



HAL
open science

Tomographie à l'aide de décalages temporels d'ondes sismiques P : développements méthodologiques et applications

Vadim Monteiller

► **To cite this version:**

Vadim Monteiller. Tomographie à l'aide de décalages temporels d'ondes sismiques P : développements méthodologiques et applications. Géophysique [physics.geo-ph]. Université de Savoie, 2005. Français. NNT: . tel-00653731

HAL Id: tel-00653731

<https://theses.hal.science/tel-00653731>

Submitted on 20 Dec 2011

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.

Université de Savoie

Thèse de doctorat

Tomographie à l'aide de décalages temporels
d'ondes sismiques P :
développements méthodologiques et applications

par
Vadim Monteiller

Jury

M. Campillo	Professeur, Université Joseph Fourier, Grenoble.
P. Bernard	Physicien, Institut de Physique du Globe de Paris.
M. Cuer	Professeur, Université Montpellier 2.
B. Valette	Ingénieur de Recherche, IRD.
Y. Cansi	Expert Senior, CEA.
G. Poupinet	Directeur de recherche, CNRS.
J. Virieux	Professeur, Université de Nice.
J.L Got	Maître de conférences, Université de Savoie.

RESUME

Connaître mieux l'intérieur de la Terre nécessite d'améliorer les méthodes d'investigation géophysique. Cette thèse s'est intéressée à une classe importante de ces méthodes : la tomographie sismique. La tomographie sismique classique utilise des temps d'arrivée d'ondes sismiques, dont la précision dépend du pas d'échantillonnage et du pointé de l'onde. Lorsque les trains d'onde sont similaires, c'est-à-dire lorsque les sources et/ou les récepteurs sont proches, on peut calculer, avec une précision inférieure au pas d'échantillonnage, le décalage temporel entre les trains d'ondes. Cette thèse a eu pour objectif l'utilisation de telles données en tomographie sismique. Utiliser pleinement la précision de ces données a nécessité une revisite complète des algorithmes de tomographie, tant du point de vue du problème direct que du point de vue du problème inverse. Une solution robuste et précise a été trouvée au problème direct (calcul du temps de propagation des ondes dans un milieu hétérogène). Une attention particulière a été portée au problème inverse : une approche Tarantola-Valette a été utilisée, de façon à trouver le modèle de norme minimale qui ajuste les données. La minimisation de cette norme est obtenue par l'introduction d'une fonction de corrélation des paramètres du modèle, dont la longueur et l'amplitude sont ajustées en suivant une procédure d'optimisation. De cette façon, toute fluctuation du modèle est contrainte par les données. L'algorithme mis au point a été utilisé avec des données provenant de séismes du volcan Kilauea (Hawaii) enregistrés par le réseau local USGS-HVO (50 stations). Il a permis de retrouver le modèle de vitesse à deux échelles : (1) une échelle régionale, où l'on retrouve le système magmatique profond, sous les calderas et les rifts du Mauna Loa et du Kilauea ; le modèle trouvé permet d'expliquer les déformations mises en évidence dans la croûte océanique, la distribution confinée de la sismicité et les déformations intenses du flanc Sud du Kilauea ; (2) une échelle locale, où l'on met en évidence un conduit magmatique sub-vertical sous la caldera et le rift Est du Kilauea, indétectable avec d'autres données. Cette méthode, précise, permet de choisir un modèle unique et stable à l'aide un critère d'optimisation objectif.

Table des matières

RESUME	2
NOTATIONS	5
INTRODUCTION.	6
Chapitre 1. Eléments théoriques.	8
1.1. Problème inverse.	8
1.2. Résultats généraux en optimisation.	26
1.3. Schémas numériques.	31
1.4. Conclusion.	40
Chapitre 2. Application à la tomographie sismique.	41
2.1. Modélisation.	41
2.2. Localisations.	45
2.3. Tomographies.	72
2.4. Conclusion.	80
Chapitre 3. Tomographie de l'île de Hawaii.	85
3.1. Tomographie à l'échelle régionale.	87
3.2. Tomographie à l'échelle locale.	98
3.3. Comparaison avec d'autres méthodes.	112
3.4. Conclusion.	122
Chapitre 4. Annexes	124
Annexe 1	125
Modélisation du milieu géophysique.	125
Annexe 2	131
Décalages temporels interspectraux.	131
Annexe 3	134
Equations en doubles différences.	134
Annexe 4	136
Eléments d'algèbre linéaire.	136
Annexe 5	139

TABLE DES MATIÈRES

4

Eléments de Calcul différentiel.	139
Annexe 6	141
Matrice de covariance.	141
Bibliographie	152

NOTATIONS

- \mathbf{A}^t , tranposée de la matrice \mathbf{A} .
- $\dot{\gamma}(\zeta)$, vecteur dérivé de la courbe paramétrée $\zeta \mapsto \gamma(\zeta)$ par rapport au paramètre ζ .
- $\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)^T$, gradient de la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$.
- $\nabla^2 f$, Hessian de la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

- $D_{\mathbf{x}_o} \mathbf{f}$, matrice jacobienne de la fonction $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, calculée en \mathbf{x}_o .
- $\langle \cdot, \cdot \rangle$, produit scalaire usuel sur \mathbb{R}^n , ($n \geq 1$).
- $\|\mathbf{x}\|$, norme L2 du vecteur \mathbf{x} de \mathbb{R}^n , ($n \geq 1$).
- $|\mathbf{x}|$, norme L1 du vecteur \mathbf{x} de \mathbb{R}^n , ($n \geq 1$).

- t_i^k temps de propagation réel de la source i à la station k (lorsqu'il n'y a qu'un seul séisme (resp - station), l'indice i (resp - j) est omis.)
- $t_i^k(\mathbf{m})$ temps de propagation théorique dans le modèle \mathbf{m} de la source i à la station k (lorsqu'il n'y a qu'un seul séisme (resp - station), l'indice i (resp - j) est omis.)

INTRODUCTION.

1) Position du problème. La connaissance des structures, tant passives qu’actives est une quête de longue haleine des sismologues, notamment après les premiers travaux de Geiger [14] concernant la localisation des séismes, et ceux de Aki et Lee [1], concernant la tomographie. De très nombreux travaux ont depuis été réalisés en tomographie, que ce soit à l’aide des temps d’arrivée des ondes de volumes, ou par inversion des courbes de dispersion des ondes de surfaces, et ceci à des échelles variées.

Dans les dernières décennies, les méthodes de localisation maintenant dites “en doubles différences” (Jordan et Sverdrup, [22]; Poupinet et coll., [38]; Ito, [21]; Fréchet, [13]; Got et coll., [15]; Slunga et coll., [44]; Shearer, [43]; Rubin et coll., [41]; Waldhauser et Ellsworth, [52]; voir Wolfe, [54] pour une revue, ainsi que Menke et Schaff, [27]) ont montré que la connaissance des structures actives de la croûte supérieure pouvait être profondément changée par la relocalisation précise de séismes. Cette précision accrue provient de la prise en compte des relations entre séismes voisins dans le processus de la localisation. Elle provient (1) de la grande précision de la mesure directe de la différence de temps d’arrivée des ondes par une méthode de traitement du signal performante lorsque les signaux enregistrés sont semblables, et (2) de la linéarisation naturelle du problème de localisation offerte par la localisation relative des séismes. Ces deux caractéristiques sont plus fortes lorsque les séismes sont plus proches et montrent des formes d’ondes similaires. Le gain en précision sur les positions relatives peut atteindre un facteur 10.

Utiliser une technique de doubles différences en tomographie paraît donc *a priori* offrir une perspective intéressante : un tel gain en précision serait-il envisageable en tomographie, si l’on utilisait des décalages temporels interspectraux calculés à l’aide d’essaims suffisamment denses et étendus de séismes aux formes d’ondes semblables ? la nécessaire proximité des événements sélectionnés n’est-elle pas contradictoire avec les conditions requises pour une tomographie ? quel est le niveau de précision requis dans les problèmes direct et inverse pour pouvoir obtenir des résultats fiables et interprétables ? peut-on améliorer la tomographie par l’inversion de simples différences de temps d’arrivée ? de façon plus générale, peut-on améliorer la connaissance des structures actives et passives dans un milieu (très) hétérogène en inversant des mesures différentielles (éventuellement très précises) ? cette thèse a pour objectif de répondre, au moins partiellement, à ces quelques questions.

2) Démarche suivie. Les résidus de relocalisation en doubles différences, lorsque les données sont des décalages temporels interspectraux (d’une précision moyenne d’environ 5 ms), sont en moyenne inférieurs à 10 ms. On comprend donc que si l’on veut prétendre interpréter les 5 à 10 ms du résidu qui peuvent être attribués à l’erreur de modélisation,

il faut être capable de produire des calculs direct et inverse très précis et stables. La précision et la robustesse nécessaires au calcul direct sera atteinte grâce à une amélioration du calcul du champ des temps de parcours fait en différences finies par l'algorithme de Podvin et Lecomte [37]. On peut donc penser que les différences de temps de parcours théoriques n'entacheront pas les décalages calculés d'une erreur qui rendraient les résidus ininterprétables.

Il est très important de traiter correctement le problème inverse, de façon à ce que la solution en soit stable. Le risque est en effet grand, en tomographie, de trouver tout une classe de modèles solutions qui ajustent également bien les données tout en montrant des variations très différentes qui peuvent sembler significatives, simplement parce qu'une combinaison linéaire des paramètres ajustés appartient au noyau et n'est pas contrainte par les données. Comment choisir un modèle parmi une classe de modèles possibles et garantir qu'il ne contient pas de fluctuations non contraintes par les données ? Répondre à cette question nécessitera une revisite de la tomographie classique, et la définition d'une stratégie d'optimisation complète. Elle se situera dans le cadre général de la théorie probabiliste développée par Tarantola et Valette [47] : il ne s'agit pas seulement de trouver un modèle qui ajuste les données, mais plutôt le modèle de norme minimale, c'est-à-dire qui comporte le moins de degrés de liberté (le plus "lisse" ou le plus simple ou le moins "haute fréquence" possible), qui ajuste les données. Il s'agit donc de déterminer le nombre de degrés de liberté optimal, ce qui sera fait en introduisant une fonction de corrélation des paramètres dans l'inversion. Inverser la matrice de covariance comportant cette fonction de corrélation est impossible de par la dimension de l'espace modèle : une approximation précise et pratique de cet inverse sera donnée, ce qui rendra possible une approche complète d'optimisation en tomographie. Procéder ainsi garantit un bon conditionnement du problème inverse, et une topologie de la fonctionnelle à minimiser favorable aux algorithmes d'optimisation (à directions de descente) qui seront utilisés. On peut ainsi espérer trouver un modèle unique, dont on sait justifier le choix, et dont les paramètres auront des valeurs interprétables.

Une application de cet algorithme sera faite à l'aide des séismes enregistrés sur le volcan Kilauea, situé sur l'île de Hawaii. Elle permet d'abord de déterminer un modèle de vitesse régional, unique et justifiable, par tomographie des temps d'arrivée ; il servira de modèle de référence pour l'étude locale plus précise, faite à l'aide de décalages temporels. Ce modèle régional est stable et montre bien les principales structures volcaniques de l'île. Une tomographie des décalages temporels est ensuite effectuée, permettant de déterminer un modèle unique, justifiable et stable toujours choisi suivant le même approche d'optimisation. Le modèle trouvé montre des variations stables et cohérentes des vitesses sismiques dans l'espace, qui peuvent être interprétées comme étant dues au système d'approvisionnement magmatique du Kilauea.

CHAPITRE 1

Eléments théoriques.

Ce chapitre introduit la notion générale de problème inverse. C'est une démarche mettant en jeu des méthodes théoriques et numériques, permettant d'estimer les paramètres ajustant au mieux un ensemble de mesures, effectuées lors d'une expérience de physique. On expose ici, des éléments théoriques utiles à la compréhension et à la formulation de méthodes tomographiques. Nous verrons qu'un problème inverse se pose en terme d'optimisation. Nous donnerons alors quelques notions concernant les méthodes numériques d'optimisation à direction de descente.

1.1. Problème inverse.

Pour commencer définissons quelques notations. Soit P l'espace des paramètres dont \mathbf{p} sera un élément générique. Soit une fonction $\mathbf{g} : P \rightarrow \mathbb{R}^n$ modélisant le phénomène physique considéré. Dans la terminologie des Sciences de la Terre, *modèle* désigne un ensemble de paramètres caractérisant le milieu, qui en général ne sont pas directement mesurables. Le terme *données*, quant à lui, est consacré aux paramètres mesurés instrumentalement. Ainsi on pourra décomposer $P = D \times M$, où D est l'espace des données et M l'espace des paramètres du modèle ; un élément générique s'écrira donc $\mathbf{p} = (\mathbf{d}, \mathbf{m})$, avec $\mathbf{d} \in D$ et $\mathbf{m} \in M$. La relation théorique modélisant le phénomène physique étudié s'écrira,

$$(1.1.1) \quad \mathbf{g}(\mathbf{d}, \mathbf{m}) = 0.$$

Par la suite nous supposons qu'elle s'exprime sous forme explicite,

$$(1.1.2) \quad \mathbf{d} = \mathbf{g}(\mathbf{m}).$$

Le *problème direct* consiste à calculer le vecteur \mathbf{d} connaissant le modèle \mathbf{m} , en utilisant la relation (1.1.2). Dans le *problème inverse*, au contraire, il s'agit de trouver un modèle $\overline{\mathbf{m}}$ correspondant à un jeu de données mesurés \mathbf{d}^{obs} ,

$$(1.1.3) \quad \mathbf{d}^{obs} = \mathbf{g}(\overline{\mathbf{m}}).$$

Le principe du problème inverse peut paraître simple mais, en général, la relation (1.1.2) ne peut pas être inversée, c'est à dire qu'il n'existe pas de formule permettant de calculer directement le modèle \mathbf{m} étant donné un vecteur \mathbf{d} . Il faut y ajouter une difficulté supplémentaire : il faut généralement tenir compte des erreurs de mesures sur \mathbf{d}^{obs} et les erreurs de modélisations de \mathbf{g} . Il est alors impossible de trouver un modèle $\overline{\mathbf{m}}$ qui satisfasse exactement (1.1.3). Pour répondre à ce problème, les géophysiciens ont défini une méthode

d'analyse de données, devenue classique en sismologie. Considérons, en premier lieu le cas le plus simple : un problème direct (1.1.2) linéaire,

$$(1.1.4) \quad \mathbf{d} = \mathbf{G}\mathbf{m}.$$

Le problème inverse revient à chercher un modèle $\bar{\mathbf{m}}$ vérifiant le système linéaire suivant,

$$(1.1.5) \quad \mathbf{d}^{obs} = \mathbf{G}\bar{\mathbf{m}}.$$

Un tel modèle, impliquant l'égalité stricte (1.1.5) n'existe que si le vecteur \mathbf{d}^{obs} est dans l'image de l'application linéaire $\mathbf{m} \mapsto \mathbf{G}\mathbf{m}$. En général cette condition n'est pas vérifiée. Il est donc nécessaire de définir une notion de solution "généralisée" d'un système linéaire. L'idée est de chercher un modèle $\bar{\mathbf{m}}$ tel que le vecteur $\mathbf{G}\bar{\mathbf{m}}$ soit le plus proche possible de \mathbf{d}^{obs} . Pour mesurer la distance entre ces deux vecteurs, plusieurs normes sont disponibles, mais le choix le plus simple est d'utiliser la norme L2 (qui correspond à l'appellation *moindres carrés*). Autrement dit, un problème aux moindres carrés consiste à trouver la (ou les) solution(s) du problème de minimisation suivant,

$$(1.1.6) \quad \min_{\mathbf{m} \in M} \|\mathbf{d}^{obs} - \mathbf{G}\mathbf{m}\|.$$

Lorsque le problème direct n'est plus linéaire, une approche classique consiste à linéariser au voisinage d'un modèle \mathbf{m}_0 ,

$$(1.1.7) \quad \mathbf{g}(\mathbf{m}_0 + \delta) \simeq \mathbf{g}(\mathbf{m}_0) + \mathbf{G}\delta.$$

Comme précédemment, l'idée est de calculer une perturbation δ telle que les données \mathbf{d}^{obs} et $\mathbf{g}(\mathbf{m}_0 + \delta)$ soient les plus proches possible au sens de la norme L2. En utilisant (1.1.7), on obtient un problème aux moindres carrés comparable à (1.1.6),

$$(1.1.8) \quad \min_{\delta \in M} \|\mathbf{d}^{obs} - \mathbf{g}(\mathbf{m}_0) - \mathbf{G}\delta\|.$$

Dans ce cas, on ajuste une quantité $\mathbf{r}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{g}(\mathbf{m}_0)$, appelée *résidu*, en fonction d'une perturbation $\bar{\delta}$ minimisant (1.1.8). Une solution du problème inverse se calcule donc en faisant $\bar{\mathbf{m}} = \mathbf{m}_0 + \bar{\delta}$. Cette méthode est utilisable lorsque le modèle \mathbf{m}_0 se trouve dans un voisinage de la solution $\bar{\mathbf{m}}$ de sorte que la relation (1.1.7) reste valide. Si ce n'est pas le cas, il faut procéder itérativement. On calcule un modèle \mathbf{m}_1 solution du problème (1.1.8), ensuite on linéarise le problème direct autour de \mathbf{m}_1 et on calcule un modèle \mathbf{m}_2 . En continuant de la sorte, on définit une suite de modèles $\{\mathbf{m}_k\}_{k=1,\dots,n}$. Dans les cas favorables, la suite converge vers une solution au problème inverse. Mais si le problème direct est fortement non linéaire, cette méthode peut diverger, il est nécessaire alors de stabiliser numériquement le problème. De plus cette méthode ne prend pas en compte la statistique des erreurs de mesures sur les données, ce qui peut entraîner des biais importants sur la solution. Pour ces raisons il est préférable d'aborder l'inversion d'un problème non-linéaire d'un point de vue probabiliste. La suite de ce chapitre reprend, de manière succincte, certaines idées défendues notamment par Tarantola et Valette [47] dans le but de définir correctement un problème inverse sismologique répondant aux préoccupations exposées en introduction.

1.1.1. Modélisation. Lorsque l'on considère un paramètre physique on s'attend à pouvoir le quantifier à l'aide d'une valeur numérique. Mais dans bien des cas l'information disponible sur ce paramètre ne permet pas de connaître précisément la valeur de celui-ci (erreurs de mesure, d'instrumentation, ou bien absence de mesure,...). Il est alors plus adapté de représenter un paramètre par une loi de probabilité, c'est à dire par une variable aléatoire réelle X . La probabilité qu'un paramètre \mathbf{p} soit dans l'ensemble I est noté $P_X(I)$. Si cette variable aléatoire admet une densité ρ par rapport à une mesure μ ,

$$(1.1.9) \quad P_X(I) = \int_I \rho d\mu.$$

Cette façon de faire très générale permet, par exemple, de considérer un paramètre \mathbf{p}_0 dont la valeur est parfaitement déterminée. La mesure μ dans ce cas est la mesure de Dirac $\delta_{\mathbf{p}_0}$,

$$(1.1.10) \quad P_X(I) = \int_I d\delta_{\mathbf{p}_0},$$

avec,

$$(1.1.11) \quad \delta_{\mathbf{p}_0}(\mathbf{p}) = 1 \text{ si } \mathbf{p}=\mathbf{p}_0, 0 \text{ sinon.}$$

Cette représentation des paramètres modélise aussi les erreurs de mesures effectuées par un opérateur. Si on effectue n mesures d'un paramètre physique, on obtient m valeurs différentes. En reportant ces résultats sur un histogramme, on peut se demander si la distribution correspond à une loi de probabilité connue. Si tel est le cas il est parfois possible de trouver une densité de probabilité décrivant ces erreurs. L'expérience de la figure (1.1.1) montre l'histogramme des valeurs mesurées d'un paramètre. Au vu de sa répartition, l'hypothèse d'une loi gaussienne semble convenir. Il suffit alors de calculer la valeur moyenne d_0 et l'écart-type σ de la fonction densité, supposée gaussienne, pour exprimer la probabilité

$$(1.1.12) \quad P_X(I) = \int_I \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-d_0)^2}{2\sigma^2}\right) dx.$$

Une densité de probabilité permet aussi de transcrire l'information que l'on obtient sur un paramètre en l'absence de mesure directe. Le contexte physique permet bien souvent de cibler une plage de valeurs admissibles pour le paramètre. Par exemple on peut s'attendre à ce qu'un paramètre reste dans un intervalle J de \mathbb{R} , dans ce cas la probabilité s'écrira,

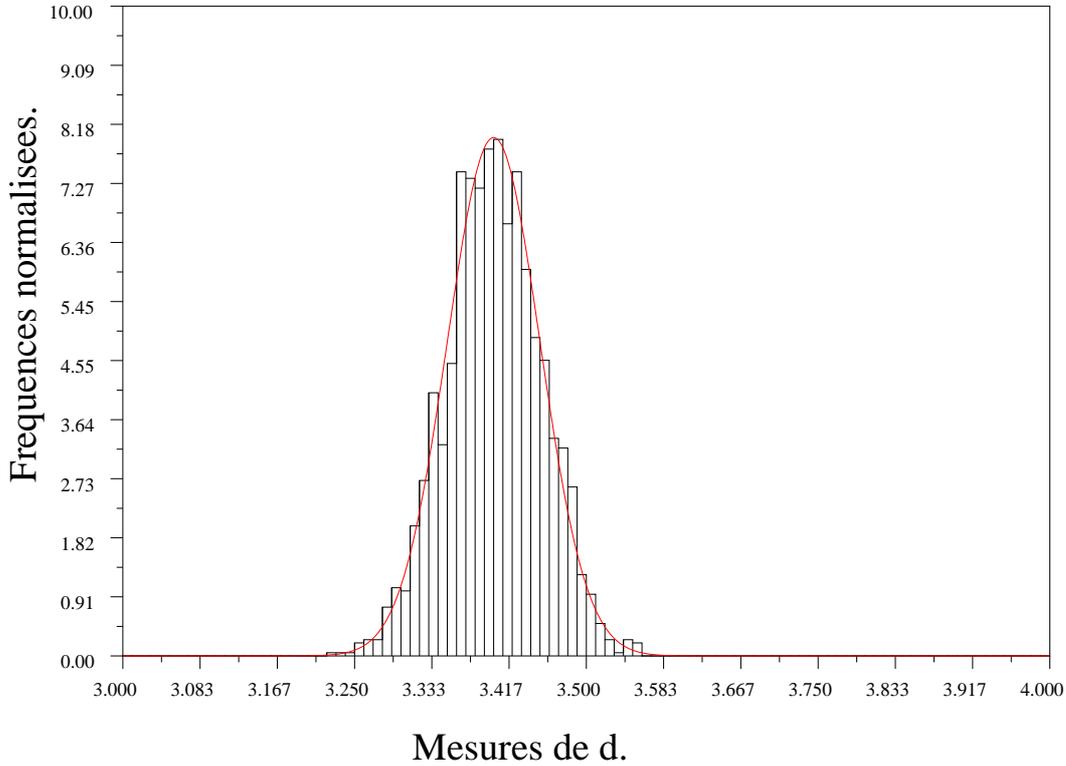
$$(1.1.13) \quad P_X(I) = \int_I \nu \chi_J d\lambda,$$

où χ_J est la fonction caractéristique de l'intervalle J et λ est la mesure de Lebesgue sur \mathbb{R} , et $\frac{1}{\nu} = \int_{\mathbb{R}} \chi_J d\lambda$.

Nous venons de voir qu'une densité de probabilité modélise en quelque sorte l'information que l'on a sur un paramètre. Si ce paramètre est mesuré instrumentalement, la densité de probabilité représente la statistique des erreurs de mesures (s'il y en a), si le paramètre n'est pas directement mesuré, la densité de probabilité décrit une plage dans laquelle on

FIG. 1.1.1. Modélisation des erreurs de mesures.

