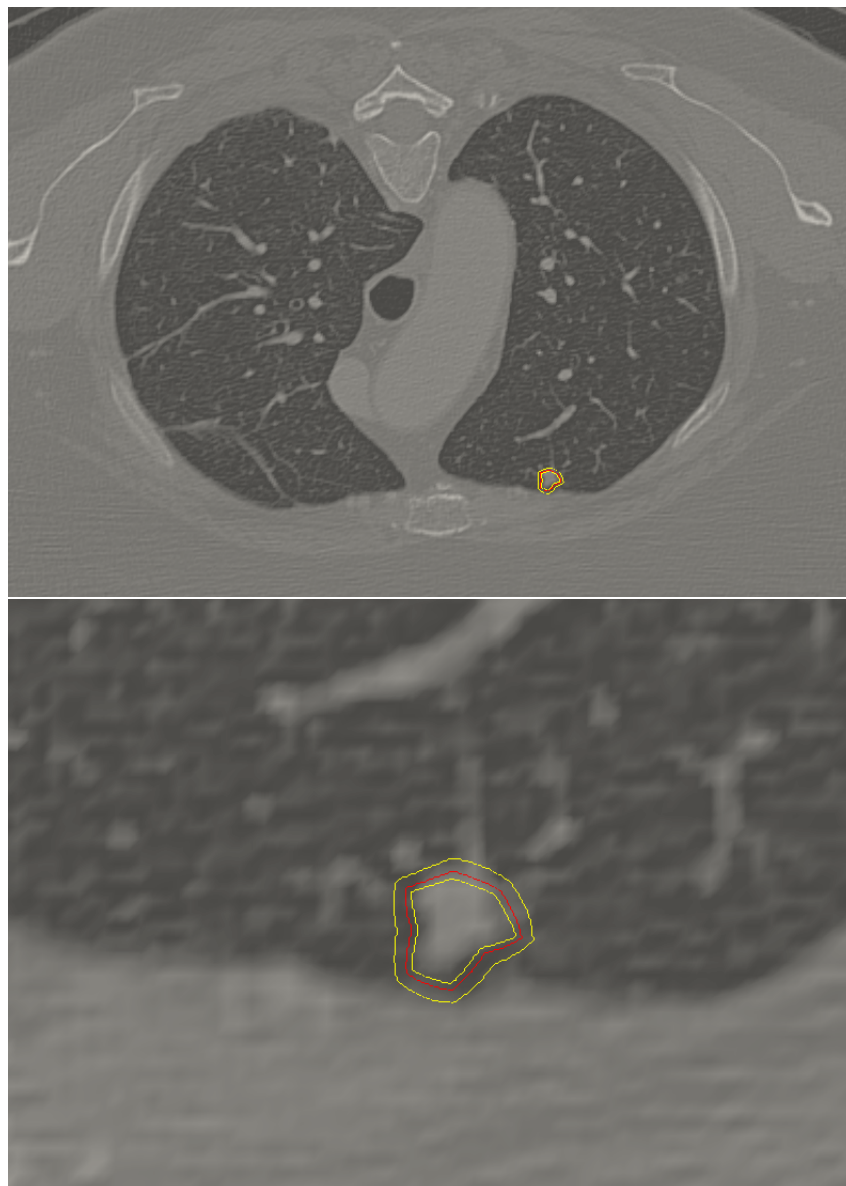


FIGURE 3.40 – Rendu spatial de la méthode de calibration scalaire : superposition des simulations sur l'image initiale de la tumeur. On retrouve à gauche l'image du poumon et à droite un zoom sur la tumeur. Le contour en rouge représente la simulation médiane et les contours jaunes délimitent la zone comprenant 90% des simulations.

Comme nous l'avons vu, le gain de temps est conséquent et permet d'obtenir des résultats dans des délais tout à fait raisonnables pour une utilisation clinique (environ 30 secondes pour la calibration et la simulation EDP des jeux de paramètres particuliers). L'inconvénient de cette approche est qu'on ne tient plus compte des aspects spatiaux pour calibrer le modèle.

### 3.8 Étude de la propagation des incertitudes de segmentation

L'une des principales difficultés inhérente aux simulations de croissances tumorales basées sur l'imagerie médicale provient du contourage des tumeurs sur les images. En effet, cette étape est hautement dépendante de l'opérateur. On estime que la variabilité inter-opérateurs peut engendrer des écarts de 20% sur la surface de la tumeur contourée. Étant donnée cette incertitude, il est important de savoir de quelle manière elle se propage dans le modèle. En effet si le modèle est peu robuste et très sensible à de petits écarts de forme, la qualité des prédictions en sera affectée.

#### 3.8.1 Introduction de perturbations de forme

Pour simuler cette incertitude, on se place dans le cas 2D d'une tumeur de forme circulaire, la frontière de celle-ci est matérialisée par  $n = 20$  points. On perturbe indépendamment les coordonnées des points en utilisant une erreur gaussienne. On peut écrire les coordonnées non perturbées des points sous forme d'un vecteur  $\mathbf{P}$  :

$$\mathbf{P} = \begin{pmatrix} (x_1; y_1) \\ \vdots \\ (x_i; y_i) \\ \vdots \\ (x_n; y_n) \end{pmatrix}.$$

Puis le contourage perturbé, noté  $\tilde{\mathbf{P}}$  est donné par :

$$\tilde{\mathbf{P}} = \begin{pmatrix} (\tilde{x}_1; \tilde{y}_1) \\ \vdots \\ (\tilde{x}_i; \tilde{y}_i) \\ \vdots \\ (\tilde{x}_n; \tilde{y}_n) \end{pmatrix},$$

où les  $\tilde{x}_i, \tilde{y}_i$  sont des réalisations des variables aléatoires gaussiennes  $\tilde{X}_i, \tilde{Y}_i$ , d'écart-type  $\sigma$  (qui ne dépend donc pas du point) et centrées sur les coordonnées non perturbées  $x_i, y_i : \forall i \in \{1; \dots; n\}$ ,

$$\tilde{X}_i \sim \mathcal{N}(x_i, \sigma),$$

$$\tilde{Y}_i \sim \mathcal{N}(y_i, \sigma).$$

Une fois obtenue la segmentation perturbée, on peut lancer une simulation du modèle. La figure ci-dessous (Fig. 3.41) compare les simulations avec ou sans perturbation.

### 3.8.2 Approche statistique

Afin de pouvoir tirer des conclusions statistiques quant à la propagation de l'incertitude de contourage dans le modèle, on effectue un grand nombre de simulations en perturbant indépendamment les coordonnées des points. Ce qui nous intéresse ici est de déterminer la fonction de répartition de l'erreur sur la forme après simulation et de la comparer à la fonction de répartition de l'erreur sur la segmentation initiale. Cette étude sera menée sur un point en particulier de la frontière  $(x_1; y_1)$  afin de vérifier que l'erreur reste centrée et symétrique par rapport à 0, puis sur l'ensemble des points.

On ne s'intéresse qu'aux abscisses des points  $x_i$ , le problème étant invariant par rotation. Le jeu de paramètres utilisé pour cette étude est donné dans la Table 3.5.

$M_0$	1,0
$M_1$	1,5
$\eta$	1,5

TABLE 3.5 – Jeu de paramètres utilisé pour l'étude de propagation des incertitudes de segmentation.

**Propagation de l'incertitude sur  $x_1$  :** On étudie et compare les variables aléatoires  $\Xi_1^0 = \frac{\tilde{X}_1(t=0) - x_1(t=0)}{r(t=0)}$  et  $\Xi_1^2 = \frac{\tilde{X}_1(t=2) - x_1(t=2)}{r(t=2)}$ , où  $r(t)$  représente le rayon moyen de la tumeur au temps  $t$ . Cela revient donc à calculer l'écart de position dû à la perturbation initiale. On divise par le volume pour ne pas tenir compte des changements d'échelles liés à la croissance de la tumeur et ne comparer que les formes. On simule un échantillon de 1000 réalisations de ces variables aléatoires et on étudie les fonctions de répartition de celles-ci, les résultats sont en Figures 3.42 et 3.43.

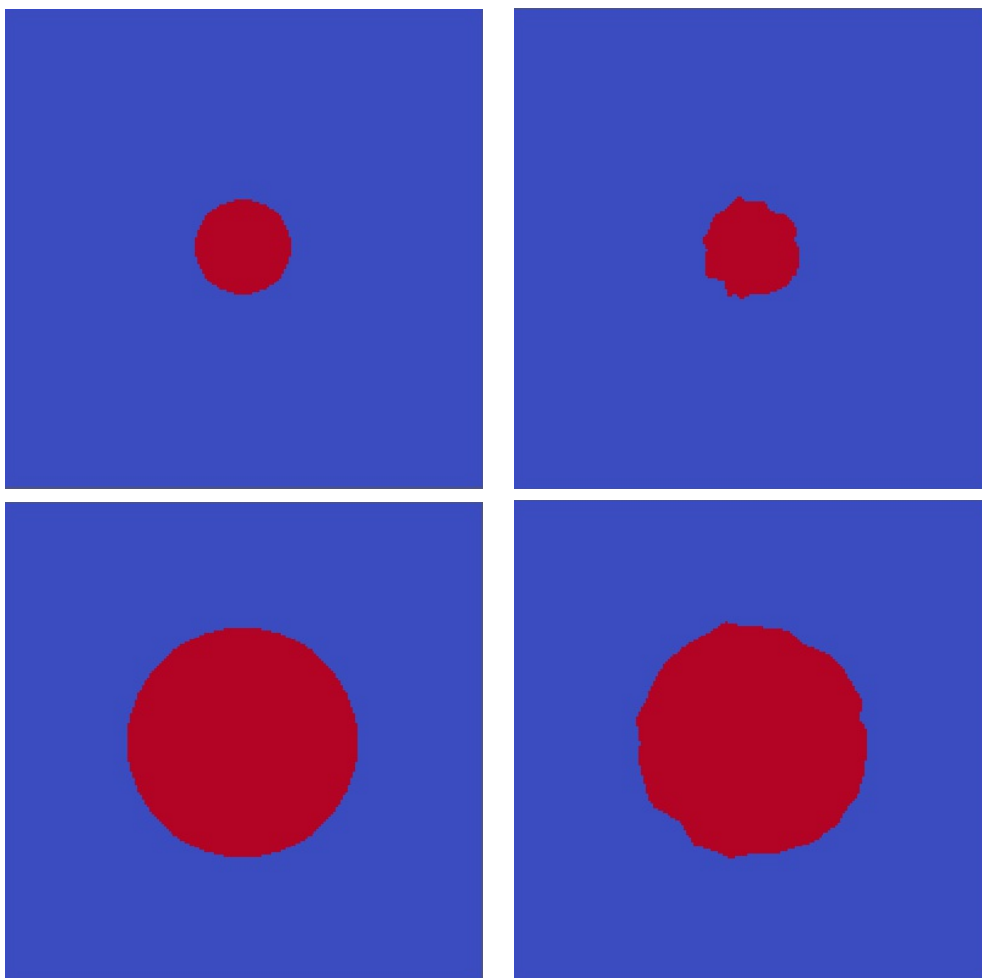


FIGURE 3.41 – Comparaison des simulations du modèle avec forme perturbée (à droite) et non perturbée (à gauche). La ligne du haut correspond aux conditions initiales ( $t=0$ ) et celle du bas aux résultats de simulation à  $t=2$ . Le rayon initial est de 0,1cm et l'écart-type est  $\sigma = 0,007$ cm.

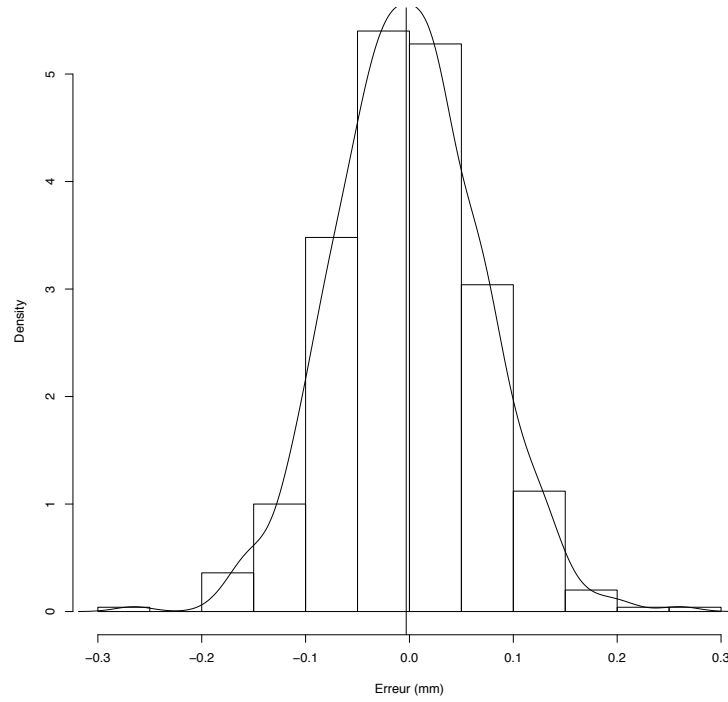


FIGURE 3.42 – Répartition des réalisations de la variable aléatoire  $\Xi_1^0$ . Les estimateurs de la moyenne et de l'écart-type sont :  $\bar{m} = -1,5 \times 10^{-3}$ ,  $\bar{\sigma} = 6,9 \times 10^{-2}$ .

**Propagation de l'incertitude globale :** Cette fois, on tient compte des erreurs cumulées sur l'ensemble des points de discrétisation de la frontière de la tumeur.

Les moyennes et écarts-types correspondant aux données présentées Figure 3.44 sont donnés Table 3.6.

L'analyse de ces différentes données statistiques issues de simulations nous permet de constater que le modèle utilisé a un effet régularisant sur les formes puisque la moyenne et l'écart-type des écarts relatifs diminuent au cours des simulations. Les incertitudes de segmentation ne vont donc pas se trouver amplifiées par le modèle de croissance tumorale.

## 3.9 Conclusion

Au cours de ce chapitre, nous avons mis au point un modèle simplifié pour la croissance des métastases pulmonaires à seulement trois paramètres. Cela nous a permis de repenser intégralement l'algorithme de calibration et de réduire la complexité des calculs, la

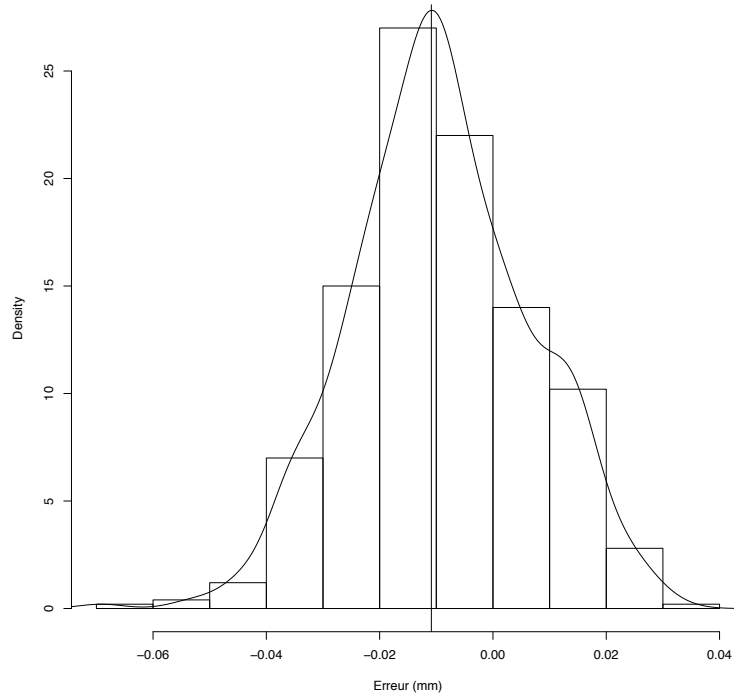
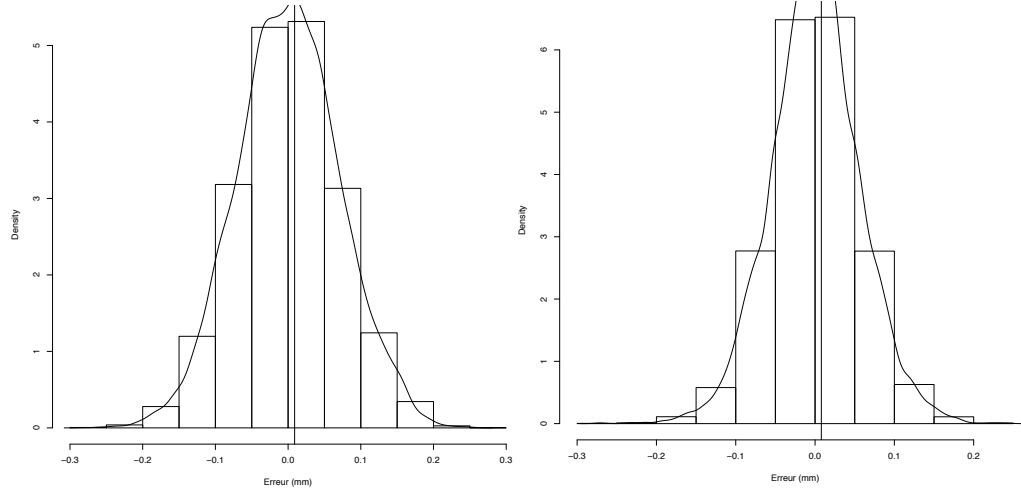
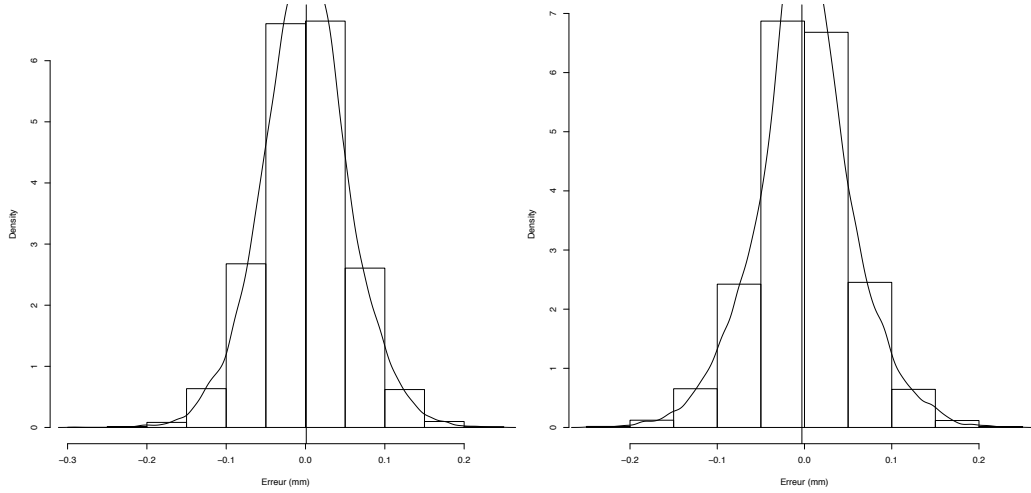


FIGURE 3.43 – Répartition des réalisations de la variable aléatoire  $\Xi_1^2$ . Les estimateurs de la moyenne et de l'écart-type sont :  $\bar{m} = -9,2 \times 10^{-3}$ ,  $\bar{\sigma} = 1,5 \times 10^{-2}$ .

calibration se faisant désormais quasiment en temps réel. D'un point de vue numérique, nous avons mis au point une méthode basée sur le suivi d'interface adaptée à notre problème pour les cas 2D et 3D. Les premiers résultats sont prometteurs et une étude plus approfondie sur de nombreux cas tests de l'ensemble de la méthode est présentée au chapitre suivant. L'intérêt ici est le développement d'une méthode utilisable par les médecins. Le logiciel développé au sein de l'équipe pour l'implémentation de cette méthode a fait l'objet d'un dépôt auprès de l'Agence de la Protection des Programmes. De plus des voies de valorisation et le transfert de ces travaux de recherche sont à l'étude.



(a)  $t = 0$  : répartition des réalisations de  $\Xi^0$  (b)  $t = 1$  : répartition des réalisations de  $\Xi^1$



(c)  $t = 1,5$  : répartition des réalisations de  $\Xi^{1,5}$  (d)  $t = 2$  : répartition des réalisations de  $\Xi^2$

FIGURE 3.44 – Propagation de l'incertitude de contourage sur l'ensemble des points : à l'instant initial, puis aux instants  $t = 1$ ,  $t = 1,5$  et  $t = 2$ .



	moyenne	écart-type
$t = 0$	$4,6 \times 10^{-4}$	$7,0 \times 10^{-2}$
$t = 1$	$2,9 \times 10^{-4}$	$5,6 \times 10^{-2}$
$t = 1.5$	$-1,6 \times 10^{-4}$	$5,5 \times 10^{-2}$
$t = 2$	$1,0 \times 10^{-4}$	$5,5 \times 10^{-2}$

TABLE 3.6 – Moyenne et écart-type de l'écart à la position non perturbée sur l'ensemble des points à l'instant initial et aux instants  $t = 1$ ,  $t = 1,5$  et  $t = 2$ .

