

## Critère quantifiant l'hétérogénéité.

**D**ANS ce chapitre, nous allons construire un critère permettant de quantifier l'hétérogénéité d'une tumeur. A chaque image, on construira l'histogramme des niveaux de gris associés aux tumeurs. Dans un premier temps, ce traitement sera effectué sur les images cliniques. Les histogrammes associés à ces scanners (que l'on appellera par abus de langage « histogrammes cliniques ») seront ensuite étudiés. Plusieurs quantités seront examinées afin de construire un quantificateur de l'hétérogénéité. Dans un second temps, nous appliquerons ce même traitement aux images produites par la simulation numérique du modèle EDP présenté précédemment. Les histogrammes des images produites numériquement (que l'on appellera aussi par abus de langage « histogramme numérique ») seront présentés et comparés aux histogrammes cliniques. De même, le quantificateur de l'hétérogénéité que l'on aura construit sera appliqué aux histogrammes numériques. Nous pourrons ainsi voir à quel point le modèle EDP est capable de reproduire les aspects homogènes et hétérogènes des tumeurs que l'on considère.

### 1.1 Construction des histogrammes de niveaux de gris.

Dans cette section, il s'agit, à partir d'une image donnée en niveaux de gris et d'un contour donné, de reconstruire l'histogramme des niveaux de gris des pixels présents à l'intérieur du contour. Une telle zone est communément appelée ROI (de l'anglais : Region Of Interest). Dans la suite, pour une image donnée, on notera  $p(\mathbf{x})$  la valeur du pixel (comprise entre 0 et 255) situé en position  $\mathbf{x}$ . Ainsi les données de l'histogramme sont représentées par la liste (ensemble avec valeurs multiples autorisées) :

$$X := \{ p(\mathbf{x}) \mid \mathbf{x} \in \text{ROI} \}, \quad (1.1)$$

et l'histogramme lui-même, que l'on normalise, est donné par la fonction :

$$H(x) = \frac{\#X_x}{\#X}, \quad \forall x \in \{0, 1, 2, \dots, 255\} \quad (1.2)$$

où  $x$  désigne un niveau de gris et  $X_x$  désigne la partition de la liste  $X$  qui ne contient que les éléments  $x$  (le symbole  $\#$  désignant le cardinal).

### 1.1.1 Histogrammes cliniques

Les données dont nous disposons sont celles produites par le scanner (cf. Chapitre ?? pour la procédure d'acquisition d'images médicales). Beaucoup plus riches qu'un agglomérat de pixels, ces données (méta-images) au format DICOM, nécessitent l'utilisation d'un outil adapté pour les visualiser. OsiriX est ainsi utilisé pour :

- Choisir une coupe pertinente sur chaque scanner et l'exporter (également au format DICOM) de sorte à avoir ensuite des données 2D à traiter.
- Contourer manuellement la métastase. OsiriX dispose d'un outil crayon adapté à ce type de manipulation. Le contourage réalisé définit alors une ROI, que l'on peut également exporter (au format .xml)

A partir de ces 2 fichiers, un code C++, s'appuyant sur la librairie ITK qui traite entre autre le format DICOM, permet de :

- Construire l'histogramme des niveaux de gris en parcourant l'ensemble des pixels du scanner contenus uniquement dans la ROI. A titre indicatif, l'ensemble des histogrammes cliniques de Patient A et de Patient B sont présentés Figure ?? et ??. Nous les commenterons plus tard.
- Produire des images sur lesquelles le contour est visible. L'ensemble des contourages effectués pour Patient A et pour Patient B sont présentés respectivement Figure ?? et ??.

Remarquons que, comme montré sur la figure ??, le logiciel OsiriX, permet de visualiser directement l'histogramme des niveaux de gris d'une ROI. Cependant, il n'y a aucune possibilité d'exporter ces histogrammes... De plus, on souhaite appliquer autant que possible un traitement similaire aux images cliniques et aux images numériques, images numériques qui ne peuvent être traitées avec OsiriX. Le développement d'un code pour réaliser ceci était donc nécessaire.

Description  
du processus  
médical pour  
les scanners :  
ajout ici ?

Commenter  
les histo-  
grammes

Commenter  
les contou-  
rages ?

### 1.1.2 Histogrammes numériques

Les fichiers de sorties de nos simulations numériques stockent la densité de chaque population de notre modèle EDP, en chacune des mailles du quadrillage que nous avons choisi. A partir de ceux-ci, en utilisant la reconstruction d'images scanners détaillée au chapitre précédent, on peut fournir une image en niveaux de gris. Il reste donc à définir un contour, pour pouvoir en

obtenir les histogrammes numériques en appliquant le même processus que sur les images cliniques. On définit le contour par seuillage sur le tissu sain : si la proportion de tissu sain est inférieure à un pourcentage donné, alors on considérera que l'on est à l'intérieur de la tumeur, sinon on est à l'extérieure. Nous avons donc également dans cas, une image en niveaux de gris et une ROI. Le code C++ pour la partie clinique peut être réutilisé et produira ainsi les histogrammes numériques.

### 1.1.3 Traitements appliqués aux histogrammes : fit par un mélange bi-gaussien

Le but de ce chapitre est de mettre en exergue l'aspect hétérogène de certaines tumeurs. Sur les scanners faisant apparaître une hétérogénéité, on voit clairement deux composantes distinctes de niveaux gris qui se dégagent au sein de la tumeur. Nous avons donc choisi de décrire chacun des histogrammes (que l'on a au préalable normalisés) à l'aide d'une somme de deux gaussiennes :

$$g(x) = g_1(x) + g_2(x) \quad \text{avec} \quad g_i(x) = h_i \exp \left( \frac{-1}{2} \left( \frac{x - c_i}{\sigma_i} \right)^2 \right), \quad (1.3)$$

où :

- $c_i$  est le centre de chacune des gaussiennes,
- $\sigma_i$  est l'écart-type de chacune des composantes,
- $h_i$  est la hauteur de chaque composantes. Elles sont données par :

$$h_i := \frac{w_i}{\sigma_i \sqrt{2\pi}} \quad i = 1; 2, \quad (1.4)$$

où  $w_i$  est le poids associé à chacune des composantes. Notez que les poids sont choisis de telle sorte que :

$$w_1 + w_2 = 1. \quad (1.5)$$

Ainsi, selon les besoins de l'écriture et pour mettre en évidence dans certains cas le nombre de paramètres, on pourra alléger les notations de la manière suivante :

$$w_1 := w \quad \text{et} \quad w_2 = 1 - w. \quad (1.6)$$

De même, lorsqu'on aura besoin d'écrire  $g$  en fonction de ces paramètres sans nécessité de distinction sur la nature de ceux-ci, on écrira  $g(x, \theta)$  avec  $\theta = (c_1, c_2, \sigma_1, \sigma_2, w)$ .

L'optimisation des 5 paramètres  $c_1, c_2, \sigma_1, \sigma_2$  et  $w$  est réalisée grâce à une librairie Python nommée Scikit-learn, qui contient un module dédié aux mélanges gaussiens. Ce module procède d'abord à un partitionnement des données par la méthodes des K-moyennes afin d'estimer les centres des composantes. Le jeu de paramètres résultant est ensuite donné comme point de

départ à une méthode de descente aléatoire qui cherche à maximiser la logvraisemblance. Par algorithme de descente aléatoire, on entend qu'étant donné un jeu de paramètres courants, on réalise les étapes suivantes :

- 1) Un nouveau jeu de paramètres est choisi dans un certain périmètre (plus ou moins grand) autour du jeu de paramètres courants.
- 2) La logvraisemblance de ce nouvel ensemble de paramètres est calculée. Si elle est meilleure que celle du jeu de paramètres courants, alors ce nouveau jeu de paramètres devient le jeu courant. Sinon le jeu de paramètres courants reste celui qu'il était.
- 3) On recommence à 1). On boucle ainsi jusqu'à ce qu'une précision donnée soit atteinte ou que nous ayons atteint le nombre maximum d'itérations voulues.

La vraisemblance, dont on maximise le logarithme selon  $\theta$ , est quant à elle donnée par :

$$\mathcal{L}(X, \theta) = \prod_{x \in X} g(x, \theta). \quad (1.7)$$

Quel que soit  $\theta$ , l'intégrale de  $g$  vaut 1 (*cf.* Propriété ??, page ?? dans l'annexe ??). Ainsi la vraisemblance est le produit des probabilités que chacun des éléments de  $X$  appartiennent à la distribution de paramètre  $\theta$ .