Histogramme des mesures d'un paramètre et modélisation de celles-ci par une densité de probabilité (courbe rouge).



s'attend à voir évoluer le paramètre. Adopter ce point de vu probabiliste nous conduit à faire quelques modifications dans la définition des données et des paramètres du modèle.

1.1.2. Données et modèles. Nous supposons que l'information que l'on possède sur les paramètres s'exprime par l'intermédiaire d'une densité de probabilité ρ définie sur P . Nous noterons les densités de probabilités marginales de ρ : ρ_D sur D et ρ_M sur M . Par exemple, en l'absence de corrélations entre les données et le modèle nous avons,

$$(1.1.14) \quad \rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d}) \rho_M(\mathbf{m}).$$

L'information contenue dans le problème direct sera aussi représentée par une densité de probabilité Θ définie sur P . Cette densité de probabilité modélise la relation physique théorique par des corrélations entre modèles et données. On définit sur D , la densité de probabilité conditionnelle noté $\Theta(\mathbf{d} | \mathbf{m})$; c'est la loi de la donnée \mathbf{d} quand on connaît le

modèle \mathbf{m} . Par exemple, si on connaît la statistique ϵ des erreurs de modélisation,

$$(1.1.15) \quad \mathbf{d} + \epsilon = \mathbf{g}(\mathbf{m}),$$

par sa densité de probabilité f_ϵ alors,

$$(1.1.16) \quad \Theta(\mathbf{d} | \mathbf{m}) = f_\epsilon(\mathbf{d} - \mathbf{g}(\mathbf{m})).$$

Dans le cas d'une théorie exacte (sans erreurs de modélisation, cela revient à considérer le problème direct comme en (1.1.2)),

$$(1.1.17) \quad f_\epsilon = \delta_{\mathbf{d} - \mathbf{g}(\mathbf{m})}.$$

Si l'information sur le modèle est représentée par la probabilité μ sur M , la densité de probabilité Θ définie sur P s'écrit alors¹,

$$(1.1.18) \quad \Theta(\mathbf{d}, \mathbf{m}) = \Theta(\mathbf{d} | \mathbf{m}) \mu(\mathbf{m}).$$

La densité de probabilité Θ est une façon générale de définir le problème direct ; elle englobe le cas classique, relation (1.1.2), et elle permet en plus de considérer, si besoin est, des erreurs de modélisation.

1.1.3. Définition du problème inverse. Pour définir le problème inverse il faut d'abord considérer de deux types d'informations sur l'espace des paramètres P :

- l'information *a priori*, $\rho(\mathbf{d}, \mathbf{m})$ qui représente les mesures des données et l'information *a priori* que l'on a sur le modèle ;
- l'information $\Theta(\mathbf{d}, \mathbf{m})$ qui modélise la relation théorique physique par des corrélations entre données et modèles (Problème direct).

L'idée est de recouper les deux informations précédentes en définissant une nouvelle densité de probabilité σ sur P . Tarantola [45] propose une méthode pour calculer cette densité de probabilité σ combinant les deux informations provenant de ρ et Θ . Dans le cas où l'espace P est de dimension finie, il pose deux postulats :

- (1) L'état de l'information sur les paramètres physiques est modélisé par une probabilité (plus généralement une mesure) sur P .

¹Si ρ est une densité de probabilité sur $A \times B$, ρ_A et ρ_B les lois marginales sur, respectivement A et B . On a les relations

$$\rho(\mathbf{a}, \mathbf{b}) = \rho(\mathbf{a} | \mathbf{b}) \rho_B(\mathbf{b}),$$

et

$$\rho(\mathbf{a}, \mathbf{b}) = \rho(\mathbf{b} | \mathbf{a}) \rho_A(\mathbf{a}).$$

Lorsque les lois ρ_A et ρ_B sont indépendantes,

$$\rho_A(\mathbf{a}) = \rho(\mathbf{a} | \mathbf{b}), \quad \forall \mathbf{b} \in B,$$

et

$$\rho_B(\mathbf{b}) = \rho(\mathbf{b} | \mathbf{a}), \quad \forall \mathbf{a} \in A,$$

ainsi

$$\rho(\mathbf{a}, \mathbf{b}) = \rho_A(\mathbf{a}) \rho_B(\mathbf{b}).$$

- (2) Il existe un état de référence pour l'information sur les paramètres, appelé aussi état de totale ignorance. Cet état est modélisé par une probabilité (ou mesure) en vertu du postulat précédent.

Tarantola [45] définit ensuite la combinaison de deux états de l'information : si P_1 et P_2 sont deux probabilités sur P décrivant deux états différents de l'information sur les paramètres, il définit la probabilité (P_1 et P_2) qui modélise l'état de l'information réunissant les informations provenant de P_1 et P_2 . Il démontre que si P_1 admet une densité de probabilité f_1 , P_2 admet une densité f_2 et que l'état de totale ignorance admet une densité μ , alors (P_1 et P_2) a pour densité,

$$(1.1.19) \quad \sigma(\mathbf{p}) = \frac{f_1(\mathbf{p}) f_2(\mathbf{p})}{\mu(\mathbf{p})}.$$

Cet outil permet de combiner les informations de ρ et Θ pour définir la densité de probabilité *a posteriori*, (figure(1.1.2)),

$$(1.1.20) \quad \sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m}) \Theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})}.$$

Utilisons cette formule dans les cas où, (hypothèses (H))

- L'information *a priori* s'écrit : $\rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d}) \rho_M(\mathbf{m})$.
- Le problème direct est : $\Theta(\mathbf{d}, \mathbf{m}) = \Theta(\mathbf{d} | \mathbf{m}) \mu_M(\mathbf{m})$.
- L'état de l'information nulle est : $\mu(\mathbf{d}, \mathbf{m}) = \mu_D(\mathbf{d}) \mu_M(\mathbf{m})$.

On peut alors exprimer les lois marginales de σ , d'abord sur D ,

$$(1.1.21) \quad \sigma_D(\mathbf{d}) = \frac{\rho_D(\mathbf{d})}{\mu_D(\mathbf{d})} \int_M \rho_M(\mathbf{m}) \Theta(\mathbf{d} | \mathbf{m}) d\mathbf{m},$$

et puis sur M ,

$$(1.1.22) \quad \sigma_M(\mathbf{m}) = \rho_M(\mathbf{m}) \int_D \frac{\rho_D(\mathbf{d}) \Theta(\mathbf{d} | \mathbf{m})}{\mu_D(\mathbf{d})} d\mathbf{d}.$$

On peut préciser un peu plus ces lois en calculant l'intégrale (1.1.22) dans des cas particuliers mais très importants pour les applications.

1.1.3.1. *Cas d'une théorie exacte avec des erreurs de mesures.* Dans le cas d'une théorie exacte,

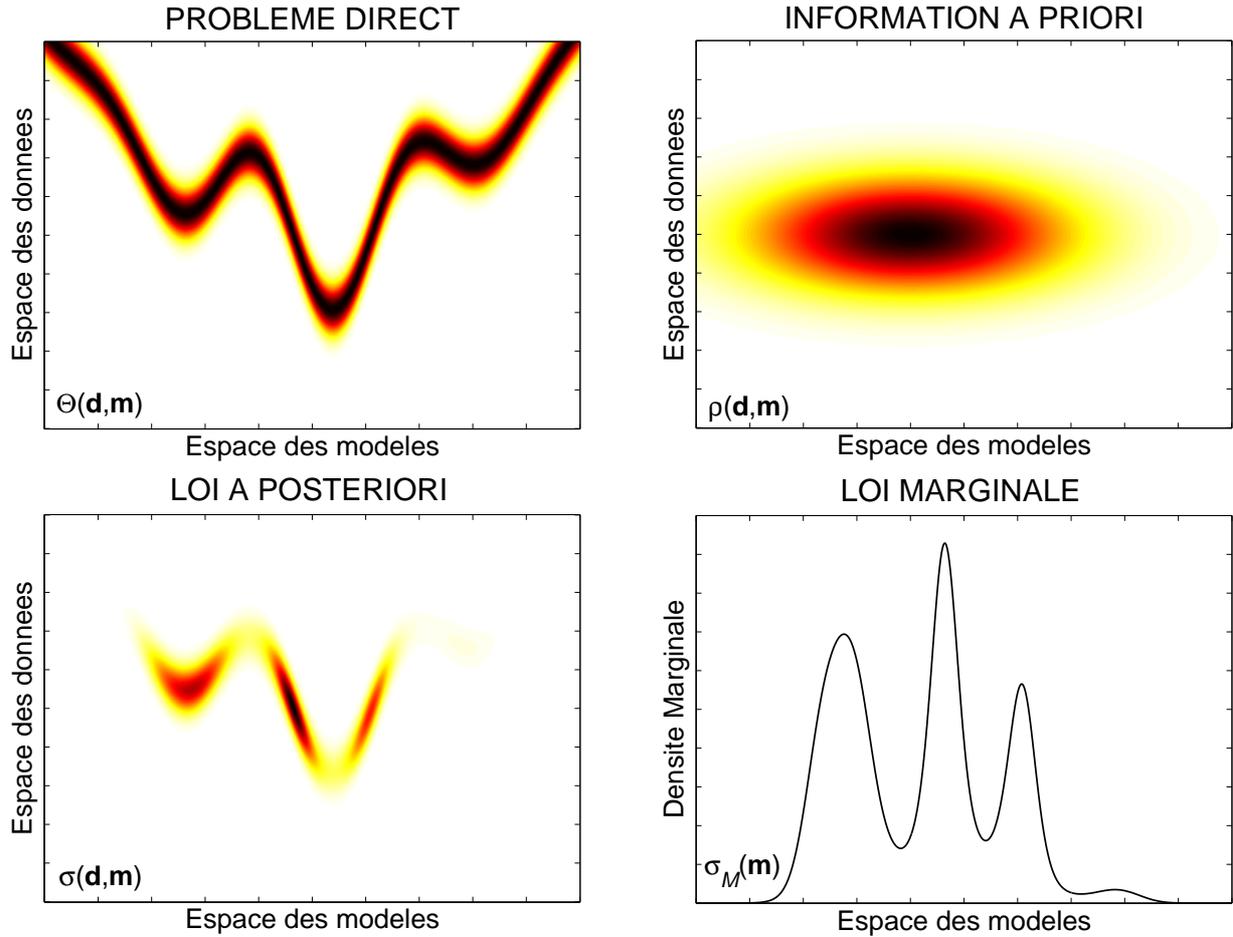
$$(1.1.23) \quad \Theta(\mathbf{d} | \mathbf{m}) = \delta_{\mathbf{d}-\mathbf{g}(\mathbf{m})},$$

et la relation (1.1.22) devient,

$$(1.1.24) \quad \sigma_M(\mathbf{m}) = \rho_M(\mathbf{m}) \frac{\rho_D(\mathbf{g}(\mathbf{m}))}{\mu_D(\mathbf{g}(\mathbf{m}))}.$$

FIG. 1.1.2. **Problème inverse général en dimension finie.**

Représentation schématique d'un problème inverse. La figure en haut à gauche montre dans l'espace $D \times M$, la densité de probabilité $\Theta(\mathbf{d}, \mathbf{m})$ modélisant le problème direct. La figure en haut à droite montre dans $D \times M$, la densité de probabilité $\rho(\mathbf{d}, \mathbf{m})$ représentant l'information a priori. La figure en bas à gauche montre dans $D \times M$, la densité de probabilité a posteriori, $\sigma(\mathbf{d}, \mathbf{m})$ regroupant les informations modélisées par $\Theta(\mathbf{d}, \mathbf{m})$ et $\rho(\mathbf{d}, \mathbf{m})$ (formule (1.1.20)). La figure en bas à droite montre dans M , la densité de probabilité marginale $\sigma_M(\mathbf{m})$ de $\sigma(\mathbf{d}, \mathbf{m})$ (formule (1.1.22)).



Si on note ρ_ϵ les erreurs de mesures sur un vecteur \mathbf{d}_o alors²,

$$(1.1.25) \quad \frac{\rho_D(\mathbf{d})}{\mu_D(\mathbf{d})} = \rho_\epsilon(\mathbf{d}_o - \mathbf{d})$$

et la formule (1.1.24) devient,

$$(1.1.26) \quad \sigma_M(\mathbf{m}) = \rho_M(\mathbf{m}) \rho_\epsilon(\mathbf{d}_o - \mathbf{g}(\mathbf{m})).$$

²Pour une démonstration de cette formule, se reporter à Tarantola [45] p37 et p38.

1.1.3.2. *Cas d'une théorie inexacte avec des erreurs de mesures.* Si on considère des erreurs de modélisation,

$$(1.1.27) \quad \Theta(\mathbf{d} \mid \mathbf{m}) = f_\epsilon(\mathbf{d} - \mathbf{g}(\mathbf{m})),$$

la formule (1.1.22) devient,

$$(1.1.28) \quad \sigma_M(\mathbf{m}) = \rho_M(\mathbf{m}) \int_D \rho_\epsilon(\mathbf{d}_o - \mathbf{d}) f_\epsilon(\mathbf{d} - \mathbf{g}(\mathbf{m})) \, d\mathbf{d}.$$

C'est une relation de convolution de la densité de probabilité ρ_ϵ par f_ϵ .

$$(1.1.29) \quad f = \rho_\epsilon \star f_\epsilon,$$

il est facile de voir que la densité de probabilité *a posteriori* sur M s'écrit alors,

$$(1.1.30) \quad \sigma_M(\mathbf{m}) = \rho_M(\mathbf{m}) f(\mathbf{d}_o - \mathbf{g}(\mathbf{m})).$$

La densité de probabilité f représente à la fois, les erreurs de mesures sur les données \mathbf{d}_o et les erreurs de modélisation du problème direct.

1.1.4. Statistique Gaussienne des erreurs de mesures. L'hypothèse Gaussienne sur l'information *a priori*, conduit à une formulation simple du problème inverse. En effet, nous verrons que cela revient à résoudre un problème aux moindres carrés. De plus, il existe des méthodes éprouvées résolvant ce type de problème. Cela explique que dans les applications, de nombreux problèmes inverses sont posés en ces termes. Dans cette partie nous supposons que l'information sur les données se ramène à une loi Gaussienne notée,

$$(1.1.31) \quad \rho_\epsilon(\mathbf{d}_o - \mathbf{d}) = ((2\pi) \det(\mathbf{C}_d))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{d}_o - \mathbf{d})^t \mathbf{C}_d^{-1}(\mathbf{d}_o - \mathbf{d})\right),$$

où \mathbf{C}_d est la matrice de variance-covariance (matrice symétrique définie positive) caractérisant la loi ρ_ϵ . Nous allons voir comment s'exprime le problème inverse quand l'information *a priori* sur le modèle est Gaussienne. Ensuite nous étudierons le cas d'une densité de probabilité *a priori* égale à une fonction caractéristique d'un sous-ensemble I de M (équiprobabilité pour tous les modèles appartenant à I).

1.1.4.1. *Information a priori Gaussienne.* Reprenons la formule (1.1.26) et supposons Gaussienne la densité de probabilité ρ_M ,

$$(1.1.32) \quad \rho_M(\mathbf{m}) = ((2\pi) \det(\mathbf{C}_m))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0)\right),$$

où \mathbf{C}_m est la matrice de variance-covariance caractérisant la loi ρ_M . La formule (1.1.26) devient alors,

$$(1.1.33) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) - \frac{1}{2}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_d^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))\right).$$

Supposons maintenant que les erreurs de modélisation f_ϵ suivent une loi Gaussienne de matrice de variance-covariance \mathbf{C}_ϵ ,

$$(1.1.34) \quad \Theta(\mathbf{d} \mid \mathbf{m}) = f_\epsilon(\mathbf{d} - \mathbf{g}(\mathbf{m})) \propto \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_\epsilon^{-1}(\mathbf{d} - \mathbf{g}(\mathbf{m}))\right).$$

La densité de probabilité *a posteriori* σ_M s'exprime par l'intermédiaire de la formule (1.1.30), il faut calculer la convolution entre ρ_ϵ et f_ϵ comme indiqué par la formule (1.1.29). On doit faire la convolution entre deux lois Gaussiennes de matrices de variance-covariance \mathbf{C}_d et \mathbf{C}_ϵ . On peut démontrer³ qu'il en résulte une loi Gaussienne de variance-covariance $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_\epsilon$. Finalement la densité de probabilité *a posteriori* s'écrit,

$$(1.1.35) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_0) - \frac{1}{2}(\mathbf{d}_o-\mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o-\mathbf{g}(\mathbf{m}))\right).$$

Les formules (1.1.33) et (1.1.35) nous montrent que la densité de probabilité *a posteriori* σ_M s'exprime de la même façon, aussi bien dans le cadre d'une théorie exacte qu'avec des erreurs de modélisation Gaussiennes. Par conséquent, la démarche pour étudier ce type de problème sera la même dans les deux cas. Nous utiliserons la formule (1.1.35), où la matrice \mathbf{C}_D sera identifiée soit à la somme $\mathbf{C}_d + \mathbf{C}_\epsilon$, dans le cadre d'une théorie inexacte, soit à la matrice \mathbf{C}_d dans le cadre d'une théorie exacte. Cet abus de notation n'aura aucune incidence sur les raisonnements généraux. Dans les applications particulières il faudra prendre garde à interpréter convenablement la matrice \mathbf{C}_D .

Il s'agit maintenant d'exploiter la fonction σ_M . Le cas le plus favorable est celui d'une densité de probabilité unimodale, ce qui arrive lorsque le problème direct n'est pas trop fortement non linéaire. On peut alors chercher le maximum de vraisemblance qui se trouve être le minimum de la fonction E_g définie par,

$$(1.1.36) \quad E_g(\mathbf{m}) = \frac{1}{2}(\mathbf{m}-\mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_0) + \frac{1}{2}(\mathbf{d}_o-\mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o-\mathbf{g}(\mathbf{m})).$$

Les matrices $\mathbf{C}_m^{-\frac{1}{2}}$ et $\mathbf{C}_D^{-\frac{1}{2}}$ existent puisque, par hypothèse, les matrices \mathbf{C}_m et \mathbf{C}_D sont symétriques définies positives, nous pouvons alors réécrire (1.1.36) sous la forme suivante,

$$(1.1.37) \quad E_g(\mathbf{m}) = \frac{1}{2} \left\langle \mathbf{C}_m^{-\frac{1}{2}}(\mathbf{m}-\mathbf{m}_0), \mathbf{C}_m^{-\frac{1}{2}}(\mathbf{m}-\mathbf{m}_0) \right\rangle + \frac{1}{2} \left\langle \mathbf{C}_D^{-\frac{1}{2}}(\mathbf{d}_o-\mathbf{g}(\mathbf{m})), \mathbf{C}_D^{-\frac{1}{2}}(\mathbf{d}_o-\mathbf{g}(\mathbf{m})) \right\rangle.$$

On peut reformuler l'expression précédente (1.1.37) en faisant apparaître le produit scalaire sur $D \times M$,

$$(1.1.38) \quad E_g(\mathbf{m}) = \frac{1}{2} \left\langle \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m} \end{bmatrix}, \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m} \end{bmatrix} \right\rangle.$$

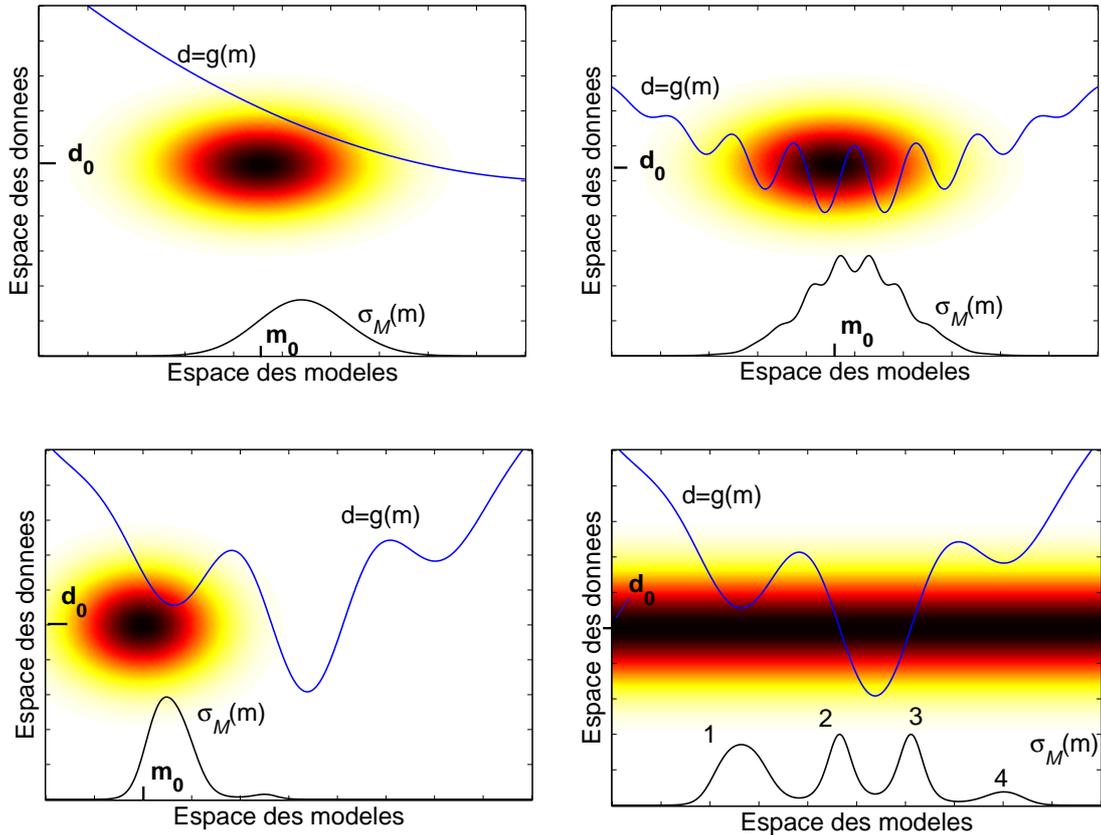
Ainsi sous l'hypothèse Gaussienne, le problème inverse revient à résoudre un problème aux moindres carrés :

$$(1.1.39) \quad \text{minimiser } \frac{1}{2} \left\| \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m} \end{bmatrix} \right\|^2 \text{ lorsque } \mathbf{m} \in M.$$

³Tarantola [45] p58 et p158 pour une démonstration complète.

FIG. 1.1.3. Information *a priori* gaussienne.

Sur les quatre figures sont représentés, dans l'espace $D \times M$, l'information *a priori* (zone en couleur) et le problème direct (Courbe bleue). La courbe noire est la loi de probabilité *a posteriori* $\sigma_M(\mathbf{m})$. Lorsque le problème direct est faiblement non linéaire (figure en haut à gauche), la loi $\sigma_M(\mathbf{m})$ est proche d'une loi gaussienne, il est alors légitime de chercher le maximum de vraisemblance par une méthode numérique appropriée, et de modéliser les erreurs par la loi gaussienne tangente. Lorsque le problème direct est fortement non linéaire (figure en haut à droite), il devient difficile de répondre au problème inverse en fournissant un modèle $\hat{\mathbf{m}}$ et en estimant des erreurs sur celui-ci. En effet, la loi $\sigma_M(\mathbf{m})$ est trop éloignée d'une loi gaussienne. Néanmoins, on peut dans certains cas, en jouant sur l'information *a priori* sur le modèle, transformer une loi $\sigma_M(\mathbf{m})$ quelconque en une loi proche d'une gaussienne. C'est ce que l'on voit sur les deux figures du bas. A droite, sans *a priori* sur le modèle, on observe quatre maxima de vraisemblance. A gauche, en donnant un *a priori* physiquement acceptable, on obtient une loi $\sigma_M(\mathbf{m})$ presque gaussienne. Dans ce cas, un algorithme classique fournit le maximum de $\sigma_M(\mathbf{m})$, et donne des erreurs sur le modèle ainsi estimé. La méthode d'ajustement des données aux moindres carrés est très fréquemment employée dans les problèmes sismologiques (figure en bas à droite). Elle devient numériquement instable dans les problèmes non linéaire à cause des minima secondaires.