# Chapitre 4

## Validation clinique

Afin de tester la méthode décrite au chapitre précédent, celle-ci a été utilisée sur 37 cas cliniques de métastases pulmonaires. Le but de cette étude est de confronter le modèle et la méthode de calibration à la réalité clinique afin d'évaluer ses qualités prédictives. Cette partie décrit la démarche mise en place pour ces essais et détaille les résultats obtenus.

### 4.1 Protocole d'étude

#### 4.1.1 Données cliniques utilisées

Grâce à une étroite collaboration avec le CHU Pellegrin et l'Institut Bergonié, centre régional de lutte contre le cancer situé à Bordeaux, il a été possible d'avoir accès à un nombre important de cas cliniques atteints de métastases pulmonaires. Nous avons restreint l'étude à la calibration du modèle à partir de seulement deux données cliniques. Il est bien sûr possible d'utiliser davantage de données pour personnaliser le modèle et nous avons pu voir que cela permet d'affiner les prédictions. Cependant, souvent nous disposons d'assez peu de données et il a donc semblé important de tester la méthode dans un cadre réaliste du point de vue clinique et donc d'utiliser le moins de données possible. Les cas auxquels nous nous sommes intéressés présentent donc au moins trois examens à des dates différentes et sans traitement. Cela nous permet d'utiliser les deux premières images pour calibrer le modèle et de comparer et évaluer ensuite la prédiction avec la troisième image.

Lorsque plus de trois images de la même lésion sont disponibles, nous avons parfois dissocié ces données en plusieurs séries traitées indépendamment. Cela nous permet d'avoir plus de cas pour tester la méthode. Ainsi, si on dispose de quatre examens aux dates  $t_1$ ,  $t_2$ ,  $t_3$  et  $t_4$ , on peut établir une prédiction au

temps  $t_3$  à partir des examens aux temps  $t_1$  et  $t_2$  d'une part, et une prédiction au temps  $t_4$  à partir des examens des temps  $t_2$  et  $t_3$ . Le pré-traitement des données a été effectué avec l'aide précieuse des cliniciens, au moins pour le choix des lésions et des coupes à segmenter.

### 4.1.2 Méthodologie

Pour cette étude, nous avons utilisé la méthode « allégée » décrite au chapitre précédent, dans la section 3.7. La méthode peut permettre de prédire le temps auquel est atteint une masse donnée ou, à l'inverse, de fournir une prédiction sur la masse (ou la surface) à une date donnée. Ces deux formulations du problème étant assez proches, on se limite ici à la prédiction de la surface occupée par la tumeur dans la couche choisie.

Pour chacun des cas, nous avons calibré le modèle sur deux gammes de paramètres distinctes (voir section 3.5). La première correspond aux cas où  $M_1 \leq M_0$ , c'est à dire que le taux de croissance de la tumeur augmente au cours du temps. Ce type de dynamique sera qualifié de croissance « **exponentielle** ». La seconde gamme de paramètres est caractérisée par  $M_1 = 0$  et le taux de croissance tend localement vers 0. Cela amène à des tumeurs dont la masse semble saturée et on qualifiera ces dynamiques de croissances de type « **plateau** ».

## 4.2 Résultats

Dans cette section, on ne présente que les courbes d'évolution des huit premiers cas. Les cas suivants apparaissent en Annexe B .

La Table 4.1 reprend les résultats de l'ensemble de cette série de tests. Pour chaque tumeur, identifiée par un nombre, et pour les gammes de paramètres correspondant aux croissances de type « plateau » d'une part et « exponentielle » d'autre part, on retrouve :

- Le volume cible  $V_c$ , c'est à dire le volume observé sur le troisième examen et que nous tentons de prédire. C'est donc à ce volume là que nous comparerons les prédictions fournies par l'algorithme.
- La moyenne des prédictions volumiques ainsi que l'écart-type  $\sigma$  calculé sur la répartition des résultats.
- La médiane ainsi que les 5<sup>e</sup> et 95<sup>e</sup> centiles, notés respectivement  $ct\ 5\%$  et  $ct\ 95\%$  permettant de situer la zone dans laquelle sont compris 90% des résultats.
- L'erreur relative calculée sur la moyenne  $\varepsilon_{rm}$  donnée par :  $\varepsilon_{rm} = 100 \times \frac{|moy - V_c|}{V_c}$ .

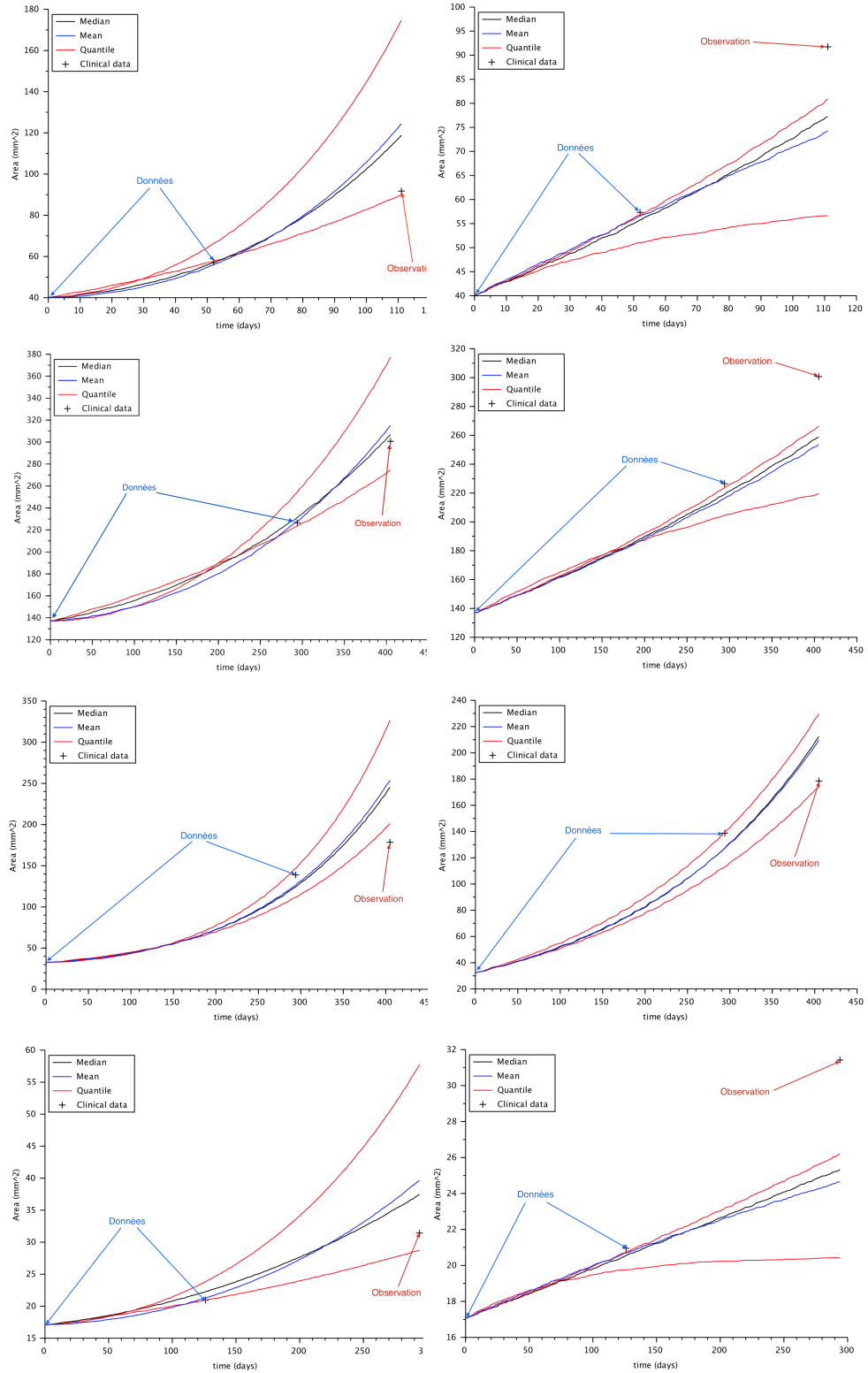


FIGURE 4.1 – Résultat de la méthode sur les cas 1, 2, 3 et 4 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

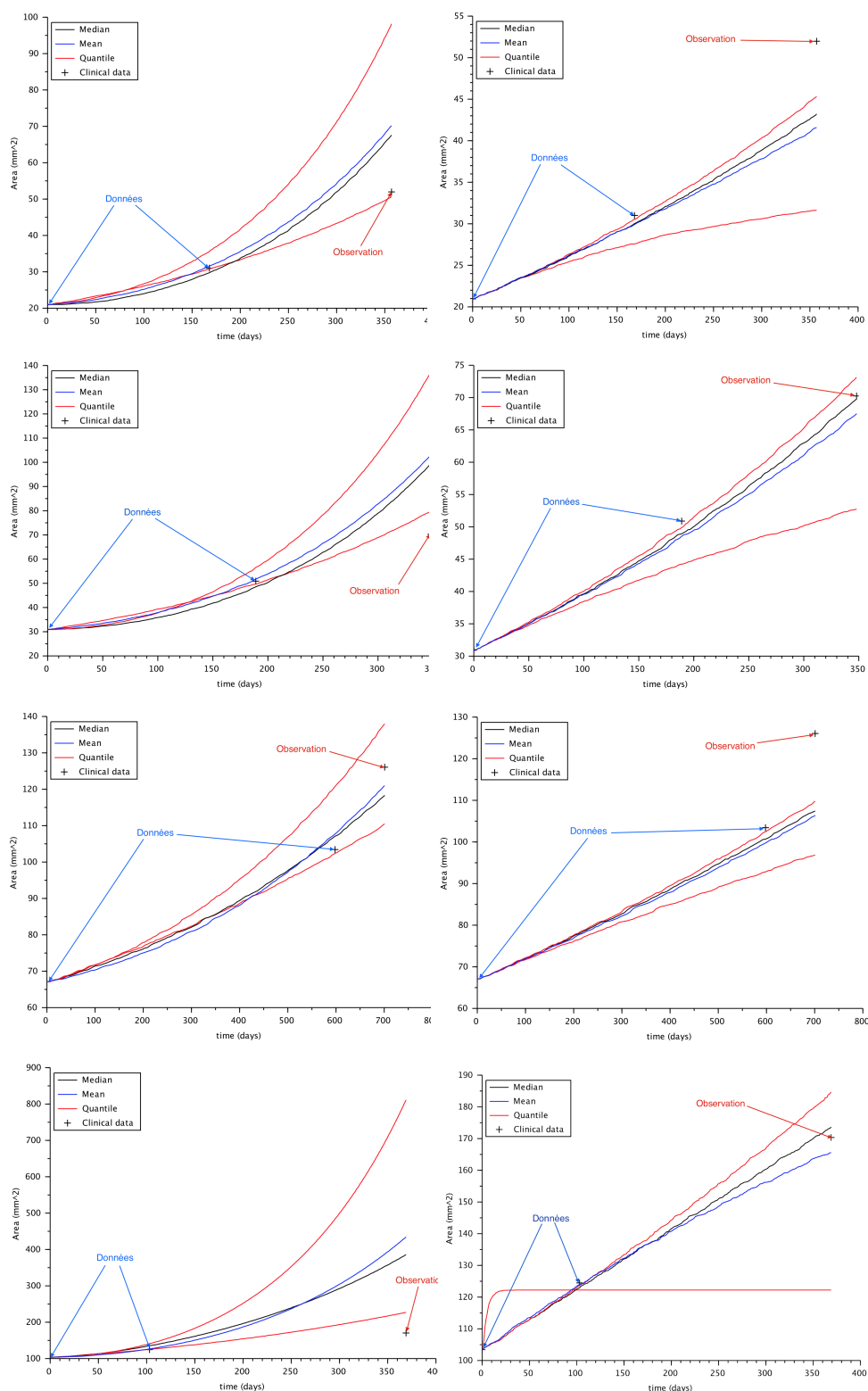


FIGURE 4.2 – Cas 5, 6, 7 et 8 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

Toutes les valeurs sont données en  $\text{mm}^2$  sauf l'erreur relative qui est donnée en pourcentages.

Id. cas	$V_c$ ( $\text{mm}^2$ )	gamme	moy.	$\sigma$	ct 5%	méd.	ct 95%	$\varepsilon_{rm}$ (%)
1	91,7	« plat »	75,7	7,7	56,6	77,3	80,9	17,4
		« exp »	126,2	26,6	90	118,8	174,6	37,6
2	300,7	« plat »	258	15,4	219,6	258,7	266,5	14,2
		« exp »	320,1	33	274,7	306,9	377,3	6,4
3	178,5	« plat »	225,9	19,5	174,3	212,6	229,6	26,5
		« exp »	262,2	40,1	201,1	245,4	326,7	46,9
4	31,4	« plat »	25	1,9	20,4	25,3	26,2	20,4
		« exp »	40,5	9,1	28,7	37,5	57,7	28,9
5	52	« plat »	42,5	4,4	31,6	43,2	45,3	8,3
		« exp »	71,5	15,1	50,6	67,6	98,2	37,5
6	70,2	« plat »	69,2	6,7	52,8	69,8	73,1	1,4
		« exp »	105,5	18,3	79,6	99,5	137,3	50,3
7	126,1	« plat »	107,6	4,5	96,8	107,4	109,8	14,7
		« exp »	122,4	9,1	110,5	118,3	138	2,9
8	170,3	« plat »	167,7	19,8	122,2	173,7	184,7	1,5
		« exp »	439	185,8	227,2	385,7	811,3	157,8
9	177,3	« plat »	183,3	9	161,9	184,5	188,5	3,4
		« exp »	226,4	23,8	194,4	218	268,8	27,7
10	184,5	« plat »	181,5	3,1	175	183,5	183,9	1,6
		« exp »	222,6	41,6	188,7	206,3	314,1	20,1
11	190,2	« plat »	187,6	2,6	181,3	188,2	189,3	1,4
		« exp »	219,1	32,5	192,7	205,9	292,4	15,2
12	106,5	« plat »	86,7	3,2	79,1	86,9	88,3	18,6
		« exp »	98,5	7,2	89,2	95,3	111,1	7,5
13	171,5	« plat »	152,2	21,2	104,1	158	170,2	11,3
		« exp »	427,5	182,4	212,5	377,2	785,6	149,3
14	188,5	« plat »	197	12,4	166,9	198,1	204,1	4,5
		« exp »	250	27,8	212	238,9	298	32,6
15	191,2	« plat »	201,1	7,4	183,9	203,4	207,2	5,2
		« exp »	281	61,3	217,8	259,4	413	47
16	190,1	« plat »	193,1	1,5	190,4	191	194,1	1,6
		« exp »	197,1	2,1	194,7	185,8	201,5	3,7
17	398,5	« plat »	347,1	16,5	305,1	344,3	353,1	12,9
		« exp »	390,1	30,4	346,5	375,7	437,9	2,1
18	18,9	« plat »	20,2	1,5	16,4	20,3	21	6,9
		« exp »	27,3	3,6	22,2	26	33,4	44,4
19	19,3	« plat »	20,8	1	18,4	21	21,4	7,8
		« exp »	28,8	5,2	22,7	27,1	39,2	49,2
20	15,2	« plat »	10,1	1,1	8,4	9,7	11,9	33,6
		« exp »	14,4	1,8	12,3	13,3	17,4	5,3
21	17,4	« plat »	22,8	2,7	15,9	23,3	24,6	31
		« exp »	38,9	8,1	27,6	36,3	53,3	123,6
22	403,9	« plat »	295,2	14,1	261,9	296,1	303,5	26,9
		« exp »	349,6	30,4	309,1	337,2	402	13,4
23	464,9	« plat »	472,9	33,6	393	479,2	492,5	1,7
		« exp »	645,5	89,8	524,2	615	808,5	38,8
24	17,7	« plat »	15,2	1	12,7	15,4	15,8	14,1
		« exp »	21,1	3,2	16,8	20	26,7	19,2
25	43,5	« plat »	31,1	2,5	25,2	31,7	32,8	28,1
		« exp »	44,9	7,1	35,2	42,7	57,5	3,2

Id. cas	$V_c$ (mm <sup>2</sup> )	gamme	moy.	$\sigma$	ct 5%	méd.	ct 95%	$\varepsilon_{rm}$ (%)
26	46,6	« plat »	52,7	4,9	40,6	53,2	55,5	13,1
		« exp »	79,2	13,4	60,7	74,5	102,4	70
27	73,8	« plat »	54,8	4,3	44,6	56,1	58,3	25,8
		« exp »	98,8	28,9	65,3	90,2	157,2	33,9
28	104,7	« plat »	452,4	52,4	314,7	435,9	473,7	56,8
		« exp »	644	129,5	476	581,4	875,7	38,5
29	18,1	« plat »	19,7	1,2	16,8	20,1	20,6	8,8
		« exp »	29,2	5,7	22,1	27,5	40,4	61,3
30	21,9	« plat »	18,2	0,2	17,7	18,2	18,3	16,9
		« exp »	19,7	1,5	18,5	19	23,2	10,1
31	21,8	« plat »	25,9	1,9	21,4	26,4	27,3	18,8
		« exp »	43	10,6	30,2	40,1	64	97
32	38,2	« plat »	34,8	0,6	33,5	35,1	35,3	8,9
		« exp »	41,9	6,9	36,1	39,2	57,7	9,7
33	39,5	« plat »	38,9	0,9	36,9	39,2	39,5	1,5
		« exp »	44,8	4,3	40,2	43,1	53,7	13,4
34	40,4	« plat »	41,7	0,8	39,4	41,4	41,8	3,2
		« exp »	51,6	10,5	42,8	47,6	75,6	27,8
35	42,7	« plat »	43,6	3	34,2	42,2	44,8	2,1
		« exp »	50,2	6,2	40	47,8	58,8	17,6
36	69,4	« plat »	60,5	3,2	52,7	61,3	62,5	12,8
		« exp »	81,8	12,6	66	77,8	106,1	17,9
37	131,7	« plat »	81,5	6,3	66,3	83,2	86,1	38,1
		« exp »	125,6	25,3	93,4	118,6	174,9	4,6

TABLE 4.1 – Résultats synthétiques de l'étude des cas cliniques (les valeurs sont données en mm<sup>2</sup> sauf pour l'erreur relative en %).

Ces premiers résultats obtenus sur 37 cas nous permettent dans un premier temps de comparer les deux modalités de calibration. Si on calibre l'ensemble des cas avec la gamme de paramètres de type « exponentiel » on obtient, en moyenne sur l'échantillon, une erreur relative de  $37,1 \pm 38,6\%$ . L'erreur relative commise en moyenne est donc élevée, et les valeurs constatées dans l'échantillon sont très dispersées (on va de plus de 150% à moins de 3%). Si on fait de même avec la gamme de paramètres de type « plat », on obtient une erreur relative moyenne de  $14,1 \pm 12,5\%$ . Cette calibration donne donc de bien meilleurs résultats en prédiction, tout à fait acceptables du point de vue d'une application clinique. De même, dans 27 cas sur les 37 la calibration « plat » donne une meilleure prédiction que la calibration « exponentielle ».

Il semble donc que si on devait conserver une seule des deux gammes de paramètres, on choisirait celle de type « plat ». Cependant, lorsque celle-ci est la moins performante, on va, en l'utilisant, sous estimer la croissance de la tumeur ce qui peut être dommageable pour le suivi du patient. On le constate notamment pour le dernier cas pour lequel l'erreur commise avec la gamme de type plat est importante (38,1%) alors que la gamme « exponentielle » fournit une bonne prédiction (4,6%). Au contraire, une surestimation de l'agressivité de la tumeur, semble moins problématique. Pour ces raisons, il semble judicieux, à ce stade, de ne pas écarter les comportements de type « exponentiel » du processus de prédiction. Nous avons donc, pour chaque cas, deux prédictions

différentes correspondant à une hypothèse optimiste d'une part, et une hypothèse pessimiste d'autre part. Si on conserve *a posteriori* la meilleure des deux pour chaque cas, la moyenne de l'erreur relative descend à  $9,6 \pm 9,2\%$  et est donc significativement meilleure que lorsqu'on ne considère que la gamme de type « plat ». De même, dans ce cas de figure, seuls 5 cas sur les 37 présentent une erreur relative supérieure à 20%

Cependant, lorsqu'un nouveau cas se présente, nous n'avons pas d'idée *a priori* de la gamme qui fournira la meilleure prédiction de croissance. Pour tenter de contourner cette difficulté et de sélectionner au cas par cas la gamme de paramètres à utiliser, nous avons mis en place une méthode d'apprentissage automatisé dont il sera question au chapitre suivant.