Cet algorithme de mélanges de gaussiennes nous fournit alors la meilleure (au sens de la logvraisemblance) combinaison de gaussiennes qui permet de décrire l'histogramme des niveaux de gris. Ainsi, à chaque image correspond un mélange bi-gaussien totalement identifié. L'ensemble des fits bi-gaussien, pour les images cliniques et numériques, sont présentés sur les Figures ?? et ?? pour les histogrammes cliniques et sur les figures REF FIG pour les histogrammes numériques. Voyons à présent, comment nous pouvons exploiter ceci pour quantifier une hétérogénéité.

Fig fit histo num, on commente les histos ? les fits ?

## 1.2 Définition d'une fonction objectif à reproduire

Img + histo numérique ?

Afin de correctement traduire l'hétérogénéité, il est nécessaire de fournir une fonction objectif que notre critère devra reproduire au mieux. Ainsi, j'ai décidé de catégoriser l'ensemble des scanners de nos patients. Le partage des scanners est ainsi fait en 5 catégories, en associant à chaque catégorie une valeur de l'hétérogénéité  $\mathcal{H}$  :

- $\mathcal{H} = 0.9$  : très hétérogène
- $\mathcal{H} = 0.7$  : plutôt hétérogène
- $\mathcal{H} = 0.5$  : cas intermédiaire ou difficile à caractériser
- $\mathcal{H} = 0.3$  : plutôt homogène
- $\mathcal{H} = 0.1$  : très homogène

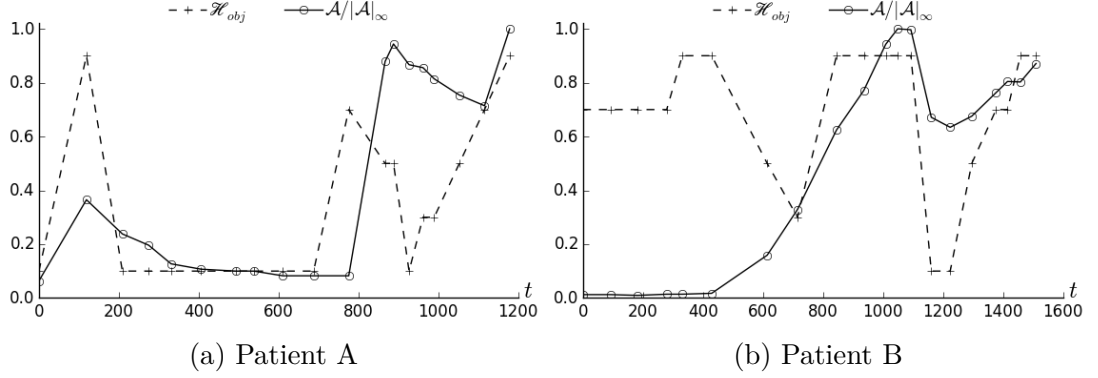


FIGURE 1.1 – Fonction objectif de l'hétérogénéité

**Comment visuellement bien apprécier l'hétérogénéité ?** Pour correctement classer les images dans les 5 catégories que nous venons de dresser, il est nécessaire de savoir précisément ce que l'on entend : à quoi correspond une hétérogénéité de 0 % ? de 100 % ? Une hétérogénéité nulle (*i.e.* homogénéité parfaite) correspond à un aplat d'une seule et unique couleur. A l'opposé, étant données 2 couleurs bien distinctes, on dira que l'hétérogénéité est maximale, si ces 2 couleurs occupent chacune la moitié de la zone considérée. Une hétérogénéité intermédiaire pourra être un cas où :

- soit les 2 couleurs sont relativement proche,
- soit l'une des deux couleurs occupe un faible espace comparé à celui occupé par l'autre couleur,
- soit un mélange des deux raisons précédentes.

En se conformant à ces règles, l'appréciation visuelle<sup>1</sup> des images cliniques nous donne les fonctions objectifs  $\mathcal{H}_{obj}$  (*cf.* Figure 1.1) que notre quantificateur de l'hétérogénéité se devra de reproduire. Notons que Patient A est encore ici un cas très représentatif de ce que l'on cherche à étudier *i.e.* qui montre bien la corrélation entre hétérogénéité et rechute imminente. En effet, ici l'hétérogénéité croît avant même que le volume tumoral ne réaugmente, signe de la reprise d'activité cellulaire sur le pourtour de la métastase. Le cœur reste nécrosé et donc l'hétérogénéité est accrue. Lorsque le volume tumorale finit par augmenter, le tissu proliférant a, en grande partie (si le centre de la tumeur est suffisamment vascularisé), recolonisé la zone nécrosée. La croissance de la métastase est alors synonyme d'homogénéisation, puisque l'ensemble de

1. Cette appréciation visuelle reste ma perception personnelle même si je me suis efforcé de rester le plus objectif possible. Mettre à contribution les membres de l'équipe de recherche par exemple, pour leur demander une catégorisation aurait pu permettre de confronter l'évolution de l'hétérogénéité au cours du temps que je perçois à celle que perçoivent les autres. La fonction objectif finale pourrait ainsi être la moyenne de celles que chacun obtient. On aurait donc un peu plus de nuances : des valeurs intermédiaires aux 5 paliers notamment ainsi que des barres d'erreurs pour chaque valeur.

la surface tumorale tend à être proliférante. Une homogénéisation a également lieu lorsque le traitement est efficace. Dans ce cas-ci, l'ensemble de la tumeur tend à être nécrosé.

Bien que nous ayons 2 patients à notre disposition, je m'efforcerai de construire un critère qui reproduira convenablement la fonction objectif pour Patient A uniquement. Le second patient, Patient B, sera gardé pour valider le ou les critère(s) retenu(s) et non pour le ou les construire. L'idéal serait bien sûr d'avoir à notre disposition une plus large cohorte de patients.

### 1.3 Premiers essais de critères

On considère désormais l'approximation en un mélange de deux gaussiennes des histogrammes de niveaux de gris provenant de nos images. La définition faite de l'hétérogénéité dans la section précédente, nous invite à prendre en compte non pas les positions des gaussiennes, mais plutôt leur écarts. Plus les gaussiennes sont similaires, et plus on tend vers un cas homogène. Notons  $\Delta$  l'opérateur de différence défini par :

$$\Delta : u \longmapsto \Delta u := u_2 - u_1, \quad (1.8)$$

et

$$Q : u \longmapsto Qu := \frac{\min(u_2, u_1)}{\max(u_2, u_1)}. \quad (1.9)$$

l'opérateur de ratio. Ainsi les quantités suivantes pourraient s'avérer intéressantes à étudier :  $\Delta c$ ,  $\Delta \sigma$ ,  $\Delta \sigma^2$ ,  $\Delta h$ ,  $\Delta w$  représentant respectivement l'écart entre les centres, la différence d'écart-type, la différence des variances, la différence des hauteurs et la différence des poids. On pourra aussi regarder leur ratio :  $Qc$ ,  $Qw$  et  $Qh$  notamment. Naïvement, on pourrait penser que ces quantités pourraient être des indicateurs directs de l'hétérogénéité. Examinons donc les informations que fournissent les quantités suivantes :

$$(1.10) \quad \mathcal{H}_1 = \frac{|\Delta c|}{256}, \quad \left| \quad \mathcal{H}'_3 = Qw, \quad (1.13) \right.$$

$$(1.11) \quad \mathcal{H}'_1 = 1 - Qc, \quad \left| \quad \mathcal{H}_8 = \sqrt{2\pi} |\Delta h|, \quad (1.14) \right.$$

$$(1.12) \quad \mathcal{H}_3 = 1 - |\Delta w|, \quad \left| \quad \mathcal{H}'_8 = Qh. \quad (1.15) \right.$$

Notons que les ratios  $Qc$ ,  $Qh$  et  $Qw$  sont par définition entre 0 et 1. Pour les quantités  $\Delta h$  et  $\Delta w$  la propriété suivante nous permet aussi d'assurer la même condition.