1.1.4.2. *Equiprobabilité.* Considérons maintenant une densité de probabilité *a priori* définie par une fonction caractéristique d'un ensemble $I \subset M$,

$$(1.1.40) \quad \rho_M(\mathbf{m}) = \begin{cases} c^{te} & \text{pour } \mathbf{m} \in I, \\ 0 & \text{pour } \mathbf{m} \notin I. \end{cases}$$

D'après la formule (1.1.30),

$$(1.1.41) \quad \sigma_M(\mathbf{m}) \propto \begin{cases} c^{te} f(\mathbf{d}_o - \mathbf{g}(\mathbf{m})) & \text{pour } \mathbf{m} \in I, \\ 0 & \text{pour } \mathbf{m} \notin I. \end{cases}$$

En utilisant les résultats vus au paragraphe précédent,

$$(1.1.42) \quad \sigma_M(\mathbf{m}) \propto \begin{cases} c^{te} \exp\left(-\frac{1}{2}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))\right) & \text{pour } \mathbf{m} \in I, \\ 0 & \text{pour } \mathbf{m} \notin I. \end{cases}$$

Supposons que le maximum de vraisemblance soit dans I , le problème se ramène alors à un problème de moindres carrés sans modèle *a priori*,

$$(1.1.43) \quad \text{minimiser } \frac{1}{2} \left\| \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o - \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \right\|^2 \text{ lorsque } \mathbf{m} \in M.$$

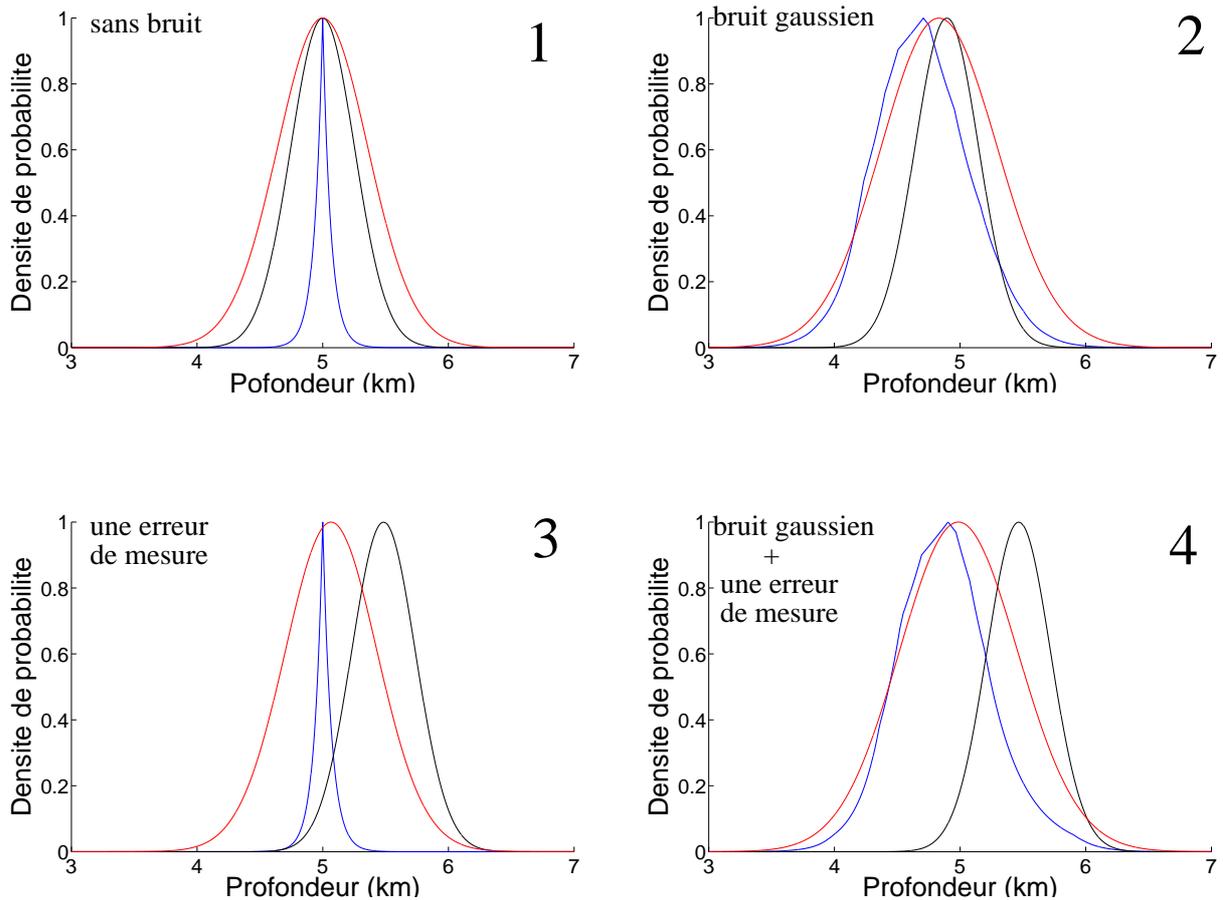
Cette formulation du problème inverse est couramment employée en géophysique pour les problèmes d'ajustement de données.

1.1.4.3. *Conclusion.* La forme de la densité σ_M dépend du problème direct : plus il est linéaire, plus on est proche d'une loi Gaussienne; à contrario, si le problème direct est fortement non linéaire, la loi *a posteriori* peut être multimodale (figure (1.1.3)). Nous verrons plus loin que les méthodes numériques cherchant les maxima de vraisemblance se comportent beaucoup mieux avec des problèmes directs proches de la linéarité. Néanmoins, avec des problèmes directs non linéaires, l'information *a priori* tend à rendre la densité *a posteriori* unimodale et presque gaussienne, donc facilement exploitable numériquement. Minimiser la fonction E_g plutôt que la fonction $\frac{1}{2}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))$ permet une régularisation du modèle; cette régularisation est effectuée en prenant en compte l'information *a priori*, grâce à la matrice de covariance *a priori* des paramètres du modèle. On peut ainsi suivre une approche numérique d'optimisation en cherchant le modèle le plus simple possible (comportant le moins de degrés de liberté) ajustant le mieux possible les données, et non pas seulement le modèle qui ajuste le mieux les données.

1.1.5. Statistique non Gaussienne des erreurs de mesures. L'hypothèse Gaussienne est couramment utilisée en sismologie, ce qui est principalement dû à l'expression mathématique du problème inverse. En effet, dans ce cas le problème inverse se ramène à un problème de moindres carrés. Il est alors possible d'utiliser des algorithmes classiques (de type Gauss-Newton) pour calculer les maxima de vraisemblance. Mais en contrepartie, cela nous amène à optimiser une fonction qui n'est pas très robuste vis à vis de grandes erreurs de mesures sur certaines données (En anglais : blunders ou outliers). Il est souvent préconisé dans ce cas d'utiliser la norme L1, comme l'illustre l'exemple de la figure(1.1.4).

FIG. 1.1.4. Influence des erreurs de mesure.

Pour illustrer l'influence des erreurs de mesure sur les données dans le cadre de problèmes inverses, nous avons conduit une expérience numérique. Dans un milieu de vitesse homogène à 5 km/s, 31 stations enregistrent le temps de première arrivée d'une onde sismique. Le paramètre cherché est la profondeur du séisme. Les figures suivantes montrent la densité de probabilité a posteriori σ_M , sans a priori sur le modèle, pour différentes lois modélisant les erreurs de mesure sur les données. En bleu, nous avons utilisé la norme L1 (formule (1.1.45)), avec une erreur caractéristique sur les données de 0,4 s. En noir, c'est l'hypothèse gaussienne qui modélise l'erreur (formule (1.1.35)), avec un écart-type de 0,4 s. Enfin en rouge, c'est la loi sech qui décrit les erreurs de mesures (formule (1.1.50)), avec une erreur caractéristique de 0,4 s. Sur la figure du haut à gauche les données sont enregistrées sans bruit, et l'optimum correspond au séisme cherché (profondeur de 5 km). Lorsque que l'on ajoute du bruit gaussien (figure en haut à droite), l'optimum se déplace peu si bien que les 3 lois semblent convenir. Sur la figure du bas à gauche, on a généré une erreur de 3 secondes sur une donnée. Dans ce cas l'hypothèse gaussienne est prise en défaut alors que les deux autres lois ne sont pas affectées par cette erreur de mesure grossière. Enfin sur le dernier dessin, on a cumulé les erreurs du deuxième et troisième cas. Là encore, l'hypothèse gaussienne ne convient plus et génère une erreur sur la solution plus importante qu'avec les autres lois. On voit qu'une distribution non-gaussienne de l'erreur peut biaiser profondément la position de l'optimum de σ_M .



1.1.5.1. *Norme L1.* L'information sur les données s'exprime de la façon suivante,

$$(1.1.44) \quad \rho_\epsilon(\mathbf{d}) \propto \exp\left(-\left(\sum_{i=1}^{i=m} \frac{|d_i - d_i^{obs}|}{\sigma_i^D}\right)\right),$$

ainsi la densité de probabilité *a posteriori* s'écrit,

$$(1.1.45) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\left(\sum_{i=1}^{i=m} \frac{|g_i(\mathbf{m}) - d_i^{obs}|}{\sigma_i^D}\right)\right) \rho_M(\mathbf{m}).$$

Une telle fonction n'étant pas partout différentiable⁴, cela nécessite d'utiliser des méthodes d'optimisation numérique qui ne vont pas calculer des dérivées partielles. Les méthodes explorant de manière statistique⁵ l'espace des modèles sont souvent utilisées mais elles supposent de gros moyens de calculs et ne sont applicables que dans les problèmes de petite taille. En tomographie l'espace des modèles est trop grand pour utiliser de telles méthodes sur les machines actuelles.

La loi L1 n'est pas totalement adaptée aux données sismologiques car elle suppose des mesures précises tout en admettant quelques données aberrantes. Il est préférable de considérer un bruit gaussien sur les données avec la possibilité de quelques données aberrantes. La loi Sech suit cette hypothèse, de plus nous verrons qu'elle est numériquement implémentable dans le cadre tomographique.

1.1.5.2. *Loi Sech.* E. Crase [8], Valette et Lesage [50] proposent de modéliser les erreurs de mesures par

$$(1.1.46) \quad \rho_\epsilon(d - d^{obs}) = \frac{1}{\pi\sigma} \frac{1}{\cosh\left(\frac{d - d^{obs}}{\sigma}\right)}.$$

Cette loi se comporte comme la norme L1 envers les *blunders* (figure (1.1.5)) et comme la norme L2 vis à vis d'un bruit gaussien,

$$(1.1.47) \quad \rho_\epsilon(d - d^{obs}) \sim \frac{2}{\pi\sigma} \exp\left(-\frac{|d - d^{obs}|}{\sigma}\right) \text{ si } \frac{|d - d^{obs}|}{\sigma} \rightarrow +\infty,$$

$$(1.1.48) \quad \rho_\epsilon(d - d^{obs}) \sim \frac{2}{\pi\sigma} \exp\left(-\frac{1}{2} \left(\frac{d - d^{obs}}{\sigma}\right)^2\right) \text{ si } \frac{|d - d^{obs}|}{\sigma} \rightarrow 0.$$

En supposant les données indépendantes, l'information sur celles-ci est modélisée par

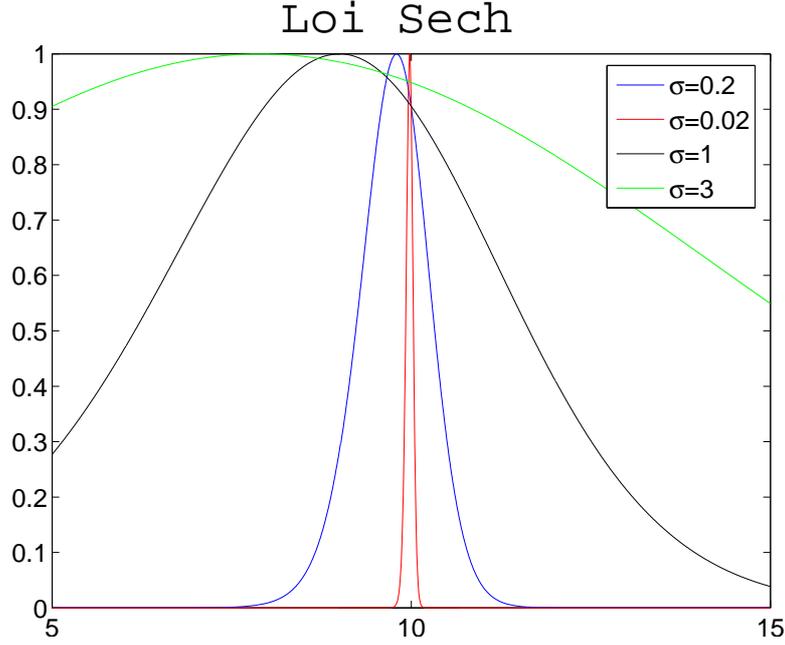
$$(1.1.49) \quad \frac{\rho_D(\mathbf{d})}{\mu_D(\mathbf{d})} = \prod_{i=1}^{i=n} \frac{1}{\pi\sigma_i^D} \frac{1}{\cosh\left(\frac{d_i - d_i^{obs}}{\sigma_i^D}\right)},$$

⁴Nous verrons plus loin que par un changement de variable, la norme L1 est malgré tout utilisable avec un algorithme classique de type Gauss-Newton (cf 1.3.2)

⁵Les méthodes les plus connues sont monte carlo, le recuit-simulé et les algorithmes génétiques.

FIG. 1.1.5. Loi Sech et données aberrantes.

L'expérience synthétique consiste à générer des temps de parcours à partir d'un séisme situé à 10 km de profondeur dans un milieu homogène à 5 km/s. Six stations en surface enregistrent les données, une erreur de mesure de 3 secondes est commise sur l'une des stations. Les courbes montrent la forme prise par la densité de probabilité a posteriori en utilisant la loi Sech selon différentes valeurs de σ (cf relation (1.1.46)). Nous remarquons que lorsque σ est petit, la loi a posteriori se rapproche d'une loi L1 (cf relation (1.1.47)) et au contraire quand σ est grand, la loi se rapproche de l'hypothèse gaussienne (cf relation (1.1.48)). La loi Sech permet par l'intermédiaire du paramètre σ de tester plusieurs hypothèses sur les erreurs de mesures : la loi L1 et la loi Gaussienne.



ce qui conduit à la densité de probabilité *a posteriori*,

$$(1.1.50) \quad \sigma_M(\mathbf{m}) \propto \prod_{i=1}^{i=n} \frac{1}{\pi \sigma_i^D} \frac{1}{\cosh\left(\frac{d_i^{obs} - g_i(\mathbf{m})}{\sigma_i^D}\right)} \rho_M(\mathbf{m}).$$

Cette formule se réécrit,

$$(1.1.51) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^{i=n} -2 \log\left(\frac{1}{\pi \sigma_i^D} \frac{1}{\cosh\left(\frac{d_i^{obs} - g_i(\mathbf{m})}{\sigma_i^D}\right)}\right)\right) \rho_M(\mathbf{m}),$$

$$(1.1.52) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^{i=n} \log\left(\pi \sigma_i^D \cosh\left(\frac{d_i^{obs} - g_i(\mathbf{m})}{\sigma_i^D}\right)\right)^2\right) \rho_M(\mathbf{m}).$$

Lorsque l'information *a priori* sur le modèle est supposée Gaussienne, la densité de probabilité *a posteriori* s'écrit, $\sigma_M(\mathbf{m}) \propto \exp(-E_s(\mathbf{m}))$, avec

$$(1.1.53) \quad E_s(\mathbf{m}) = \frac{1}{2} \left(\sum_{i=1}^{i=n} \log \left(\pi \sigma_i^D \cosh \left(\frac{d_i^{obs} - g_i(\mathbf{m})}{\sigma_i^D} \right) \right)^2 + (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) \right).$$

Sur la figure (1.1.4), on peut voir la forme que prend le graphe de σ_M , en fonction de différentes hypothèses sur la modélisation des erreurs de mesure. L'exemple choisi est simple : il s'agit d'ajuster la profondeur⁶ d'un séisme dans un milieu dont le modèle de vitesse est homogène. Le but de l'expérience est de tester l'adéquation entre la loi utilisée pour modéliser les erreurs de mesure et les erreurs effectives sur les données. Nous avons généré des données d'abord sans bruit, ensuite avec un bruit gaussien. Dans ces deux cas les normes L1, L2 et la loi Sech se comportent de la même façon. Par contre, si on génère une erreur de mesure grossière, l'hypothèse gaussienne n'est plus valide et le maximum de vraisemblance est plus éloigné de la solution qu'avec les deux autres lois. La loi sech convient dans tous les cas, il est donc intéressant de l'utiliser en tomographie. D'autant plus que dans un jeu de données constitué par des temps d'arrivée pointés manuellement sur des sismogrammes, un petit pourcentage de pointés sont entachés d'une importante erreur de mesure. Cela arrive quand la station se situe dans un plan nodal pour les ondes P : dans ce cas il est possible de prendre la phase S pour la première arrivée P. Une telle erreur peut entraîner un biais important dans la position du maximum de vraisemblance sous l'hypothèse gaussienne alors qu'il n'en est rien pour la norme L1 et la loi Sech. On peut avancer un autre argument en faveur de la loi Sech : celle-ci peut par un changement de variable dans l'espace des données, se ramener à un problème de moindres carrés. Il est alors possible d'utiliser les méthodes numériques classiques de moindres carrés pour résoudre le problème inverse. Nous verrons plus loin comment procéder en pratique (cf partie 1.3).

1.1.6. Résoudre le problème inverse. Résoudre le problème inverse, c'est étudier la fonction $\mathbf{m} \mapsto \sigma_M(\mathbf{m})$ (figure (1.1.6)). Le cas le plus favorable est lorsque la dimension de l'espace des paramètres M est inférieure ou égale à 4. On peut alors explorer entièrement l'espace des modèles admissibles physiquement en construisant le graphe de σ_M . Il est ainsi possible de calculer la probabilité de présence du paramètre dans un sous ensemble donné. Par exemple, la probabilité de présence du modèle \mathbf{m} dans l'ensemble $A \subset M$, est donnée par⁷ ;

$$(1.1.54) \quad P_{\sigma_M}(\mathbf{m} \in A) = \frac{\int_A \sigma_M(\mathbf{m}) \, d\mathbf{m}}{\int_M \sigma_M(\mathbf{m}) \, d\mathbf{m}}.$$

On peut ainsi cartographier dans M les zones où la solution au problème inverse a le plus de chance d'être présente. On comprend aisément que pour utiliser cette méthode on doit pouvoir calculer $\sigma_M(\mathbf{m})$ très rapidement (le calcul du problème direct doit donc être très rapide).

⁶Les autres paramètres sont connus.

⁷Il est nécessaire de normaliser par $\int_M \sigma_M(\mathbf{m}) \, d\mathbf{m}$, car dans la pratique on calcule une fonction proportionnelle à σ_M .

Malheureusement, les problèmes sismologiques réels sont souvent de dimension trop grande⁸ pour qu'il soit possible de représenter σ_M . On peut alors faire appel à des méthodes statistiques d'exploration de σ_M , et ainsi se faire une idée de la forme du graphe. Là encore ces méthodes ne sont vraiment utilisables que dans le cas d'espaces de dimension relativement réduite.

Néanmoins, si la fonction σ_M n'a pas une forme trop compliquée ; si elle possède un unique maximum, appelé aussi *maximum de vraisemblance*, on peut chercher à le calculer et faire une approximation de σ_M au voisinage de celui-ci. Ainsi on peut se faire une idée de la forme de la densité de probabilité *a posteriori*.

Dans les applications qui nous intéressent, la dimension de l'espace des modèles est telle qu'il faudra utiliser cette dernière démarche. Elle est particulièrement adaptée lorsque ρ suit une loi Gaussienne.

1.1.7. Etude de la loi gaussienne tangente : ellipsoïde des erreurs. La densité de probabilité *a posteriori* dans l'espace des modèles s'écrit,

$$(1.1.55) \quad \sigma_M(\mathbf{m}) \propto \exp(-E(\mathbf{m}))$$

aussi bien sous l'hypothèse gaussienne (1.1.36) qu'avec la loi sech (1.1.53). Notons $\hat{\mathbf{m}}$ un maximum de vraisemblance de (1.1.55), écrivons le développement de Taylor à l'ordre deux de la fonction E au voisinage de $\hat{\mathbf{m}}$, en utilisant le fait que $\nabla E(\hat{\mathbf{m}}) \equiv 0$,

$$(1.1.56) \quad E(\mathbf{m}) = E(\hat{\mathbf{m}}) + \frac{1}{2} \langle \nabla^2 E(\hat{\mathbf{m}})(\mathbf{m} - \hat{\mathbf{m}}), \mathbf{m} - \hat{\mathbf{m}} \rangle + o(\|\mathbf{m} - \hat{\mathbf{m}}\|^2).$$

En insérant (1.1.56) dans (1.1.55) la densité de probabilité *a posteriori* devient,

$$(1.1.57) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2}E(\hat{\mathbf{m}})\right) \exp(o(\|\mathbf{m} - \hat{\mathbf{m}}\|^2)) \exp\left(-\frac{1}{2}(\mathbf{m} - \hat{\mathbf{m}})^t \nabla^2 E(\hat{\mathbf{m}})(\mathbf{m} - \hat{\mathbf{m}})\right).$$

Ainsi au voisinage d'une solution $\hat{\mathbf{m}}$, la densité de probabilité *a posteriori* se comporte comme une loi gaussienne de moyenne $\hat{\mathbf{m}}$ et de matrice de variance-covariance $\nabla^2 E(\hat{\mathbf{m}})^{-1}$,

$$(1.1.58) \quad \sigma_M(\mathbf{m}) \simeq C^{te} \exp\left(-\frac{1}{2}(\mathbf{m} - \hat{\mathbf{m}})^t \nabla^2 E(\hat{\mathbf{m}})(\mathbf{m} - \hat{\mathbf{m}})\right).$$

Cette dernière relation induit une méthode pour modéliser les erreurs sur le modèle estimé par le biais de l'étude de la matrice $\nabla^2 E(\hat{\mathbf{m}})^{-1}$. Il est très coûteux numériquement de calculer cette matrice ; en pratique nous allons faire l'approximation suivante,

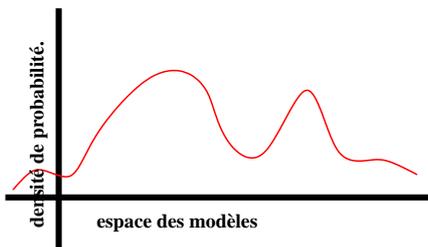
$$(1.1.59) \quad \nabla^2 E(\hat{\mathbf{m}}) \simeq \mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1},$$

où \mathbf{G} est la matrice jacobienne de l'application $\mathbf{m} \mapsto \mathbf{g}(\mathbf{m})$ calculée en $\hat{\mathbf{m}}$. Cela revient à linéariser le problème direct autour de $\hat{\mathbf{m}}$; la densité de probabilité *a posteriori* est une loi gaussienne qui a pour matrice de covariance $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})^{-1}$. La relation (1.1.59) est valable sous l'hypothèse Gaussienne mais aussi avec la loi Sech comme le justifie l'équivalent

⁸Sauf dans le cas de la localisation d'un unique séisme où l'espace des modèles est de dimension 4, il est alors possible de représenter σ_M , tout du moins ses lois marginales.