## 4.3 Conclusion

Nous avons présenté, dans ce chapitre, une validation clinique du modèle et de la méthode de calibration sur un ensemble de 37 cas tests pour deux modalités de calibration (dynamique de type « plat » ou de type « exponentiel »). Pour la plupart (33 sur 37) de ces cas cliniques, une modalité au moins donne un résultat satisfaisant, c'est à dire une erreur de prédiction de moins de 20%. Ces résultats sont très encourageants. Cependant, il est pour l'instant impossible de déterminer *a priori* laquelle des modalités utiliser pour un nouveau cas donné. Afin de traiter ce problème, le chapitre suivant décrit la mise en place d'un algorithme d'apprentissage automatisé permettant d'exploiter la base de données de cas cliniques dont nous disposons.





## Chapitre 5

# Apprentissage automatisé pour la classification de l'agressivité tumorale

L'approche basée sur l'utilisation d'un système d'équations aux dérivées partielles calibré au cas par cas sur les images médicales donne des résultats tout à fait satisfaisants. Cependant, elle ne permet pas de tirer parti des informations collectées sur les cas précédents et, même si les temps de calcul restent tout à fait acceptables compte tenu des applications, ne donne pas un résultat instantané. Il apparaît donc complémentaire de mettre en place une procédure permettant d'analyser un nouveau cas de métastase pulmonaire en le comparant à une base de données d'anciens cas, pour lesquels le comportement est connu. Les méthodes dites de Machine Learning, ou apprentissage automatisé, semblent parfaitement adaptées à ce type de problématiques.

### 5.1 Cadre de l'étude et vocabulaire

Les méthodes d'apprentissage supervisé ou automatisé visent à prédire de manière systématique un phénomène concernant un système à partir d'un grand nombre de mesures ou de données connues sur ce système. Le but est d'élaborer un modèle qui, à partir de ces variables d'entrée, nous permettra d'obtenir la valeur de la variable de sortie qui nous intéresse. Une condition nécessaire est évidemment que la variable de sortie dépende réellement, c'est à dire de façon non totalement aléatoire, des variables d'entrée. Autrement dit, le modèle doit être sensible aux variations des variables d'entrée, ou au moins à un certain nombre d'entre elles. La différence avec les travaux explicités dans les parties précédentes de cette thèse est que le modèle utilisé n'est plus un

modèle EDP déterministe mais un modèle statistique permettant d'assimiler des informations sur des cas passés pour anticiper l'évolution de nouveaux cas. Dans le cadre de notre étude, on va donc se servir des nombreuses données récoltées sur les tumeurs de patients antérieurs pour tenter de prédire certains comportements chez de nouveaux patients.

Chaque cas peut être représenté par ses variables d'entrée, que l'on nommera aussi caractéristiques  $x_1, \dots, x_n$  où  $x_i \in X_i$  peut être un réel comme le volume de la tumeur ou un entier comme le nombre de métastases détectées. Chacune de ces données  $x_i$  peut être vue comme une réalisation d'une variable aléatoire  $\hat{x}_i$  de loi dans notre cas inconnue. Ces informations peuvent être compactées dans un vecteur des caractéristiques  $\mathbf{x} = (x_1, \dots, x_n)$  de dimension  $n$  et appartenant à l'espace des caractéristiques  $\mathbf{X} = X_1 \times \dots \times X_n$ .

La variable de sortie ou variable cible est notée  $y \in Y$  et peut, elle aussi, être vue comme la réalisation d'une variable aléatoire  $\hat{y}$ . Chercher un modèle pour prédire la valeur de la variable cible à partir des valeurs connues des caractéristiques revient donc à chercher un moyen d'approcher la fonction de densité de probabilité de la variable aléatoire  $\hat{y}$  sachant les valeurs prises par les  $\hat{x}_i$ ,  $f(\hat{y}|\hat{x}_1, \dots, \hat{x}_n)$ .

On appellera base de données ou échantillon d'apprentissage un ensemble de  $N$  cas pour lesquels sont à la fois connues les valeurs des  $n$  caractéristiques  $(x_1, \dots, x_n)$  et la valeur de la cible  $y$ . Le modèle induit par le processus d'apprentissage automatisé correspond alors à une fonction  $\phi : \mathbf{X} \rightarrow Y$ , le but étant que la prédiction  $\phi(\mathbf{x})$  soit la plus précise possible. Si la cible ne peut prendre qu'un nombre fini de valeurs (on parle alors de classes), le problème est dit de **classification**. Dans le cas contraire, si  $Y$  est un intervalle, il s'agit d'un problème de **régression**.

Si on connaît la forme de la fonction  $\phi$ , on peut tenter de la déterminer explicitement. Par exemple, si  $\phi$  est une fonction linéaire des caractéristiques,  $\phi(\mathbf{x}) = \sum_{i=1}^n a_i x_i$ , les coefficients  $a_i$  peuvent être déterminés à partir de l'échantillon d'apprentissage en utilisant une méthode des moindres carrés. Il s'agit là d'un des cas les plus simples d'apprentissage automatisé. Si, au contraire, on n'a aucune idée de la forme que prend la fonction  $\phi$ , on peut la considérer comme une boîte noire et chercher d'autres moyens d'approcher sa valeur à partir des réalisations connues sur l'échantillon d'apprentissage. Il est notamment possible d'interpoler cette fonction au point de l'espace des caractéristiques correspondant au nouveau cas. Pour cela, une possibilité est d'utiliser la méthode des plus proches voisins. Cette méthode consiste à déterminer, parmi la base d'apprentissage, les cas d'apprentissage les plus proches du nouveau cas au sens d'une certaine distance à définir. On peut ensuite en déduire que, par continuité de la dépendance entre les variables d'entrée et la variable

de sortie, la valeur cible cherchée se trouve dans un voisinage des valeurs de sortie des cas d'apprentissage les plus proches. Le principal défaut de cette méthode est justement que la définition de la fonction distance suppose une connaissance *a priori* de la sensibilité de la variable cible aux variations des caractéristiques. Or justement, dans notre cas, on ne connaît pas l'importance relative des caractéristiques donc ce type de méthode semble difficile à mettre en oeuvre. Nous disposons d'un grand nombre d'informations sur chaque cas clinique, qu'il s'agisse de données propres au patient (âge, sexe...) ou de caractéristiques géométriques de la tumeur pouvant être extraites des images médicales. Pour cela, on choisit d'utiliser une technique basée sur l'utilisation d'arbres de décision, l'algorithme des forêts aléatoires.

## 5.2 Introduction aux forêts aléatoires

Le principe des forêts aléatoires [30, 10, 46] se fonde sur une utilisation intensive d'arbres binaires de décision, que l'on ait affaire à un problème de régression ou de classification. Au lieu de simplement construire un arbre de décision utilisant l'ensemble de la base d'apprentissage et des caractéristiques, on construit plusieurs arbres partiels en n'utilisant qu'un sous échantillon des patients ainsi qu'un nombre limité de caractéristiques. Le fait de construire ainsi plusieurs arbres sur des sous-ensembles de données vise, entre autres, à limiter l'influence d'éventuelles données incohérentes et à ne pas se restreindre à l'utilisation des caractéristiques les plus influentes.

L'algorithme des forêts aléatoires a été utilisé avec succès dans plusieurs travaux liés à des problématiques médicales notamment pour analyser des images (voir [25, 24, 45, 31]).

### 5.2.1 Construction d'un arbre binaire aléatoire

Un arbre binaire est un graphe connexe orienté  $\mathcal{A} = (N, B)$ . Les sommets  $N$  sont appelés noeuds de l'arbre, et les arcs  $B$  sont aussi nommés branches de l'arbre. Chaque noeud est de degré entrant (nombre d'arcs pointants vers le noeud) au plus 1, et de degré sortant (nombre d'arcs partant du noeud) au plus 2. La racine de l'arbre est le seul noeud de degré entrant 0 et les feuilles de l'arbre sont les noeuds de degré sortant 0.

Dans notre cas, chaque noeud représente un sous ensemble de l'ensemble d'apprentissage de départ (qui lui-même est un sous ensemble de la base de données d'apprentissage). Les sous ensembles correspondant aux noeuds fils issus d'un même noeud parent forment une partition du sous ensemble lié au noeud parent. Cette découpe est faite à partir du choix d'une caractéristique

et d'une valeur seuil : tous les cas d'apprentissage dont la valeur de la caractéristique est inférieure à la valeur seuil sont propagés à un noeud fils (par exemple celui de gauche), et les autres cas sont propagés à l'autre noeud fils (celui de droite). En procédant ainsi de façon itérative, on construit, à partir d'un noeud racine, un arbre binaire partitionnant notre ensemble de départ en fonction des valeurs des caractéristiques.

### Choix du nombre de variables et de cas

Les choix du nombre de cas d'apprentissage utilisés ainsi que du nombre de caractéristiques sont déterminants. Si l'on utilise tous les cas et toutes les caractéristiques, on est dans le cas classique de l'arbre binaire de décision simple, et on ne parle plus alors de forêt aléatoire. En pratique, on va utiliser un nombre  $N' \leq N$  de cas pour construire chaque arbre. De même, on a utilisé pour chaque arbre, une caractéristique sur deux et construit ainsi une forêt comptant un nombre d'arbres de l'ordre de grandeur du nombre de cas d'apprentissage  $N$ . Le tirage des cas d'entraînement ainsi que des caractéristiques utilisés pour la construction de l'arbre est fait aléatoirement de façon uniforme.

### Croissance de l'arbre et choix des noeuds

Étant donné le sous ensemble d'apprentissage sur lequel va être construit l'arbre, c'est à dire l'ensemble de cas d'apprentissage que l'on va trouver à la racine de l'arbre, il faut définir la procédure de croissance de l'arbre. Cette procédure de croissance correspond à la méthode utilisée pour partitionner un noeud parent donné en deux noeuds fils et il convient donc de définir une méthode permettant de déterminer la caractéristique et la valeur seuil associée donnant la meilleure partition du noeud parent. Dans un problème de classification comme le notre, une « bonne » partition produit deux sous ensembles chacun homogène du point de vue de la variable cible. Pour illustrer ces notions, nous allons nous appuyer sur un cas très simple d'une base de données comptant  $N = 50$  cas d'apprentissage avec deux caractéristiques réelles  $x$  et  $y$  chacune variant dans l'intervalle  $[0; 1]$ . La variable cible est binaire (croix ou rond). La base d'apprentissage est représentée Figure 5.1.

Le but est ici d'utiliser nos connaissances *a priori* de la base de données pour déterminer au cas par cas la valeur de la cible pour tout nouveau cas, c'est à dire s'il s'agit d'une croix ou d'un rond. Pour cela, on procède à des découpes itératives de l'espace des caractéristiques en fonction des valeurs de la variable cible. Étant donné l'espace des caractéristiques de départ  $\Omega$ , ici  $\Omega = [0; 1] \times [0; 1]$ , nous allons détailler le processus menant à la première découpe et donc à la formation des deux premières branches. La croissance de

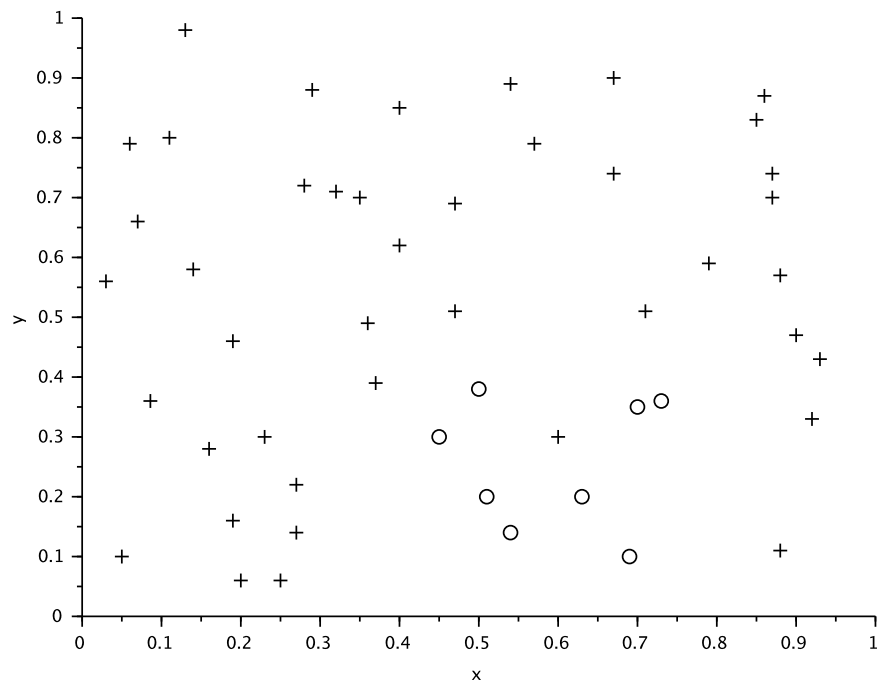


FIGURE 5.1 – Représentation de la base d'apprentissage test, les cas d'apprentissage dont la variable cible vaut 1 sont représentés par une croix, ceux dont la variable cible vaut 0 sont marqués d'un rond. Les deux caractéristiques utilisées sont les coordonnées  $x$  et  $y$ .

l'arbre (c'est à dire les découpes successives de l'espace  $\Omega$ ) se faisant de manière itérative, on procèdera de même par la suite. Lors de la construction d'un arbre binaire, la découpe de l'espace des caractéristiques revient à déterminer un hyperplan de cet espace défini par une valeur seuil  $d$  imposée à une des caractéristiques  $c^j$  (dans notre cas simple  $c_j = x$  ou  $y$ ). Ainsi, chaque sous-espace ou noeud résultant de cette découpe va être défini par une inégalité :  $c^j < d$  pour le premier sous-espace, et  $c^j \geq d$  pour le second. La partition d'un noeud implique donc de définir quelle variable  $c^j$  doit être utilisée, mais aussi quel seuil  $d$  optimise la découpe.

**critère de calcul du seuil optimal :** Le choix du seuil s'appuie sur l'utilisation de critères quantifiant l'homogénéité des noeuds fils résultant de la découpe. En effet, on veut petit à petit isoler les différents comportements de notre variable cible et donc obtenir des noeuds finaux (feuilles de l'arbre) les plus homogènes possibles du point de vue de la cible. Un critère largement utilisé pour les problèmes de classification est le critère de Gini qui évalue l'homogénéité de chacun des noeuds fils. Pour le noeud fils  $j$ , l'indice de Gini est donné par :  $G^j = \sum_{i=1}^n p_i^j(1 - p_i^j)$  où  $p_i^j$  représente la proportion, dans le sous-espace correspondant au noeud  $j$ , de la population appartenant à la  $i^{\text{ème}}$  classe. Dans notre cas, on n'a que deux classes donc  $i = 2$  et  $p_o^j = (1 - p_+^j)$ . On a finalement :

$$G^j = 2p_+^j(1 - p_+^j).$$

On voit donc que cet indice est nul si  $p_+^j$  (ou bien  $p_o^j$ ) vaut 0 ou 1, c'est à dire si l'on n'a que des croix ou que des ronds dans le noeud fils  $j$ . Ainsi, dans notre souci de produire une découpe homogène, on a intérêt à minimiser l'indice de Gini de chacun des deux noeuds fils. On peut donc prendre comme indicateur la moyenne, ou encore le maximum, des indices de Gini des deux noeuds fils. Étant donnée une caractéristique, par exemple  $x$ , on cherche donc le seuil  $d \in [0; 1]$  qui minimise notre critère, ici la moyenne des indices de Gini des noeuds fils. Pour cela on parcourt l'intervalle de recherche  $[0; 1]$ . Le plus simple est d'utiliser une méthode de type Monte Carlo : on tire un certain nombre de valeurs de  $d$  et on garde le seuil qui minimise notre critère. Certes cette méthode est rapide à mettre en oeuvre mais elle est bien loin d'être optimale en terme d'efficacité. Prenons l'exemple simple où une grande partie des cas serait concentrée dans une petite zone de l'espace comme  $x \in [0; 0,1]$ . Alors seulement 10% en moyenne des tirages seraient réellement utiles et il serait difficile d'être sûr d'avoir la meilleur découpe, à moins d'augmenter considérablement le nombre de tirages. Pour éviter ce genre de difficultés, nous avons choisi d'ordonner les cas selon leur valeur de la caractéristique concernée (ici  $x$ ) puis de tester une découpe entre chaque paire de points successifs. On est ainsi

certain de tester toutes les découpes possibles de notre espace en un nombre raisonnable de tentatives  $N - 1$ . On procède donc de la sorte afin de trouver la meilleure découpe en  $x$  de notre ensemble de départ selon l'indice de Gini. On obtient la séparation suivante (Fig. 5.2) :

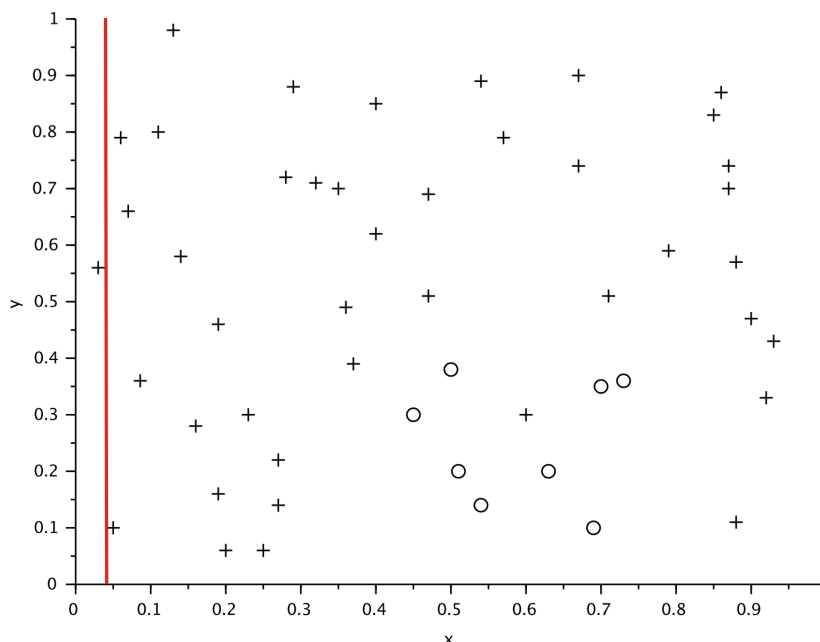


FIGURE 5.2 – Découpe optimale en  $x$  de l'espace des caractéristiques pour la base d'apprentissage du cas test selon le critère de Gini,  $d = 0,04$ .

On constate que le résultat obtenu ne correspond pas à ce à quoi on pouvait s'attendre. La découpe optimale selon ce premier critère isole un cas dans un noeud fils mettant les autres cas d'apprentissage dans le second noeud fils. Ce choix de la valeur seuil permet bien d'homogénéiser la découpe. Le noeud de gauche, correspondant à l'inégalité  $x < 0,04$  est parfaitement homogène puisqu'il ne comporte qu'un cas d'apprentissage : on a  $p_o^1 = 0$  donc  $G^1 = 0$ . Le noeud de droite ( $x \geq 0,04$ ) est relativement homogène avec  $p_o^2 = \frac{8}{49}$  soit  $G \simeq 0,272$ . Si l'on déplace notre découpe vers la droite, c'est à dire dans le sens croissant des  $x$ , dans un premier temps (tant qu'on ne dépasse pas  $d = 0,43$ ) on conserve un noeud fils de gauche parfaitement homogène mais on rend le fils de droit de plus en plus hétérogène puisque le nombre de croix diminue tout en restant plus important que le nombre de ronds. Lorsque  $d$  est compris entre 0,43 et 0,75, les deux noeuds sont plus ou moins hétérogènes et l'indice de Gini est grand. Au delà de 0,75 le noeud de gauche est plus ou moins homogène et celui de droite parfaitement homogène.