**Propriété 1.3.1.** Soit l'intervalle  $I = ]0, a[$  avec  $a > 0$ . Alors :

$$\forall (x, y) \in I^2, |x - y| < a - \min(x, y). \quad (1.16)$$

En particulier, on a :  $\forall (x, y) \in I^2, |x - y| < a$ .

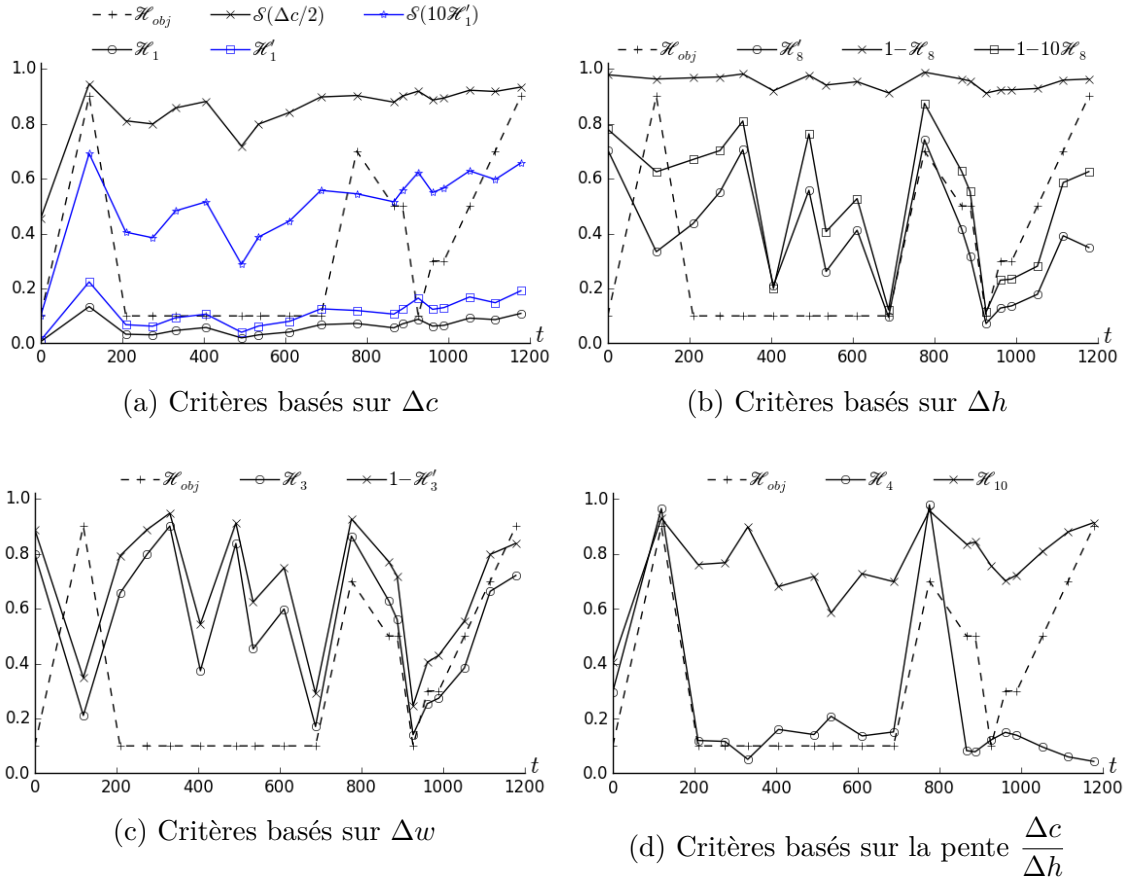


FIGURE 1.2 – Premiers critères

La démonstration de cette propriété est quasiment immédiate par disjonction des cas  $x = y$ ,  $x < y$  et  $x > y$ .

Ainsi, par cette propriété, comme les poids  $w_i$  sont compris entre 0 et 1, alors  $\Delta w$  l'est aussi. De même, comme  $\sigma_i \geq 1$  alors,  $h_i < 1/\sqrt{2\pi}$  et donc  $\Delta h < 1/\sqrt{2\pi}$ . En ce qui concerne  $\Delta c$ , rien n'assure qu'il soit dans l'intervalle  $[0; 1]$ . On le divisera par 256, pour également l'y ramener. De manière plus générale, pour garantir l'appartenance d'une quantité positive à l'intervalle  $[0, 1]$ , on pourra lui appliquer la fonction de saturation :

$$\mathbb{S} : x \mapsto \frac{x}{1+x}. \quad (1.17)$$

La Figure 1.2.abc montre l'évolution des quantités (1.10)-(1.15) au cours du temps, avec application d'une saturation pour certaines. Tout d'abord, on peut remarquer les équivalences suivantes :

$$\mathcal{H}_1 \simeq \mathcal{H}'_1 \quad \text{et} \quad \mathcal{H}_3 \simeq \mathcal{H}'_3 \quad (1.18)$$

Ce comportement était attendu puisque les ratios varient de manière inverse aux différences. Pour  $\mathcal{H}_8 (\sim |\Delta h|)$  et  $\mathcal{H}'_1 (\sim 1 - Qh)$ , on a visiblement pas

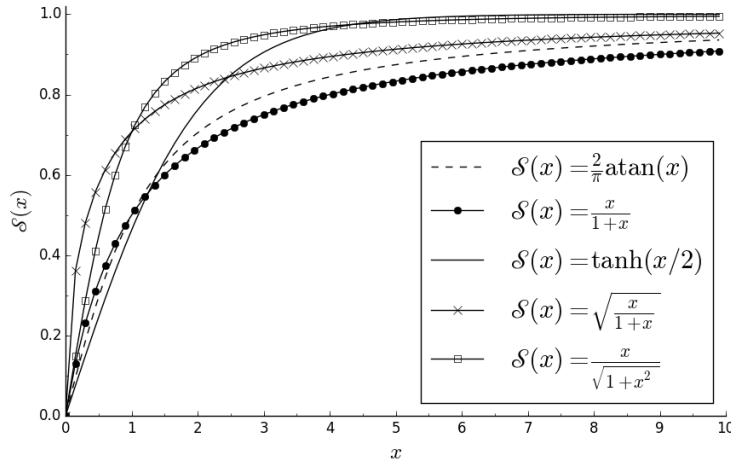


FIGURE 1.3 – Equivalence des différentes saturations

d'équivalence stricte mais les variations du ratio semblent être une dilatation de celles de la différence comme le montre la courbe représentant  $1 - 10\mathcal{H}_8$  sur la Figure 1.2b.

On peut noter également que  $Qh$  et  $Qw$  sont très similaires et reproduisent assez bien la partie sur laquelle le patient est sous antiangiogénique.

Outre ces similitudes, on remarque qu'aucune de ces quantités n'est pertinente pour décrire l'hétérogénéité. Les quantités  $\Delta c/256$  et  $Qc$  n'excèdent pas 20 %. Sur la Figure 1.2a sont aussi présentées deux dilatations de ces courbes. Même après dilatation, l'écart des centres ne semble pas pertinent pour décrire l'hétérogénéité : mis à part le premier pic qui semble capté, le reste de la courbe varie trop peu. Pour ce qui est des quantités  $\mathcal{H}_8$ ,  $\mathcal{H}'_8$ ,  $\mathcal{H}_3$  et  $\mathcal{H}'_3$  l'hétérogénéité sur la partie angiogénique (après le jour 700) semble correctement captée. On ne peut pas en dire autant sur la première partie de la courbe.

Etudions donc de nouveaux critères. Les prochains critères que nous allons étudier sont basés sur l'angle de la pente décrite entre le sommet des deux gaussiennes :

$$\mathcal{H}_4 = \frac{1}{\pi} \operatorname{atan} \left( \frac{\Delta c/256}{\Delta h} \right) + \frac{1}{2} \quad \text{et} \quad \mathcal{H}_{10} = \frac{2}{\pi} \operatorname{atan} \left( \left| \frac{\Delta c/256}{\Delta h} \right| \right). \quad (1.19)$$

Il s'agit de l'inverse de la pente

Notons que l'arctangente, n'est ni plus ni moins qu'une autre manière de saturer une quantité. En effet, comme le montre la Figure 1.3, l'arctangente est proche de la saturation  $\mathcal{S}$  définie par (1.17). Ce graphique montre d'ailleurs également que d'autres saturations sont également possibles et équivalentes à celles utilisées.

l'arctangente ?