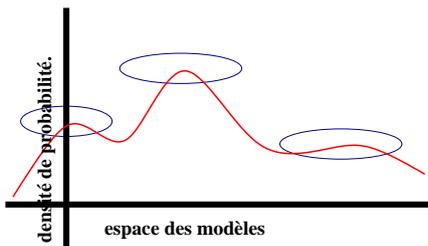
FIG. 1.1.6. Résoudre un problème inverse.

La résolution d'un problème inverse passe par l'étude de la fonction σ_M . Le cas le plus favorable est le cas 1, où on peut tracer entièrement la fonction. Si ce n'est pas possible, on peut explorer σ_M de façon statistique (Monte-carlo, recuit-simulé, algorithmes génétiques). On ne connaîtra la fonction qu'au voisinages des maxima de vraisemblance (cas 2). En général les deux méthodes précédentes ne sont pas applicables, pour des raisons de coût en temps de calcul. Néanmoins, si σ_M est unimodale on peut chercher le maximum de vraisemblance par des méthodes d'optimisations numériques et étudier la fonction au voisinage de l'optimum. Cette technique est utilisable dans certains problèmes de grande taille (cas 3).



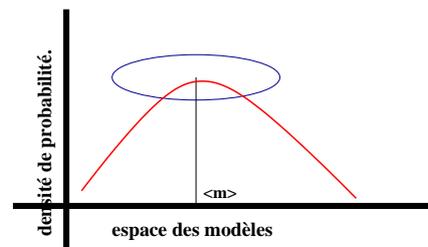
①

Lorsque la dimension de l'espace des modèles est faible, on peut dessiner la densité de probabilité a posteriori.



②

Lorsque l'on ne peut pas dessiner la densité de probabilité entièrement, il est possible d'avoir recours à une exploration statistique de la fonction pour se faire une idée de son comportement au voisinage des maxima de vraisemblance.



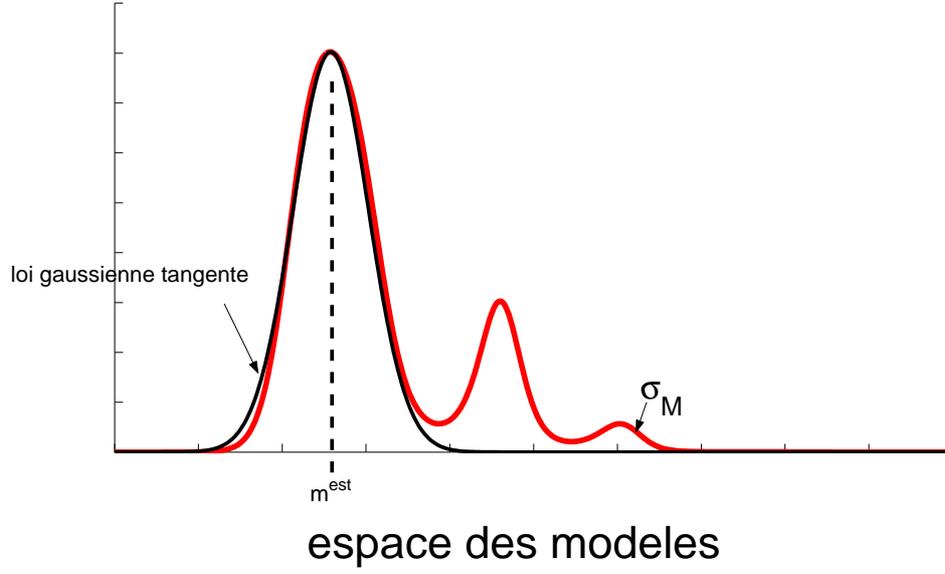
③

Si la forme de la densité de probabilité a posteriori est unimodale, on peut chercher le maximum de vraisemblance par une méthode d'optimisation et étudier le comportement de la fonction au voisinage de celui-ci.

(1.1.48). Nous étudierons donc le comportement local de σ_M autour d'une solution \hat{m} en exploitant l'expression,

FIG. 1.1.7. Loi gaussienne tangente.

Pour étudier le comportement la loi de probabilité à postériori σ_M (courbe rouge) autour d'un maximum de vraisemblance \mathbf{m}^{est} , il est pratique de l'approximer par sa loi gaussienne tangente (courbe noire).



$$(1.1.60) \quad \sigma_M(\mathbf{m}) \simeq C^{te} \exp\left(-\frac{1}{2}(\mathbf{m} - \hat{\mathbf{m}})^t (\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})(\mathbf{m} - \hat{\mathbf{m}})\right).$$

La loi (1.1.60) est appelée loi gaussienne tangente (figure (1.1.7)). On peut, en premier lieu, interpréter la diagonale de $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})^{-1}$ en terme d'erreur sur les paramètres estimés. En effet, on lit la variance de chaque paramètre sur cette diagonale. Pour une analyse plus fine, il est possible de décomposer,

$$(1.1.61) \quad (\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1}) = \mathbf{U} \mathbf{S} \mathbf{U}^t,$$

où \mathbf{S} est une matrice diagonale qui contient les valeurs propres de $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})$ et \mathbf{U} est une matrice orthogonale dont les colonnes sont composées des vecteurs propres de $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})$. Ainsi on détermine un ellipsoïde d'erreur dont les directions des demi-axes sont les vecteurs propres ; les valeurs propres associées nous informent sur la valeur de l'incertitude suivant ces directions. Lorsque une valeur propre est faible vis à vis des autres, cela indique que les paramètres du vecteur propre associé sont moins bien déterminés que les autres. La notion de conditionnement d'une matrice permet aussi de faire une analyse dans le même sens. En effet, le conditionnement de la matrice $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})$ est par définition le rapport entre la valeur propre la plus élevée et la plus faible, une forte valeur de conditionnement nous indique qu'il existe des paramètres moins bien déterminés, a

contrario une valeur de conditionnement proche de 1 indique que tous les paramètres sont déterminés de la même façon. Notons que l'on peut décomposer la matrice,

$$(1.1.62) \quad (\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1}) = \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix}^t \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix},$$

nous pourrons donc mener sur la matrice $\begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix}$ un raisonnement identique à celui fait sur $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})$ en utilisant la décomposition SVD (cf Annexe 4),

$$(1.1.63) \quad \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix} = \mathbf{V} \mathbf{S} \mathbf{U}^t.$$

Le choix d'étudier l'une ou l'autre matrice est fait en fonction du problème inverse considéré, dans certains cas il est plus facile d'étudier $\begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix}$ alors que dans d'autres types de problèmes il sera préférable de raisonner avec $(\mathbf{G}^t \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1})$. Dans les applications tomographiques, nous utiliserons la matrice $\begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{pmatrix}$ car elle a la propriété d'être creuse.

1.2. Résultats généraux en optimisation.

Le problème inverse tel que nous l'avons défini conduit à optimiser une fonctionnelle. On énonce dans cette partie des résultats mathématiques généraux, dans le but de définir les notions essentielles utilisées dans les algorithmes d'optimisation. Les résultats ne seront pas démontrés. Le lecteur intéressé par une approche mathématique rigoureuse pourra se reporter à l'ouvrage [6].

Considérons le problème d'optimisation,

$$(P) \quad \begin{cases} \min f(\mathbf{x}) \\ \mathbf{x} \in \mathbb{R}^n, \end{cases}$$

dans lequel on minimise une fonction coût (ou de mérite) $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Un minimum global de ce problème est un point $\mathbf{x}_* \in \mathbb{R}^n$ tel que,

$$(1.2.1) \quad f(\mathbf{x}_*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n.$$

On dit que \mathbf{x}_* est un minimum local s'il existe un voisinage V de \mathbf{x}_* tel que,

$$(1.2.2) \quad f(\mathbf{x}_*) \leq f(\mathbf{x}), \forall \mathbf{x} \in V.$$

Résoudre le problème (P) c'est trouver \mathbf{x}_* . Il faut d'abord se persuader de l'existence d'un optimum (local ou global), ce qui dépend des propriétés de f .

1.2.1. Question d'existence. Une première condition pour que le problème (P) ait un sens est que f soit minorée, c'est à dire qu'il existe une constante c telle que,

$$(1.2.3) \quad f(\mathbf{x}) \geq c, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Mais cela ne suffit pas pour conclure à l'existence d'un minimum global (considérer la fonction $x \mapsto \exp(x)$). Il existe une propriété qui est habituellement vérifiée dans les problèmes d'optimisation, f continue et $f(\mathbf{x}) \rightarrow +\infty$ si $|\mathbf{x}| \rightarrow +\infty$. Une fonction vérifiant cette propriété est dite inf-compacte. Dans ce cas on peut conclure à l'existence d'un minimum global (pas forcément unique).

1.2.2. Conditions d'optimalité. Les conditions d'optimalité sont des propriétés que vérifient les solutions de (P). Ces conditions sont utiles pour :

- Vérifier l'optimalité éventuelle d'un point \mathbf{x} .
- Mettre en oeuvre des méthodes pour résoudre numériquement (P)

On parle de conditions du premier ordre lorsque celles-ci ne font intervenir que les dérivées premières de f . Les conditions du second ordre font intervenir les dérivées du premier et du second ordre. Il y a des conditions nécessaires, et des conditions suffisantes :

▷ Conditions nécessaires : si \mathbf{x}_* est optimal alors :

- (1) Condition nécessaire du premier ordre : le gradient $f'(\mathbf{x}_*)$ est nul.
- (2) Condition nécessaire du second ordre : le hessien $f''(\mathbf{x}_*)$ est semi-défini positif.

▷ Condition suffisante : si \mathbf{x}_* est tel que $f'(\mathbf{x}_*) = 0$ et $f''(\mathbf{x}_*)$ est défini positif, alors \mathbf{x}_* est un minimum local.

Les algorithmes d'optimisation que nous verrons ici visent à trouver un point stationnaire (ie : vérifiant la condition nécessaire (1)). On doit résoudre un système d'équations non linéaire,

$$(1.2.4) \quad f'(\mathbf{x}) = g(\mathbf{x}) = 0.$$

Remarquons que si \mathbf{x}_* est solution de la relation précédente (1.2.4) alors il vérifie,

$$(1.2.5) \quad \mathbf{x}_* = \mathbf{x}_* - M^{-1}g(\mathbf{x}_*)$$

où M est une matrice carré inversible quelconque. On utilise la méthode itérative suivante pour résoudre (1.2.5),

$$(1.2.6) \quad \begin{cases} \mathbf{x}_0, \text{ point initial connu,} \\ \mathbf{x}_{k+1} = \mathbf{x}_k - M_k^{-1}g(\mathbf{x}_k). \end{cases}$$

Dans les cas favorables⁹, on peut appliquer le théorème du point fixe et conclure à la convergence de la suite $\{\mathbf{x}_k\}$ vers \mathbf{x}_* . Par exemple, la méthode de Newton pour résoudre un système d'équations non linéaires consiste à définir $M_k = g'(\mathbf{x}_k)$. La formulation (1.2.5) nous laisse une importante latitude pour choisir les matrices M_k . Or, nous devons optimiser la fonction coût f (avec $f'=g$), nous allons utiliser cette information pour définir la suite $\{M_k\}$. C'est l'objet d'une classe d'algorithmes que l'on nomme méthodes à direction de descente.

⁹On doit avoir certaines propriétés sur g , et la suite $\{M_k\}$ pour assurer la convergence (cf [6]).

1.2.3. Méthodes à direction de descente.

1.2.3.1. *Définition de la méthode.* Il s'agit de trouver le minimum d'une fonction, $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Dans le cas où l'on ne connaît pas de formules explicites donnant le résultat, il faut appliquer une méthode numérique calculant (bien souvent de manière approchée) le vecteur cherché. Une idée possible est de choisir un vecteur \mathbf{x}_0 et de définir une suite $\{\mathbf{x}_k\}$ telle que

$$(1.2.7) \quad f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k).$$

On génère une suite qui à chaque étape fait diminuer la fonction, dans les cas favorables on peut espérer que cette suite converge vers le minimum. Il convient de définir une méthode permettant de calculer l'itéré $k+1$ étant donné l'itéré k . Pour cela écrivons le développement de Taylor au premier ordre en \mathbf{x}_k ,

$$(1.2.8) \quad f(\mathbf{x}_k + \delta) = f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \delta \rangle + o(|\delta|),$$

supposons que δ soit assez petit pour que l'on néglige le terme $o(|\delta|)$,

$$(1.2.9) \quad f(\mathbf{x}_k + \delta) - f(\mathbf{x}_k) = \langle \nabla f(\mathbf{x}_k), \delta \rangle.$$

Étant donné la relation (1.2.7), la perturbation δ à apporter à \mathbf{x}_k doit vérifier,

$$(1.2.10) \quad \langle \nabla f(\mathbf{x}_k), \delta \rangle < 0.$$

Un vecteur δ vérifiant la formule (1.2.10) est par définition appelé direction de descente. Un algorithme d'optimisation doit donc calculer pour chaque itéré k , une direction de descente δ qui perturbe l'itéré courant pour trouver le point suivant de la suite. Une fois la direction de descente déterminée, il s'agit alors de choisir le meilleur pas possible le long de celle-ci. Cette étape est effectuée par un algorithme appelé recherche linéaire. Concrètement on écrit $\delta = t\mathbf{d}$ où t est un réel positif et \mathbf{d} la direction de descente. La recherche linéaire choisit un réel t qui vise à améliorer les propriétés de la suite.

Précisons un peu plus la structure d'un algorithme d'optimisation à direction de descente. En utilisant les résultats précédents on peut décrire schématiquement un tel algorithme. Notons \mathbf{x} l'itéré courant,

- (1) Si \mathbf{x} vérifie le critère d'arrêt¹⁰ alors stop, sinon aller en (2).
- (2) Calculer la direction de descente \mathbf{d} .
- (3) Recherche linéaire : calculer t .
- (4) faire $\mathbf{x} = \mathbf{x} + t\mathbf{d}$ et retourner en (1).

Il existe plusieurs manières (classiques) de définir les directions de descentes en (2), nous verrons plus loin les méthodes du gradient, du gradient conjugué, de Newton, de quasi-Newton et de Gauss-Newton. Ces méthodes définissent localement les directions de descente, en utilisant une approximation de la fonction à optimiser. Cela implique en général qu'il n'y a que convergence locale. C'est à dire qu'il faut que le premier itéré soit proche de la solution pour définir une suite convergente. Ces méthodes sont nettement améliorées

¹⁰Les conditions d'optimalité permettent de définir un critère d'arrêt. On utilise les propriétés de la solution pour tester si \mathbf{x} en est une.

lorsqu'on leur adjoint une recherche linéaire (3) qui a pour rôle de forcer la convergence globalement. Les premiers itérés peuvent être loin de la solution, dans ce cas la recherche linéaire doit définir des pas de grande taille. Lorsqu'on se rapproche de la solution, l'approximation de la fonction est valable, alors on va privilégier des pas restant autour de 1.

1.2.3.2. *Quelques méthodes classiques à direction de descente.* Il existe plusieurs méthodes classiques pour calculer une direction de descente. En général on minimise localement une approximation de la fonction coût au voisinage de l'itéré courant. L'idée la plus naturelle est de remplacer f par son application linéaire tangente, cela conduit à définir comme direction de descente, l'opposé du gradient.

Méthode du gradient. Ecrivons l'approximation linéaire de f au voisinage d'un itéré \mathbf{x}_k ,

$$(1.2.11) \quad f(\mathbf{x}_k + \delta) \simeq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \delta \rangle,$$

notons

$$(1.2.12) \quad \phi(\delta) = f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \delta \rangle.$$

Comme nous l'avons vu plus haut (1.2.7), on veut définir δ tel que

$$(1.2.13) \quad f(\mathbf{x}_k + \delta) < f(\mathbf{x}_k),$$

c'est à dire que l'on veut minimiser la fonction $\delta \mapsto f(\mathbf{x}_k + \delta)$. Pour simplifier remplaçons $f(\mathbf{x}_k + \delta)$ par ϕ et calculons δ tel qu'il minimise ϕ (au lieu de f) dans un voisinage de \mathbf{x}_k . Il alors suffit de poser $\delta = -t \nabla f(\mathbf{x}_k)$, où t est un réel positif. C'est une direction de descente puisque $-\nabla f(\mathbf{x}_k)$ vérifie la condition (1.2.10). Une direction de descente définie ainsi est mauvaise car en général un tel algorithme converge en un temps infini ou pas du tout, et même la convergence locale n'est pas garantie (lorsque le premier itéré est proche de la solution).

Méthode du gradient conjugué. L'algorithme du gradient conjugué s'inspire de la méthode du gradient en améliorant les directions de descente par les formules,

$$(1.2.14) \quad \mathbf{d}_k = \begin{cases} -\nabla f(\mathbf{x}_0) & \text{si } k = 1, \\ -\nabla f(\mathbf{x}_k) + \beta_k \mathbf{d}_{k-1} & \text{si } k > 1. \end{cases}$$

où le scalaire β_k peut prendre différentes valeurs.

Méthode de Newton. La méthode de Newton utilise une approximation quadratique de f , cela permet de prendre en considération la courbure de la fonction, ce que l'approximation linéaire ne fait pas. Cette façon de faire a pour but d'obtenir une meilleure convergence de la suite des approximations successives de la solution. On écrit le développement de Taylor de f au second ordre,

$$(1.2.15) \quad f(\mathbf{x}_k + \delta) \simeq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \delta \rangle + \frac{1}{2} \langle \delta, \nabla^2 f(\mathbf{x}_k) \delta \rangle,$$

on calcule alors δ de manière à minimiser l'approximation quadratique de f . On obtient,

$$(1.2.16) \quad \delta = -\nabla^2 f(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k).$$

Cette méthode construit une suite qui possède des propriétés de convergence plus favorables que celle de la méthode du gradient. Mais elle est beaucoup plus difficile à mettre en oeuvre. Il faut d'abord calculer le gradient $\nabla f(\mathbf{x}_k)$ et ensuite le hessien $\nabla^2 f(\mathbf{x}_k)$ et enfin résoudre le système linéaire,

$$(1.2.17) \quad \nabla^2 f(\mathbf{x}_k) \delta = -\nabla f(\mathbf{x}_k),$$

pour calculer δ . Dans certains problèmes, surtout lorsqu'ils sont de grande taille, la méthode de Newton peut devenir très lourde, voire impossible à mettre en oeuvre. Il est alors nécessaire de l'alléger pour définir un algorithme praticable, c'est ainsi qu'ont été introduites les méthodes dites de quasi-Newton.

Méthodes de quasi-Newton. Le principe reste le même qu'avec la méthode de Newton, remplacer f localement par une approximation quadratique,

$$(1.2.18) \quad f(\mathbf{x}_k + \delta) \simeq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \delta \rangle + \frac{1}{2} \langle \delta, M_k \delta \rangle,$$

la direction de descente est alors,

$$(1.2.19) \quad \mathbf{d}_k = -M_k^{-1} \nabla f(\mathbf{x}_k).$$

Au lieu de calculer explicitement le Hessien de f , on préfère ici utiliser une matrice M_k qui est plus facile à calculer et à utiliser. Mais cette matrice doit vérifier certaines propriétés pour conserver la convergence de la suite $\{\mathbf{x}_k\}$. Il y a deux idées à respecter, M_k est censée approximer le Hessien de f , on doit alors imposer à M_k d'être symétrique définie positive. Ensuite il faut que les directions de descente définies en (1.2.19) soient proches de celles définies par la formule (1.2.16). Cela laisse beaucoup de latitude quant au choix de M_k . C'est pourquoi il existe de nombreuses méthodes de quasi-Newton, citons pour mémoire les méthodes¹¹ BFGS, DFP,...

Méthode de Gauss-Newton. C'est une méthode de quasi-Newton appliquée au problème particulier des moindres carrés. La fonction coût s'écrit,

$$(1.2.20) \quad f(\mathbf{x}) = \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2,$$

où $\mathbf{x} \mapsto \mathbf{r}(\mathbf{x})$ est une fonction de \mathbb{R}^n valeurs dans \mathbb{R}^m . En notant $\mathbf{G}(\mathbf{x})$ la matrice jacobienne de \mathbf{r} en \mathbf{x} , On peut facilement voir que (cf Annexe 5)

$$(1.2.21) \quad \nabla f(\mathbf{x}) = \mathbf{G}(\mathbf{x})^t \mathbf{r}(\mathbf{x})$$

et

$$(1.2.22) \quad \nabla^2 f(\mathbf{x}) = \mathbf{G}(\mathbf{x})^t \mathbf{G}(\mathbf{x}) + \sum_{i=1}^{i=m} r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}).$$

¹¹Dans ces méthodes (BFGS, DFP) on approxime directement M_k^{-1} , cela évite d'avoir à résoudre un système linéaire à chaque itération.

Pour éviter de calculer les dérivées secondes de r , la méthode de Gauss-Newton définit les directions de descente par

$$(1.2.23) \quad \mathbf{d}_k = - \left(\mathbf{G}(\mathbf{x}_k)^t \mathbf{G}(\mathbf{x}_k) \right)^{-1} \mathbf{G}(\mathbf{x}_k)^t r(\mathbf{x}_k).$$

En reprenant les notations du paragraphe précédent, $M_k = \mathbf{G}(\mathbf{x}_k)^t \mathbf{G}(\mathbf{x}_k)$.

1.2.4. Conclusion. Un algorithme d'optimisation à directions de descente définit une suite $\{\mathbf{x}_k\}_k$ qui, dans les cas favorables, converge vers un vecteur \mathbf{x}_* vérifiant $f'(\mathbf{x}_*) = 0$. Cette condition ne suffit pas pour conclure que \mathbf{x}_* est le minimum cherché. Il faut étudier le comportement de f au voisinage de \mathbf{x}_* , en utilisant par exemple une approximation au second ordre de celle-ci, pour déterminer si la solution est un minimum local. Il n'existe pas de méthode générale permettant de définir des directions de descente garantissant la convergence de la suite. En effet, la pratique montre que le choix des directions de descente dépend du problème posé. Une méthode éprouvée sur un type de problème ne fonctionnera pas forcément dans un autre cas. Toute la difficulté consiste à trouver des directions de descente adaptées afin de faire converger la suite. Dans les problèmes réels, la convergence ne garantit absolument pas que la solution obtenue soit la solution au sens physique. En effet, il faut distinguer plusieurs sources d'erreurs susceptibles de biaiser la solution :

- (1) Erreurs d'approximation ou de discrétisation de la fonction f .
- (2) Erreurs dues aux approximations d'un algorithme itératif (erreurs de la méthode).
- (3) Erreurs de calcul dues à l'ordinateur (approximation des réels par les nombres flottants ; précision finie.).
- (4) Erreurs sur les données.

Définir un algorithme d'optimisation performant pour résoudre un problème physique, passe d'abord par la prise en compte de ces quatre sources d'erreurs. L'étude de celles-ci permet d'apporter une réponse adaptée à la nature du problème et de contrôler la qualité de la solution. La démarche suivie dans cette thèse va dans ce sens ; nous nous sommes particulièrement attachés aux points (1), (3) et (4) dans le cadre de problèmes tomographiques. Ainsi, nous avons implémenté un problème direct précis et robuste¹² (point (1)). Nous avons utilisé une méthode d'inversion de matrice adaptée aux cas de matrices de grande taille et mal conditionnées¹³ (point (3)). Enfin nous avons utilisé des données adaptées et précises¹⁴ (point (4)).

1.3. Schémas numériques.

Nous allons définir les schémas numériques permettant de calculer les maxima de vraisemblance des densités de probabilité *a posteriori* définies dans la partie 1.1. Ces méthodes sont de simples applications de l'algorithme de Gauss-Newton aux cas particuliers qui nous intéressent.

¹²cf partie [2.2], nous utilisons l'algorithme Podvin-Lecomte [37] pour calculer les temps de première arrivée. Ensuite, nous améliorons le temps de trajet source-station par un calcul d'intégrale.