Le problème inhérent à l'utilisation de l'indice de Gini comme critère de dichotomie, vient du fait qu'il évalue séparément l'homogénéité de chaque noeud fils. Dans notre cas, chaque noeud fils est homogène du point de vue de la proportion de cas en croix. Or si un des noeuds fils compte une grande proportion de cas de type croix, on aimerait que l'autre noeud fils présente une forte proportion de cas de type rond afin de vraiment distinguer les deux comportements. Le critère de Gini ne prend pas en compte ce genre de considérations. Afin de pouvoir tirer parti de cette remarque, nous avons mis au point un critère spécifique  $\Gamma$ . Introduisons tout d'abord quelques notations. On notera  $p_+^<$  la proportion de cas de type croix dans le noeud de gauche ( $x < d$ ). De même  $p_+^>$  sera la proportion de croix dans le noeud fils de droite. On adopte des notations similaires pour les proportions de cas de type rond dans chaque noeud fils. On part de l'hypothèse selon laquelle les cas de type croix sont concentrés à gauche de l'espace des caractéristiques (donc pour des valeurs faibles de  $x$ ). On veut donc maximiser la proportion de croix dans le noeud de gauche  $p_+^<$ , mais aussi la proportion de cas de type rond dans le noeud droit  $p_o^>$ . Dans le même temps, on veut minimiser la proportion de cas mal classés, c'est à dire sous notre hypothèse, de cas croix dans le noeud fils de droite  $p_+^>$  et de cas ronds dans le noeud de gauche  $p_o^<$ . On aboutit donc au critère suivant :

$$\gamma = \frac{p_+^< + p_o^>}{p_o^< + p_+^>},$$

que l'on peut écrire, en remarquant que  $p_o^< = 1 - p_+^<$  et  $p_+^> = 1 - p_o^>$  :

$$\gamma = \frac{p_+^< + p_o^>}{2 - (p_+^< + p_o^>)}. \quad \gamma = \frac{p_+^< + p_o^>}{2 - (p_+^< + p_o^>)}$$

On a donc un critère qui semble intéressant sous l'hypothèse que les cas de type croix sont concentrés dans le noeud de gauche. Que dire dans le cas contraire ? On peut re-écrire notre critère dans hypothèse où ce sont les cas de type rond qui se concentrent dans le noeud de gauche. On obtient alors en suivant le même raisonnement :

$$\gamma' = \frac{p_o^< + p_+^>}{p_+^< + p_o^>},$$

et finalement  $\gamma' = \frac{1}{\gamma}$ . Afin d'obtenir une découpe homogène, on peut donc minimiser le nombre de cas mal classés à travers le critère  $\Gamma = \min(\gamma, \frac{1}{\gamma})$ . On remarque que  $\Gamma \in [0; 1]$  et que  $\Gamma = 0$  quand chaque noeud est parfaitement homogène et que l'un contient les cas croix et l'autre les ronds. Au contraire, si  $\Gamma = 1$ , on a les mêmes proportions de croix que de ronds dans chaque noeud fils et donc on a deux noeuds fils très hétérogènes. On a reproduit le test établi

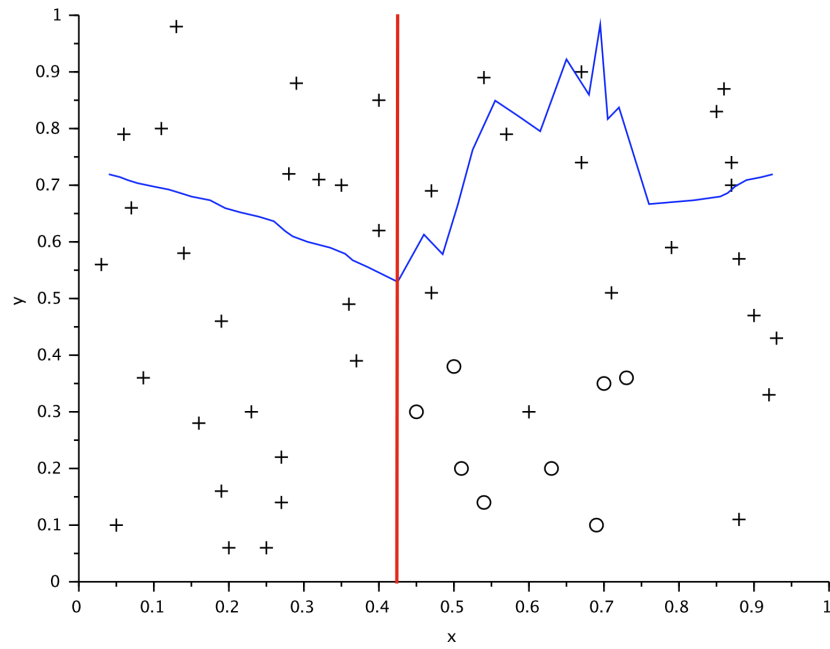


FIGURE 5.3 – Découpe optimale en  $x$  de l'espace des caractéristiques pour la base d'apprentissage du cas test selon le critère  $\Gamma$ ,  $d = 0,425$ . La droite rouge représente la découpe optimale selon le critère, la courbe bleue représente l'évolution en fonction de  $x$  du critère  $\Gamma$ .

avec le coefficient de Gini consistant à chercher la meilleure découpe en  $x$  de notre ensemble selon le critère  $\Gamma$ . On obtient le résultat donné en Figure 5.3.

La courbe bleue sur la Figure 5.3 représente la fonction  $x \rightarrow \Gamma(x)$ . On voit que le minimum est atteint pour  $x = 0,425$  et vaut 0,52. La découpe ainsi obtenue semble plus intéressante que celle donnée par le critère de Gini puisque les cas sont bien répartis entre les deux noeuds fils et les sous espaces engendrés sont suffisamment homogènes. Nous utiliserons donc ce critère pour la suite des travaux. Cependant, il convient de préciser que ce critère n'est utilisable en l'espèce que pour des problèmes de classification avec une variable cible binaire.

**Choix de la caractéristique de partition :** Nous avons vu comment, à une caractéristique donnée, déterminer un seuil de découpe optimale. Il reste donc à savoir laquelle des deux caractéristiques utiliser pour la première découpe. En pratique, on calcule les seuils optimaux correspondant à une découpe en  $x$  et en  $y$ , respectivement  $d_x$  et  $d_y$ . On choisit la variable qui minimise le critère  $\Gamma$  c'est à dire  $x$  si  $\Gamma(d_x) < \Gamma(d_y)$ ,  $y$  sinon. L'avantage est que l'on connaît ainsi d'ores et déjà la valeur du seuil correspondant. En procédant ainsi, on obtient la découpe donnée en Figure 5.4.

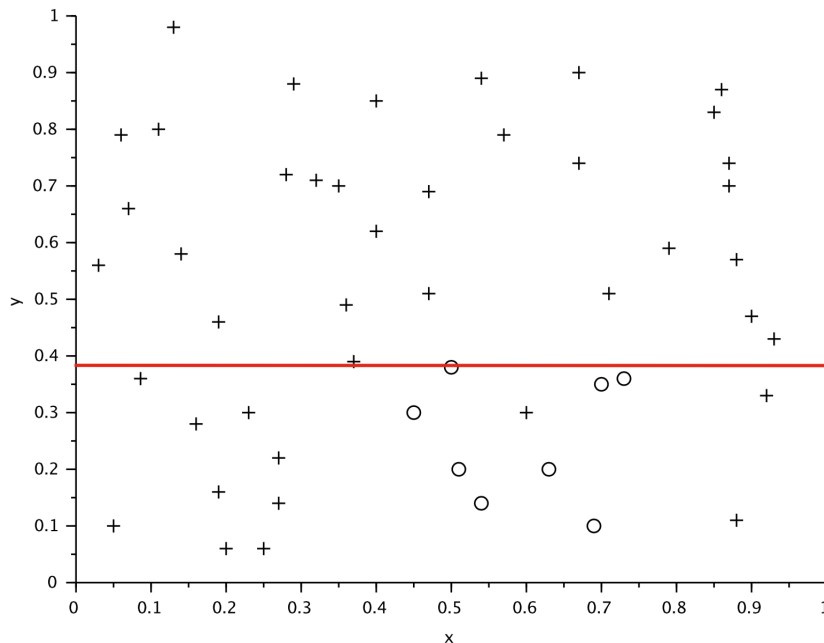


FIGURE 5.4 – Première découpe : formation des deux premières branches de l'arbre. Critère de séparation  $y = 0,385$ .

Finalement, la meilleure découpe se fait en discriminant sur la valeur de  $y$ , les deux noeuds fils obtenus correspondent donc aux inégalités  $y < 0,385$  d'une part, et  $y \geq 0,385$  d'autre part.

Ce choix automatique de la meilleure caractéristique de découpe est un des atouts majeurs de la méthode. En effet, il permet de sélectionner les variables réellement influentes. Ainsi, si on utilise une caractéristique non explicative, c'est à dire n'ayant aucune influence sur la valeur de la variable cible, cette caractéristique ne sera jamais sélectionnée par l'algorithme et restera muette lors de la construction de l'arbre. Cela procure un réel avantage à cette méthode par rapport à d'autres, plus élémentaires, comme la méthode des plus proches voisins. On peut donc utiliser un grand nombre de caractéristiques sans savoir *a priori* si elles ont un vrai rôle quant à la valeur de la variable cible sachant que l'algorithme les trie automatiquement. Nous avons ainsi classé les cas de la base d'apprentissage dans deux noeuds et construit le premier niveau de l'arbre. La croissance de l'arbre se fait itérativement en suivant le même protocole que précédemment : à chaque étape, on divise les noeuds fils issus de l'étape précédente. Si l'on procède ainsi sur un nombre suffisant d'itérations, les noeuds finaux, c'est à dire les feuilles de l'arbre, ne vont contenir chacun qu'un seul cas d'apprentissage. Un tel arbre n'est pas intéressant pour classer car la méthode devient alors très sensible aux cas potentiellement anormaux et ne tient plus compte de la tendance globale de la base de données. Il est donc important de définir des critères d'arrêt pour l'algorithme de construction de l'arbre.

**Hauteur de l'arbre et critère d'arrêt :** La profondeur ou hauteur d'un noeud de l'arbre correspond à sa distance à la racine, calculée comme le nombre d'arcs entre le noeud et la racine. La hauteur de l'arbre correspond, quant à elle, à la hauteur de la feuille la plus haute. Afin de déterminer la profondeur idéale de chaque branche, on doit fixer certains critères arrêtant la croissance de celle-ci.

- Le premier critère que nous utilisons est le critère d'homogénéité. Lorsqu'un noeud atteint une homogénéité suffisante (par exemple 90% de cas identiques), on décide de ne plus faire croître la partie de l'arbre correspondante, c'est à dire que l'on ne divise plus le noeud concerné.
- On peut décider d'imposer un nombre minimum de cas dans chaque noeud. En effet, comme on l'a vu, un noeud regroupant trop peu de cas d'apprentissage risque de ne pas être représentatif. On peut par exemple imposer que chaque feuille de l'arbre contienne au minimum 5% des cas de la base.
- Une autre possibilité est de limiter la hauteur de l'arbre, c'est à dire le

nombre maximum de subdivisions ou d'itérations lors de la construction.

En combinant ces critères et en utilisant l'algorithme de construction de l'arbre décrit précédemment, on obtient la découpe de l'espace des caractéristiques représentée en Figure 5.5.

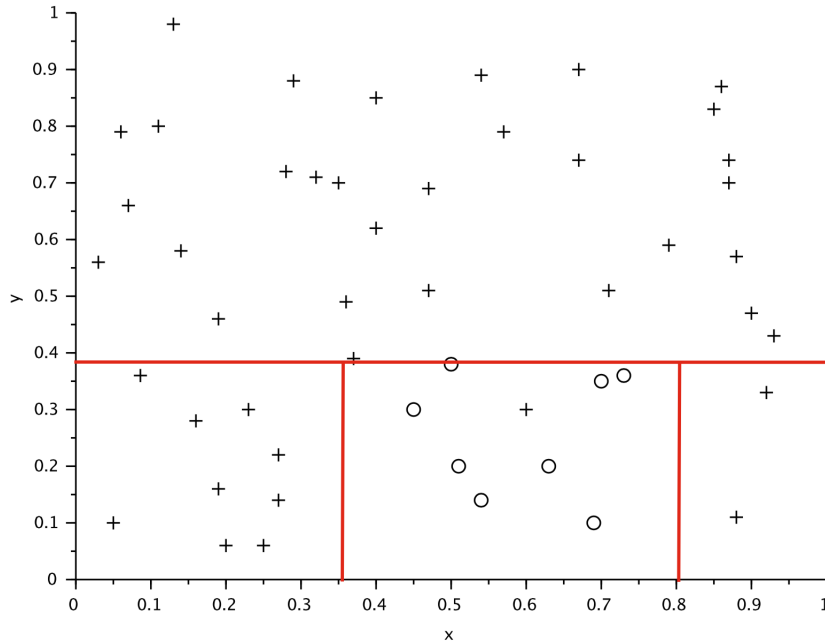


FIGURE 5.5 – Découpe de l'espace des caractéristiques du cas test correspondant à l'arbre binaire construit.

### 5.2.2 Classification du cas et prédiction

Une fois l'arbre construit à partir de la base de données d'apprentissage, notre but est bien de pouvoir classer *a priori* un nouveau cas et donc de tenter de prédire la valeur de la variable cible pour ce nouveau cas connaissant uniquement ses caractéristiques. La procédure pour cela est très simple puisqu'il suffit de positionner le nouveau cas dans l'arbre. La valeur de la variable cible est alors celle majoritairement représentée dans la feuille de l'arbre dans laquelle se trouve notre nouveau cas.

Reprenons l'exemple de notre cas simple. On considère un nouveau cas A tel que  $x_A = 0,1$  et  $y_A = 0,2$ . On parcourt l'arbre et positionne notre cas, ce qui revient à le placer dans l'espace des caractéristiques. En se reportant à la Figure 5.5, on constate que le point A appartient à la partie de l'espace définie par les inégalités  $y < 0,385$  et  $x < 0,36$  et les cas de la base d'apprentissage

correspondant à ce sous espace sont tous de types croix ( $p_+ = 1$ ). On en déduit que le cas A est, d'après notre arbre, de type croix.

### Sauvegarde des arbres

Le stockage des arbres construits sur la base d'apprentissage est un point important. En effet, le but de la méthode des forêts aléatoires est de donner un résultat en temps réel. On n'envisage donc pas de reconstruire les arbres, opération qui peut être assez coûteuse selon la taille de la base d'apprentissage, à chaque utilisation. La forêt est construite une fois pour toutes et éventuellement enrichie par l'apport de nouveaux cas d'apprentissage ponctuellement et se pose donc le problème du stockage des données et des arbres aléatoires. On remarque que l'arbre binaire ainsi construit est entièrement défini par le sous-ensemble de départ de la base d'apprentissage ainsi que la liste ordonnée des couples (caractéristique ; valeur seuil) utilisée pour la formation des différents noeuds. Il suffit donc de stocker ces couples (caractéristique ; seuil) pour chaque noeud, ainsi que les probabilités d'appartenance aux différentes classes de chaque feuille. Cela permet d'envisager des procédures de stockage plus fines que la simple copie des données présentes au niveau de chaque noeud.

### 5.2.3 Limites de l'approche de classification par un arbre binaire

La première limitation liée à l'utilisation d'un arbre binaire vient du fait que la construction de l'arbre peut être sensible à la présence dans la base de cas mal classés. Cela peut donc biaiser les résultats de prédiction. Si l'on reprend la Figure 5.5, on peut aussi constater que le classement d'un nouveau cas dépend beaucoup des découpes effectuées. Regardons par exemple un potentiel nouveau cas B, tel que  $x_B = 0,8$  et  $y_B = 0,2$ . Tel qu'est construit l'arbre, on prédit que notre cas B est de type rond. Cependant, avec exactement la même base d'apprentissage, on aurait pu obtenir une découpe légèrement différente modifiant également la classification de B. Cela est illustré en Figure 5.6.

Dans la configuration présentée dans la Figure 5.6b, l'arbre a été légèrement perturbé au niveau de la dernière découpe. Initialement, on utilisait comme seuil  $d_x = 0,805$  mais désormais on utilise  $d'_x = 0,75$ . Ce changement est parfaitement transparent du point de vue de la base d'apprentissage puisqu'aucun cas d'apprentissage est dans le sous espace défini par  $(x,y) \in [0,75; 0,805] \times [0; 0,385]$ . Les deux arbres sont donc équivalents du point de vue de la base de données. Cependant ils classent différemment le point B : dans le premier cas on prédit pour B un comportement de type rond alors que dans le second cas, on lui prédit un comportement de type croix.

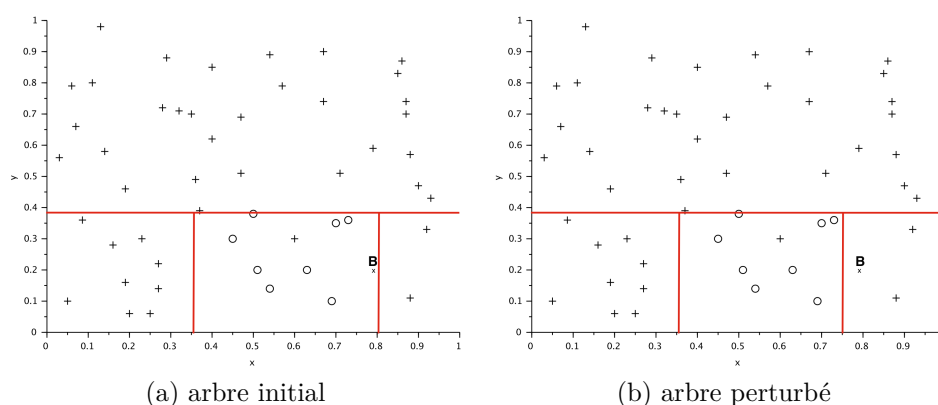


FIGURE 5.6 – Positionnement du cas B dans l’espace des caractéristiques dans le cas de la découpe correspondant à l’arbre initial, puis de la découpe correspondant à un arbre perturbé.

Ainsi l’utilisation d’un arbre binaire présente des limites pouvant biaiser la classification de nouveaux cas et nous induire en erreur.

### 5.2.4 De l’arbre à la forêt

La méthode des forêts aléatoires consiste en l’utilisation, non plus d’un unique arbre binaire, mais d’un grand nombre d’arbres construits sur la base de données d’apprentissage. Pour chaque arbre, on utilise un certain nombre  $N' < N$  de cas. De même, seul un nombre restreint de caractéristiques est sélectionné pour chaque arbre. Ces choix, tant pour les cas d’apprentissage que pour les caractéristiques retenues pour chaque arbre, sont faits de façon aléatoire uniforme dans l’ensemble de départ. En procédant ainsi, chaque arbre est différent, puisque construit à partir de données partielles différentes. Une fois la forêt construite, on peut l’utiliser pour classer chaque nouveau cas et déterminer la valeur de sa variable cible (dans notre cas, un booléen). Pour cela, on demande à chaque arbre de classer le cas comme vu précédemment, puis on renvoie une réponse moyenne. Par exemple, dans le cas d’une valeur cible booléenne 0 ou 1, si, en réponse de classification d’un nouveau cas la forêt renvoie la valeur 0,72, cela signifie que 72% des arbres de celle-ci ont classifié le cas comme étant de type 1. Si on établit, ce qui est le cas en général, que la majorité l’emporte, on classe donc ce cas comme probablement de type 1.

L’avantage de cette méthode, par rapport à l’utilisation d’un unique arbre binaire, est donc de donner un résultat moyen sur un grand nombre d’arbres. Le fait d’utiliser, pour chaque arbre, une partie seulement des données, permet de limiter l’influence de cas mal classés pouvant correspondre à des erreurs de

mesures ou simplement à un bruit dans les données. De même, l'utilisation de caractéristiques différentes permet de ne pas se focaliser sur l'importance des quelques caractéristiques les plus influentes et de tenir compte d'effets plus subtils.

## 5.3 Utilisation des forêts aléatoires sur les cas de métastases pulmonaires

Les différents essais faits sur le modèle EDP nous ont amenés à constituer une base de données de différents cas cliniques. Notre but est de tirer parti de cette quantité d'informations afin de tenter d'affiner les prédictions de croissance tumorale.

### 5.3.1 Protocole d'étude et base d'apprentissage

Comme nous l'avons vu précédemment, l'espace de paramètres du modèle EDP simplifié peut être divisé en trois sous espaces auxquels correspondent différentes dynamiques de croissance. Notre problème vient du fait que, *a priori*, nous ne savons pas quelle gamme de paramètres utiliser lorsqu'un nouveau cas se présente. **L'intérêt de l'utilisation d'une méthode d'apprentissage automatisé est donc de comparer chaque nouveau cas à notre base de données afin de pré-sélectionner la gamme de paramètres qui semble la plus appropriée pour prédire l'évolution de la tumeur.** En pratique, nous avons vu que seuls deux sous-espaces de calibration suffisent à décrire convenablement les dynamiques de croissance de l'ensemble des cas de notre base de données de patients. Pour la suite de l'étude, nous distinguerons ces deux types de croissances en les nommant « croissance lente » d'une part, et « croissance rapide » d'autre part. Le problème étudié est donc un problème de classification dont la variable cible  $v_c$  peut être vue comme une variable booléenne. On définit cette variable de la façon suivante :  $v_c = 0$  si le cas est bien décrit par une dynamique de croissance lente,  $v_c = 1$  si au contraire la croissance de la tumeur est de type rapide.