Sur la Figure 1.2d, on peut voir que le critère  $\mathcal{H}_4$  recolle relativement bien sur la partie correspondante au premier traitement, rechute incluse (jusqu'au jour 900). Cependant la recroissance finale n'est pas captée. En fait,  $\mathcal{H}_4$  ne décrit pas l'hétérogénéité. Il décrit le ratio proliférant sur nécrose. En effet :



- une pente positive va traduire qu'on a une majorité de proliférantes,
- une pente négative va traduire qu'on a une majorité de tissu nécrosé.

Ce ratio est donc corrélé à l'hétérogénéité, mais pas de manière linéaire puisqu'une tumeur avec 25 % de proliférant et 75 % de nécrose est aussi hétérogène qu'une tumeur présentant 75 % de proliférant et 25 % de nécrose. Nous avons donc besoin d'un critère indépendant du signe de la pente, tel  $\mathcal{H}_{10}$ . Comme on peut le remarquer sur la Figure 1.2d, ce critère n'est encore pas très convaincant en tant que quantificateur de l'hétérogénéité... Les 3 pics à 90 % sont correctement captés, mais les parties basses traduisant des tumeurs bien homogènes ne sont pas bien décrites et ne diffèrent que trop peu des valeurs aux pics. D'autres critères doivent encore être explorés.

## 1.4 Critères basés sur la manière dont s'intersectent les gaussiennes

Avant de proposer divers autres critères, étudions de manière plus précise, la façon dont peuvent s'intersecter deux gaussiennes.

### 1.4.1 Ensemble des configurations d'un mélange bi-gaussien.

Comme le montre la Figure 1.4, deux gaussiennes ne s'intersectent pas nécessairement. De plus, il n'est pas obligatoire d'avoir un point d'intersection dont l'abscisse est située entre  $c_1$  et  $c_2$ . Pour cela, résolvons l'équation suivante :

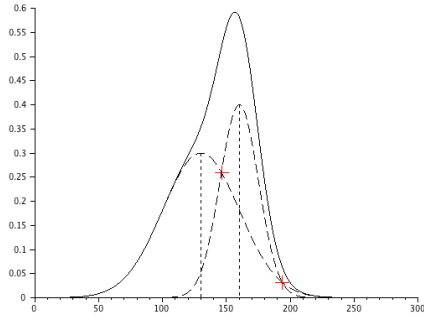
$$\begin{aligned} g_1(x) = g_2(x) &\Leftrightarrow h_1 \exp\left(\frac{-1}{2} \left(\frac{x - c_1}{\sigma_1}\right)^2\right) = h_2 \exp\left(\frac{-1}{2} \left(\frac{x - c_2}{\sigma_2}\right)^2\right), \\ &\Leftrightarrow \ln h_1 - \frac{1}{2} \left(\frac{x - c_1}{\sigma_1}\right)^2 = \ln h_2 - \frac{1}{2} \left(\frac{x - c_2}{\sigma_2}\right)^2, \\ &\stackrel{\sigma_i \neq 0}{\Leftrightarrow} 0 = \sigma_2^2(x - c_1)^2 - \sigma_1^2(x - c_2)^2 + 2\sigma_1^2\sigma_2^2 \ln(h_2/h_1), \end{aligned}$$

qui amène à la résolution d'un polynôme du second degré en  $x$  :

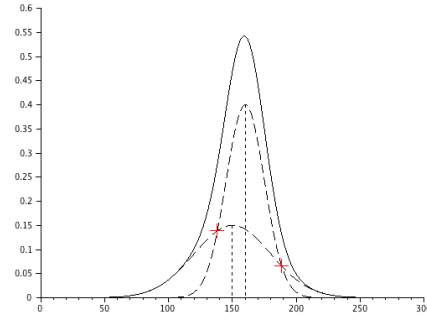
$$\begin{aligned} g_1(x) = g_2(x) &\iff Ax^2 + 2B'x + C = 0 \\ \text{avec : } A &= \sigma_2^2 - \sigma_1^2, \\ B' &= c_2\sigma_1^2 - c_1\sigma_2^2, \\ C &= c_1^2\sigma_2^2 - c_2^2\sigma_1^2 + 2\sigma_1^2\sigma_2^2 \ln(h_2/h_1). \end{aligned} \tag{1.20}$$

**Cas particulier.** Ecartons tout de suite le cas particulier  $\sigma_1 = \sigma_2$ . Dans ce cas, l'équation (1.20) se réécrit :

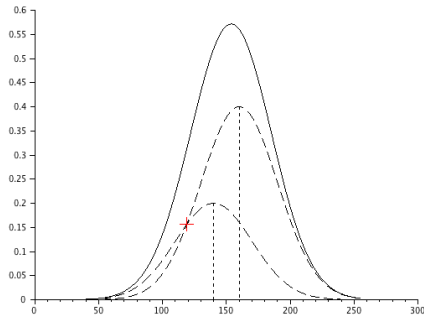
$$2\Delta c x + c_1^2 - c_2^2 + 2\sigma^2 \ln((1-w)/w) = 0 \tag{1.21}$$



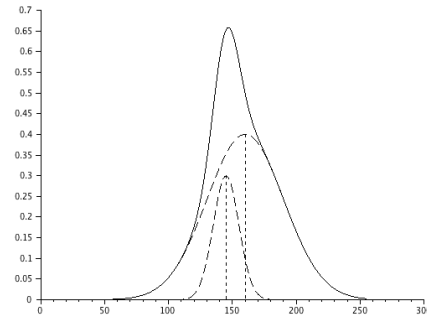
(a) Cas avec 2 intersections, l'une se situant entre les centres des gaussiennes.



(b) Cas avec 2 intersections, toutes les deux en dehors de l'intervalle défini par le centre des gaussiennes.



(c) Cas avec un seul et unique point d'intersection (ici  $\sigma_1 = \sigma_2$ ).



(d) Cas sans aucun point d'intersection.

FIGURE 1.4 – Ensemble des configurations avec 2 gaussiennes.

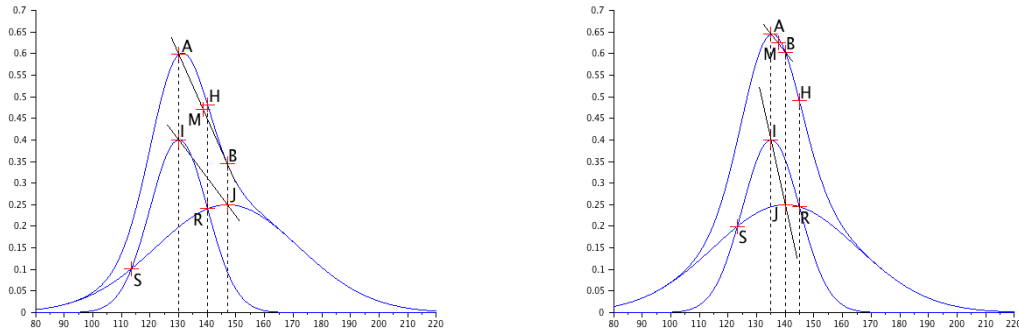
- Si de plus  $\Delta c = 0$ , alors (1.21) implique que  $h_1 = h_2$ , et donc les deux gaussiennes sont absolument identiques et superposées.
- Si  $\Delta c \neq 0$ , alors on a un seul et unique point de croisement, dont l'abscisse est :

$$x = \frac{c_1 + c_2}{2} - \frac{\sigma^2}{\Delta c} \ln \left( \frac{1-w}{w} \right). \quad (1.22)$$

**Cas général.** Il convient ici de calculer le discriminant réduit :

$$\Delta' := B'^2 - AC = \sigma_1^2 \sigma_2^2 [(\Delta c)^2 - 2(\sigma_2^2 - \sigma_1^2) \ln(h_2/h_1)]. \quad (1.23)$$

Ce discriminant n'est pas nécessairement positif! Donc il existe des cas où les gaussiennes ne s'intersectent pas. L'annexe ?? page ?? présente une étude détaillée du signe de ce discriminant (et donc de la manière dont peuvent s'intersecter 2 gaussiennes). On y montre que celui-ci est positif dans une large majorité de cas, ce qui appuie le fait qu'en pratique nous n'ayons jamais



(a) Cas où l'une des intersections se situe entre les centres des gaussiennes ( $R_x \in [c_1; c_2]$ ). (b) Cas où les deux points d'intersection sont à l'extérieur ( $R_x \notin [c_1; c_2]$ ).