¹³cf partie [1.3], nous avons choisi la méthode LSQR de Paige et Saunders [36].

¹⁴Un point important est l'utilisation de décalages temporels interspectraux [13].

1.3.1. Information *a priori* gaussienne. Reprenons la formule (1.1.35) de la densité de probabilité *a posteriori* sous l'hypothèse gaussienne,

$$(1.3.1) \quad \sigma_M(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0)\right).$$

Nous cherchons à calculer un maximum de vraisemblance de σ_M , c'est à dire minimiser la fonction coût,

$$(1.3.2) \quad E_g(\mathbf{m}) = \frac{1}{2}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}))^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0).$$

Il est possible d'utiliser l'algorithme de Gauss-Newton (1.2.23) pour effectuer cette optimisation. Pour cela, il faut en premier lieu calculer le gradient de la fonction E_g , (cf Annexe 5),

$$(1.3.3) \quad \nabla E_g(\mathbf{m}_k) = -\mathbf{G}_k^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}_k)) - \mathbf{C}_m^{-1}(\mathbf{m}_0 - \mathbf{m}_k),$$

en notant \mathbf{G}_k la matrice jacobienne de l'application $\mathbf{m} \mapsto \mathbf{g}(\mathbf{m})$ calculée en \mathbf{m}_k . Comme nous l'avons vu au paragraphe 1.1.4, équation (1.1.39), la fonction coût s'écrit comme un problème aux moindres carrés non linéaire,

$$(1.3.4) \quad E_g(\mathbf{m}) = \frac{1}{2} \left\| \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m} \end{bmatrix} \right\|^2,$$

la matrice jacobienne de l'application $\mathbf{m} \mapsto \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{d}_o \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{g}(\mathbf{m}) \\ -\mathbf{C}_m^{-\frac{1}{2}} \mathbf{m} \end{bmatrix}$, calculée en \mathbf{m}_k s'écrit,

$$(1.3.5) \quad \mathbf{A}_k = \begin{bmatrix} -\mathbf{C}_D^{-\frac{1}{2}} \mathbf{G}_k \\ \mathbf{C}_m^{-\frac{1}{2}} \end{bmatrix}.$$

Tous les éléments sont réunis pour appliquer l'algorithme de Gauss-Newton (1.2.23),

$$(1.3.6) \quad \mathbf{m}_{k+1} = \mathbf{m}_k + (\mathbf{A}_k^t \mathbf{A}_k)^{-1} \{-\nabla E_g(\mathbf{m}_k)\},$$

ainsi, en injectant (1.3.5) et (1.3.3) dans (1.3.6) on obtient la relation

$$(1.3.7) \quad \mathbf{m}_{k+1} = \mathbf{m}_k + (\mathbf{G}_k^t \mathbf{C}_D^{-1} \mathbf{G}_k + \mathbf{C}_m^{-1})^{-1} (\mathbf{G}_k^t \mathbf{C}_D^{-1}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}_k)) + \mathbf{C}_m^{-1}(\mathbf{m}_0 - \mathbf{m}_k)).$$

Nous remarquons aussi que,

$$(1.3.8) \quad \nabla E_g(\mathbf{m}_k) = \mathbf{A}_k^t \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}}(\mathbf{d}_o - \mathbf{g}(\mathbf{m}_k)) \\ \mathbf{C}_m^{-\frac{1}{2}}(\mathbf{m}_k - \mathbf{m}_0) \end{bmatrix},$$

ainsi en reformulant (1.3.6), la perturbation δ_k à apporter au modèle \mathbf{m}_k vérifie,

$$(1.3.9) \quad (\mathbf{A}_k^t \mathbf{A}_k) \delta_k = \mathbf{A}_k^t \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}}(\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ \mathbf{C}_m^{-\frac{1}{2}}(\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix},$$

ce sont les équations normales du problème de moindres carrés linéaire,

$$(1.3.10) \quad \text{minimiser } \frac{1}{2} \left\| \mathbf{A}_k \delta - \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} (\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix} \right\|^2.$$

Nous écrivons ce problème de manière formelle de la façon suivante,

$$(1.3.11) \quad \mathbf{A}_k \delta \sim \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} (\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix}.$$

Nous aboutissons enfin au schéma numérique qui sera appliqué aux méthodes tomographiques,

$$(\text{schéma gn}) \quad \begin{bmatrix} -\mathbf{C}_D^{-\frac{1}{2}} \mathbf{G}_k \\ \mathbf{C}_m^{-\frac{1}{2}} \end{bmatrix} \delta_k \sim \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} (\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix}, \text{ au sens des moindres carrés.}$$

1.3.2. Loi Sech. Le schéma numérique de Gauss-Newton est facile à implémenter quand on sait calculer les dérivées partielles du problème direct $\mathbf{m} \mapsto \mathbf{g}(\mathbf{m})$. Cette méthode est définie pour des erreurs de mesure Gaussiennes sur des données $\mathbf{d}_o = (d_i^o)_{i \in \mathcal{S}}$. Or, en pratique cette hypothèse n'est pas toujours valide. On peut malgré tout adapter un algorithme de Gauss-Newton au cas d'erreurs de mesure non Gaussiennes. En effet, en théorie on peut toujours ramener une densité de probabilité *a posteriori* σ à une loi Gaussienne. L'idée est de définir une fonction φ ,

$$(1.3.12) \quad \begin{array}{ccc} D \times M & \xrightarrow{\sigma} & \mathbb{R}^+ \\ \varphi \downarrow & \nearrow & \bar{\sigma} = \sigma \circ \varphi^{-1} \\ D_a \times M & & \end{array}$$

telle que la densité *a posteriori* $\bar{\sigma}$ suive une loi Gaussienne dans l'espace $D_a \times M$. Plus précisément, supposons que la donnée d_i suive une loi quelconque de densité ρ_i et de médiane d_i^o . Définissons une nouvelle donnée \bar{d}_i par la formule suivante,

$$(1.3.13) \quad \bar{d}_i(d_i) = \text{Erf}^{-1} \left(2 \int_{d_i^o}^{d_i} \rho_i(x) dx \right),$$

où $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-\xi^2) d\xi$. On peut montrer que la variable \bar{d}_i suit une loi Gaussienne de moyenne 0 et de variance $\frac{1}{2}$. Ainsi, par ce changement de variable, le problème d'ajustement des données d_i revient à ajuster les nouvelles données \bar{d}_i au sens des moindres carrés. Notons que l'on ne travaille plus dans l'espace des données physiques D , mais dans un espace de données auxiliaires D_a qui n'a pas de signification physique. Le problème direct est alors différent, il s'exprime en fonction du problème direct physique,

$$(1.3.14) \quad \bar{g}_i(\mathbf{m}) = \bar{d}_i(g_i(\mathbf{m})).$$

Pour appliquer la méthode de Gauss-Newton au nouveau problème il faut calculer les dérivées partielles du problème direct (1.3.14). En dérivant (1.3.13) on exprime les dérivées partielles en fonction de la donnée physique d_i ,

$$(1.3.15) \quad \frac{\partial \bar{d}_i}{\partial d_i}(d_i) = \sqrt{\pi} \exp\left(\bar{d}_i^2\right) \rho_i(d_i),$$

cette dernière expression est utilisée pour calculer les dérivées partielles du nouveau problème direct,

$$(1.3.16) \quad \frac{\partial \bar{d}_i}{\partial m_k}(g_i(\mathbf{m})) = \frac{\partial \bar{d}_i}{\partial d_i}(g_i(\mathbf{m})) \frac{\partial g_i}{\partial m_k}(\mathbf{m}) = \sqrt{\pi} \exp\left(\bar{d}_i^2\right) \rho_i(d_i) \frac{\partial g_i}{\partial m_k}(\mathbf{m}).$$

Les formules (1.3.13) et (1.3.16) permettent d'implémenter la méthode de Gauss-Newton sur des données qui ne satisfont pas l'hypothèse Gaussienne. Par exemple, dans le cas où ρ_i est une loi Sech, la formule (1.3.13) se réécrit,

$$(1.3.17) \quad \bar{d}_i(d_i) = \text{Erf}^{-1}\left(\frac{2}{\pi} \arctan\left(\sinh\left(\frac{d_i - d_i^o}{\sigma_i}\right)\right)\right),$$

et la formule (1.3.15) devient,

$$(1.3.18) \quad \frac{\partial \bar{d}_i}{\partial d_i}(d_i) = \frac{\exp\left(\bar{d}_i^2\right)}{\sigma \sqrt{\pi} \cosh\left(\frac{d_i - d_i^o}{\sigma_i}\right)}.$$

Dans ce cas, la fonction φ mentionnée ci-dessus s'écrit,

$$(1.3.19) \quad \varphi(\mathbf{d}, \mathbf{m}) = (\bar{\mathbf{d}}, \mathbf{m}),$$

où $\bar{\mathbf{d}} = \left(\text{Erf}^{-1}\left(\frac{2}{\pi} \arctan\left(\sinh\left(\frac{d_i - d_i^o}{\sigma_i}\right)\right)\right)\right)_{i \in \mathfrak{S}}$. L'algorithme de Gauss-Newton appliqué aux données $\bar{\mathbf{d}}$ s'écrit,

$$(1.3.20) \quad \mathbf{m}_{k+1} = \mathbf{m}_k + \left(\mathbf{C}_m^{-1} + \left(\mathbf{G}'_k\right)^t \mathbf{G}'_k\right)^{-1} \left\{ \left(\mathbf{G}'_k\right)^t (\bar{\mathbf{g}}(\mathbf{m}_k)) + \mathbf{C}_m^{-1} (\mathbf{m}_0 - \mathbf{m}_k) \right\},$$

où $\bar{\mathbf{g}}(\mathbf{m}_k) = \left(\text{Erf}^{-1}\left(\frac{2}{\pi} \arctan\left(\sinh\left(\frac{g_i(\mathbf{m}_k) - d_i^o}{\sigma_i}\right)\right)\right)\right)_{i \in \mathfrak{S}}$. En utilisant la formule (1.3.16) et (1.3.14) on voit que la matrice $\bar{\mathbf{G}}_k$ se décompose comme produit de matrices,

$$(1.3.21) \quad \bar{\mathbf{G}}_k = (\mathbf{D}_{\mathbf{g}(\mathbf{m}_k)} \bar{\mathbf{d}}) \mathbf{G}_k$$

où $\mathbf{D}_{\mathbf{g}(\mathbf{m}_k)} \bar{\mathbf{d}}$ est la matrice (diagonale dans ce cas précis) jacobienne de l'application $\mathbf{d} \mapsto \bar{\mathbf{d}}(\mathbf{d})$, calculée en $\mathbf{g}(\mathbf{m}_k)$. Autrement dit, on doit calculer la perturbation δ_k vérifiant,

$$(1.3.22) \quad \begin{bmatrix} \bar{\mathbf{G}}_k \\ \mathbf{C}_m^{-\frac{1}{2}} \end{bmatrix} \delta_k \sim \begin{bmatrix} \bar{\mathbf{g}}(\mathbf{m}_k) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix},$$

c'est à dire,

$$\text{(schéma se)} \quad \begin{bmatrix} (D_{\mathbf{g}(\mathbf{m}_k)} \bar{\mathbf{d}}) \mathbf{G}_k \\ \mathbf{C}_m^{-\frac{1}{2}} \end{bmatrix} \delta_k \sim \begin{bmatrix} \bar{\mathbf{g}}(\mathbf{m}_k) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}_k) \end{bmatrix}, \text{ au sens des moindres carrés.}$$

Ce schéma s'implémente facilement à partir de l'algorithme (**gn**) : il suffit de pondérer chaque ligne de \mathbf{G}_k par la matrice (diagonale) $D_{\mathbf{g}(\mathbf{m}_k)} \bar{\mathbf{d}}$, et de transformer les résidus par la relation **non linéaire** (1.3.17). Le point de vue probabiliste permet de définir des transformations plus générales et mieux adaptés (cf Tarantola [45] pour d'autres exemples) du système d'équation que la simple pondération généralement utilisée dans la conception "classique" du problème inverse (cf début de la partie 1.1). En effet, les géophysiciens pondèrent les équations linéarisée de manière à favoriser les "bonnes" données (affectées par de forts coefficients) par rapport aux "mauvaises" données (affectées par des faibles coefficients) dans le but d'éliminer les données aberrantes qui risquent d'avoir une forte influence en norme L2. Il est parfois difficile de savoir *a priori* si une donnée est bonne ou mauvaise. Le risque est alors grand d'affecter des coefficients erronés à certaines données. Cela risque de biaiser la solution, d'autant plus que la norme L2 n'est pas robuste vis à vis des mauvaises données. Cette difficulté est contournée en utilisant une statistique appropriée des erreurs de mesures telle que la loi Sech.

Dans la suite de ce document les schémas numériques (**gn**) et (**se**) serviront de base à la conception de méthodes d'inversions de données. Nous ferons donc régulièrement référence aux schémas (**gn**) et (**se**) dans les problèmes particuliers issus de cas réels.

1.3.3. Calcul de la perturbation. Comme nous venons de le voir, à chaque itération l'algorithme (schémas (**gn**) et (**se**)) doit calculer une perturbation vérifiant un système linéaire au sens des moindres carrés,

$$(1.3.23) \quad \mathbf{Ax} \sim \mathbf{b}.$$

Cela revient à poser le problème d'optimisation suivant,

$$(1.3.24) \quad \text{trouver } \bar{\mathbf{x}} \text{ minimisant la fonction } \mathbf{x} \mapsto \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2,$$

Nous avons choisi l'algorithme LSQR développé par Paige et Saunders [36] pour résoudre ce problème. C'est une méthode de gradients conjugués avec un schéma numérique plus favorable dans le cas où les matrices sont mal conditionnées. L'implémentation du code LSQR convient particulièrement bien aux cas où la matrice \mathbf{A} est creuse, ce qui est le cas pour les matrices inversées lors d'une tomographie sismique. LSQR génère une suite $(\mathbf{x}_n)_{n=1, \dots, n_{\max}}$ d'approximations successives de la solution ; l'algorithme suit schématiquement les étapes :

- (1) . Initialisation : $\mathbf{x} = 0$.
- (2) . Construction d'un vecteur δ , à partir de l'itéré courant \mathbf{x} , \mathbf{A} et \mathbf{b} .
- (3) . Calcul de l'itéré suivant : $\mathbf{x} = \mathbf{x} + \delta$.
- (4) . Test : \mathbf{x} vérifie les critères d'arrêt alors stop, sinon retour en (2).

C'est une méthode d'optimisation à direction de descente, la perturbation est calculée avec des formules de récurrence décrites en détail dans [36]. Le test d'arrêt (étape (4)) se fait selon deux critères pour les problèmes de moindres carrés. Le premier fait intervenir la précision relative $atol$ obtenue sur la matrice \mathbf{A} ,

$$(1.3.25) \quad atol = \frac{\|\mathbf{A} - \tilde{\mathbf{A}}\|}{\|\mathbf{A}\|},$$

où $\tilde{\mathbf{A}}$ représente la vraie matrice. La suite d'itérations de LSQR s'arrête lorsque,

$$(1.3.26) \quad \frac{\|\mathbf{A}^t \mathbf{r}_k\|}{\|\mathbf{A}\| \|\mathbf{r}_k\|} \leq atol,$$

où $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$. D'après [36] ce critère est suffisant pour garantir que \mathbf{x}_k est une solution acceptable du problème aux moindres carrés (1.3.24). Le second critère d'arrêt est utile dans le cas de problèmes mal conditionnés. Pour chaque itéré \mathbf{x}_k , on définit une valeur c_k qui tend vers la valeur $\text{cond}(\mathbf{A})$ lorsque \mathbf{x}_k tend vers une solution du problème. La solution retenue pour le problème (1.3.24) est la solution à l'itéré k si $c_k > conlim$, où $conlim$ est une valeur fournie par l'utilisateur (voir ci-dessous). Ce critère constitue une régularisation du problème. On peut se représenter ce critère en utilisant la décomposition SVD de \mathbf{A} . Supposons que la matrice \mathbf{A} possède r valeurs singulières μ_i non nulles, la décomposition SVD s'écrit, $\mathbf{A} = \mathbf{V}\tilde{\mathbf{S}}\mathbf{U}^*$. Soit \mathbf{A}_k la matrice $\mathbf{A}_k = \mathbf{V}\tilde{\mathbf{S}}_k\mathbf{U}^*$, où $\tilde{\mathbf{S}}_k$ contient les k premières valeurs singulières de $\tilde{\mathbf{S}}$. On définit \mathbf{x}_k^s la solution de norme minimale du problème aux moindres carrés, $\mathbf{A}_k\mathbf{x} \sim \mathbf{b}$. Ici la valeur c_k peut se définir comme le conditionnement de \mathbf{A}_k ,

$$(1.3.27) \quad c_k = \frac{\mu_1}{\mu_k} \rightarrow \frac{\mu_1}{\mu_r} = \text{cond}(\mathbf{A}).$$

Si $c_k > conlim < \text{cond}(\mathbf{A})$, la solution retenue est calculée en "coupant" les valeurs singulières les plus faibles, ce qui tend à régulariser les problèmes mal conditionnés. En réalité l'algorithme LSQR ne calcule pas les itérés en fonction des valeurs singulières et vecteurs propres (figure (1.3.1)). La valeur c_k est calculée par l'algorithme en fonction de \mathbf{A} , \mathbf{b} et \mathbf{x}_k , la relation (1.3.27) n'est pas utilisée dans l'algorithme mais on peut s'y référer pour définir un ordre de grandeur de la valeur $conlim$.

Nous allons à présent explorer quelques propriétés de LSQR sur des exemples numériques. Considérons une matrice $\mathbf{A}_n^0 \in \mathcal{M}_n(\mathbb{R})$ de structure tri-diagonale, composée de 4 sur la diagonale, de -1 sur la diagonale supérieure et de -2 sur la diagonale inférieure. Par exemple pour $n=5$,

$$(1.3.28) \quad \mathbf{A}_5^0 = \begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -2 & 4 & -1 & 0 & 0 \\ 0 & -2 & 4 & -1 & 0 \\ 0 & 0 & -2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 4 \end{pmatrix}.$$

Soit la matrice $\mathbf{A}_n \in \mathcal{M}_{n+1,n}(\mathbb{R})$ constituée par la matrice \mathbf{A}_n^0 à laquelle on ajoute une ligne constituée de 0, et un dernier élément égal à 1 ; par exemple pour $n=5$,

$$\mathbf{A}_5 = \begin{pmatrix} 4 & -1 & 0 & 0 & 0 \\ -2 & 4 & -1 & 0 & 0 \\ 0 & -2 & 4 & -1 & 0 \\ 0 & 0 & -2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Soit $\mathbf{x}_n \in \mathbb{R}^n$ un vecteur composé de 1. Notons $\mathbf{b}_n = \mathbf{A}_n \mathbf{x}_n$; pour $n=5$, nous avons,

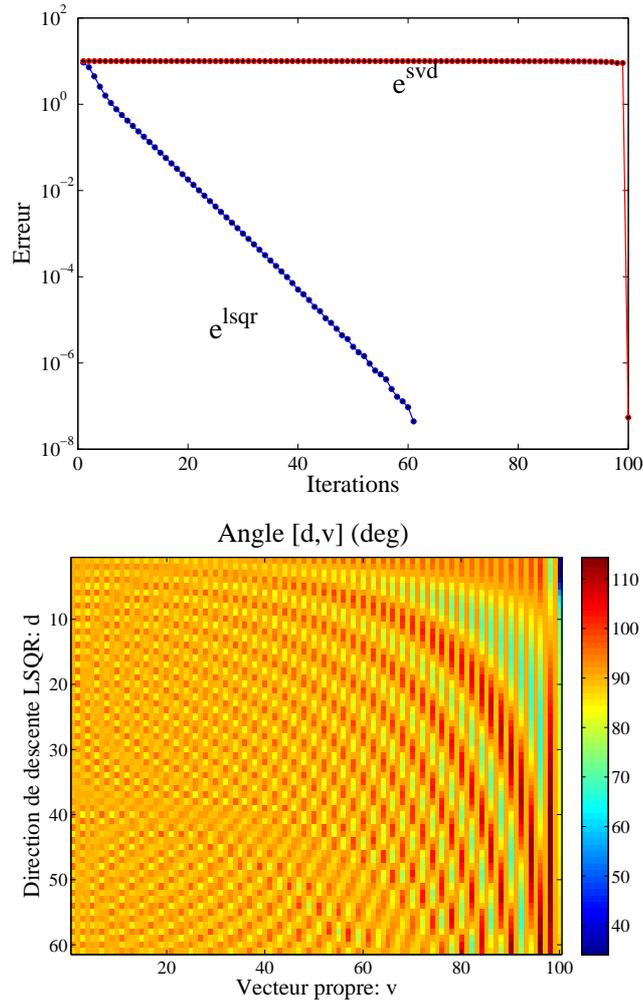
$$(1.3.29) \quad \mathbf{x}_5 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{b}_5 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}.$$

Nous allons résoudre numériquement le système $\mathbf{A}_n \mathbf{x} = \mathbf{b}_n$, dont nous connaissons la solution \mathbf{x}_n . Nous utiliserons LSQR et une décomposition SVD afin de comparer les deux méthodes. L'algorithme LSQR génère une suite $(\mathbf{l}_k)_k$ d'approximations successives de la solution du système. La figure (1.3.1) montre la quantité $e_k^{lsqr} = \|\mathbf{l}_k - \mathbf{x}_n\|$, en fonction de l'itération k . Décomposons la matrice \mathbf{A}_n en valeurs singulières, $\mathbf{A}_n = \mathbf{V}\mathbf{S}\mathbf{U}^t$, et définissons une famille de pseudo-inverses $\mathbf{A}_{n,k}^\dagger = \mathbf{U}\mathbf{S}_k\mathbf{V}^t$, où \mathbf{S}_k est une matrice diagonale constituée des inverses des k premières valeurs singulières de \mathbf{A}_n . Soit $\mathbf{s}_k = \mathbf{A}_{n,k}^\dagger \mathbf{b}_n$, la solution du système calculée par SVD en utilisant les k premières valeurs singulières de \mathbf{A}_n . La figure (1.3.1) montre la quantité $e_k^{svd} = \|\mathbf{s}_k - \mathbf{x}_n\|$. Nous remarquons que les courbes correspondant à LSQR et à SVD sont différentes, ce qui montre que les itérations de LSQR ne progressent pas dans la direction des vecteurs propres de la matrice \mathbf{A}_n . De plus, dans cet exemple précis, pour obtenir une solution satisfaisante par SVD, il faut utiliser toutes les valeurs singulières : comme nous le voyons sur la courbe, l'erreur e_{99}^{svd} est encore importante alors que $s_{100} = x_{100}$. Quant à l'algorithme LSQR, il trouve la solution en 60 itérations. L'exploration de l'espace des paramètres du modèle par LSQR, ne se fait pas vecteur propre après vecteur propre mais de manière plus efficace, en utilisant une combinaison linéaire de vecteurs propres¹⁵ à chaque itération, ce qui permet de converger plus vite vers la solution.

¹⁵Sans les calculer individuellement ce qui est trop long et trop complexe.

FIG. 1.3.1. Résolution d'un système linéaire.

Pour $n=100$, nous avons résolu le système linéaire $A_{100}x = b_{100}$ (cf texte) par LSQR (en bleu) et par SVD (en rouge). LSQR converge plus rapidement vers la solution, alors que dans cet exemple la méthode SVD a besoin de toutes les valeurs singulières pour trouver la solution. La figure du bas montre l'angle (en degrés) que fait chaque direction de descente d générée par LSQR avec chacun des vecteurs propres. Les directions de descentes de LSQR ne coïncident pas avec les vecteurs propres puisque aucun des angles n'est nul. L'avantage de LSQR réside dans le temps de calcul nécessaire pour trouver la solution : il ne calcule pas de vecteurs propres (ce qui est trop long et trop complexe), il converge vite vers la solution en utilisant des directions de descentes inspirées par la méthode des gradients conjugués, il particulièrement efficace avec des matrices creuses (comme dans cet exemple).