La base de données d'apprentissage que nous utilisons comporte 25 métastases pulmonaires. C'est assez peu pour l'utilisation d'algorithmes d'apprentissage de type forêts aléatoires mais rappelons qu'il ne s'agit là que d'une preuve de concept et que la base de données est alimentée régulièrement par les médecins et est donc amenée à s'étoffer. Pour chacune de ces tumeurs, nous avons trois examens à des dates différentes donc rendant compte de la dynamique de croissance de celle-ci. Nous avons (voir partie précédente) utilisé à chaque fois les deux premiers examens afin de calibrer le modèle et d'établir



une prédiction de croissance pour les deux types de comportements qui nous intéressent (croissance rapide ou lente). Enfin, pour chaque cas, nous avons comparé les deux prédictions obtenues avec la troisième image de la tumeur pour déterminer lequel des deux comportements décrit au mieux la croissance de la tumeur. Ainsi, pour chacune des 25 tumeurs de la base d'apprentissage, nous connaissons la valeur 0 ou 1 prise par la variable cible  $v_c$ .

Afin d'utiliser cette base de données pour construire une forêt aléatoire, il faut encore définir un certain nombre de caractéristiques qui nous semblent potentiellement en lien avec l'agressivité de la tumeur. Ces caractéristiques doivent pouvoir être calculées uniquement à partir des deux premiers examens et éventuellement de données intrinsèques au patient. En effet, des nouveaux cas de tumeurs que nous aurons besoin de classer à des fins prédictives nous ne connaissons que deux examens. De plus, ces caractéristiques doivent pouvoir être mises sous la forme de scalaires (réels ou éventuellement entiers) pour être utilisables dans l'algorithme des forêts aléatoires.

### 5.3.2 Caractéristiques utilisées

Les caractéristiques que nous avons choisies pour définir chaque cas peuvent être rangées dans trois catégories différentes. Certaines d'entre-elles sont liées à la dynamique volumique de la croissance de la tumeur. On y retrouve notamment des indicateurs scalaires provenant de modèles de croissance EDO. Une deuxième classe de caractéristiques tient compte de la géométrie de la tumeur. Il n'est en effet pas exclu que des critères tels que l'allongement de la tumeur ou encore la régularité des contours soient des indicateurs de l'agressivité tumorale. Enfin, d'autres facteurs liés à l'histoire du patient et de la lésion sont pris en compte. Ces caractéristiques ne sont pas directement accessibles dans les images et viennent enrichir le processus d'apprentissage.

#### Critères de dynamique de croissance volumique :

- Les premières caractéristiques que nous avons ajoutées à la base de données sont la surface de la coupe sélectionnée (car on travaille en 2D) aux premier et deuxième examens  $S_0$  et  $S_1$  ainsi que le temps séparant ces examens  $t_1$ .
- On utilise de même le **taux de croissance linéaire surfacique** entre les deux examens :  $\tau = \frac{S_1 - S_0}{t_1}$ .
- On calcule le **taux de croissance exponentielle** du modèle. Cela correspond au taux de croissance  $\alpha$  de la surface tumorale dans l'hypothèse où celle-ci est régie par une loi exponentielle :  $S(t) = S_0 e^{\alpha t}$ . On a donc  $\alpha = \frac{1}{t} \ln\left(\frac{S_1}{S_0}\right)$ .

- On utilise de même un paramètre du modèle de croissance basé sur la fonction puissance ou modèle « power growth » (voir [8]). Il s'agit du modèle EDO suivant :  $\frac{\partial S}{\partial t} = \alpha S^\gamma$ . En première approximation, on peut fixer le paramètre  $\gamma = 0,75$  afin de n'avoir plus qu'un degré de liberté. On calibre alors le **paramètre**  $\alpha$  à partir des valeurs  $S_0$ ,  $S_1$  et  $t_1$  et on l'utilise comme caractéristique.
- On procède de même avec le modèle de Gompertz [8] qui peut s'écrire, sous sa forme intégrée :  $S(t) = S_0 e^{\frac{a}{b}(1-e^{-bt})}$ . Malheureusement, ce modèle présente deux paramètres que l'on ne peut donc pas utiliser tels quels. On trouve cependant dans la littérature des travaux portant sur la corrélation entre les paramètres  $a$  et  $b$  du modèle de Gompertz dans des applications liées à la croissance tumorale. Nous avons calibré le modèle de Gompertz sur un ensemble de cas de métastases pulmonaires pour lesquels on disposait de trois données en temps et positionné les jeux de paramètres de chaque cas dans un plan  $(a,b)$  (voir Figure 5.7). *La corrélation observée n'est malheureusement pas significative.* On uti-

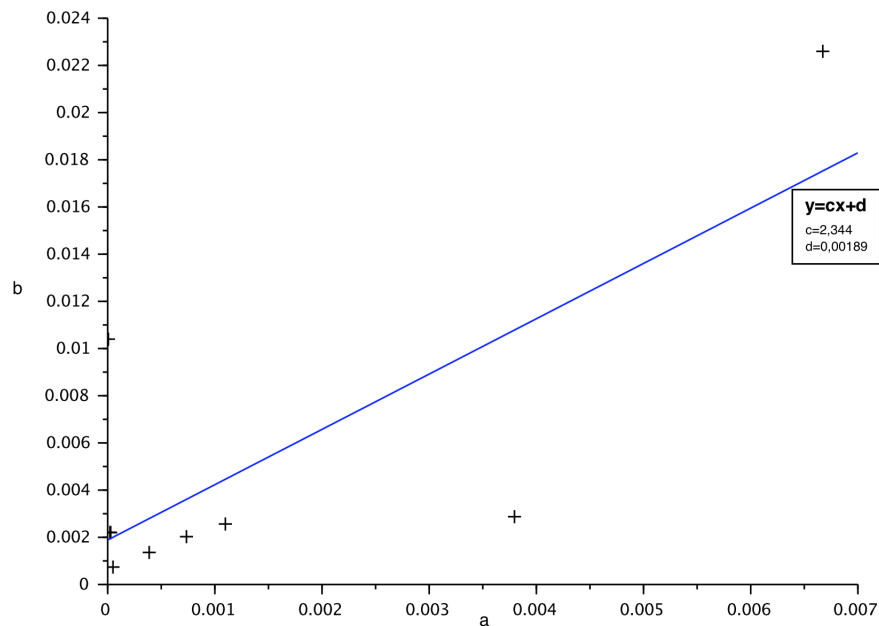


FIGURE 5.7 – Position dans le plan  $(a,b)$  de jeux de paramètres du modèle de Gompertz pour une population de métastases pulmonaires.

lise tout de même la régression affine entre  $a$  et  $b$  issue de la Figure 5.7 afin de réduire à 1 le nombre de degrés de libertés du modèle. On ajoute alors comme caractéristique le **paramètre**  $a$  du modèle de Gompertz.

### Critères liés à la géométrie de la tumeur :

La taille de la tumeur est prise en compte de façon plus ou moins explicite par les caractéristiques précédentes. On souhaite ici étudier uniquement les aspects liés à la forme de la tumeur et pour cela, on adimensionne chaque donnée afin que la tumeur correspondante soit de surface unitaire.

- Les premières caractéristiques de forme que l'on utilise sont les périmètres de la tumeur à l'instant initial et  $t_1$  respectivement  $\mathbf{p}_0$  et  $\mathbf{p}_1$ . Pour une surface donnée (ici 1) cela permet d'avoir un premier renseignement sur la régularité du bord de la tumeur. De même, on ajoute le rapport  $\frac{\mathbf{p}_1 - \mathbf{p}_0}{t_1}$  qui nous indique si la forme de la tumeur tend à se régulariser.
- On voudrait aussi pouvoir ajouter, dans la liste des caractéristiques, un critère permettant de quantifier l'allongement de la tumeur. Nous avons, dans un premier temps, voulu utiliser les moments d'inertie d'ordre 2 de la tumeur. Cela dit, ils dépendent de l'orientation de la tumeur par rapport aux axes du repère. Pour s'affranchir de cela, nous avons ajouté un moment moyenné en rotation d'ordre 2 :

$$m_2(t) = \frac{2}{\Pi} \int_{\theta=0}^{\frac{\Pi}{2}} \int_x \int_y x^2 \mathcal{R}_\theta(T)(x,y) dx dy d\theta,$$

où  $\mathcal{R}_\theta(T)$  est le champ de densité de cellules tumorales auquel on applique la rotation d'angle  $\theta$  autour du centre de gravité de la tumeur. Nous utilisons donc comme caractéristiques  $\mathbf{m}_2(\mathbf{t} = \mathbf{0})$ ,  $\mathbf{m}_2(\mathbf{t} = \mathbf{t}_1)$  et  $\frac{\mathbf{m}_2(\mathbf{t}_1) - \mathbf{m}_2(\mathbf{0})}{t_1}$ .

### Facteurs intrinsèques au patient :

D'autres facteurs propres à la lésion ou au patient peuvent être liés à l'agressivité tumorale.

- Le **sexe du patient** ou encore son **âge** sont ainsi ajoutés à la liste des caractéristiques.
- De même, on sait que la lignée cellulaire a de l'importance et conditionne en partie la gravité de la maladie. On s'intéresse donc à la **localisation de la tumeur primitive**. Pour transcrire ce type d'information en un format de données utilisable par l'algorithme des forêts aléatoires, on fait correspondre une caractéristique booléenne à chaque organe possible pour la tumeur primitive. On met à 1 la caractéristique correspondant effectivement à la localisation de la tumeur primitive à l'origine de la métastase étudiée (les autres sont donc à 0).

Pour d'autres pathologies, on pourrait intégrer des données provenant par exemple de biopsies qui sont difficiles à intégrer directement dans le modèle. De plus, ces caractéristiques jouent sur les paramètres du modèle mais la corrélation ne peut pas être déterminée explicitement.

### 5.3.3 Utilisation des forêts aléatoires sur la base de données de cas cliniques

Nous disposons donc d'une base de données de 25 cas de métastases pulmonaires. Pour chaque cas, nous avons trois images issues de trois CT-scans à des dates distinctes. **En utilisant les deux premières images, on est en mesure de calculer toutes les caractéristiques définies précédemment. La troisième et dernière image nous permet de savoir quelle est la zone de paramètres qui prédit le mieux la croissance de la tumeur et donc de déterminer la valeur de la variable cible, 0 ou 1.** Nous avons utilisé cette base de données pour étudier la pertinence de l'utilisation de la méthode des forêts aléatoires pour notre problème de choix de la zone de paramètres.

Afin de tirer parti au mieux de la base de cas dont nous disposons, nous avons tenté de prédire, pour chaque cas et sans utiliser le troisième scan, quelle serait la meilleure zone de paramètres. Pour cela, nous avons isolé de la base de données le cas étudié et construit une forêt aléatoire composée de 200 arbres en utilisant l'ensemble des 24 autres cas. L'inconvénient de cette façon de procéder est que pour chacun des cas, nous avons dû construire une forêt différente alors que l'un des atouts de la méthode est de permettre, lorsqu'on a déjà une base de données conséquente, de ne pas recalculer la forêt systématiquement. Cela dit, afin de ne pas biaiser le résultat de prédiction, il était indispensable de supprimer de la forêt le cas que nous voulions tester et c'est pourquoi cette démarche nous a semblé être la plus judicieuse.

Pour chacun des cas, on connaît la valeur attendue et on peut donc évaluer, *a posteriori* l'efficacité de la méthode sur un nombre intéressant de cas. On obtient les résultats présentés dans la Table 5.1.

Le résultat renvoyé par la méthode (deuxième colonne de la Table 5.1) correspond à la proportion d'arbres de la forêt ayant voté pour la valeur 1 (correspondant à la croissance rapide). On fait voter les arbres à la majorité : si cette proportion est supérieure à 0,5 on considère que la variable vaut 1 (croissance rapide), sinon on la considère à 0 (croissance lente).

On peut donc comparer, pour chaque cas, la prédiction de comportement donnée par l'algorithme des forêts aléatoires et la valeur réelle de la cible (c'est à dire les troisième et quatrième colonnes de la Table 5.1). Dans 18 cas sur les

tumeur	résultat FA	prédiction	cible
1	0,64	1	1
2	0,65	1	0
3	0,07	0	0
4	0,12	0	0
5	0,19	0	0
6	0,86	1	0
7	0,11	0	0
8	0,34	0	0
9	0,42	0	0
10	0,86	1	1
11	0,62	1	1
12	0,75	1	1
13	0,14	0	0
14	0,25	0	1
15	0,34	0	1
16	0,63	1	0
17	0,72	1	1
18	0,16	0	0
19	0,21	0	1
20	0,17	0	0
21	0,16	0	0
22	0,60	1	1
23	0,36	0	0
24	0,41	0	1
25	0,66	1	1

TABLE 5.1 – Résultat du test de la méthode des forêts aléatoires sur la base de données de métastases pulmonaires. De la gauche vers la droite : identifiant de la tumeur, résultat de l'algorithme, prédiction (arrondi du résultat), valeur cible.

25, la méthode donne la bonne réponse.

Afin de valider la qualité de ce résultat, il est pertinent de calculer la probabilité d'obtenir ce résultat par hasard. L'hypothèse du test est donc la suivante : on a bien classé 18 cas sur 25 totalement par hasard. Si on choisit pour chaque cas un comportement au hasard 0 ou 1 sans *a priori* particulier, on a une chance sur deux de se tromper. La probabilité du nombre de succès lors d'une série de tirages aléatoires est donnée par la loi binomiale. Dans notre cas on calcule donc  $P(X = 18) = \binom{25}{18} p^{18} (1 - p)^7$  où  $p$  est la probabilité d'un succès, c'est à dire  $p = 0,5$ , donc  $P(X = 18) = \binom{25}{18} 0,5^{25} = 0,014$ . On aurait donc moins de 2% de chances d'obtenir un tel résultat par hasard et on rejette cette hypothèse ce qui démontre ici l'utilité de la méthode à des fins prédictives.

De plus, on calcule l'erreur moyenne entre la valeur de la cible et celle renvoyée par le vote des arbres :  $\varepsilon = \frac{1}{25} \sum_{i=1}^{25} |(\text{résultat FA})_i - \text{cible}_i|$  et on trouve  $\varepsilon = 0,38$ . Cela signifie que sur l'ensemble des 5000 arbres construits pour chacune des 25 forêts utilisées dans ce test, 62% ont donné la bonne réponse.

Pour 6 des 7 cas mal classés par la méthode, on constate qu'ils s'agit des cas pour lesquels on a dans la base de données d'autres tumeurs du même patient présentant le comportement opposé. On comprend donc que cela biaise leur évaluation puisque certaines des caractéristiques sont intrinsèques au patient. Le cas restant correspond à une tumeur très agressive difficile à étudier avec notre modèle.

L'utilisation de la méthode des forêts aléatoires semble donc prometteuse pour apporter une solution au problème de détermination de la zone de paramètres à explorer. Les travaux présentés dans cette partie sont préliminaires et représentent davantage une preuve de concept qu'une procédure aboutie et rodée. Il semble notamment important de pousser les tests en constituant une base de données de cas plus importante et en ne conservant dans celle-ci qu'une lésion par patient afin d'éviter de biaiser (en notre défaveur d'ailleurs) les résultats. De plus d'autres types d'informations sur la tumeur devront être ajoutés à la liste des caractéristiques. Il s'agit principalement d'indicateurs permettant de quantifier les hétérogénéités de texture au sein de la lésion.

## 5.4 Conclusion

Dans cette dernière partie, nous avons mis en place un algorithme d'apprentissage automatisé pour exploiter au mieux la base de données de cas cliniques dont nous disposons. Le but était de classer les tumeurs selon leur agressivité et donc d'obtenir une information,

*a priori*, sur l'espace de paramètres du modèle EDP le plus adapté pour fournir une prédiction de la croissance tumorale. Un essai sur une petite base de données a montré l'intérêt de cette approche.

# Chapitre 6

## Conclusion et perspectives

Le but de ces travaux de thèse était de **définir une méthode permettant d'analyser et éventuellement de prédire l'évolution de métastases pulmonaires à partir d'images médicales**. L'ensemble de la démarche a été axé sur les besoins et les habitudes de travail des médecins afin de s'orienter, à terme, vers une utilisation des résultats obtenus dans un cadre clinique.

Afin d'utiliser et de tirer parti de la richesse des informations disponibles dans les images médicales, nous avons utilisé des **modèles d'équations aux dérivées partielles**. Ces modèles tiennent compte des **spécificités mécaniques** des métastases pulmonaires (équation d'advection, loi de Darcy) mais aussi de phénomènes, comme **l'angiogénèse**, qui influencent la dynamique de croissance de la tumeur. De même, ce type de modèle permet de tenir compte de **l'effet cytotoxique** de certains traitements (voir Figure 2.15). Notre objectif étant d'utiliser ces modèles sur des données cliniques, il est important de garder à l'esprit, lors de leur conception, le type et la quantité d'informations disponibles pour la calibration.

Un premier modèle est étudié au Chapitre 2. Il comporte un micro modèle d'angiogénèse tenant compte de la sécrétion d'un agent pro-angiogénique de type VEGF par les cellules tumorales et de l'action de cet agent sur la formation de néo-vascularisation. Nous avons défini un algorithme de calibration adapté, basé sur une fouille de l'espace des paramètres. **L'utilisation d'un modèle réduit par décomposition orthogonale aux valeurs propres (POD) nous a permis des gains considérables en temps de calcul.** Au cours du Chapitre 3, nous avons présenté un modèle simplifié permettant de s'affranchir de la POD et de **définir une procédure de personnalisation à la fois plus fine, notamment dans le choix de l'espace de paramètres, plus robuste et plus rapide**. Ce modèle simplifié permet de reproduire des comportements de croissance tumorale observés cliniquement, mais aussi de fournir une **estimation de l'évolution future** de la tumeur.



Ainsi, la méthode et sa capacité prédictive ont été testées sur un panel de **37 lésions** avec des résultats tout à fait satisfaisants. Cette base de données de différents cas cliniques est une vraie mine d'information pour nous puisqu'elle contient les différentes dynamiques de croissance que l'on peut observer sur des cas de métastases pulmonaires. **Un algorithme d'apprentissage automatisé** (Chapitre 4) a été mis en place pour exploiter au mieux l'ensemble de ces données et en tirer partie lors de l'étude d'un nouveaux cas. Il devrait nous permettre d'affiner la calibration du modèle en tenant compte de données supplémentaires comme l'âge ou le sexe du patient.

La qualité des résultats obtenus sur ces premiers cas test, souligne l'intérêt que pourrait avoir l'utilisation d'outils basés sur cette technique dans un cadre clinique. **Une estimation précoce de l'agressivité d'une tumeur, ou de sa réponse à un traitement, serait une aide conséquente pour le médecin contribuant à l'amélioration du suivi des patients.** En terme d'applications, on peut imaginer à long terme, que cela permettrait d'ajuster plus finement la planification des examens de suivi, voire éventuellement des traitements et interventions. De même, il peut être pertinent d'utiliser ce genre d'outil pour la surveillance d'événements : si l'observation s'écarte de la courbe prédite, il s'est passé quelque chose (mutation, échappement, ...) et le médecin peut réagir plus vite. On peut aussi penser que ce type de méthode pourrait trouver une utilité dans le cadre d'essais cliniques de nouvelles thérapies. Une fois le modèle calibré sur des examens effectués avant traitement, on simule l'évolution future de la tumeur sans traitement. Puis on compare cette évolution à celle, réelle, de la tumeur sous le traitement à l'essai. On pourrait en déduire l'efficacité de la thérapie.

L'originalité de ces travaux provient de l'utilisation d'un **modèle spatial à la fois simple et pertinent d'un point de vue biologique** (prolifération cellulaire, angiogenèse, prise en compte des traitements). De plus **la calibration du modèle sur peu de données** (deux examens suffisent) est une force de cette approche et un atout dans l'optique d'une utilisation clinique.