FIGURE 1.5 – Points caractéristiques d'un mélange de deux gaussiennes.

été confronté à des cas de gaussiennes ne s'intersectant pas (ou s'intersectant en un seul et unique point). Outre le sommet des gaussiennes, ces points d'intersections sont de nouveaux points caractérisant le mélange. Voyons à présent s'ils peuvent conduire à un critère pertinent.

## 1.4.2 Etudes de différents critères

Différents critères seront étudiés dans cette section. Ils sont basés sur

- l'intégrale commune aux deux gaussiennes
- la valeur de certains angles

### Critères basés sur la valeur d'angles particuliers.

Plutôt que de considérer seulement 2 points (le sommet de chaque gaussienne), élargissons notre éventail de points caractéristiques. La Figure 1.5 présente l'ensemble des points utilisés dans les critères d'évaluation de l'hétérogénéité de cette section. Cet ensemble de points n'existe que dans le cas où les gaussiennes possèdent 2 points d'intersections. Comme montré dans l'annexe ??, les cas avec aucun ou un seul point d'intersection sont relativement marginaux (plus ou moins selon la valeur de  $\Delta c$  notamment). En plus des points  $I$  et  $J$  représentant le sommet de chaque composante ( $I$  représentera toujours le sommet de la plus haute composante), on se servira de  $A$  et  $B$  qui représentent les valeurs du mélange gaussien en les centres des composantes :  $c_1$  et  $c_2$  ( $A$  ayant même abscisse que  $I$ , et  $B$  même abscisse que  $J$ ). On notera également  $R$  et  $S$  les points d'intersections des composantes,  $R$  étant le point le plus haut. On considère également  $H$  positionné sur la courbe du mélange gaussien à la même abscisse que le point  $R$ . Enfin,

on note  $M$  le milieu de  $[AB]$ .

On regardera ici les informations que peuvent fournir l'étude de différents angles. Un large spectre sera examiné :  $\widehat{ARB}$ ,  $\widehat{MRB}$ ,  $\widehat{MRA}$ ,  $\widehat{HRB}$ ,  $\widehat{HRA}$ ,  $\widehat{IRJ}$ ,  $\widehat{MRJ}$ ,  $\widehat{MRI}$ ,  $\widehat{HRJ}$ ,  $\widehat{HRI}$ ,  $\widehat{IRB}$  et  $\widehat{ARJ}$ . On s'attend cependant à ce que certain d'entre eux soit équivalents, notamment ceux qui font intervenir des points sur la même verticale comme  $\widehat{HRA}$  et  $\widehat{HRI}$  par exemple avec *a priori* une variation un peu plus importante du critère d'hétérogénéité qui sera basé sur  $\widehat{HRA}$  que celui basé sur  $\widehat{HRI}$ .

Tous ces angles ne seront pas calculés. On ne s'intéressera uniquement qu'à leur cosinus qui se calcule aisément de la manière suivante avec le produit scalaire :

$$\cos(\widehat{BAC}) = \frac{\vec{AB} \cdot \vec{AC}}{\|\vec{AB}\| \|\vec{AC}\|}. \quad (1.24)$$

Notons que :

- Plus les angles définis ci-dessus sont petits, plus on s'attend à une tumeur homogène. Le critère d'hétérogénéité doit donc avoir des variations inversées par rapport à celle du cosinus.
- Le critère doit être indépendant du signe de l'angle orienté. Tout critère basé sur le cosinus de l'angle respectera ceci, puisque le cosinus est une fonction paire.
- Le cosinus est à valeur dans  $[-1; 1]$ . Le critère doit quant à lui être entre 0 et 1. Il faut donc adapter. Mais attention à la manière d'adapter. L'idée triviale de la valeur absolue ou du carré ne peut pas être appliquée ici. En effet, le signe de l'angle est sans importance mais le signe de son cosinus l'est ! Si l'angle est obtus (donc grand, ce qui traduirait une hétérogénéité) le cosinus de l'angle est négatif et donc il ne faut surtout pas le ramener à son équivalent aigu ! De même on ne veut pas qu'un angle droit traduise un cas limite ( $\mathcal{H} = 0$  ou 1). L'angle droit doit être le cas de transition entre l'angle obtus et l'angle aigu.

On souhaite construire ici un critère qui varie de manière monotone en fonction de l'angle. Nous examinerons deux types d'adaptation d'échelle :

$$\mathcal{H}_5(\theta) = \frac{1 - \cos \theta}{2} \quad (1.25)$$

$$\mathcal{H}_6(\theta) = \sqrt{1 - \left( \frac{1 + \cos \theta}{2} \right)^2} \quad (1.26)$$

Pour beaucoup d'angles les résultats fournis sont insatisfaisants : certains sont quasiment constant ( $\mathcal{H}_{5,6}(\widehat{HRJ})$  par exemple), d'autres sont très chaotiques (comme  $\mathcal{H}_{5,6}(\widehat{ARB})$ ). Sur la Figure 1.6 sont présentés les critères d'hétérogénéité restants (*i.e.* non écartés pour les raisons précédentes). Les critères basés sur l'angle  $\widehat{MRI}$  ou sur l'angle  $\widehat{MRA}$  semblent ne pas bien capturer le premier pic d'hétérogénéité même si le second pic est capté ainsi que la

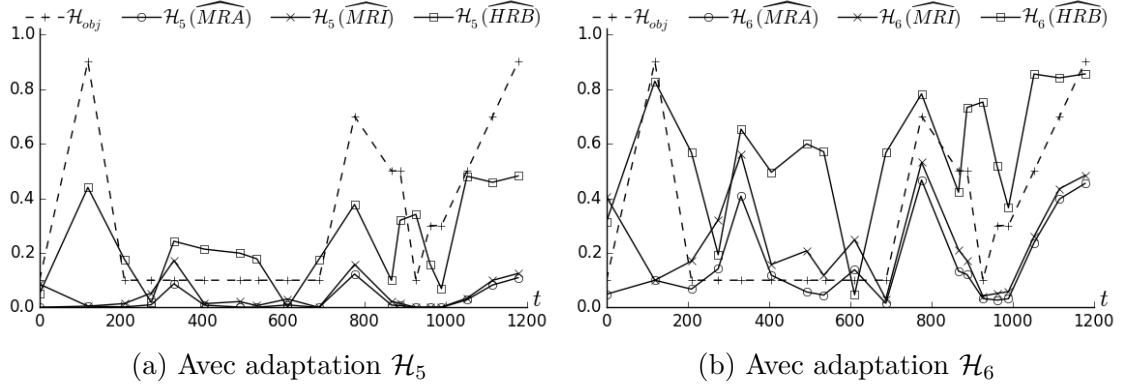


FIGURE 1.6 – Critères basés sur des angles entre points particuliers de l'histogramme des niveaux de gris.

remontée finale. Les critères basés sur l'angle  $\widehat{HRB}$  semblent eux bien capter les moments hétérogènes. Les pics d'hétérogénéité sont cependant assez peu différents de certains passages homogènes... Explorons à présent d'autres critères encore.

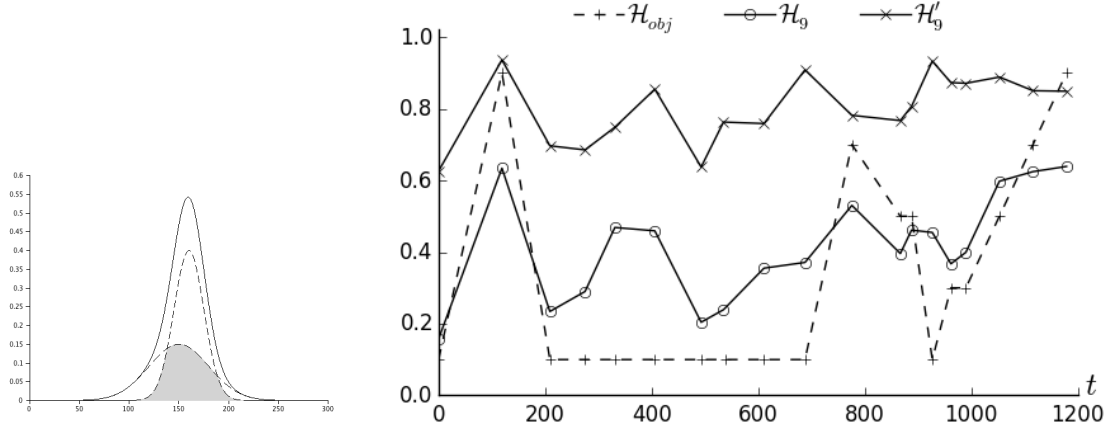
### Critères basés sur des intégrales.