Considérons maintenant un autre exemple,

$$(1.3.30) \quad \mathbf{A} = \begin{pmatrix} -7 & -6 & 6 & 7 & 1 & -1 \\ -5 & -8 & 3 & 9 & 1 & -1 \\ -2 & -9 & 1 & 9 & 1 & -1 \\ -1 & -9 & -1 & 9 & 1 & -1 \\ -1 & -9 & -3 & 8 & 1 & -1 \\ 3 & -9 & -5 & 8 & 1 & -1 \\ 6 & -7 & -7 & 6 & 1 & -1 \end{pmatrix}, \quad \mathbf{x}_s = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ -1 \\ -1 \\ -2 \\ -4 \\ -3 \\ -2 \end{pmatrix},$$

\mathbf{x}_s vérifie le système

$$(1.3.31) \quad \mathbf{A}\mathbf{x}_s = \mathbf{b}.$$

la matrice \mathbf{A} possède une seule valeur singulière nulle, le noyau est donc une droite vectorielle avec, par exemple, comme vecteur de base,

$$(1.3.32) \quad \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},$$

nous avons donc

$$(1.3.33) \quad \mathbf{A}\mathbf{v} = 0.$$

Il en s'ensuit que quelque soit le réel λ , nous avons la relation suivante,

$$(1.3.34) \quad \mathbf{A}(\mathbf{x}_s + \lambda\mathbf{v}) = \mathbf{b},$$

en effet, si on développe,

$$(1.3.35) \quad \mathbf{A}\mathbf{x}_s + \mathbf{A}(\lambda\mathbf{v}) = \mathbf{A}\mathbf{x}_s + \lambda\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{x}_s = \mathbf{b}.$$

Les solutions du système

$$(1.3.36) \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

sont

$$(1.3.37) \quad S = \{\mathbf{x}_s + \lambda\mathbf{v}, \lambda \in \mathbb{R}\}.$$

Nous sommes en présence d'un système sous-déterminé, c'est à dire que certains paramètres ne sont pas contraints par les équations, ce qui explique l'infinité de solutions vérifiant le système. Dans les applications tomographiques, l'algorithme itératif calcule une perturbation à apporter aux paramètres pour trouver le modèle suivant. Si certaines perturbations ne sont pas contraintes par les équations, il faut que la résolution du système laisse ces valeurs à zéro. En effet il n'est pas nécessaire d'affecter une valeur arbitraire aux paramètres indéterminés, il suffit de les laisser à la valeur *a priori*. Lorsqu'on résout le système avec le code LSQR, la solution produite est :

$$(1.3.38) \quad \mathbf{x}_{\text{lsqr}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

c'est la solution de norme minimale puisque sa composante est nulle sur le noyau. LSQR laisse donc les paramètres non contraints à zéro. C'est un avantage dans les applications

sismologiques, par exemple, lors de la localisation d'un essaim de séismes en ajustant les décalages temporels. Dans certains cas, le barycentre de l'essaim en question peut ne pas être contraint par les équations, dans ce cas, l'algorithme LSQR laisse le barycentre invariant et ajuste les positions relatives entre les séismes. Il serait problématique que l'algorithme d'inversion affecte une valeur quelconque au barycentre. En effet il est préférable, en l'absence d'information sur celui-ci de le laisser à la valeur *a priori*, qui est issue d'une localisation préliminaire avec les temps de première arrivée.

L'algorithme LSQR est particulièrement adapté aux applications sismologiques, en effet,

- (1) Il est rapide.
- (2) Il calcule la solution de norme minimale.

De plus il permet le stockage de matrices creuses en évitant de surcharger la capacité mémoire des machines avec des zéros inutiles.

1.4. Conclusion.

Dans ce chapitre, nous avons défini le problème inverse en tant que problème de minimisation d'une fonction coût, et nous avons montré que la prise en compte des erreurs de mesures ainsi que des erreurs sur les paramètres contrôlait la forme de la fonction coût : la fonction coût doit permettre de trouver le modèle de norme minimale qui ajuste le mieux possible les données. Ainsi définie, cette fonction coût possède des caractéristiques optimales pour pouvoir être explorée à l'aide d'une méthode à direction de descente. Un algorithme de Gauss-Newton a été choisi pour réaliser cette minimisation de la fonction coût. La résolution du système linéaire qui résulte de cet algorithme à chaque étape est effectuée à l'aide du moteur d'inversion LSQR.

Nous allons voir dans le chapitre suivant, comment se formule notre problème d'inversion particulier (*tomographie des temps de première arrivée* et *tomographie des décalages temporels*) et quelle forme particulière prendra la fonction coût à minimiser et l'algorithme d'inversion correspondant.

CHAPITRE 2

Application à la tomographie sismique.

Ce chapitre présente en détail la méthodologie des tomographies utilisant les temps de première arrivée et les décalages temporels. Ces méthodes implémentées seront appliquées ultérieurement à l'île de Hawaïi. Les méthodes sont présentées de manière formelle et nous étudierons quelques cas synthétiques. En effet, il est pratiquement impossible de mener une étude mathématique complète de ces problèmes. Les tests numériques permettent la maîtrise de tous les paramètres expérimentaux afin de caractériser le comportement des méthodes. Les cas synthétiques présentés dans ce chapitre sont, à cet effet, volontairement simples afin d'exclure d'éventuels artefacts dus à une conjonction de plusieurs facteurs qui risqueraient de se produire sur des exemples trop complexes.

2.1. Modélisation.

2.1.1. Données et terminologie. Dans le cadre de ce travail, nous aurons à traiter des données enregistrées par un réseau sismique local implanté sur l'île de Hawaïi. Nous avons accès à deux types de données : les temps de première arrivée des ondes P pour chaque séisme, et les décalages temporels interspectraux (pour les ondes P) calculés pour deux séismes semblables. Ces données seront inversées afin de décrire les structures du sous-sol. Le chapitre précédent décrit de manière générale une démarche possible pour étudier un problème inverse. Dans cette partie, nous allons appliquer cette démarche sur les données considérées. Nous avons choisi d'utiliser un processus spécifique selon la nature des données. Nous aurons à traiter séparément les temps de première arrivée et les décalages temporels interspectraux. Pour éviter toute ambiguïté, nous allons définir une terminologie faisant référence aux méthodes. Le terme *tomographie* désignera une méthode visant à retrouver simultanément le modèle de vitesse des ondes P et les hypocentres des séismes. Lorsque les données inversées seront des temps de première arrivée d'onde P, nous parlerons de *tomographie des temps de première arrivée*. Lorsque nous inverserons les décalages temporels, nous utiliserons le terme *tomographie des décalages temporels*. Dans ce document nous étudierons aussi les méthodes de *localisation* ; elles visent à calculer les hypocentres des séismes à partir des temps de première arrivées ou bien à partir de décalages temporels. Comme en tomographie, pour ces méthodes, nous parlerons de *localisation à l'aide des temps de première arrivée* et de *localisation à l'aide des décalages temporels*.

2.1.2. Discrétisation du milieu géophysique. Afin de mener des calculs numériques, le milieu doit être discrétisé. Considérons n_x mailles dans la direction de la latitude espacées de d_x mètres, n_y mailles selon la longitude espacées de d_y mètres et enfin n_z mailles selon la profondeur espacées de d_z mètres, nous obtenons ainsi un maillage fini

\mathbf{m}_i , de dimension $((n_x - 1)d_x \times (n_y - 1)d_y \times (n_z - 1)d_z)$. Etant donné une fonction (par ex : vitesse ou lenteur) $\mathbf{x} \mapsto f(\mathbf{x})$ où \mathbf{x} est un point appartenant au volume délimité par le maillage, étant donné une famille de n_φ fonctions de base $(\varphi_k)_{k=1,\dots,n}$ définies sur les pavés du maillage, on peut discrétiser le modèle f en l'exprimant comme une combinaison linéaire des fonctions de base,

$$(2.1.1) \quad \tilde{f}(\mathbf{x}) = \sum_{l=1}^{l=n_\varphi} \lambda_l \varphi_l(\mathbf{x}).$$

Nous choisirons ici d'utiliser les formules d'interpolation trilineaires ; pour cela nous devons connaître la valeur du modèle en chaque noeud (x_i, y_j, z_k) du maillage. Nous pouvons calculer la valeur en un point (x, y, z) situé dans le parallélépipède rectangle $\wp_{i,j,k}$ délimité par les deux noeuds (x_i, y_j, z_k) et $(x_{i+1}, y_{j+1}, z_{k+1})$ en utilisant la relation [39]

$$(2.1.2) \quad \begin{aligned} \tilde{f}(x, y, z) = & \dots \\ & \dots \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right) f(x_i, y_j, z_k) + \dots \\ & \dots \left(\frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right) f(x_{i+1}, y_j, z_k) + \dots \\ & \dots \left(\frac{x-x_i}{x_{i+1}-x_i}\right) \left(\frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right) f(x_{i+1}, y_{j+1}, z_k) + \dots \\ & \dots \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(\frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right) f(x_i, y_{j+1}, z_k) + \dots \\ & \dots \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(\frac{z-z_k}{z_{k+1}-z_k}\right) f(x_i, y_j, z_{k+1}) + \dots \\ & \dots \left(\frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(\frac{z-z_k}{z_{k+1}-z_k}\right) f(x_{i+1}, y_j, z_{k+1}) + \dots \\ & \dots \left(\frac{x-x_i}{x_{i+1}-x_i}\right) \left(\frac{y-y_j}{y_{j+1}-y_j}\right) \left(\frac{z-z_k}{z_{k+1}-z_k}\right) f(x_{i+1}, y_{j+1}, z_{k+1}) + \dots \\ & \dots \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(\frac{y-y_j}{y_{j+1}-y_j}\right) \left(\frac{z-z_k}{z_{k+1}-z_k}\right) f(x_i, y_{j+1}, z_{k+1}). \end{aligned}$$

On définit ainsi une famille de fonctions de base φ_{ijk} indexée par les noeuds (i, j, k) de la grille,

$$(2.1.3) \quad \begin{aligned} \varphi_{ijk}(x, y, z) = & \dots \\ & \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right), \quad \text{dans } \wp_{i,j,k} \\ & \left(1 - \frac{x-x_i}{x_{i-1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right), \quad \text{dans } \wp_{i-1,j,k} \\ & \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j-1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right), \quad \text{dans } \wp_{i,j-1,k} \\ & \left(1 - \frac{x-x_i}{x_{i-1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j-1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k+1}-z_k}\right), \quad \text{dans } \wp_{i-1,j-1,k} \\ & \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k-1}-z_k}\right), \quad \text{dans } \wp_{i,j,k-1} \\ & \left(1 - \frac{x-x_i}{x_{i-1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j+1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k-1}-z_k}\right), \quad \text{dans } \wp_{i-1,j,k-1} \\ & \left(1 - \frac{x-x_i}{x_{i+1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j-1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k-1}-z_k}\right), \quad \text{dans } \wp_{i,j-1,k-1} \\ & \left(1 - \frac{x-x_i}{x_{i-1}-x_i}\right) \left(1 - \frac{y-y_j}{y_{j-1}-y_j}\right) \left(1 - \frac{z-z_k}{z_{k-1}-z_k}\right), \quad \text{dans } \wp_{i-1,j-1,k-1}. \end{aligned}$$

$$\varphi_{ijk}(x, y, z) = 0 \text{ dans les autres cellules.}$$

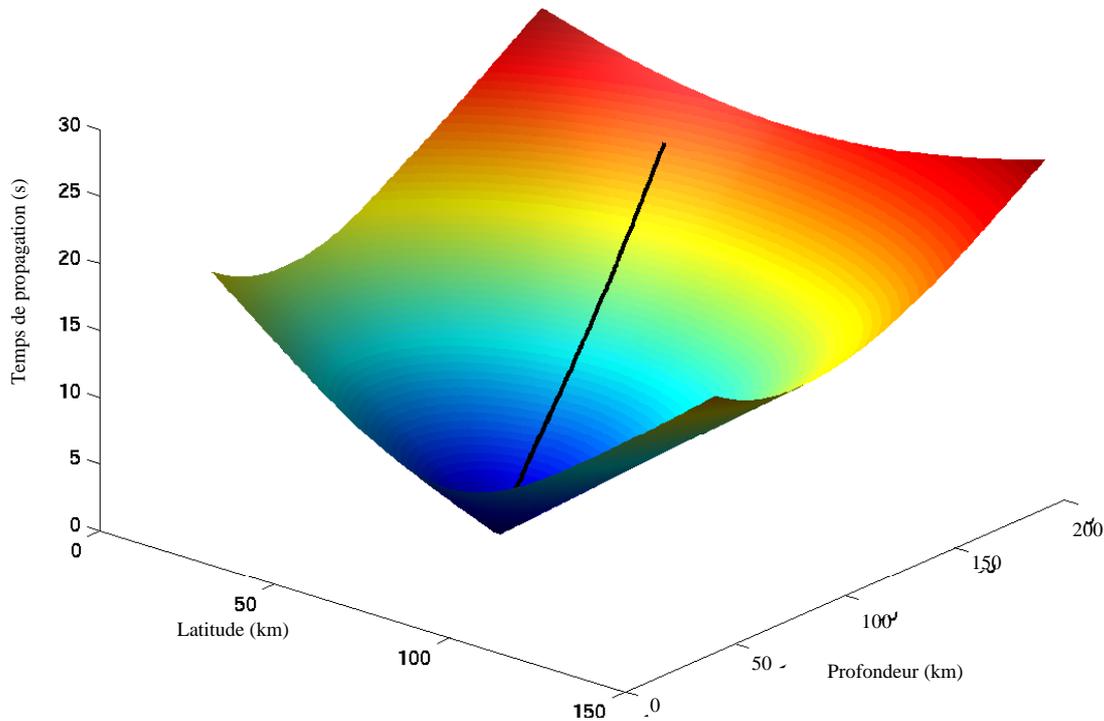
Ainsi nous utiliserons la formule de discrétisation,

$$(2.1.4) \quad \tilde{f}(x, y, z) = \sum_{i,j,k} f(x_i, y_j, z_k) \varphi_{ijk}(x, y, z).$$

Il faut prendre en compte les effets de bord du domaine discrétisé : dans les formules (2.1.2) et (2.1.3) si le noeud (u, v, w) [resp. le pavé (u, v, w)] n'existe pas, il faut éliminer les lignes qui n'ont pas de sens. Par exemple pour le noeud $i=1, j=1, k=1$, les lignes où apparaissent les termes $i-1, j-1$ et $k-1$ ne doivent pas être prises en compte. Ces considérations permettent de représenter un modèle f comme un vecteur $(f(x_1, y_1, z_1), \dots, f(x_i, y_j, z_k), \dots, f(x_{n_x}, y_{n_y}, z_{n_z}))^t$. Dans la suite de ce document, les coordonnées seront noté λ_{ijk} [ou bien λ_i] pour les lenteurs et ν_{ijk} [ou bien ν_i] pour les vitesses. Lorsque qu'il ne sera pas nécessaire de préciser explicitement l'indice du noeud i, j, k nous pourrons pour simplifier indiquer seulement un indice $\mathbf{i} = (i, j, k)$.

FIG. 2.1.1. Tracé d'un rai.

Une fois la fonction temps de parcours calculée, on trace le rai joignant la source et le récepteur en utilisant une méthode d'optimisation de gradient à pas constant.



2.1.3. Problème direct : calcul du temps de propagation. La première étape consiste à calculer la fonction temps de parcours de l'onde sismique $\mathbf{x} \mapsto T(\mathbf{x})$, émise par une source \mathbf{x}_s . Cette tâche est dévolue à l'algorithme de Podvin-Lecomte [37] résolvant l'équation de l'Eikonale par la méthode des différences finie. Ensuite, il s'agit de tracer le rai joignant la source \mathbf{x}_s à un récepteur \mathbf{x}_r . Pour cela on va utiliser une méthode d'optimisation :

la méthode du gradient. Nous allons implémenter cette méthode sur la fonction T , en utilisant \mathbf{x}_r comme point initial. Nous définissons ainsi la suite,

$$(2.1.5) \quad \begin{cases} \mathbf{x}_0 = \mathbf{x}_r, \\ \mathbf{x}_{k+1} = -\lambda \frac{\nabla T(\mathbf{x}_k)}{\|\nabla T(\mathbf{x}_k)\|} \end{cases}$$

où λ est un réel positif définissant le pas de la suite. La fonction T est :

- (1) inf-compacte car plus on s'éloigne de la source plus le temps de propagation est long.
- (2) Elle est minorée car le temps de propagation est toujours positif ou nul.
- (3) Elle possède un minimum unique \mathbf{x}_s car c'est le seul point où le temps de propagation est nul.

La suite converge nécessairement vers le point \mathbf{x}_s . Si nous relient tous les points de la suite nous obtenons un trajet allant de \mathbf{x}_r à \mathbf{x}_s . De plus, en chaque point, ce chemin est tangent au gradient de la fonction T ; le chemin ainsi défini est donc le rai $[\mathbf{x}_r, \mathbf{x}_s]$.

La méthode de différences finies de Podvin-Lecomte [37] nécessite une maille de discrétisation cubique la plus fine possible pour minimiser les erreurs de calcul sur la fonction T . Nous allons donc en premier lieu raffiner le maillage \mathbf{m}_i pour définir un maillage \mathbf{m}_f de pas d_f et de dimension $((n_{fx} - 1)d_f \times (n_{fy} - 1)d_f \times (n_{fz} - 1)d_f)$. Sur chacun des noeuds nous définissons une valeur de vitesse par l'intermédiaire de la formule d'interpolation trilineaire (2.1.2). L'algorithme Podvin-Lecomte [37] calcule le temps de parcours de l'onde sismique en chaque noeud du maillage \mathbf{m}_f . Pour tracer un rai avec la méthode (2.1.5), il est nécessaire de pouvoir calculer le gradient de la fonction T en tout point $\mathbf{x} = (x, y, z)^t$ du domaine. Pour cela, nous repérons le pavé qui contient ce point et nous effectuons le calcul des dérivées partielles de T aux 8 noeuds définissant les sommets du cube en question par les formules suivantes,

$$(2.1.6) \quad \frac{\partial T}{\partial x}(x_i, y_j, z_k) \simeq \frac{T(x_{i+1}, y_j, z_k) - T(x_{i-1}, y_j, z_k)}{2d_f},$$

$$(2.1.7) \quad \frac{\partial T}{\partial y}(x_i, y_j, z_k) \simeq \frac{T(x_i, y_{j+1}, z_k) - T(x_i, y_{j-1}, z_k)}{2d_f},$$

$$(2.1.8) \quad \frac{\partial T}{\partial z}(x_i, y_j, z_k) \simeq \frac{T(x_i, y_j, z_{k+1}) - T(x_i, y_j, z_{k-1})}{2d_f},$$

où les indices i, j et k parcourent les 8 sommets du pavé considéré. Le gradient au point \mathbf{x} est calculé en utilisant la formule d'interpolation trilineaire (2.1.2) et en utilisant les dérivées partielles aux 8 noeuds calculées par les formules (2.1.6)-(2.1.7)-(2.1.8). L'expérience nous a montré que cette méthode de tracé de rai donne de bons résultats lorsque le pas de marche λ le long du rai est égal à $\frac{d_f}{\sqrt{2}}$.

Nous connaissons le temps de parcours de l'onde sismique aux noeuds de la grille \mathbf{m}_f , néanmoins il est possible d'accroître la précision du temps de parcours le long d'un rai γ

en effectuant numériquement l'intégration suivante

$$(2.1.9) \quad t = \int_{\gamma} \frac{d\sigma}{\tilde{v}}.$$

La formule de Simpson permet le calcul de (2.1.9),

$$(2.1.10) \quad t \simeq \sum_{k=1}^{n_r-1} \delta\sigma \left(\frac{1}{3\tilde{v}(\gamma_k)} + \frac{4}{3\tilde{v}(\gamma_{k+\frac{1}{2}})} + \frac{1}{3\tilde{v}(\gamma_{k+1})} \right),$$

lorsque le rai γ est subdivisé en n_r points $(\gamma_k)_{k=1,\dots,n_r}$ distants de $2\delta\sigma$, où $\gamma_{k+\frac{1}{2}}$ désigne le milieu du segment $[\gamma_k, \gamma_{k+1}]$. Dans la pratique, nous utilisons $\delta\sigma = \frac{\lambda}{2}$, où λ est défini dans la méthode (2.1.5).

Exemple dans un milieu à 2 dimensions. Considérons un milieu à gradient à deux dimensions, dans lequel on place une source et un récepteur (figure (2.1.2)). Il existe des formules analytiques [12] donnant la fonction temps de parcours pour de tels milieux, nous pouvons donc les comparer aux résultats numériques. La figure (2.1.2) montre en fonction de la taille de la maille de discrétisation l'erreur, commise sur le temps de parcours le long du rai joignant la source et le récepteur : (1) par la méthode de Podvin-Lecomte seule (P&L) et (2) par la méthode de Podvin-Lecomte suivie de l'intégration numérique le long du rai (INT). La figure (2.1.3) montre que la précision du rai dépend de la taille de maille utilisée, le rai étant mieux calculé lorsque que la maille est fine. Cette erreur de trajet du rai se répercute sur l'intégrale (2.1.10) et fausse le calcul du temps de parcours. Nous remarquons que dans les deux méthodes la précision sur le calcul du temps de propagation s'accroît lorsque la maille diminue, et que la méthode INT est toujours meilleure. Il convient donc de choisir une maille adaptée en tenant compte des ressources matérielles dont on dispose. En effet le temps de calcul sur machine peut devenir très long si la maille est trop fine¹. Dans notre exemple une maille de 250 mètres semble tout indiquée car nous obtenons une erreur de l'ordre de 10^{-3} seconde et le temps de calcul est de 0,3 seconde sur notre machine, alors qu'avec une maille de 50 m le temps de calcul est de 6 s pour une erreur de l'ordre de 10^{-4} s. Pour gagner un ordre de grandeur sur la précision, nous avons multiplié le temps de calcul par 20. Pour finir, la figure (2.1.4) montre l'erreur pour les deux méthodes, pour une maille de 250 m ; l'erreur n'est pas distribuée de façon homogène mais la méthode INT est toujours plus performante.

2.2. Localisations.

Le problème direct défini, nous allons à présent l'utiliser dans un processus d'inversion de données. Dans un premier temps, nous allons supposer que l'on connaît le modèle de vitesse des ondes P. On définit alors des méthodes visant à localiser les hypocentres des séismes. Nous allons détailler deux méthodes de localisation ; une première méthode consiste à ajuster les temps de première arrivée des ondes P, dans une deuxième méthode nous verrons qu'il est possible de calculer les hypocentres à partir des décalages temporels.

¹Surtout dans les milieux à trois dimensions.

FIG. 2.1.2. Calcul du temps de parcours.

La figure de gauche montre le milieu dans lequel est calculé le temps de parcours d'une onde sismique. En premier lieu on calcule le temps de parcours par résolution de l'équation de l'eikonale par différences finies (Podvin-Leconte [37], cette méthode est désignée ici par :P&L). Ensuite, on trace le rai joignant la source et le récepteur et on calcule le temps de parcours par intégration le long du rai (cette méthode est désignée ici par :INT). Tout cela est fait pour différentes tailles de mailles de discrétisation. La figure de droite montre les erreurs numériques des deux méthodes en comparant le calcul avec des formules analytiques [12].