Dans la continuité de ces travaux, il sera intéressant de poursuivre les essais de la méthode sur de nouveaux cas cliniques. Cela permettra d'évaluer les performances du modèle et de l'algorithme de calibration sur davantage de cas test, mais aussi d'enrichir la base de données d'apprentissage utilisée par la méthode des forêts aléatoires. De même, nous avons assez peu étudié la modélisation des effets de traitements malgré un premier cas test encourageant. Lors de la conception des modèles utilisés pour l'évolution des métastases pulmonaires, nous avons fait en sorte de pouvoir intégrer différents types de traitements : cytotoxiques, anti-angiogéniques... L'évaluation précoce de la réponse aux traitements serait un atout certain pour une utilisation dans

un cadre clinique. Les techniques médicales utilisées en cancérologie évoluent constamment. L'imagerie en est le parfait exemple : amélioration de la définition des images, conception de nouvelles modalités, algorithmes de traitement de l'image (recalage, segmentation automatique...). Étant donné que l'image est le point de départ et l'élément central de nos travaux, une veille sur ces techniques d'imagerie est indispensable pour évoluer et adapter notre approche aux pratiques des médecins. On peut penser, par exemple utiliser des images de la vascularisation de la tumeur, via des I.R.M. de perfusion par exemple, pour initialiser notre modèle et ainsi, peut-être, expliquer des changements de la forme de la lésion dans le temps. Pour cela, la collaboration étroite avec un institut de lutte contre le cancer tel que l'Institut Bergonié est un atout certain. Nous avons démontré ici l'intérêt de la méthode développée pour les métastases pulmonaires pour améliorer le suivi des patients. Il serait intéressant d'étendre ce type d'approche à d'autres pathologies. Cela passe par la conception d'autres modèles adaptés aux différents types de cancers, mais aussi d'algorithmes de personnalisation compatibles avec ces modèles. Des travaux au sein de l'équipe vont dans ce sens puisque d'autres types de cancers sont étudiés (méningiomes, cancers digestifs...).



# Annexe A

## Implémentation d'un schéma de transport de Lax-Wendroff TVD

Les différents modèles EDP de croissance tumorale que j'ai utilisés tout au long de ma thèse sont basés sur le transport des cellules. Il est donc important d'avoir à disposition un schéma de transport adapté à nos besoins, c'est à dire suffisamment précis et peu coûteux en calculs. Les schémas de transport amont et WENO 5 étant déjà codés dans la librairie Cadmos, je me suis intéressé au schéma de Lax-Wendroff (voir [67]).

### A.1 Schéma de Lax-Wendroff dans un cas simplifié

Dans un premier temps, nous allons nous intéresser à une équation de transport à une dimension, à vitesse constante et sans second membre :

$$\frac{\partial u}{\partial t}(t,x) + a \frac{\partial u}{\partial x}(t,x) = 0, \quad (\text{A.1})$$

où  $u(x,t)$  est la grandeur transportée et  $a \in \mathbb{R}$  est la vitesse de transport suivant l'axe  $x$ . À des fins simplificatrices et pour le confort du lecteur, on adopte pour la suite les notations suivantes :  $u = u(x,t)$  et  $u_X = \frac{\partial u}{\partial X}$ . Ainsi, l'équation (A.1) devient :

$$u_t + a \times u_x = 0.$$

On part du développement de Taylor à l'ordre 2 de la fonction  $u$  par rapport à la variable temporelle :

$$u(t+k,x) = u(t,x) + ku_t(t,x) + \frac{k^2}{2}u_{tt}(t,x) + o(k^2), \quad (\text{A.2})$$

où  $k$  représente un pas de temps. Puis, en dérivant l'équation de transport (A.1), on obtient :

$$u_{tt} = -a \times u_{xt}.$$

Or, d'après le théorème de Schwarz et l'équation (A.1),  $u_{xt} = u_{tx} = -a \times u_{xx}$ . On déduit des deux précédentes expressions :

$$u_{tt} = a^2 \times u_{xx}.$$

On peut donc réécrire l'équation (A.2) de la manière suivante :

$$u(t+k, x) = u(t, x) - ak u_x(t, x) + \frac{a^2 k^2}{2} u_{xx}(t, x) + o(k^2).$$

On remplace ensuite les dérivées spatiales par des différences finies centrées d'ordre 2 :

$$\begin{aligned} u(t+k, x) = & u(t, x) - ak \frac{u(t, x+h) - u(t, x-h)}{2h} \\ & + \frac{a^2 k^2}{2} \frac{u(t, x+h) - 2u(t, x) + u(t, x-h)}{h^2} + o(k^2) + o(h^2), \end{aligned}$$

où  $h$  correspond à un pas d'espace. On obtient ainsi le schéma de Lax-Wendroff, en utilisant la notation classique  $u(t_n, x_m) = u_m^n$  :

$$u_m^{n+1} = u_m^n - ak \frac{u_{m+1}^n - u_{m-1}^n}{2h} + \frac{a^2 k^2}{2} \frac{u_{m+1}^n - 2u_m^n + u_{m-1}^n}{h^2}. \quad (\text{A.3})$$

## A.2 Schéma de Lax-Wendroff 2D avec vitesses non constantes

On a vu le principe du schéma de Lax-Wendroff dans un cas très simplifié. Pour pouvoir l'utiliser dans les cas qui nous intéressent, il faut l'adapter au transport bi-dimensionnel avec des champs de vitesse non constants. On notera d'ailleurs dans la suite,  $v^x(t, x, y)$  et  $v^y(t, x, y)$  les vitesses de transport respectivement suivant l'axe des abscisses et des ordonnées. L'équation de transport que l'on considère dans ce paragraphe est donnée par :

$$u_t + \mathbf{v} \cdot \nabla u = 0, \quad (\text{A.4})$$

ou encore,

$$u_t + v^x u_x + v^y u_y = 0.$$

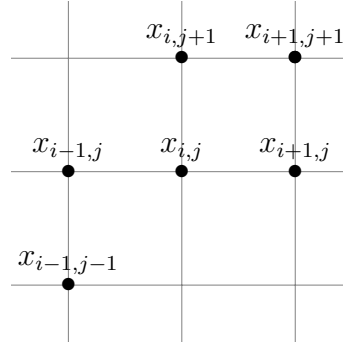


FIGURE A.1 – Grille cartésienne bidimensionnelle. Les noeuds du maillage correspondent avec le centre des pixels.

Puis on procède de même que pour le cas simplifié, on dérive l'expression précédente par rapport aux variables  $t$ ,  $x$  et  $y$  :

$$u_{tt} = -v_t^x u_x - v^x u_{xt} - v_t^y u_y - v^y u_{yt},$$

$$u_{xt} = -v_x^x u_x - v^x u_{xx} - v_x^y u_y - v^y u_{yx},$$

$$u_{yt} = -v_y^x u_x - v^x u_{xy} - v_y^y u_y - v^y u_{yy}.$$

Si on recombine les lignes précédentes, on obtient :

$$\begin{aligned} u_{tt} = & [-v_t^x + v^x v_x^x + v^y v_y^x] \times u_x \\ & + [-v_t^y + v^x v_x^y + v^y v_y^y] \times u_y \\ & + (v^x)^2 u_{xx} + (v^y)^2 u_{yy} + 2v^x v^y u_{xy}. \end{aligned}$$

Ce qui donne, lorsqu'on injecte cette expression dans le développement de Taylor (A.2) :

$$\begin{aligned} u(t+k, x, y) = & u(t, x, y) + [-kv^x + \frac{k^2}{2}(-v_t^x + v^x v_x^x + v^y v_y^x)] \times u_x \\ & + [-kv^y + \frac{k^2}{2}(-v_t^y + v^x v_x^y + v^y v_y^y)] \times u_y \\ & + \frac{k^2}{2}(v^x)^2 u_{xx} + \frac{k^2}{2}(v^y)^2 u_{yy} + k^2 v^x v^y u_{xy} + o(k^2). \end{aligned}$$

En remplaçant les dérivées partielles de  $u$  par les différences finies centrées, on obtient le schéma de transport de Lax-Wendroff en dimension 2 et avec

vitesse non constante, pour  $(i, j) \in [1; N_x - 2] \times [1; N_y - 2]$  :

$$\begin{aligned}
u_{i,j}^{n+1} = & u_{i,j}^n + [-kv^x + \frac{k^2}{2}(-v_t^x + v^x v_x^x + v^y v_y^x)] \times \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2h_x} \\
& + [-kv^y + \frac{k^2}{2}(-v_t^y + v^x v_x^y + v^y v_y^y)] \times \frac{u_{i,j+1}^n - u_{i,j-1}^n}{2h_y} \\
& + \frac{k^2}{2}(v^x)^2 \frac{u_{i-1,j}^n - 2u_{i,j}^n + u_{i+1,j}^n}{h_x^2} + \frac{k^2}{2}(v^y)^2 \frac{u_{i,j-1}^n - 2u_{i,j}^n + u_{i,j+1}^n}{h_y^2} \\
& + k^2 v^x v^y \frac{u_{i-1,j-1}^n - u_{i-1,j+1}^n - u_{i+1,j-1}^n + u_{i+1,j+1}^n}{4h_x h_y},
\end{aligned}$$

où les vitesses et leurs dérivées partielles sont prises au point  $(i, j)$ . Les dérivées partielles en espace des vitesses sont approchées par des différences finies centrées (par exemple  $v_x^x(n, i, j) = \frac{v^x(n, i+1, j) - v^x(n, i-1, j)}{2h_x}$ ) et les dérivées en temps sont approchées par des différences finies décentrées du type  $v_t^x(n, i, j) = \frac{v^x(n, i, j) - v^x(n-1, i, j)}{k}$ .

Aux bords du domaine de calcul, on utilise des différences finies décentrées, par exemple, pour la dérivée partielle par rapport à  $x$  en  $i = 0$ ,  $j$  quelconque, on utilise  $\frac{u_{1,j}^n - u_{0,j}^n}{h_x}$  et en  $i = N_x - 1$ ,  $j$  quelconque, on utilise  $\frac{u_{N_x-1,j}^n - u_{N_x-2,j}^n}{h_x}$ . On procède de même pour les dérivées partielles en  $y$  et pour les dérivées partielles d'ordre 2.

Le schéma ainsi obtenu est d'ordre 2 en temps et en espace. Cela dit, il n'est pas TVD (à variation totale décroissante, voir [26]) et peut faire apparaître des oscillations parasites. Cela est d'autant plus vrai que les champs que nous transportons sont principalement des densités cellulaires (dont les valeurs varient entre 0 et 1) et peuvent présenter des bords raides. Afin de rendre le schéma numérique TVD, on utilise des limiteurs de flux.

### A.3 Limiteurs de flux

Cette partie consiste en une adaptation 2D des travaux présentés dans [26]. Pour que le schéma numérique de Lax-Wendroff ne produise pas d'oscillations parasites, on ajoute des termes correctifs appelés limiteurs de flux, qui prennent la forme suivante :

$$\begin{aligned}
LF_{i,j} = & [K_{i+\frac{1}{2},j}^+(r_i^+) + K_{i+\frac{1}{2},j}^-(r_{i+1}^-)](u_{i+1,j}^n - u_{i,j}^n) \\
& - [K_{i-\frac{1}{2},j}^+(r_{i-1}^+) + K_{i-\frac{1}{2},j}^-(r_i^-)](u_{i,j}^n - u_{i-1,j}^n) \\
& + [K_{i,j+\frac{1}{2}}^+(r_j^+) + K_{i,j+\frac{1}{2}}^-(r_{j+1}^-)](u_{i,j+1}^n - u_{i,j}^n) \\
& - [K_{i,j-\frac{1}{2}}^+(r_{j-1}^+) + K_{i,j-\frac{1}{2}}^-(r_j^-)](u_{i,j}^n - u_{i,j-1}^n),
\end{aligned}$$

avec,

$$r_i^- = \frac{u_{i+1,j}^n - u_{i,j}^n}{u_{i,j}^n - u_{i-1,j}^n}, \quad r_j^- = \frac{u_{i,j+1}^n - u_{i,j}^n}{u_{i,j}^n - u_{i,j-1}^n},$$

$$r_i^+ = \frac{1}{r_i^-}, \quad r_j^+ = \frac{1}{r_j^-},$$

et

$$K_{i-\frac{1}{2},j}^+(r) = \frac{\nu^x(i,j)}{2}(1 - \nu^x(i,j)[1 - \Phi(r)]) \text{ si } v^x(i,j) > 0; 0 \text{ sinon,}$$

$$K_{i,j-\frac{1}{2}}^+(r) = \frac{\nu^y(i,j)}{2}(1 - \nu^y(i,j)[1 - \Phi(r)]) \text{ si } v^y(i,j) > 0; 0 \text{ sinon,}$$

$$K^- = -K^+,$$

où  $\nu^x(i,j) = v^x(i,j)\frac{k}{h_x}$ ,  $\nu^y(i,j) = v^y(i,j)\frac{k}{h_y}$  et  $\Phi(r) = \max(\min(2r,1),0)$ .

## A.4 Essais numériques

Quelques tests comparatifs ont été réalisés sur le schéma numérique avec et sans limiteurs de flux. Pour le test présenté dans ce paragraphe, la vitesse a été prise constante avec  $v_x = 0,6$  et  $v_y = 0,9$ . La condition initiale est l'indicatrice d'une ellipse pour se placer dans un cadre proche de celui du transport de cellules tumorales. Les résultats sont présentés en figures A.2 pour le schéma non TVD et A.3 pour le schéma corrigé TVD.

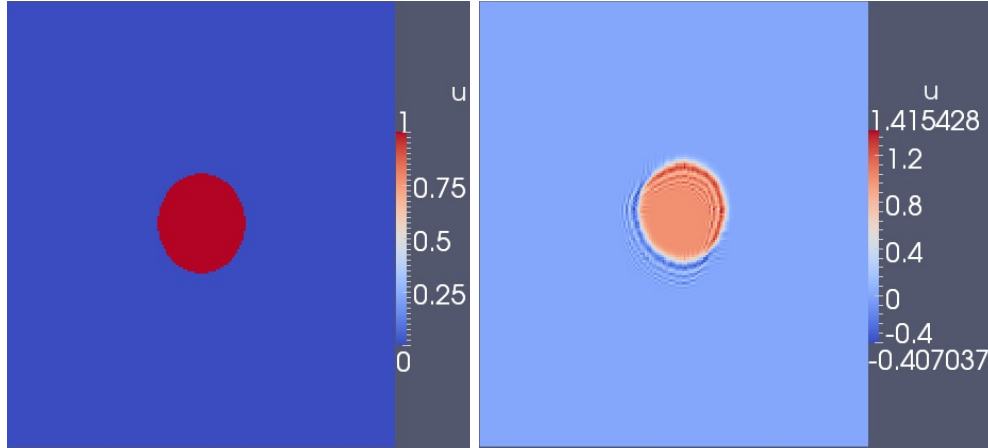


FIGURE A.2 – Simulation numérique de transport d'une ellipse à vitesse constante utilisant le schéma de Lax-Wendroff 2D non TVD. À gauche la condition initiale, à droite le champ transporté.



Le transport d'une quantité discontinue pose manifestement problème au schéma sans limiteurs de flux (voir Figure A.2). Comme on le redoutait, on voit bien apparaître des oscillations non souhaitées. Le champ scalaire dont les valeurs sont initialement comprises entre 0 et 1 prend localement des valeurs en dehors de ces bornes qui ne devraient pourtant pas être franchies, illustrant le caractère non TVD du schéma.

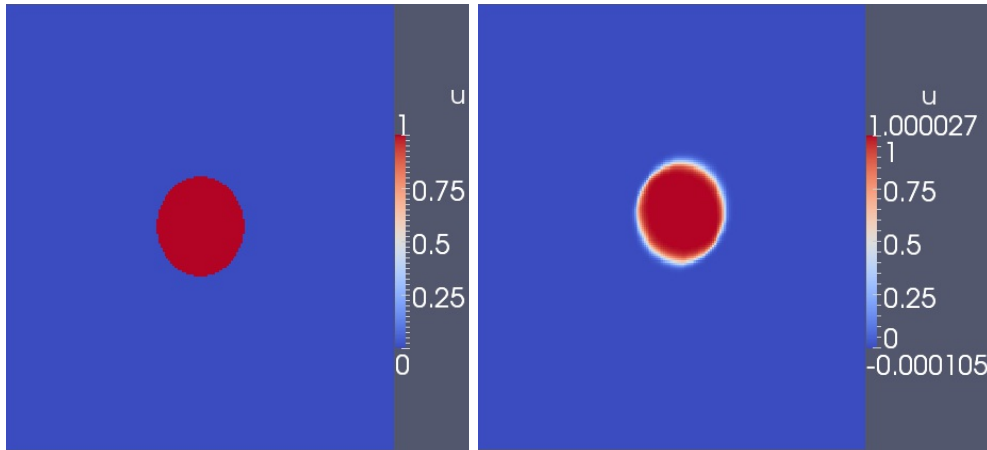


FIGURE A.3 – Simulation numérique de transport d'une ellipse à vitesse constante utilisant le schéma de Lax-Wendroff 2D TVD. À gauche la condition initiale, à droite le champs transporté.

L'ajout des limiteurs de flux (voir Figure A.3) permet d'éviter l'apparition des oscillations, et de conserver un champ scalaire borné entre 0 et 1. Comme on peut le voir, ce schéma d'ordre 2 en temps et en espace produit toutefois de la diffusion numérique.

## Annexe B

### Suite des résultats de l'étude de cas cliniques

Les figures qui suivent présentent les courbes d'évolution de la surface tumorale pour les cas cliniques traités mais non présentés à la section 4.2. Ces résultats sont néanmoins inclus dans la Table 4.1.

## 194 ANNEXE B. SUITE DES RÉSULTATS DE L'ÉTUDE DE CAS CLINIQUES

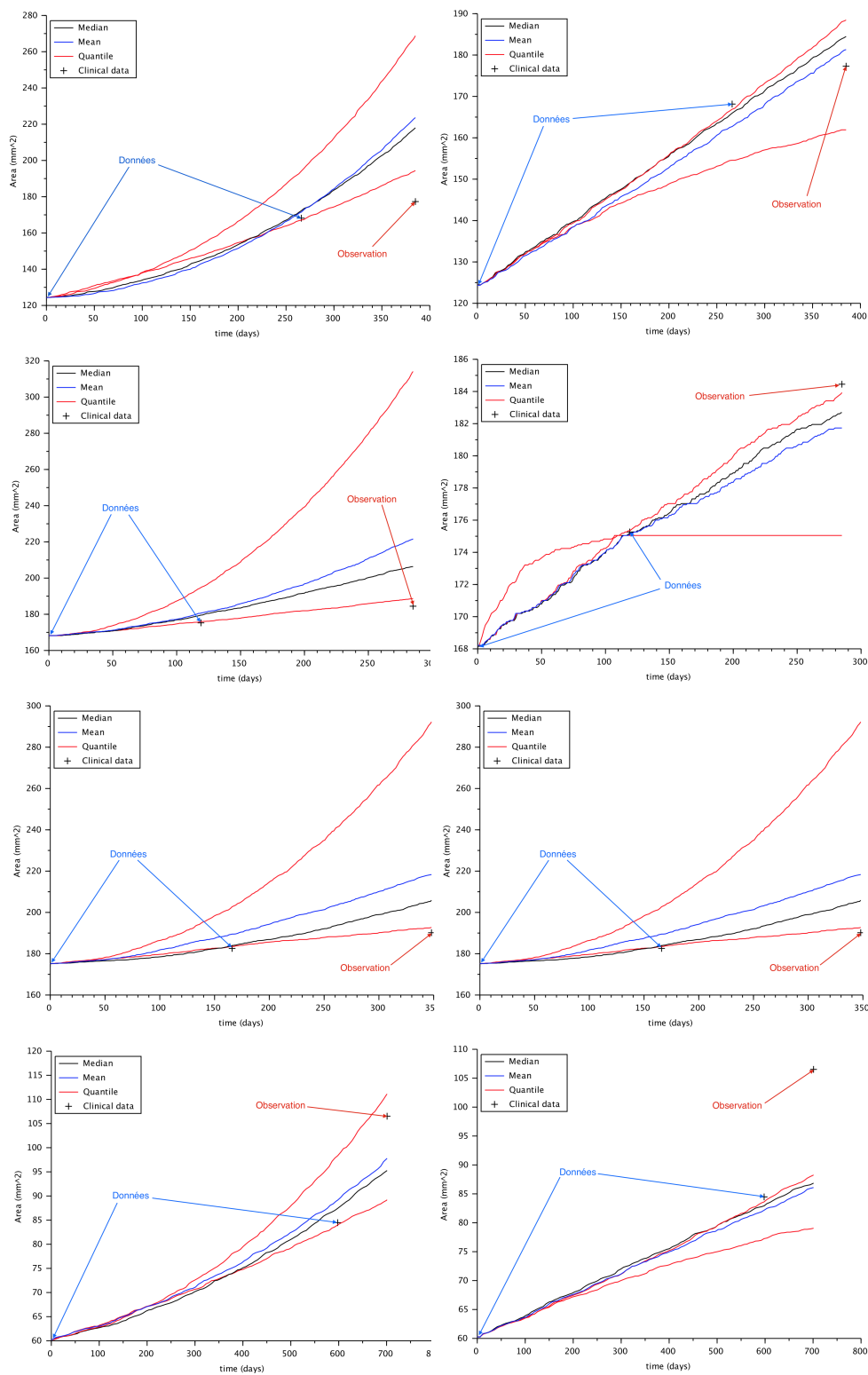


FIGURE B.1 – Cas 9, 10, 11 et 12 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