Dans cette section, on va s'intéresser à construire des critères basés sur des comparaisons d'aires. Dans la section 1.3, on a déjà examiné un critère qui compare l'aire des composantes entre elles. Il s'agit du critère  $Qw$ , l'aire (l'intégrale) de la  $i$ -ème composante valant  $w_i$ . Ici, en observant les gaussiennes produites pour Patient A de plus près, on pourrait penser que l'aire commune aux deux gaussiennes (*cf.* schéma représentatif Figure 1.7a) pourrait être un indicateur. Plus l'aire commune aux deux gaussiennes est élevée, plus les deux composantes seraient d'une certaine manière proche, et ainsi plus l'image serait homogène. On s'intéresse ainsi aux critères suivants :

$$\mathcal{H}_9 = 1 - \frac{1}{\min(w_1, w_2)} \int \min(g_1(x), g_2(x)) \, dx, \quad (1.27)$$

$$\mathcal{H}'_9 = 1 - \int \min(g_1(x), g_2(x)) \, dx. \quad (1.28)$$

Les résultats pour ces deux critères sont présentés Figure 1.7b. Considérer l'aire commune relativement à l'aire de la plus petite des gaussiennes ( $\mathcal{H}_9$ ) semble plus pertinent que de considérer l'aire commune seulement ( $\mathcal{H}'_9$ ). Ce critère n'est pas des plus mauvais : l'ensemble des moments hétérogènes est capturé (premier pic avant traitement jour 119, second et troisième pics avant les rechutes (jour 776 et 1116)). Cependant les homogénéisations ne sont pas très bien capturées : il y a notamment le plateau lors du premier traitement et le pic descendant lors du second traitement.



(a) Schéma représentatif de l'aire commune à deux gaussiennes.

(b) Hétérogénéité fournie avec les critères  $\mathcal{H}_9$  et  $\mathcal{H}'_9$ .

FIGURE 1.7 – Critères basés sur l'aire commune aux deux gaussiennes (donnée par :  $\int_0^{255} \min(g_1(x), g_2(x)) dx$ ).

## 1.5 Critère retenu

L'idée de ce dernier critère m'est venue de la constatation suivante, en repartant du critère  $\mathcal{H}_{10}$  basé sur la pente décrite entre le sommet des gaussiennes. Sur la Figure 1.8 sont présentées deux configurations très différentes, mais présentant la même pente. Pourtant la Figure 1.8a est très clairement représentative d'une image hétérogène alors que la Figure 1.8a serait plutôt représentative de quelque chose d'homogène puisque l'approximation par une seule et unique gaussienne ne serait pas des plus mauvaise. Comment différencier ces deux cas ? Cet exemple mis en exergue nous invite à dire que  $\Delta c$  doit avoir plus de poids que  $\Delta h$  dans le calcul du critère de l'hétérogénéité. Ainsi, j'ai décidé de regarder le critère suivant :

$$\mathcal{H}_2 = \left| \frac{(\Delta c / 256)^2}{\Delta h} \right| \quad (1.29)$$

L'hétérogénéité clinique, pour Patient A, fournie par le critère  $\mathcal{H}_2$  est présentée sur la Figure 1.9a. Ici, le critère reproduit bien les pics d'hétérogénéité (jour 119 et jour 776). La regain d'hétérogénéité final, qui démarre avant la rechute au second traitement, et qui se poursuit pendant la rechute, est également bien capturé. Les phases homogènes sont également relativement bien reproduites. L'évaluation visuelle qui amenait à considérer l'hétérogénéité constante du jour 209 au jour 688 est ainsi précisée par le critère. Manifestement le jour 331 est donc légèrement plus hétérogène. Notons que bien qu'écartés, plusieurs autres critères (notamment  $\mathcal{H}_{10}$  et  $\mathcal{H}_6(\widehat{MRA})$ ) avaient aussi capté cette subtilité. Enfin la forte homogénéisation causée par le second

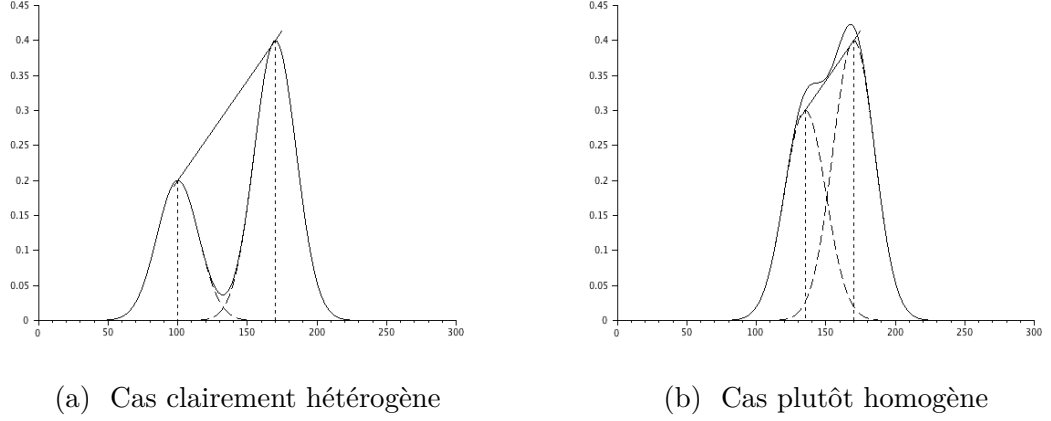


FIGURE 1.8 – Deux configurations très différentes mais fournissant la même pente entre les gaussiennes.

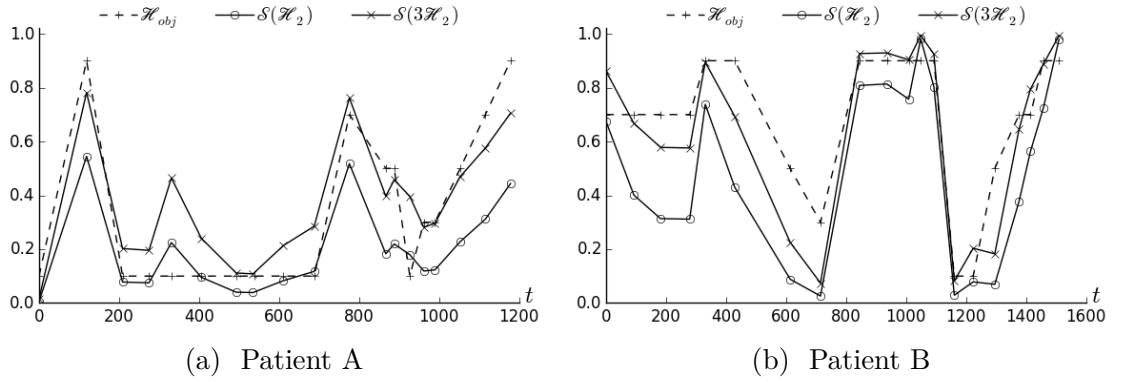


FIGURE 1.9 – Hétérogénéité clinique fournie par le critère  $\mathcal{H}_{12}$  dans lequel  $\Delta c$  joue un rôle prépondérant.

traitement est également bien traduite. Le seul bémol réside dans le fait que ce critère semble avoir des difficultés à monter vers les grandes hétérogénéité. Multiplier ce critère par 3 semble corriger le problème. Mais pourquoi 3 ? Si l'on regarde de plus près les histogrammes clinique de Patient A, on peut remarquer que les histogrammes sont grossièrement tous compris dans l'intervalle  $[75; 220]$ . La longueur de cet intervalle est de 145, et non 256. Ainsi si au lieu de diviser  $\Delta c$  par 256, on le divise par 145, on remarque alors :

$$3 \left( \frac{\Delta c}{256} \right)^2 = \left( \frac{\Delta c}{256\sqrt{3}} \right)^2 \approx \left( \frac{\Delta c}{145} \right)^2. \quad (1.30)$$

Nous avons donc trouvé un critère qui semble quantifier de manière acceptable l'hétérogénéité clinique de Patient A. Cela dit ce critère a été construit dans ce but ! Afin de le valider, regardons ce qu'il en est sur notre second patient : Patient B.