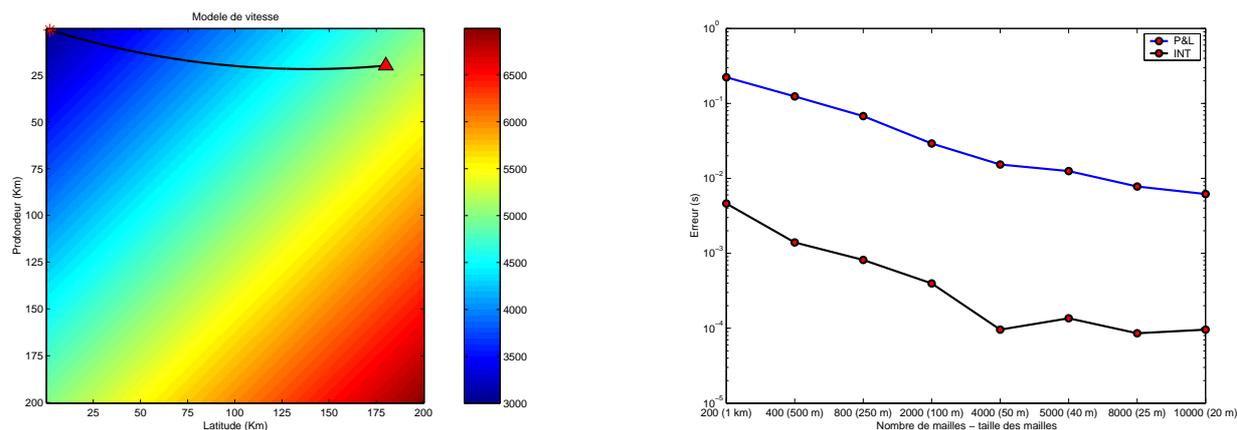
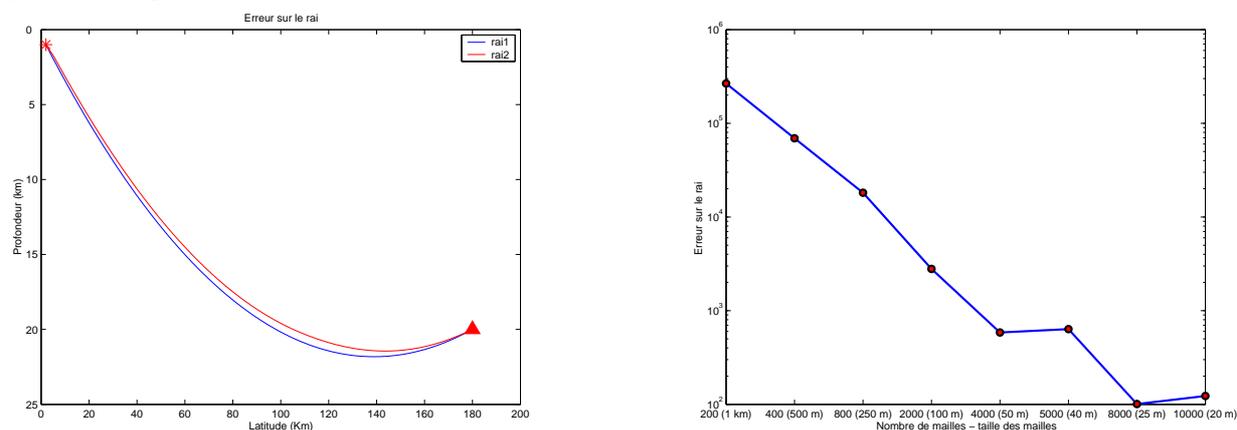


FIG. 2.1.3. Rais selon le maillage.

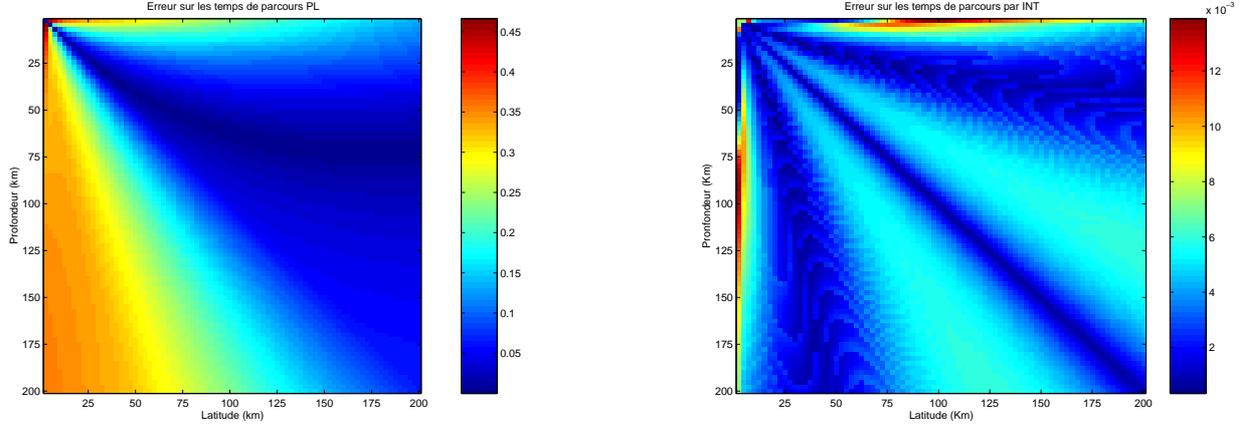
Sur la figure de gauche, le rai en rouge est celui calculé avec une maille de 20 m, le rai en bleu est calculé avec une maille de 1 Km. Sur la figure de droite la courbe représente l'erreur sur le rai en fonction du pas de discrétisation.



2.2.1. Localisation avec les temps de première arrivée. Considérons un ensemble de p stations et un ensemble de q sources sismiques, notons $J=\{1,\dots,p\}$ et $I=\{1,\dots,q\}$. Chaque séisme émet des ondes qui se propagent dans le milieu jusqu'aux stations. Chaque

FIG. 2.1.4. Précision des temps de parcours.

Erreurs sur les temps de parcours des deux méthodes (les échelles de couleur sont en secondes), Podvin-Lecomte (P \mathcal{E} L) à gauche et intégration à droite (INT).



station sismique enregistre les mouvements du sol en fonction du temps, sous la forme d'un sismogramme. Un opérateur pointe alors sur celui-ci le temps de première arrivée du train d'onde émis par un séisme. Notons t_{ij}^{obs} le temps de première arrivée de l'onde, pointé à la station $j \in J$ et produite par la source $i \in I$. Les enregistrements du réseau sismique permettent d'exploiter un catalogue $\mathbf{t}^{obs} = (t_{ij}^{obs})_{(i,j) \in \mathfrak{S}}$, avec $\mathfrak{S} \subset I \times J$. L'instant d'arrivée théorique est la somme de l'instant de déclenchement du séisme et du temps de propagation le long du rai allant de la source à la station,

$$(2.2.1) \quad t_i^j(s, \mathbf{x}_i, t_i^o) = t_i^o + \int_{\gamma_{ij}} s d\sigma,$$

où t_i^o est le temps origine du séisme i , $\mathbf{x}_i = (x_i, y_i, z_i)$ sa position et γ_{ij} est le rai joignant la source i à la station j dans le modèle de lenteur s . Il est maintenant possible de définir un problème inverse ajustant les données \mathbf{t}^{obs} . Le vecteur \mathbf{m} est un ensemble de paramètres, considérés comme inconnus, pris parmi l'ensemble de tous les paramètres $\{t_i^o, \mathbf{x}_i, s\}_{i \in I}$. En faisant l'hypothèse gaussienne, la densité de probabilité *a posteriori* s'écrit, (ref (1.1.35))

$$(2.2.2) \quad \sigma_M(\mathbf{m}) \propto \exp(-E_g(\mathbf{m})),$$

en notant $\mathbf{t}(\mathbf{m}) = (t_i^j(\mathbf{m}))_{(i,j) \in \mathfrak{S}}$,

$$(2.2.3) \quad E_g(\mathbf{m}) = \frac{1}{2} (\mathbf{t}^{obs} - \mathbf{t}(\mathbf{m}))^t \mathbf{C}_D^{-1} (\mathbf{t}^{obs} - \mathbf{t}(\mathbf{m})) + \frac{1}{2} (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0).$$

Nous pouvons aussi donner la formulation de la densité *a posteriori* lorsque les erreurs de mesures suivent une loi Sech,

$$(2.2.4) \quad \sigma_M(\mathbf{m}) \propto \exp(-E_s(\mathbf{m})),$$

avec (en utilisant (1.1.53))
(2.2.5)

$$E_s(\mathbf{m}) = \frac{1}{2} \left(\sum_{(i,j) \in \mathfrak{S}} \log \left(\pi \sigma_{ij} \cosh \left(\frac{t_{ij}^{obs} - t_i^j(\mathbf{m})}{\sigma_{ij}} \right) \right) \right)^2 + (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_0),$$

où σ_{ij} représente l'erreur caractéristique sur la donnée correspondant au séisme i et à la station j .

Nous allons étudier en détail cette classe de problèmes en commençant par le plus simple, la recherche de l'hypocentre d'un séisme, lorsque l'information *a priori* sur le modèle de vitesse est modélisée par une mesure de Dirac. Ensuite nous verrons le cas d'une information *a priori* moins contraignante, ce qui nous amènera naturellement à considérer les hypocentres, les temps origines et la vitesse des ondes du milieu de propagation comme des paramètres à ajuster.

2.2.1.1. *Localisation d'un séisme : faisabilité.* Dans ce problème, les paramètres inconnus sont les coordonnées (x, y, z) d'un séisme et son instant de déclenchement t^o . En général, une station enregistre une trace du séisme, permettant de connaître l'instant d'arrivée du premier train d'onde avec une certaine incertitude (typiquement de l'ordre de quelques centièmes à quelques dixièmes de seconde, pour un séisme local enregistré par un réseau local). Nous allons d'abord voir s'il est possible de retrouver les paramètres du séisme à partir de telles données.

Commençons par considérer le cas le plus simple ; supposons que l'on connaisse l'instant de déclenchement du séisme et la loi de vitesse des ondes P en profondeur. Il est alors possible de déterminer la zone dans laquelle doit se trouver le séisme. En effet, si l'onde parcourt le trajet source-station en t_p secondes avec une incertitude de δt_p secondes, le séisme se trouve entre les deux fronts d'ondes, $T^{-1}(t_p - \delta t_p)$ et $T^{-1}(t_p + \delta t_p)$ où $(x, y, z) \mapsto T(x, y, z)$ est la fonction temps de propagation, en considérant la station comme source émettrice d'ondes sismiques fictives se propageant dans le sous sol. En utilisant plusieurs stations, on détermine le domaine géographique qui contient le séisme (figure (2.2.1)). Lorsque le modèle de vitesse est lisse, les fronts d'ondes sont convexes, il est alors certain que le séisme se trouve dans une zone connexe. Mais si les fronts d'ondes sont trop complexes, c'est le cas dans des modèles de vitesse très hétérogènes, il n'est pas possible de garantir qu'il existe une zone unique susceptible de contenir le séisme. La connaissance du temps origine et la régularité du modèle de vitesse constituent une condition suffisante pour assurer l'existence et l'unicité de la solution du problème de localisation. On peut ajouter qu'un modèle de vitesse lisse est une condition nécessaire pour bien poser le problème.

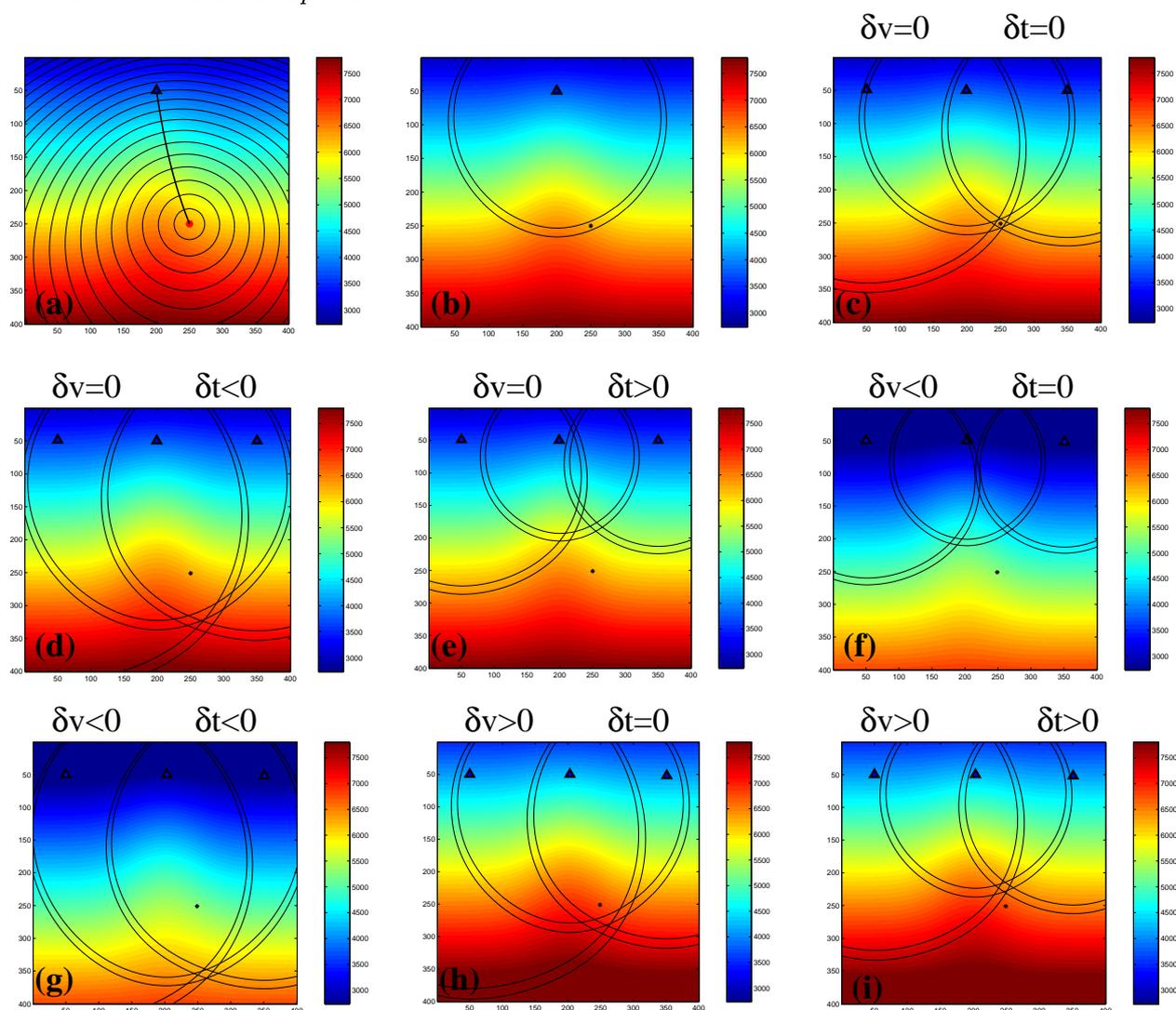
Est-il possible de déterminer un hypocentre simultanément avec son temps origine dans un modèle de vitesse connu et suffisamment régulier ? Comme nous venons de le voir, si le temps origine est connu les fronts des ondes fictives issues des stations se croisent sur la position du séisme. Supposons que l'on déclenche le séisme avant le temps origine vrai. Les temps de propagation résultants sont alors sur-estimés. L'intersection (figure (2.2.1) (d)) des front d'ondes issues des stations est alors vide. En effet, le temps de propagation étant trop grand, les fronts d'ondes ont dépassé la position du séisme et ils ne se croisent

FIG. 2.2.1. Localisation d'un séisme (1).

La figure (a) montre un séisme et les fronts d'ondes qui en sont issus. Connaissant l'instant de déclenchement du séisme, on peut en déduire le temps t_p de propagation le long du rai. Si l'on considère une onde fictive (figure (b)) partant de la station et se propageant dans le milieu, le séisme doit obligatoirement se trouver entre les deux fronts d'ondes $T^{-1}(t_p - \delta t_p)$ et $T^{-1}(t_p + \delta t_p)$. En recommençant l'opération pour différentes stations (figure (c)), il est possible de localiser le séisme dans la zone formée de l'intersection des volumes délimités par les fronts d'ondes.

Lorsque le temps origine est faux, l'intersection des fronts d'ondes est vide. La figure (d) montre le cas où le séisme est déclenché trop tôt (le temps de propagation est surestimé), et la figure (e) montre les fronts d'ondes lorsqu'on déclenche le séisme trop tard (le temps de propagation est sous-estimé). Il est donc possible dans un milieu de vitesse connu de déterminer les coordonnées d'un évènement en simultanéité avec le temps origine.

Lorsque le modèle de vitesse est mal connu, la position du séisme en est affectée. La figure (f) montre les fronts d'ondes pour le temps origine exact dans un modèle de vitesse trop lent, leur intersection est vide. Pour arriver à localiser, il faut faire une erreur négative sur le temps origine (figure (g)) qui a pour effet de placer le séisme trop bas. L'effet inverse se produit pour un modèle de vitesse trop rapide (figures (h) et (j)), on doit déclencher le séisme plus tard pour arriver à faire coïncider les fronts d'ondes : l'évènement est localisé trop haut.



pas simultanément. Lorsque le déclenchement du séisme se fait trop tard (figure (2.2.1) (e)), les temps de propagation sont alors trop court, les fronts d'ondes n'ont pas encore atteint la zone du séisme et leur intersection est vide. Finalement, l'intersection est non vide seulement quand le temps origine est correct. Cela prouve qu'il existe une solution unique au problème de la localisation d'un séisme dans un modèle de vitesse suffisamment régulier. Ainsi on peut espérer trouver la position et le temps origine d'un séisme à l'aide d'une méthode numérique, du moins lorsque les stations sont bien réparties et lorsque l'on connaît le modèle de vitesse.

Nous allons voir maintenant l'effet de la méconnaissance du modèle de vitesse sur la localisation du séisme. La figure (2.2.1) ((f) (g) (h) (i)) montre des expériences numériques de localisation d'un séisme dans un modèle de vitesse trop lent ou trop rapide. Dans le cas d'un modèle de vitesse trop lent, avec le temps origine exact, les fronts d'ondes ne se sont pas encore croisés. On obtient l'intersection en augmentant le temps de parcours, c'est à dire en déclenchant le séisme plus tôt. De cette manière les fronts d'ondes se croisent mais trop profondément par rapport à la position exacte. Lorsque le modèle de vitesse est trop rapide, il se produit l'effet inverse, on parvient à faire coïncider les fronts d'ondes en déclenchant le séisme après le temps origine exact, ce qui positionne l'évènement moins profondément qu'il ne l'est en réalité. On peut résumer ainsi,

$$\begin{cases} \delta v > 0 \Rightarrow \delta t > 0 \text{ et } \delta p < 0, \\ \delta v < 0 \Rightarrow \delta t < 0 \text{ et } \delta p > 0. \end{cases}$$

Ainsi une erreur sur le modèle de vitesse induira une erreur sur le calcul de la localisation d'un séisme, quelle que soit la méthode numérique utilisée. Il existe également un artefact de localisation dû à la géométrie du réseau. La figure (2.2.2) montre une géométrie qui place le séisme loin du réseau, la distance entre les stations est faible par rapport à la distance au séisme. Comme précédemment si on utilise le vrai temps origine, les fronts d'ondes fictifs issus des stations se croisent sur le séisme. Or, en faisant une erreur (aussi bien positive que négative) sur le temps origine de l'évènement, contrairement à ce qui se passait pour un réseau correctement réparti (figure (2.2.1)), nous observons une intersection des fronts d'onde non vide. La zone géographique des solutions potentielles est plus vaste que précédemment, l'incertitude sur la localisation est plus importante. Cela se traduit ici par un phénomène de dépendance (ce que les géophysiciens appellent un "trade-off") entre la direction source-réseau et le temps origine ; une erreur sur le temps origine induit une erreur de localisation suivant la direction source-réseau et réciproquement.

Il est donc envisageable de déterminer numériquement les paramètres d'un séisme, position et temps origine. Il faut pour cela vérifier certaines conditions,

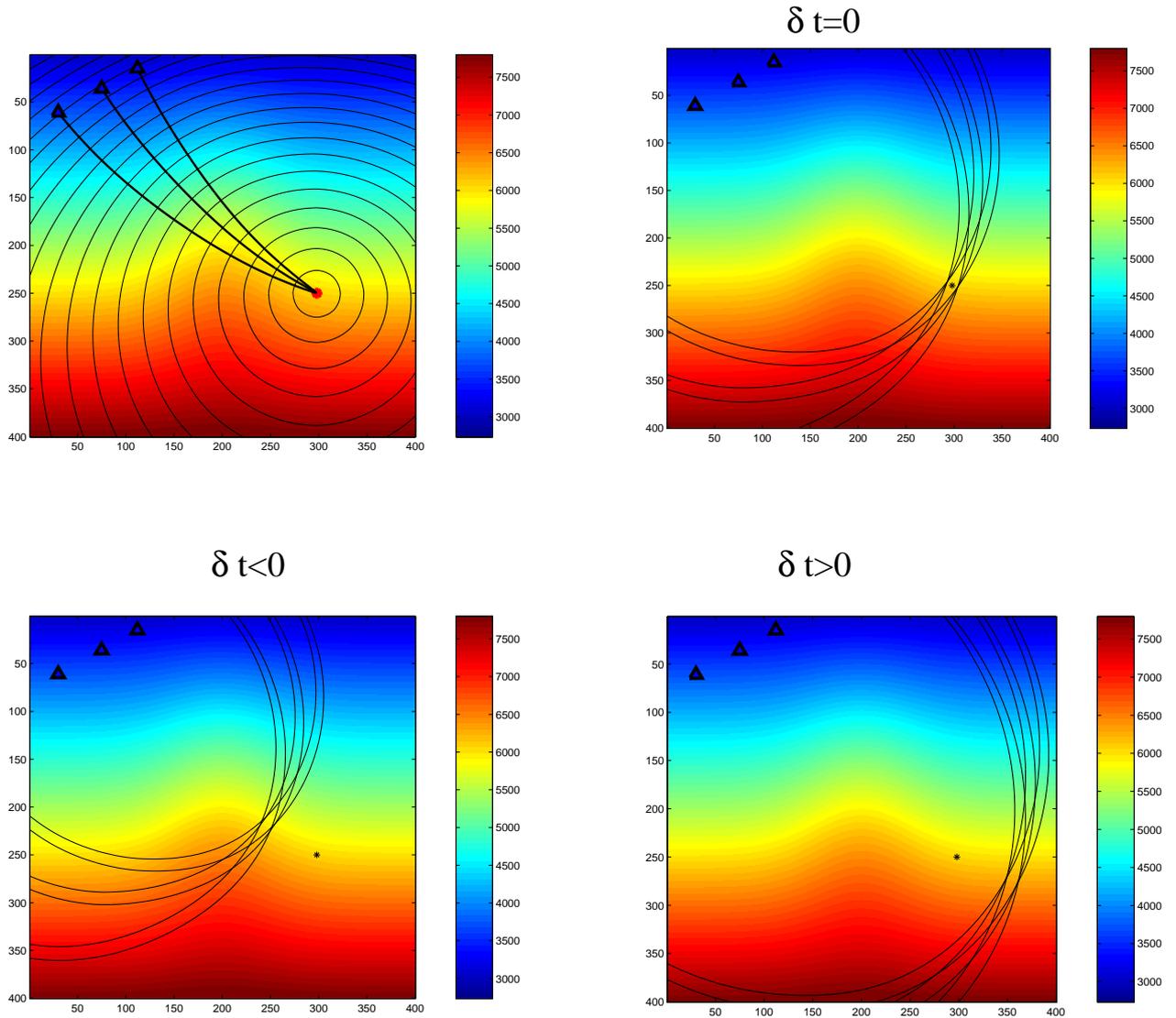
- le modèle de vitesse doit être régulier et lisse,
- le modèle de vitesse doit être exact,
- la géométrie du réseau doit être favorable.