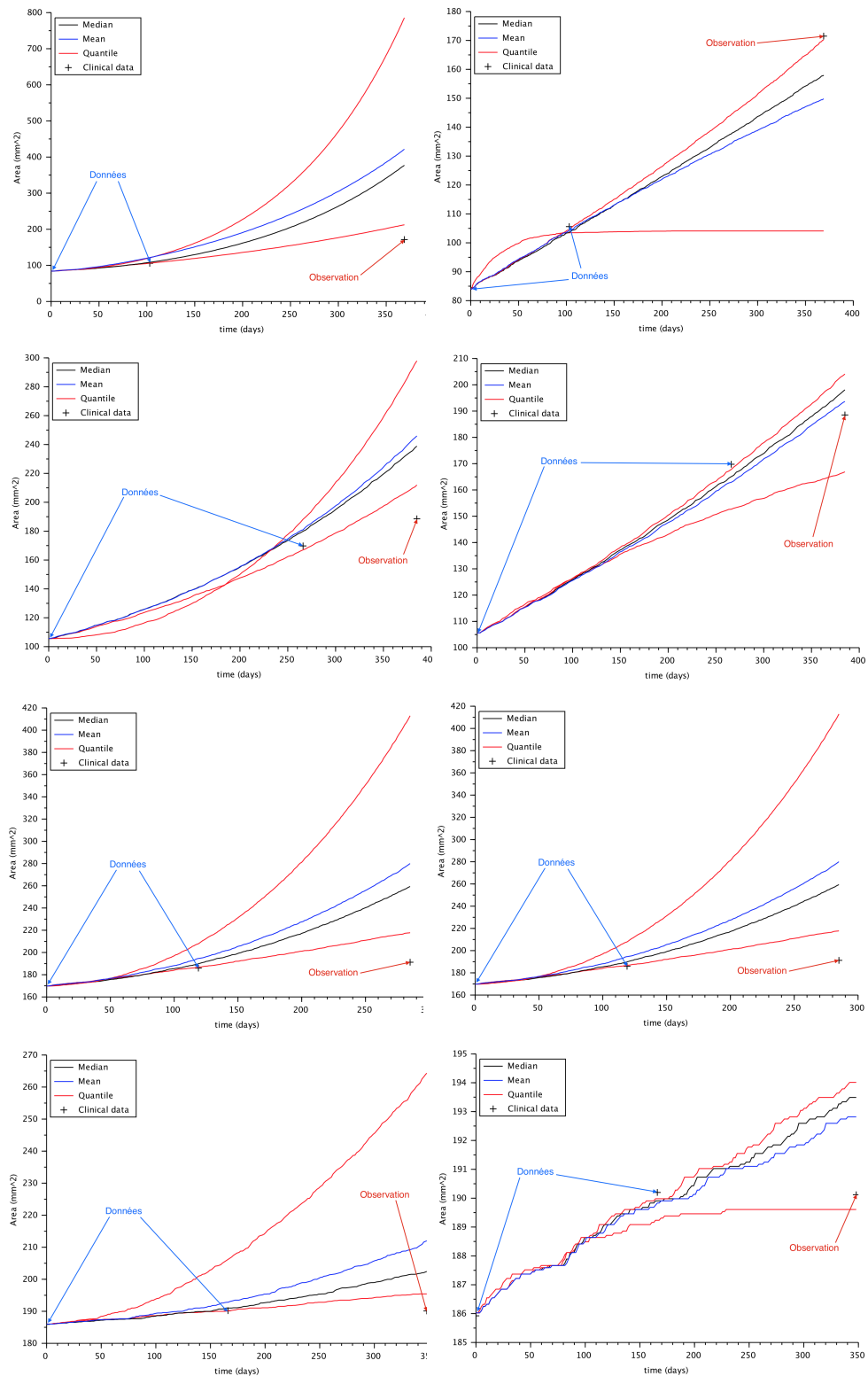


FIGURE B.2 – Cas 13, 14, 15 et 16 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

## 196 ANNEXE B. SUITE DES RÉSULTATS DE L'ÉTUDE DE CAS CLINIQUES

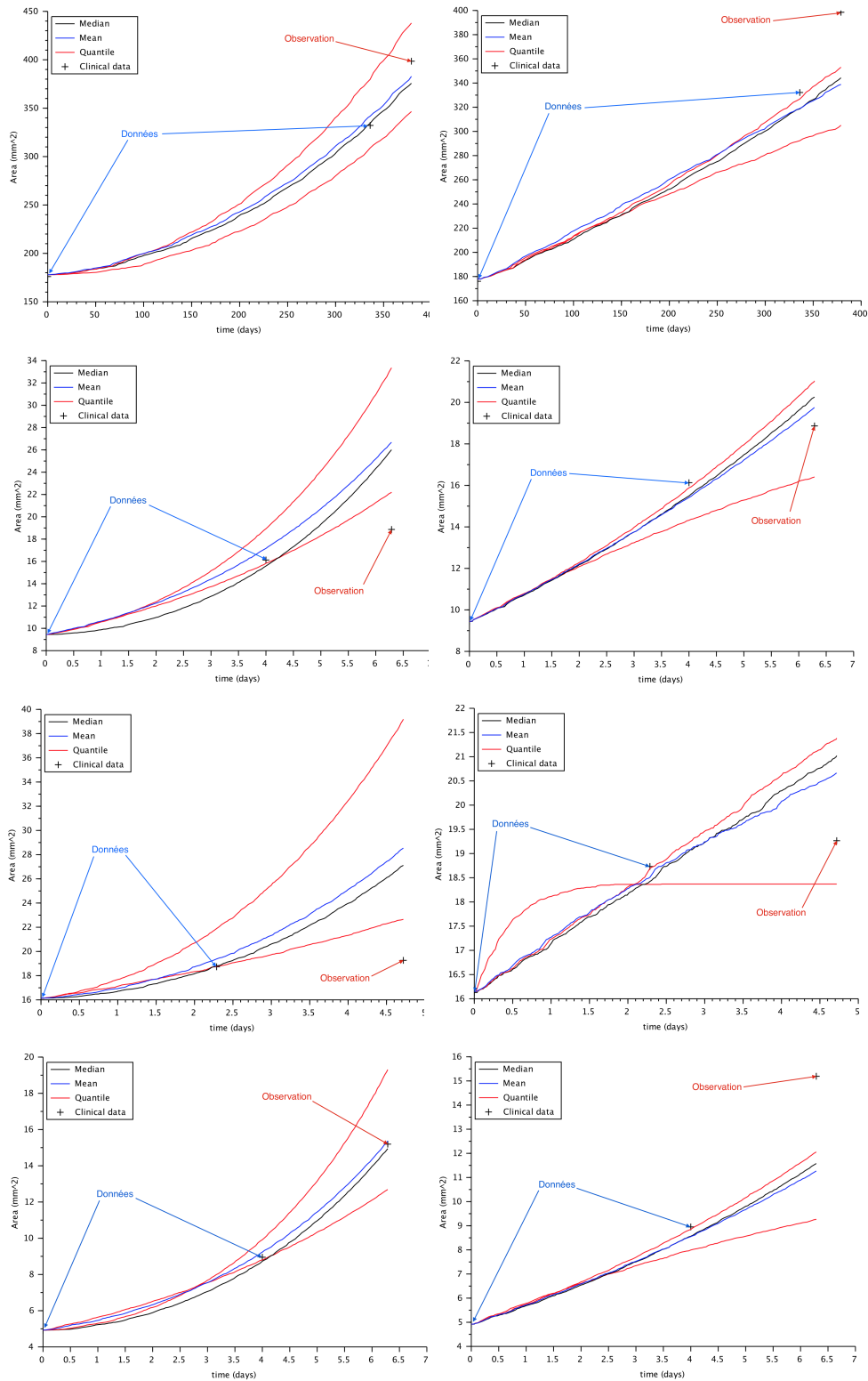


FIGURE B.3 – Cas 17, 18, 19 et 20 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

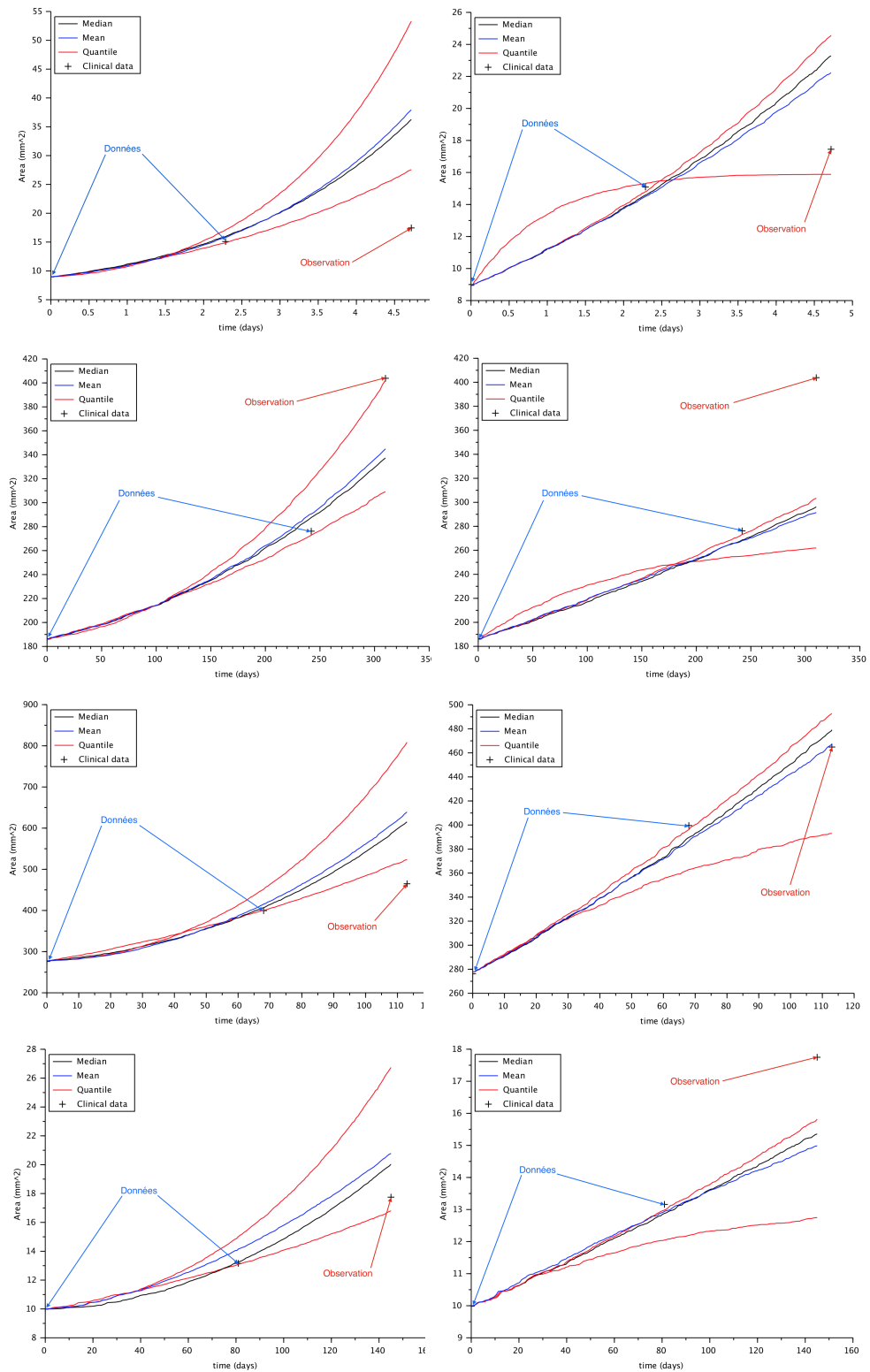


FIGURE B.4 – Cas 21, 22, 23 et 24 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

## 198 ANNEXE B. SUITE DES RÉSULTATS DE L'ÉTUDE DE CAS CLINIQUES

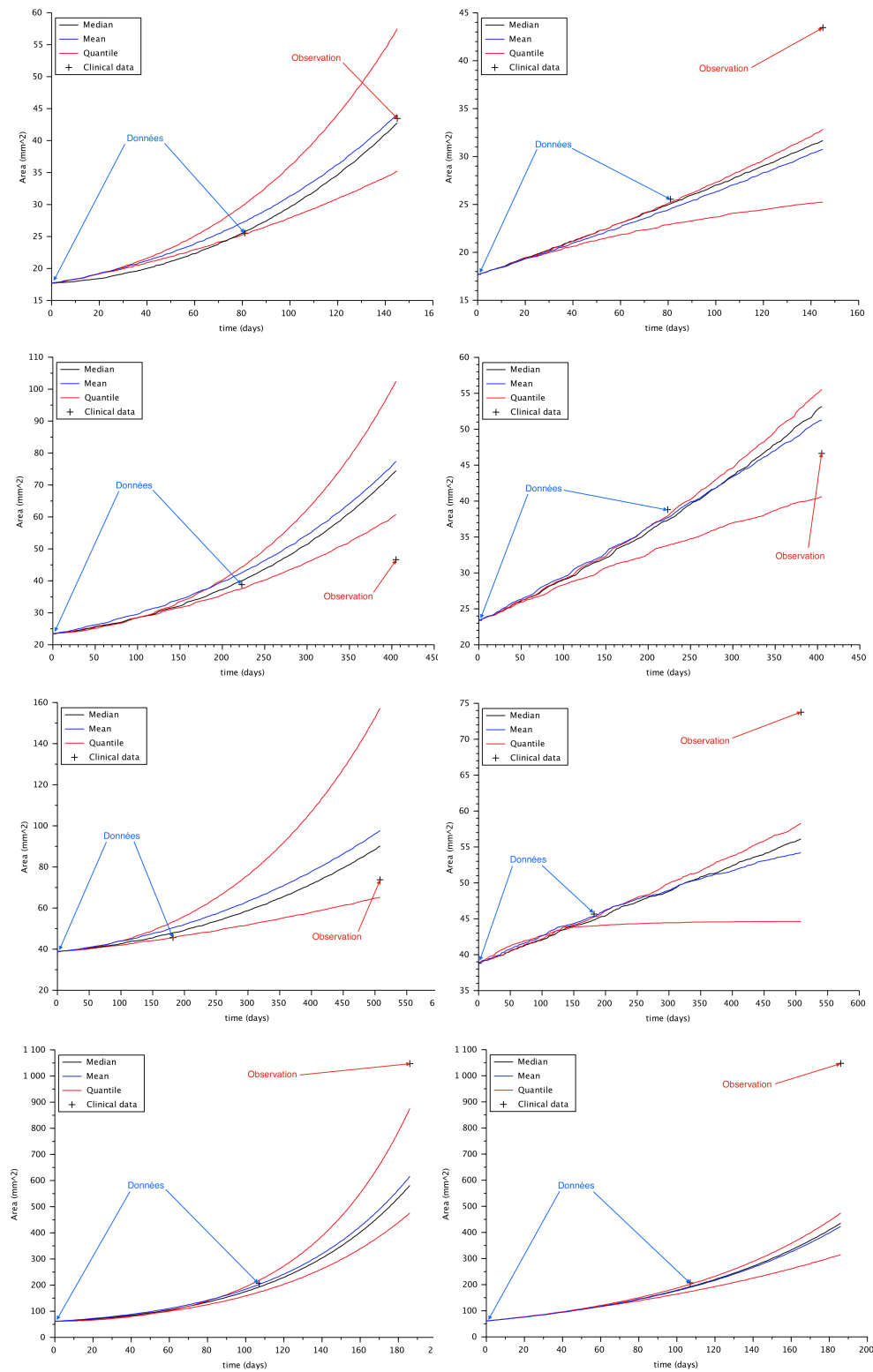


FIGURE B.5 – Cas 25, 26, 27 et 28 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

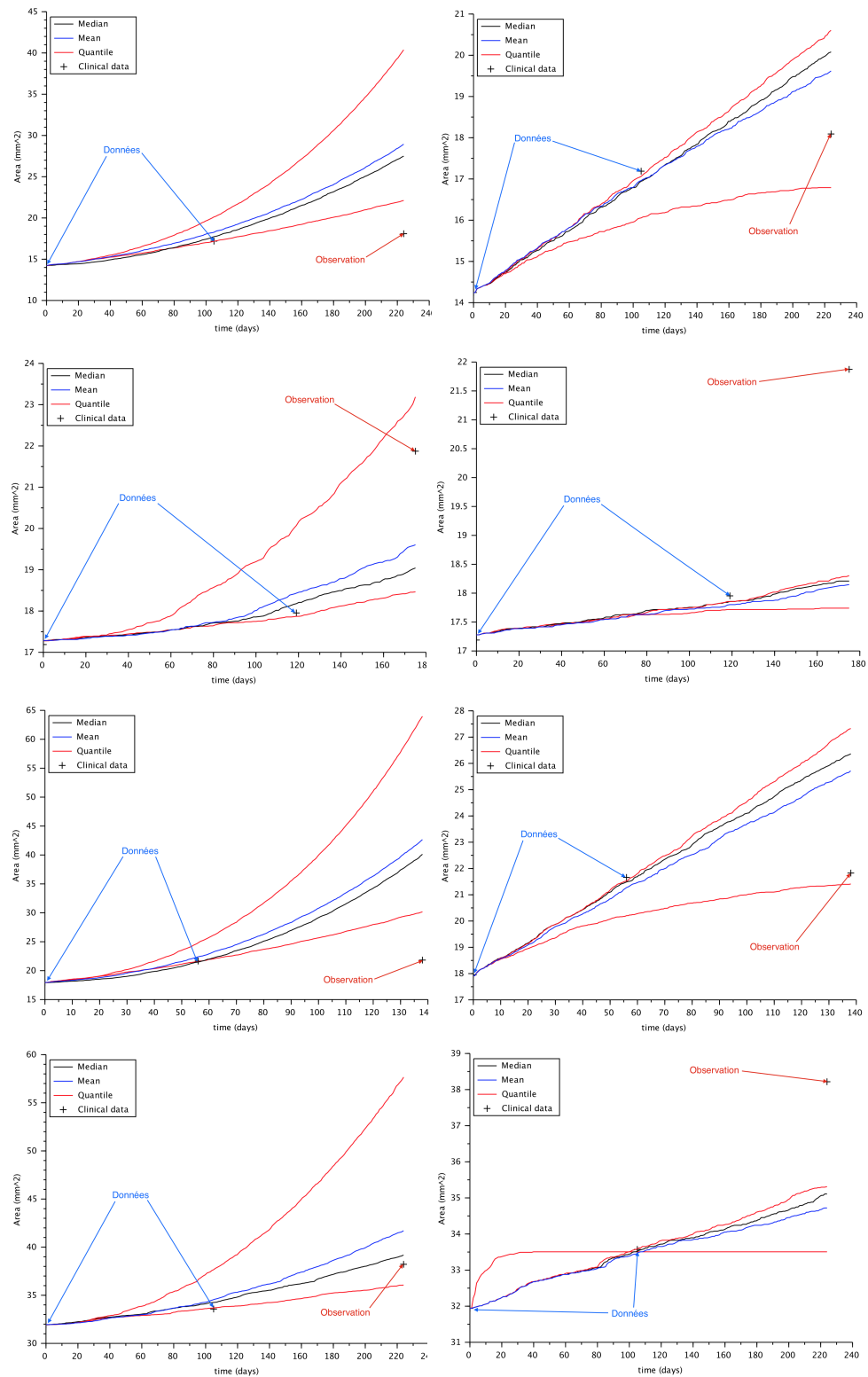


FIGURE B.6 – Cas 29, 30, 31 et 32 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».