L'hétérogénéité clinique, pour Patient B, fournie par le critère  $\mathcal{H}_2$  est présentée sur la Figure 1.9b. Ici aussi, la retranscription est aussi tout à fait acceptable. Les phases où il y a regain d'hétérogénéité sont toutes correctement décrites. Au tout début (jusqu'au jour 279) la métastase est très petite. Il est donc difficile d'apprécier visuellement l'hétérogénéité. De plus, le quantificateur sera aussi plus sensible aux erreurs dues au contourage. La tendance réelle à une baisse de l'hétérogénéité est tout à fait plausible. D'autant plus que c'est le comportement attendu lorsqu'on traite de manière efficace : l'ensemble de la métastase tend à se nécroser, et donc l'ensemble s'homogénéise. L'homogénéisation lors du second traitement est également très bien capturée. Les pics d'hétérogénéité (du jour 800 à 1100 puis remontée finale) sont aussi bien traduits. De même que pour Patient A, le critère  $\mathcal{S}(3\mathcal{H}_2)$  s'approche plus de l'objectif que le critère  $\mathcal{S}(\mathcal{H}_2)$ . Bien qu'étant plus larges, les histogrammes cliniques de Patient B n'occupent pas l'ensemble de l'intervalle  $[0; 256]$ . Si l'on se restreint à l'intervalle  $[75; 220]$ , ce qui est en dehors reste très marginal et la multiplication par 3 se justifie également ici.

## 1.6 L'hétérogénéité sur les simulations numériques

Maintenant que nous avons un critère qui décrit correctement l'hétérogénéité clinique (d'une métastase à partir de l'imagerie médicale), faisons parler ce critère sur nos simulations numériques. En ce qui concerne cet aspect, les images résultantes (gouvernées par EQREF ) des simulations numériques dépendent de 3 paramètres :  $\tau_N, \tau_P$  et  $\tau_S$  qui représentent les niveaux de gris associés à chacune de nos populations de notre modèle EDP. Ainsi, pour une simulation numérique donnée, il n'y a pas unicité de l'image produite en niveau de gris, et donc non unicité de l'histogramme. Tout dépend de ces 3 paramètres. Dans un premier temps, on examinera ce que cela donne avec les valeurs heuristiques considérées dans la première partie de ce manuscrit :  $\tau_N = 38, \tau_P = 166$  et  $\tau_S = 204$ . Dans un second temps, on pourra faire varier ces paramètres pour examiner l'influence de ceux-ci sur l'hétérogénéité numérique. On examinera notamment ce qui est produit avec les niveaux de gris optimaux du chapitre précédent.

eqref

### 1.6.1 Patient A

L'hétérogénéité numérique de Patient A est présentée la Figure 1.10. La fonction objectif pour l'hétérogénéité clinique est ici rappelée sur ce graphique à titre comparatif. La phase avec imatinib est correctement décrite :

- Présence d'un pic d'hétérogénéité jour 119.



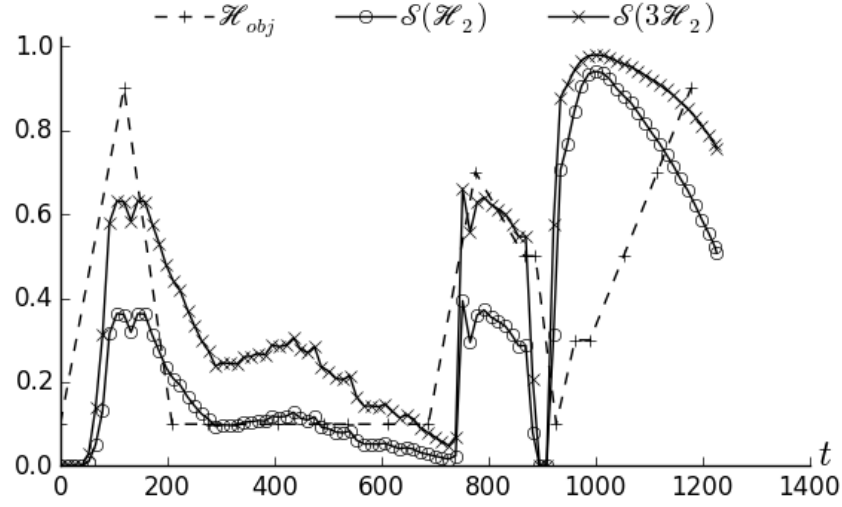


FIGURE 1.10 – Hétérogénéité numérique pour Patient A –  $\tau_N = 38$ ,  $\tau_P = 166$  et  $\tau_S = 215$ . La fonction objectif de l'hétérogénéité clinique ( $\mathcal{H}_{obj}$ ) est rappelée ici à titre de comparaison.

- Décroissance de l'hétérogénéité lorsque l'imatinib agit de manière efficace du jour 200 au jour 700.
- Saut important de l'hétérogénéité qui grandit, autour du jour 800, juste avant la recroissance de l'aire tumorale.

En ce qui concerne la partie avec sunitinib, au début de l'administration du traitement l'hétérogénéité décroît. Cependant :

- La recroissance de l'hétérogénéité numérique a lieu un peu tôt par rapport à celle constatée cliniquement.
- Sur la partie finale (lors de la rechute au sunitinib, après le jour 1116), l'hétérogénéité numérique décroît alors que celle clinique continue d'augmenter.

En ce qui concerne le deuxième point, cela peut venir soit de la manière dont on calcule l'hétérogénéité, soit du modèle EDP lui-même qui ne retranscrirait pas bien l'évolution de l'hétérogénéité. La Figure 1.11 tend à dire que c'est plutôt le modèle EDP qui est responsable puisque le scanner numérique du jour 1120 est beaucoup plus hétérogène que celui du jour 1227. En effet, le contraste entre les deux masses dominantes (pourtour et intérieur de la tumeur) est beaucoup plus important jour 1120 que jour 1227. De plus le rapport du volume de ces dominantes est beaucoup plus proche de 1 au jour 1120 qu'au jour 1227 (si le ratio est égal à 1 alors les 2 nuances de gris sont de volume égal). Ces impressions visuelles sont confirmées par les histogrammes également présentés sur la Figure 1.11. Tout ceci renforce donc l'idée que l'image numérique du jour 1120 est plus hétérogène que celle du jour 1227. Le modèle EDP semblerait donc bien retranscrire globalement les différentes phases hétérogènes/ homogènes. Après le jour 950, l'hétérogénéité numérique

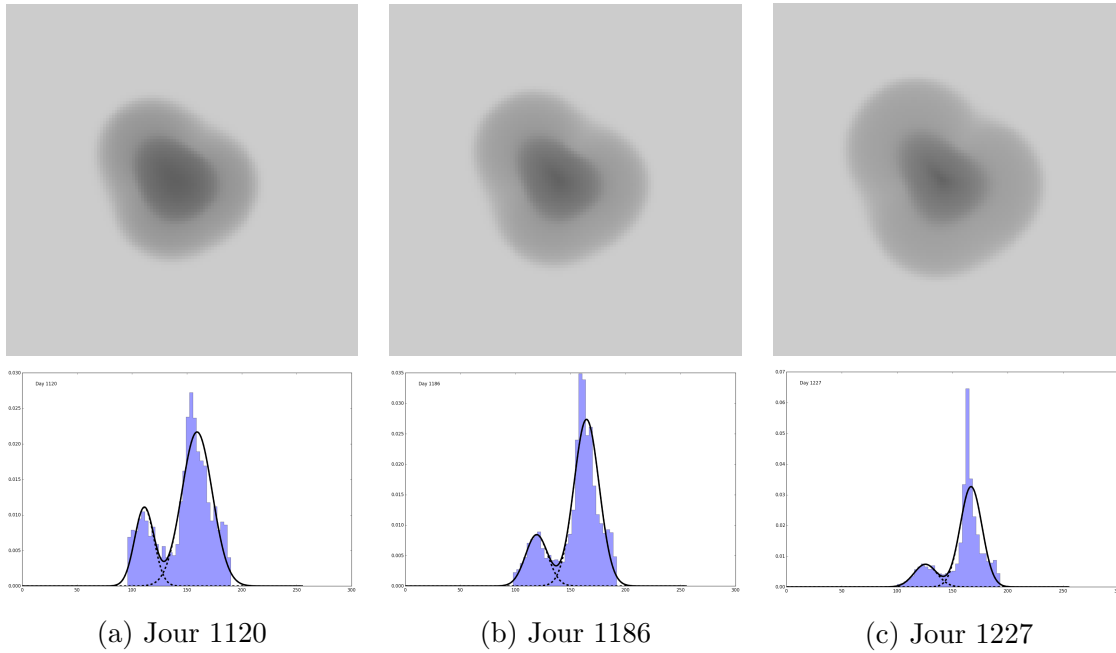


FIGURE 1.11 – Hétérogénéité numérique de Patient A.

ne semble plus correctement décrire l'hétérogénéité clinique. Ceci n'est guère surprenant. En effet, un grand nombre d'itérations est effectué dans le calcul numérique pour parvenir au temps final. Les erreurs se cumulant au fil des itérations, il n'est donc pas étonnant qu'à un moment donné l'hétérogénéité numérique ne recolle pas complètement à l'hétérogénéité clinique.

De plus les images numériques sont dépendantes du choix des niveaux de gris  $\tau_N$ ,  $\tau_P$  et  $\tau_S$ . Ce choix pourrait aussi être une source d'écart entre le numérique et les données.

### 1.6.2 Influence des niveaux de gris sur l'hétérogénéité numérique

La principale conséquence du changement des niveaux de gris  $\tau_N$ ,  $\tau_P$  et  $\tau_S$  est la dilatation de l'histogramme des niveaux de gris. Les variations de l'hétérogénéité ne sont donc que peu dépendantes de ces paramètres, comme le montre la Figure 1.12, sur laquelle toutes les courbes sont comparables. Comme différence, on pourra relever tout de même que plus  $\tau_N$  est écarté de  $\tau_P$ , plus les variations de l'hétérogénéité numérique sont importantes. Ceci est notamment visible lors de la rechute à l'imatinib, entre les jour 776 et 888 où le pic descendant de l'hétérogénéité numérique est plus prononcé si  $\tau_P - \tau_N$  est grand. Ceci est conforme à ce que l'on pouvait attendre, puisque cette différence va impacter directement la position des gaussiennes sur l'histogramme, position relative en grande partie donnée par  $\Delta c$  qui intervient dans le calcul de notre critère de l'hétérogénéité numérique.

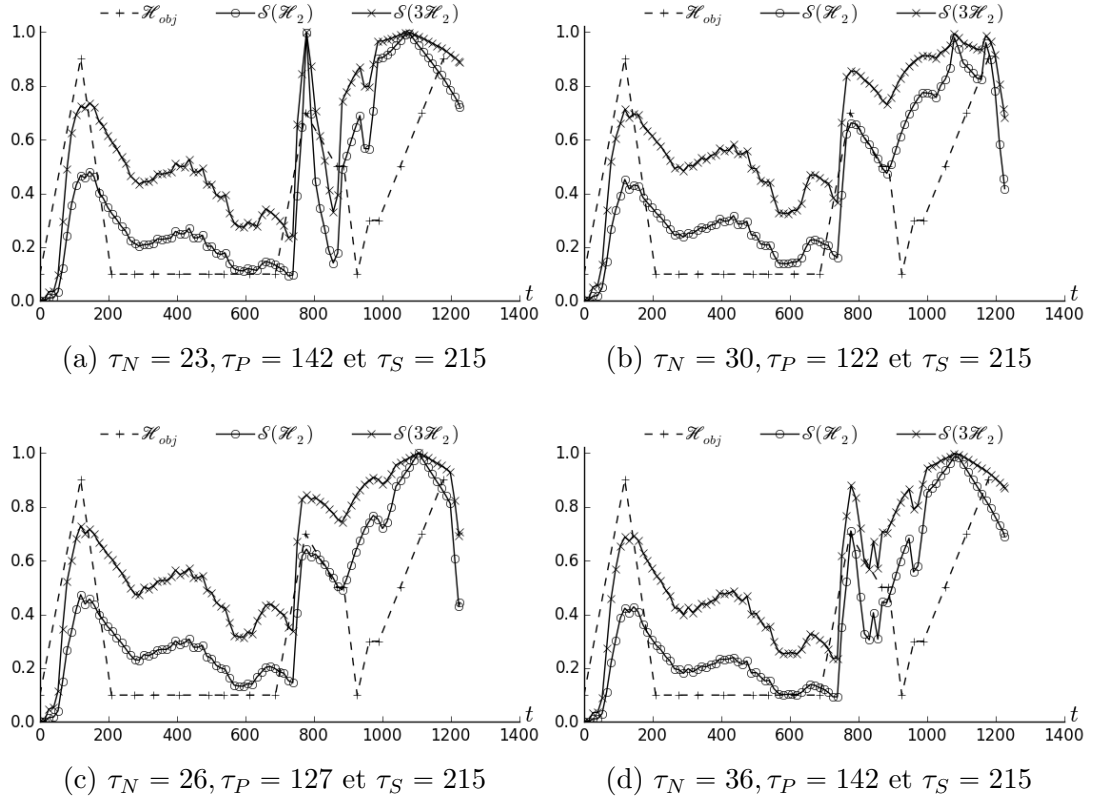


FIGURE 1.12 – Influence du choix des niveaux de gris  $\tau_N, \tau_P$  et  $\tau_S$  sur l'hétérogénéité numérique.

### 1.6.3 Patient B

L'hétérogénéité des scanners de Patient B n'a aucune chance d'être numériquement reproduite correctement comme le confirme la Figure 1.13. En effet, le premier scanner est très hétérogène, alors que notre condition initial dans le modèle numérique est complètement homogène. Il faudrait prendre une condition initiale plus en relation avec l'image médicale, à minima une condition initiale qui présenterait le même niveau d'hétérogénéité pour pouvoir poursuivre l'étude avec ce patient.

On notera que le critère semble tout de même bien décrire l'hétérogénéité numérique puisque comme le montrent les images numériques présentées dans la Figure ?? de l'annexe ?? (page ??), la simulation produit une tumeur quasi homogène jusqu'au jour 1100, puis une apparition brutale d'hétérogénéité entre les jours 1100 et 1200 et enfin une baisse progressive de l'hétérogénéité après le jour 1200. Le quantificateur de l'hétérogénéité est donc aussi correct ici.

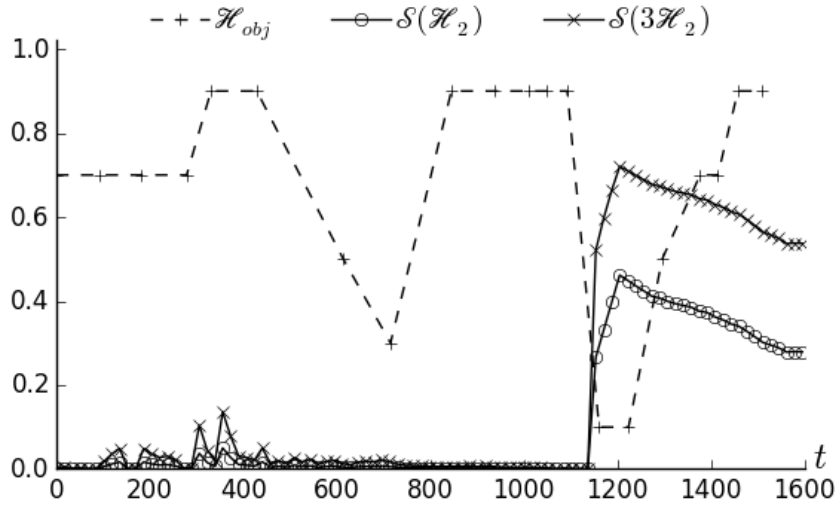


FIGURE 1.13 – Hétérogénéité numérique pour Patient B –  $\tau_N = 38$ ,  $\tau_P = 166$  et  $\tau_S = 215$ . La fonction objectif de l'hétérogénéité clinique ( $\mathcal{H}_{obj}$ ) est rappelée ici à titre de comparaison. La simulation numérique ne reproduit ici pas du tout l'hétérogénéité clinique.