## 200 ANNEXE B. SUITE DES RÉSULTATS DE L'ÉTUDE DE CAS CLINIQUES

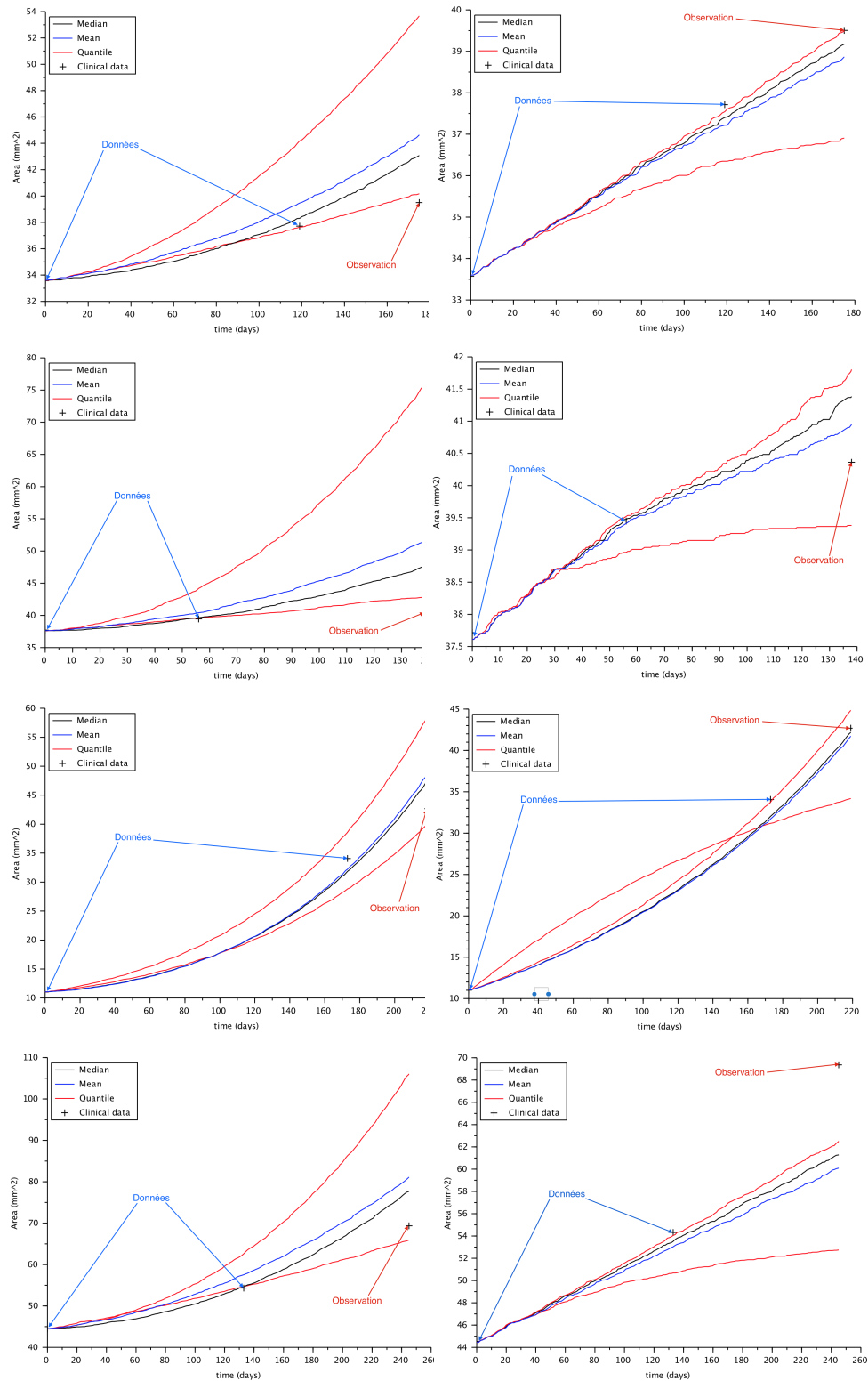


FIGURE B.7 – Cas 33, 34, 35 et 36 (de bas en haut), à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».

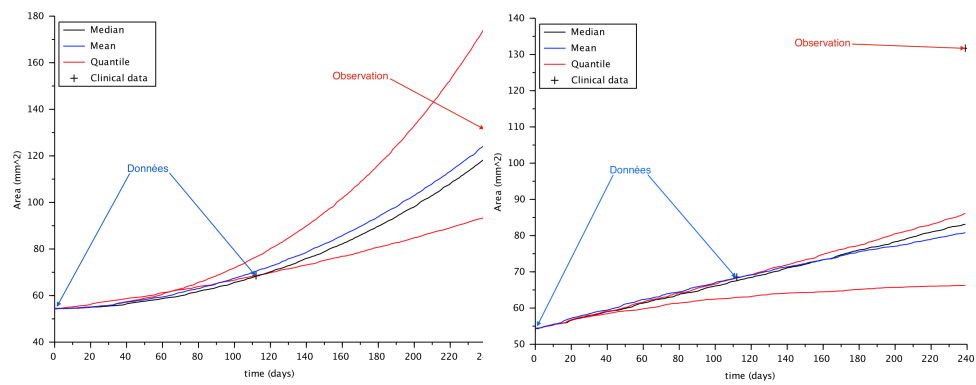


FIGURE B.8 – Cas 37, à gauche la dynamique de type « exponentielle » et à droite la dynamique « plateau ».



# Bibliographie

- [1] Tomás Alarcón, Helen M Byrne, and Philip K Maini. A cellular automaton model for tumour growth in inhomogeneous environment. *Journal of Theoretical Biology*, 225(2) :257–274, 2003.
- [2] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2013.
- [3] Davide Ambrosi and L Preziosi. On the closure of mass balance models for tumor growth. *Mathematical Models and Methods in Applied Sciences*, 12(05) :737–754, 2002.
- [4] Alexander RA Anderson, Alissa M Weaver, Peter T Cummings, and Vito Quaranta. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell*, 127(5) :905–915, 2006.
- [5] M Andreeff, D.W. Goodrich, and H.P. Koeffler. *Cell proliferation and differentiation*. Cancer Medicine, Holland-Frei, 2006.
- [6] Dominique Barbolosi, Assia Benabdallah, Florence Hubert, and Federico Verga. Mathematical and numerical analysis for a model of growing metastatic tumors. *Mathematical biosciences*, 218(1) :1–14, 2009.
- [7] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John ML Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8) :e1003800, 2014.
- [8] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John ML Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8) :e1003800, 2014.

- [9] Michel Bergmann. *Optimisation aérodynamique par réduction de modèle POD et contrôle optimal : application au sillage laminaire d'un cylindre circulaire*. PhD thesis, 2004.
- [10] Leo Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.
- [11] Didier Bresch, Thierry Colin, Emmanuel Grenier, Benjamin Ribba, and Olivier Saut. A viscoelastic model for avascular tumor growth. 2009.
- [12] Didier Bresch, Thierry Colin, Emmanuel Grenier, Benjamin Ribba, and Olivier Saut. Computational modeling of solid tumor growth : The avascular stage. *SIAM Journal on Scientific Computing*, 32(4) :2321–2344, January 2010.
- [13] Fadia Bekkal Brikci, Jean Clairambault, Benjamin Ribba, and Benoît Perthame. An age-and-cyclin-structured cell population model for healthy and tumoral tissues. *Journal of mathematical biology*, 57(1) :91–110, 2008.
- [14] Helen Byrne and Dirk Drasdo. Individual-based and continuum models of growing cell populations : a comparison. *Journal of mathematical biology*, 58(4-5) :657–687, 2009.
- [15] Peter Carmeliet and Rakesh K Jain. Angiogenesis in cancer and other diseases. *Nature*, 407(6801) :249–257, 2000.
- [16] Jean Clairambault, Béatrice Laroche, Stéphane Mischler, Benoît Perthame, et al. A mathematical model of the cell cycle and its control. 2003.
- [17] Olivier Clatz, Pierre-Yves Bondiau, Hervé Delingette, Grégoire Malandain, Maxime Serresant, Simon K Warfield, and Nicholas Ayache. In silico tumor growth : application to glioblastomas. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*, pages 337–345. Springer, 2004.
- [18] Olivier Clatz, Maxime Serresant, P-Y Bondiau, Hervé Delingette, Simon K Warfield, Grégoire Malandain, and Nicholas Ayache. Realistic simulation of the 3-d growth of brain tumors in mr images coupling diffusion with biomechanical deformation. *Medical Imaging, IEEE Transactions on*, 24(10) :1334–1346, 2005.
- [19] T Colin, A Iollo, D Lombardi, and O Saut. Prediction of the evolution of thyroidal lung nodules using a mathematical model. *ERCIM News*, 2010 :37–8, 2010.

- [20] Thierry Colin, François Cornelis, Julien Jouganous, Jean Palussière, and Olivier Saut. Patient-specific simulation of tumor growth, response to the treatment, and relapse of a lung metastasis : a clinical case. *Journal of Computational Surgery*, 2(1) :1–17, 2015.
- [21] Thierry Colin, François Cornelis, Julien Jouganous, Marie Martin, and Olivier Saut. Patient specific image driven evaluation of the aggressiveness of metastases to the lung. In *MICCAI (1)*, pages 553–560, 2014.
- [22] Thierry Colin, Angelo Iollo, Damiano Lombardi, and Olivier Saut. System identification in tumor growth modeling using semi-empirical eigenfunctions. *Mathematical Models and Methods in Applied Sciences*, 22(06), 2012.
- [23] L Corrias, B Perthame, and H Zaag. A chemotaxis model motivated by angiogenesis. *Comptes Rendus Mathématique*, 336(2) :141–146, 2003.
- [24] Antonio Criminisi and Jamie Shotton. *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.
- [25] Antonio Criminisi, Jamie Shotton, and Stefano Bucciarelli. Decision forests with long-range spatial context for organ localization in ct volumes. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 69–80. Citeseer, 2009.
- [26] Stephen F Davis. *TVD finite difference schemes and artificial viscosity*. Number 84. National Aeronautics and Space Administration, Langley Research Center, 1984.
- [27] Thomas S Deisboeck, Zhihui Wang, Paul Macklin, and Vittorio Cristini. Multiscale cancer modeling. *Annual review of biomedical engineering*, 13, 2011.
- [28] D.Lombardi. *Problèmes inverses pour les modèles de croissance tumorale*. PhD thesis, Université Bordeaux 1, 2011.
- [29] Judah Folkman. Role of angiogenesis in tumor growth and metastasis. In *Seminars in oncology*, volume 29, pages 15–18. Elsevier, 2002.
- [30] Robin Genuer. *Forêts aléatoires : aspects théoriques, sélection de variables et applications*. PhD thesis, Université Paris Sud-Paris XI, 2010.

- [31] Ezequiel Geremia, Darko Zikic, Olivier Clatz, BH Menze, Ben Glocker, Ender Konukoglu, Jamie Shotton, OM Thomas, SJ Price, Tilak Das, et al. Classification forests for semantic segmentation of brain lesions in multi-channel mri. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 245–260. Springer, 2013.
- [32] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, pages 513–583, 1825.
- [33] Gaël Guennebaud and Markus Gross. Algebraic point set surfaces. In *ACM Transactions on Graphics (TOG)*, volume 26, page 23. ACM, 2007.
- [34] Philip Hahnfeldt, Dipak Panigrahy, Judah Folkman, and Lynn Hlatky. Tumor development under angiogenic signaling a dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer research*, 59(19) :4770–4775, 1999.
- [35] Hana LP Harpold, Ellsworth C Alvord Jr, and Kristin R Swanson. The evolution of mathematical modeling of glioma proliferation and invasion. *Journal of Neuropathology & Experimental Neurology*, 66(1) :1–9, 2007.
- [36] Cosmina Hoguea, Christos Davatzikos, and George Biros. An image-driven parameter estimation problem for a reaction–diffusion glioma growth model with mass effects. *Journal of mathematical biology*, 56(6) :793–825, 2008.
- [37] Guang-Shan Jiang and Cheng-chin Wu. A high-order weno finite difference scheme for the equations of ideal magnetohydrodynamics. *Journal of Computational Physics*, 150(2) :561–594, 1999.
- [38] K.Kunisch and S.Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90 :117–148, 2001.
- [39] K.Kunisch and S.Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40 :492–515, 2002.
- [40] Ender Konukoglu, Olivier Clatz, Pierre-Yves Bondiau, Hervé Delingette, and Nicholas Ayache. Extrapolating glioma invasion margin in brain magnetic resonance images : Suggesting new irradiation margins. *Medical image analysis*, 14(2) :111–125, 2010.

- [41] Ender Konukoglu, Olivier Clatz, Bjoern H Menze, Bram Stieltjes, Marc-André Weber, Emmanuel Mandonnet, Hervé Delingette, and Nicholas Ayache. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. *Medical Imaging, IEEE Transactions on*, 29(1) :77–95, 2010.
- [42] J Kucwaj. Delaunay triangulation of surfaces. *Zeitschrift für angewandte Mathematik und Mechanik*, 76 :487–488, 1996.
- [43] Karl Kunisch and Stefan Volkwein. Control of the burgers equation by a reduced-order approach using proper orthogonal decomposition. *Journal of Optimization Theory and Applications*, 102(2) :345–371, 1999.
- [44] Jean-Baptiste Lagaert. *Modélisation de la croissance tumorale : estimation de paramètres d’un modèle de croissance et introduction d’un modèle spécifique aux gliomes de tout grade*. PhD thesis, Université Sciences et Technologies-Bordeaux I, 2011.
- [45] Matthieu Lê, Hervé Delingette, Jayashree Kalpathy-Cramer, Elisabeth Gerstner, Helen A Shih, Tracy Batchelor, Jan Unkelbach, and Nicholas Ayache. Multimodal analysis of vasogenic edema in glioblastoma patients for radiotherapy planning. In *Workshop Image-Guided Adaptive Radiation Therapy*.
- [46] Victor Lempitsky, Michael Verhoeck, J Alison Noble, and Andrew Blake. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In *Functional Imaging and Modeling of the Heart*, pages 447–456. Springer, 2009.
- [47] YC Liang, HP Lee, SP Lim, WZ Lin, KH Lee, and CG Wu. Proper orthogonal decomposition and its applications—part i : Theory. *Journal of Sound and Vibration*, 252(3) :527–544, 2002.
- [48] Y Lin, WY Lin, CH Kao, KY Yen, SW Chen, and JJ Yeh. Prognostic value of preoperative metabolic tumor volumes on pet-ct in predicting disease-free survival of patients with stage i non-small cell lung cancer. *Anticancer Res*, 32(11) :5087–91, November 2012.
- [49] Nikos V Mantzaris, Steve Webb, and Hans G Othmer. Mathematical modeling of tumor-induced angiogenesis. *Journal of Mathematical Biology*, 49(2), February 2004.
- [50] Gary H Meisters. Polygons have ears. *American Mathematical Monthly*, pages 648–651, 1975.



- [51] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2) :161–174, 1991.
- [52] Larry Norton. A gompertzian model of human breast cancer growth. *Cancer research*, 48(24 Part 1) :7067–7071, 1988.
- [53] Y Ohno. New applications of magnetic resonance imaging for thoracic oncology. *Seminars in respiratory and critical care medicine*, 35(1) :27–40, February 2014.
- [54] Hahnfeldt P, Panigraphy D, Folkman J, and Hlatky L. Tumor development under angiogenic signaling : a dynamical theory of tumor growth, treatment, response and postvascular dormancy. *Cancer Research*, 59, 1999.
- [55] Aalpen A Patel, Edward T Gawlinski, Susan K Lemieux, and Robert A Gatenby. A cellular automaton model of early tumor growth and invasion : the effects of native tissue vascularity and increased anaerobic tumor metabolism. *Journal of Theoretical Biology*, 213(3) :315–331, 2001.
- [56] Benoît Perthame, Fernando Quirós, and Juan Luis Vázquez. The heleshaw asymptotics for mechanical models of tumor growth. *Archive for Rational Mechanics and Analysis*, 212(1) :93–127, 2014.
- [57] William H Press. *Numerical recipes 3rd edition : The art of scientific computing*. Cambridge university press, 2007.
- [58] M Prokop. Lung cancer screening : The radiologist’s perspective. *Seminars in respiratory and critical care medicine*, 35(1) :91–8, February 2014.
- [59] RM Quiros and WJ Scott. Surgical treatment of metastatic disease to the lung. *Semin Oncol*, 35(2) :134–46, April 2008.
- [60] Benjamin Ribba, Gentian Kaloshi, Mathieu Peyre, Damien Ricard, Vincent Calvez, Michel Tod, Branka Čajavec-Bernard, Ahmed Idbah, Dimitri Psimaras, Linda Dainese, et al. A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clinical Cancer Research*, 18(18) :5071–5080, 2012.
- [61] Jack E Riggs and Ronald J Millecchia. Using the gompertz—strehler model of aging and mortality to explain mortality trends in industrialized countries. *Mechanisms of ageing and development*, 65(2) :217–228, 1992.

- [62] R Rockne, EC Alvord Jr, JK Rockhill, and KR Swanson. A mathematical model for brain tumor response to radiation therapy. *Journal of mathematical biology*, 58(4-5) :561–578, 2009.
- [63] RK Sachs, LR Hlatky, and P Hahnfeldt. Simple ode models of tumor growth and anti-angiogenic or radiation treatment. *Mathematical and Computer Modelling*, 33(12) :1297–1305, 2001.
- [64] Olivier Saut, Jean-Baptiste Lagaert, Thierry Colin, and Hassan M Fathallah-Shaykh. A multilayer grow-or-go model for gbm : effects of invasive cells and anti-angiogenesis on growth. *Bulletin of mathematical biology*, 76(9) :2306–2333, 2014.
- [65] Yi-qing Shen, Ru-quan Wang, and Hong-zhi Liao. A fifth-order accurate weighted enn difference scheme and its applications. *JOURNAL OF COMPUTATIONAL MATHEMATICS-INTERNATIONAL EDITION-*, 19(5) :531–538, 2001.
- [66] CJ Simon, DE Dupuy, TA DiPetrillo, HP Safran, CA Grieco, T Ng, et al. Pulmonary radiofrequency ablation : long-term safety and efficacy in 153 patients. *Radiology*, 243(17392258) :268–75, 2007.
- [67] John C Strikwerda. *Finite difference schemes and partial differential equations*. Siam, 2004.
- [68] Kristin R Swanson, Ellsworth C Alvord Jr, and JD Murray. Quantifying efficacy of chemotherapy of brain tumors with homogeneous and heterogeneous drug delivery. *Acta biotheoretica*, 50(4) :223–237, 2002.
- [69] Kristin R Swanson, Carly Bridge, JD Murray, and Ellsworth C Alvord. Virtual and real brain tumors : using mathematical modeling to quantify glioma growth and invasion. *Journal of the neurological sciences*, 216(1) :1–10, 2003.
- [70] Maurice Tubiana. *Le cancer*. 2003.
- [71] Jan Unkelbach, Bjoern H Menze, Ender Konukoglu, Florian Dittmann, Matthieu Le, Nicholas Ayache, and Helen A Shih. Radiotherapy planning for glioblastoma based on a tumor growth model : improving target volume delineation. *Physics in medicine and biology*, 59(3) :747, 2014.
- [72] EM von Meyenfeldt, GA Gooiker, W van Gijn, PN Post, CJ van de Velde, RA Tollenaar, et al. The relationship between volume or surgeon specialty and outcome in the surgical treatment of lung cancer : a systematic review and meta-analysis. *J Thorac Oncol*, 7(7) :1170–8, July 2012.

- [73] EM von Meyenfeldt, MW Wouters, NL Fat, W Prevoo, SA Burgers, JW van Sandick, et al. Local treatment of pulmonary metastases : from open resection to minimally invasive approach ? less morbidity, comparable local control. *Surgical endoscopy*, 26(8) :2312–21, August 2012.
- [74] Ken CL Wong, Ronald Summers, Electron Kebebew, and Jianhua Yao. Tumor growth prediction with hyperelastic biomechanical model, physiological data fusion, and nonlinear optimization. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 25–32. Springer, 2014.
- [75] Ken CL Wong, Ronald M Summers, Electron Kebebew, and Jianhua Yao. Tumor growth prediction with reaction-diffusion and hyperelastic biomechanical model by physiological data fusion. *Medical image analysis*, 2015.
- [76] SL Wood, M Pernemalm, PA Crosbie, and Whetton AD. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. *Cancer Treat Rev*, October 2013.
- [77] Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, C Demiralp, Jamie Shotton, OM Thomas, T Das, R Jena, and SJ Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 369–376. Springer, 2012.
- [78] Charalambos Zisis, Kosmas Tsakiridis, Ioanna Kougioumtzi, Paul Zarogoulidis, Kaid Darwiche, Nikolaos Machairiotis, Bojan Zaric, Nikolaos Katsikogiannis, Georgios Kesisis, Aikaterini Stylianaki, et al. The management of the advanced colorectal cancer : management of the pulmonary metastases. *Journal of thoracic disease*, 5(Suppl 4) :S383, 2013.