Ces conditions sont rarement réunies dans les cas réels, on doit alors avoir recours à de l'information *a priori* pour régulariser le problème afin de pouvoir le traiter numériquement.

Pour appliquer les schémas numériques généraux (**gn**) et (**se**), dans le cadre de la localisation d'un séisme à l'aide des temps de première arrivée, nous supposons la matrice de

FIG. 2.2.2. Stations trop proches.

Si la distance entre les stations est faible vis à vis de la distance du réseau au séisme, il est impossible de localiser celui-ci avec précision. Nous avons représenté les fronts d'ondes pour différents temps origines, en haut à droite on a utilisé le temps origine exact, en bas à gauche le séisme est déclenché avec une avance de 1 seconde et enfin en bas à droite le séisme explose avec un retard de 1 secondes. Dans les trois cas l'intersection des fronts d'ondes est non vide. Les données admettent plusieurs solutions cela se traduit par la dépendance de certains paramètres (trade off); ici entre la direction source réseau et le temps origine.



variance-covariance sur les données \mathbf{C}_D diagonale. Cela revient à considérer les données non corrélées entre elles. Nous supposons aussi diagonale la matrice de variance-covariance \mathbf{C}_m décrivant la densité de probabilité *a priori* sur les hypocentres et les temps origine. Cela revient à supposer qu'*a priori* il n'y a aucune corrélation entre les différents paramètres. Nous allons à présent définir la matrice \mathbf{G}_k .

2.2.1.2. *Matrice jacobienne dans le cas gaussien.* Nous devons calculer les dérivées partielles du problème direct. Dérivons l'expression (2.2.6) suivante,

$$(2.2.6) \quad t_i^j(t_i^o, \mathbf{x}_i) = t_i^o + \int_{\gamma_{ij}} s d\sigma,$$

nous en déduisons immédiatement la dérivée partielle par rapport au temps origine,

$$(2.2.7) \quad \frac{\partial t}{\partial t_i^o}(t_i^o, \mathbf{x}_i) = 1.$$

Considérons maintenant la fonction $\mathbf{x} \mapsto \tau(\mathbf{x})$, le temps de propagation de l'onde du récepteur j au point \mathbf{x} du domaine. Nous avons

$$(2.2.8) \quad \tau(\mathbf{x}_i) = t_i^j(0, \mathbf{x}_i),$$

nous pouvons alors écrire,

$$(2.2.9) \quad \nabla \tau(\mathbf{x}_i) = \begin{pmatrix} \frac{\partial t_i^j}{\partial x_i^1}(0, \mathbf{x}_i) \\ \frac{\partial t_i^j}{\partial x_i^2}(0, \mathbf{x}_i) \\ \frac{\partial t_i^j}{\partial x_i^3}(0, \mathbf{x}_i) \end{pmatrix}.$$

$\nabla \tau(\mathbf{x}_i)$ est un vecteur tangent au rai γ_{ij} au point \mathbf{x}_i ; d'après l'équation de l'Eikonale, le module de ce vecteur est la valeur de la lenteur $s(\mathbf{x}_i)$. Nous savons aussi que $\nabla \tau(\mathbf{x}_i)$ suit la direction de plus grande pente dans le sens "montant", c'est à dire le sens récepteur-source (on "tire" à partir du récepteur). Nous en déduisons alors que $\nabla \tau(\mathbf{x}_i)$ est l'opposé du vecteur lenteur $\mathbf{s}_i^j = (s_1^{ij}, s_2^{ij}, s_3^{ij})^t$ au point \mathbf{x}_i ,

$$(2.2.10) \quad \begin{aligned} \frac{\partial t_i^j}{\partial x_i^1}(0, \mathbf{x}_i) &= -s_1^{ij}, \\ \frac{\partial t_i^j}{\partial x_i^2}(0, \mathbf{x}_i) &= -s_2^{ij}, \\ \frac{\partial t_i^j}{\partial x_i^3}(0, \mathbf{x}_i) &= -s_3^{ij}. \end{aligned}$$

Nous avons ainsi tous les éléments composant la matrice \mathbf{G}_k pour le problème de localisation. La ligne de la matrice correspondant au rai γ_{ij} s'écrit,

$$(2.2.11) \quad (\mathbf{G}_k \delta \mathbf{m})_{ij} = -s_1^{ij} \delta x_i - s_2^{ij} \delta y_i - s_3^{ij} \delta z_i + \delta t_0^i.$$

2.2.1.3. *Matrice jacobienne avec la loi sech.* Comme nous l'avons vu plus haut, si les erreurs de mesure se distribuent suivant une loi Sech, il suffit d'une modification minime de l'algorithme de Gauss-Newton pour prendre en compte cette distribution. Il suffit en effet de pondérer chaque ligne de la matrice \mathbf{G}_k ; on obtient ainsi

$$(2.2.12) \quad (\overline{\mathbf{G}_k \delta \mathbf{m}})_{ij} = \frac{\exp \left(\operatorname{Erf}^{-1} \left(\frac{2}{\pi} \arctan \left(\sinh \left(\frac{t_i^j(\mathbf{m}_k) - t_{ij}^{obs}}{\sigma_{ij}} \right) \right) \right) \right)}{\sqrt{\pi} \cosh \left(\frac{t_i^j(\mathbf{m}_k) - t_{ij}^{obs}}{\sigma_{ij}} \right)} (-s_1^{ij} \delta x_i - s_2^{ij} \delta y_i - s_3^{ij} \delta z_i + \delta t_0^i).$$

2.2.1.4. *Loi gaussienne tangente.* La méthode de Gauss-Newton permet de calculer un maximum de vraisemblance de la loi de probabilité *a posteriori*. Elle fournit un modèle moyen, l'hypocentre du séisme et son temps origine. Pour compléter l'étude du problème inverse, il convient de quantifier l'erreur sur ce modèle. On pourrait cartographier la loi de probabilité *a posteriori*, ce qui requiert un volume de calcul important. On peut simplifier ce calcul en utilisant la loi Gaussienne tangente. Lorsqu'on ne considère pas de modèle *a priori* elle s'écrit,

$$(2.2.13) \quad \sigma_t(\mathbf{m}) = \exp \left(-\frac{1}{2} (\mathbf{m} - \widehat{\mathbf{m}})^t \mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{G} (\mathbf{m} - \widehat{\mathbf{m}}) \right),$$

c'est une loi gaussienne de matrice de variance-covariance $(\mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{G})^{-1}$. La base des vecteurs propres de cette matrice définit les directions principales de l'ellipsoïde d'erreur sur les paramètres. Les valeurs propres quantifient l'erreur caractéristique sur chaque axe de l'ellipsoïde. La figure (2.2.3) montre un test synthétique de problème de localisation. La géométrie du réseau impose les vecteurs d'ondes contenus dans la matrice de variance-covariance *a posteriori* (figure (2.2.3)-(b)) qui une fois diagonalisée, nous permet de tracer un ellipsoïde d'erreur. Afin de vérifier expérimentalement la pertinence de cette démarche, nous avons dessiné les lois marginales de la densité de probabilité *a posteriori* (figure (2.2.3)-(c)-(e)-(g)) que l'on peut comparer aux projections de l'ellipsoïde d'erreur sur les différents plans (figure (2.2.3)-(d)-(f)-(h)). Le problème direct étant non linéaire, la densité *a posteriori* n'est évidemment pas gaussienne mais la loi gaussienne tangente peut convenir au calcul d'erreur. En effet, les deux lois (Figure 2.2.3) sont suffisamment proches l'une de l'autre pour permettre cette approximation.

L'étude de la matrice de covariance *a posteriori* permet de voir l'influence de la géométrie du réseau sur les erreurs de localisation. La figure (2.2.4) (a1-a2) montre que des stations équi-réparties azimutalement ne privilégient aucun paramètre, tous sont déterminés avec la même précision, alors qu'un réseau de surface (figure (2.2.4) (b1-b2)) détermine moins bien la profondeur que la latitude et la longitude. Il en est de même avec un séisme superficiel (figure (2.2.4) (c1-c2)), la profondeur n'est pas très bien contrainte. Lorsque deux stations sont trop proches (figure (2.2.4) (d1-d2)), la profondeur est encore moins bien contrainte. Lorsque l'ellipsoïde d'erreur est très allongé dans certaines directions, le conditionnement de la matrice des dérivées partielles \mathbf{G} est mauvais (figure (2.2.5)). Cela

peut entraîner des instabilités numériques dans le processus de Gauss-Newton. On peut y remédier en définissant un modèle *a priori*.

2.2.1.5. *Information a priori gaussienne.* Nous supposons qu'*a priori* les paramètres hypocentraux et le temps origine ne sont pas corrélés. L'information *a priori* se ramène donc à considérer des valeurs de paramètres et des erreurs sur celles-ci. En notant (x, y, z, t_o) les paramètres d'un séisme, nous noterons σ_x σ_y σ_z et σ_{t_o} les écart-types *a priori*. La matrice de variance-covariance *a priori* s'écrit donc,

$$(2.2.14) \quad \mathbf{C}_m = \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_z^2 & 0 \\ 0 & 0 & 0 & \sigma_{t_o}^2 \end{pmatrix},$$

la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ utilisée dans les schémas (**gn**) et (**se**) s'écrit alors,

$$(2.2.15) \quad \mathbf{C}_m^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sigma_x} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_y} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_z} & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_{t_o}} \end{pmatrix}.$$

2.2.2. Localisation avec les décalages temporels. Quant on connaît la position d'un séisme, il est possible de positionner relativement un autre évènement par rapport à ce dernier. Lorsque les deux évènements sont proches, le problème direct posé par la localisation relative est quasi-linéaire, ce qui d'un point de vue numérique est très favorable. Les hypothèses sur la physique du problème sont alors (figure (2.2.6)) :

– Les deux évènements sont supposés suffisamment proches pour que :

- (1) La vitesse au voisinage des foyers soit constante.
- (2) Les rais partent dans la même direction aux deux points sources.

– La différence de temps de propagation est uniquement due à la différence de position entre les deux séismes.

Si on approxime le parcours de l'onde du point B au point M par le segment [BM], le temps de parcours sur ce segment est $\frac{1}{V} \|\overrightarrow{BM}\|$, ce qui se réécrit $\mathbf{s}^t \mathbf{u}$ (en notant $\mathbf{u} = \overrightarrow{AB}$ et \mathbf{s} le vecteur lenteur commun aux deux rais). En notant γ_1 la courbe paramétrée du rai R1 et γ_2 celle du rai R2, on peut écrire

$$(2.2.16) \quad \int_{\gamma_2} s d\sigma \simeq \int_{\gamma_1} s d\sigma + \mathbf{s}^t \mathbf{u}.$$

Pour définir le problème direct, on utilise cette dernière relation (2.2.16),

$$(2.2.17) \quad \int_{\gamma_2} s d\sigma - \int_{\gamma_1} s d\sigma = \mathbf{s}^t \mathbf{u}.$$

FIG. 2.2.3. Localisation d'un séisme : loi gaussienne tangente.

Des données sont générés synthétiquement dans le modèle (a). La matrice de covariance a posteriori permet de définir l'ellipsoïde d'erreur (b). Cette façon de faire est justifiée car la loi a posteriori (c)-(e)-(g) et la loi Gaussienne tangente (d)-(f)-(h) sont très proches.

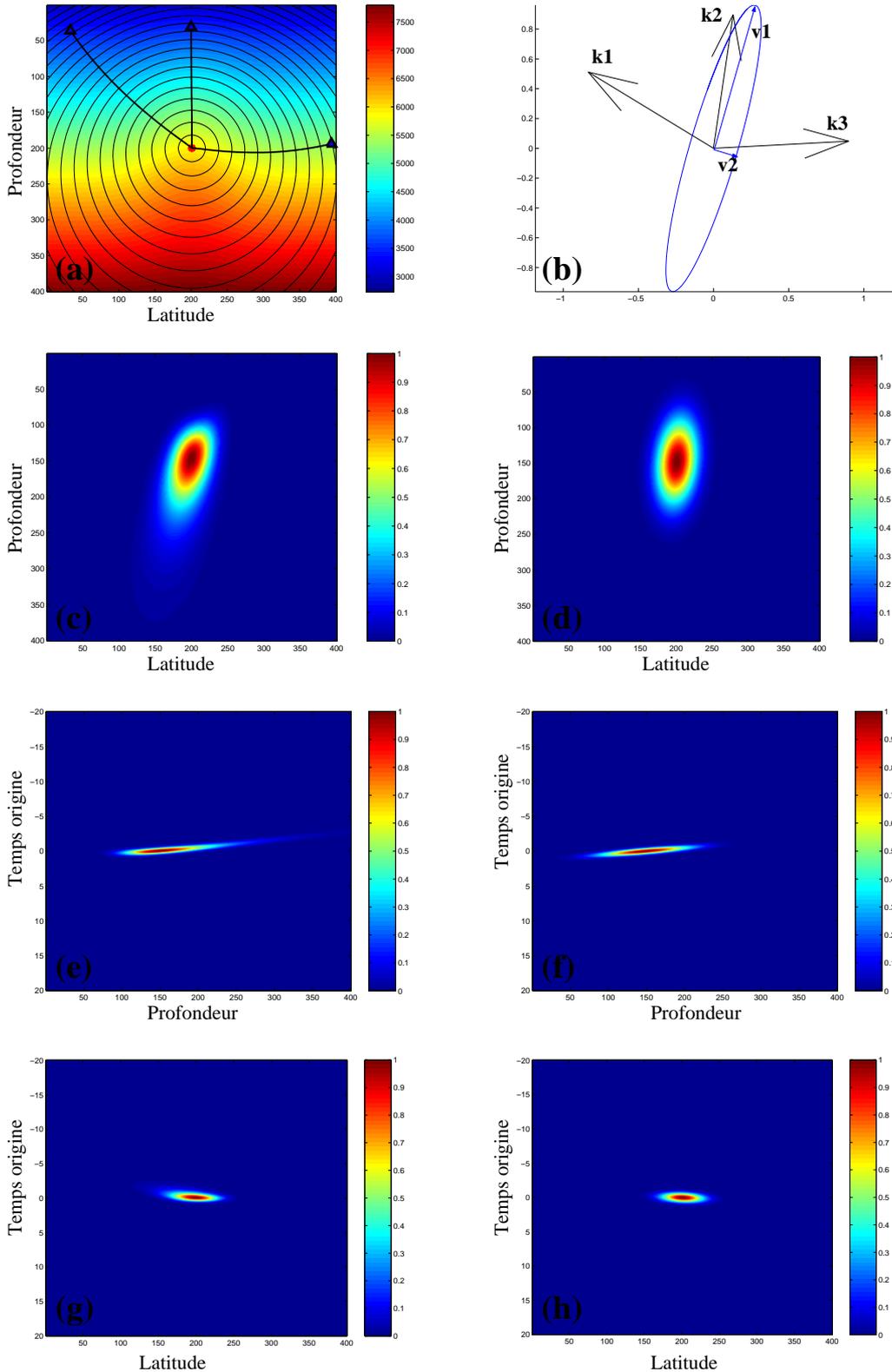


FIG. 2.2.4. Localisation d'un séisme : influence du réseau.

La répartition du réseau sismique influence la forme de l'ellipsoïde d'erreur par l'intermédiaire de la matrice de covariance *a posteriori*. Voici quelques géométries et les erreurs correspondantes. Nous avons dessiné la forme de l'ellipsoïde en normalisant le grand axe puisque l'étendue de celui-ci ne dépend que du bruit sur les données.

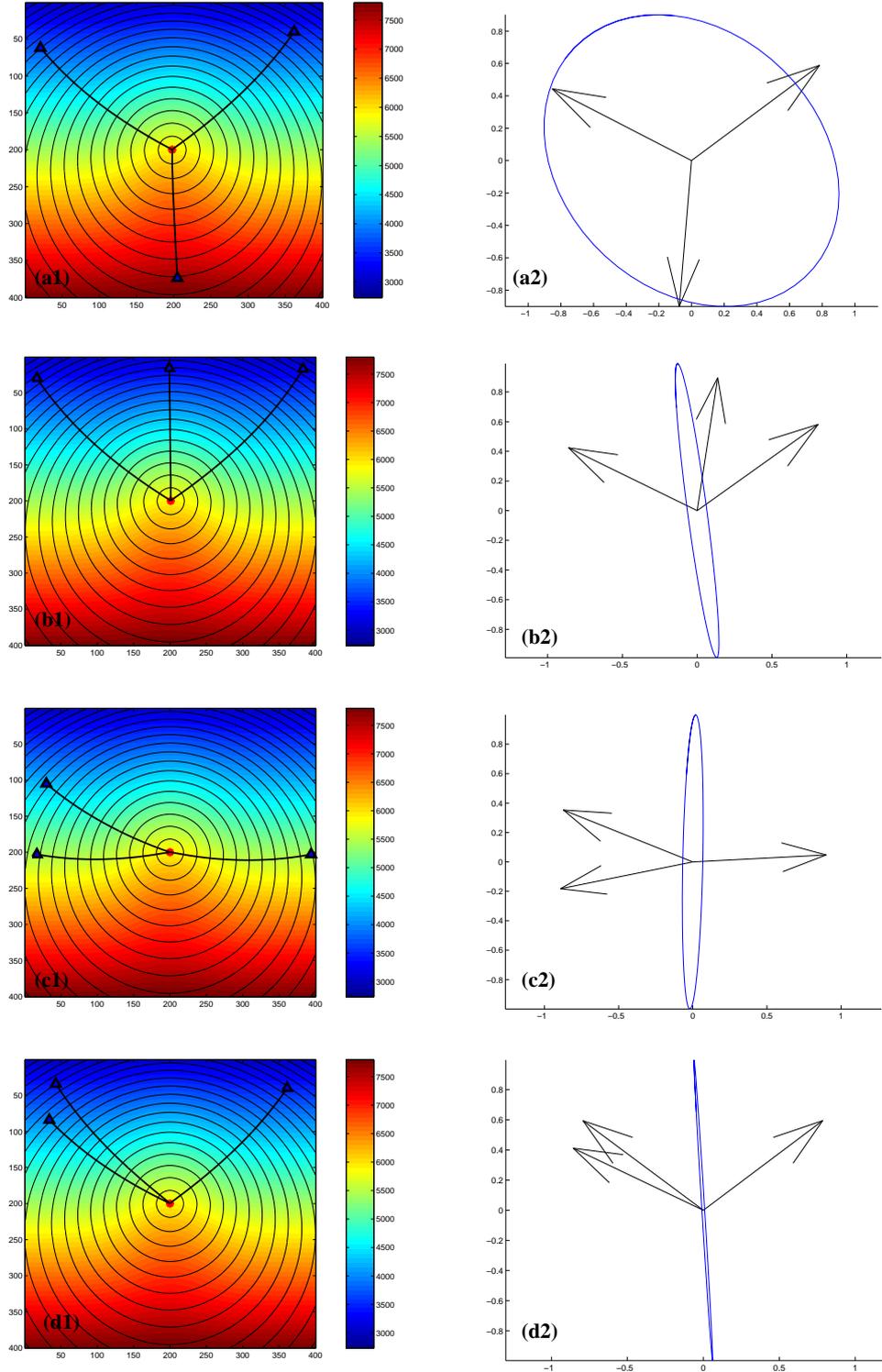


FIG. 2.2.5. Localisation d'un séisme : conditionnement de G .

Le conditionnement de la matrice G est le rapport entre la longueur du plus grand axe de l'ellipsoïde d'erreur et le plus petit. Nous voyons qu'il dépend de l'ouverture azimutale du réseau.

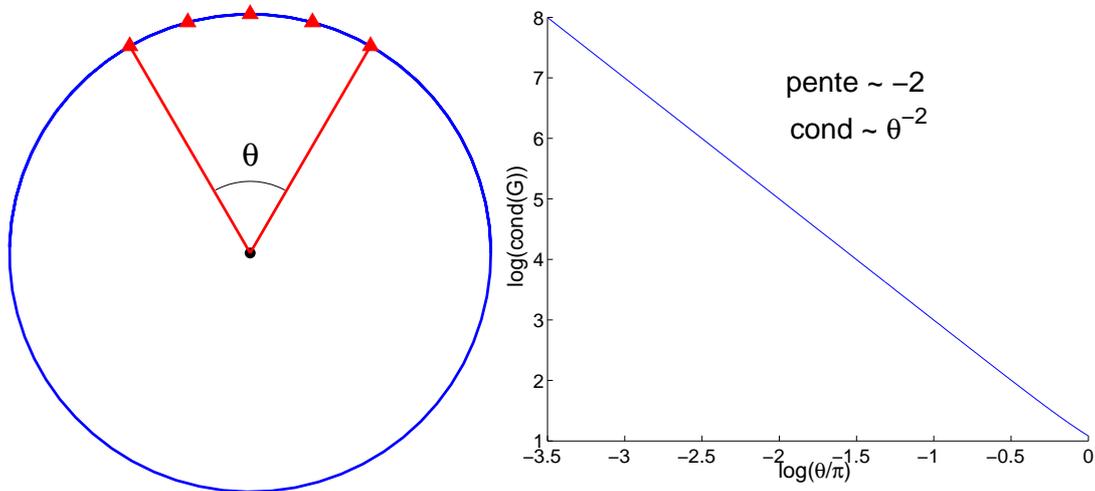
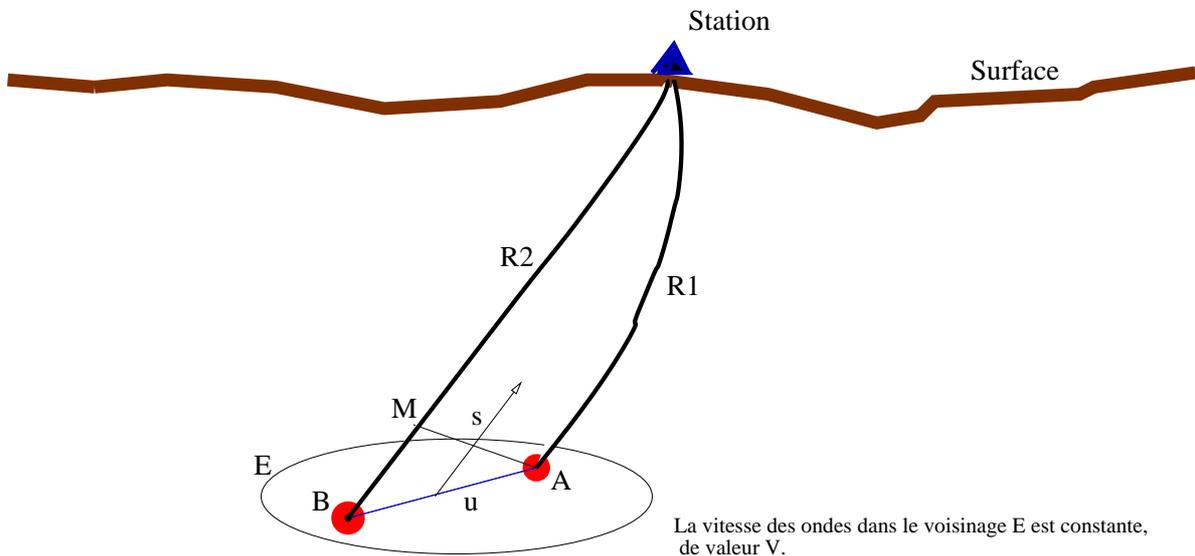


FIG. 2.2.6. Localisation relative de deux séismes.

Un séisme à lieu au point A , les ondes se propagent jusqu'à la station le long du rai $R1$. Il s'agit de retrouver la position de l'évènement $B (x_B, y_B, z_B, t_o^B)^t$ relativement à $A (x_A, y_A, z_A, t_o^A)^t$ en utilisant la différence des temps de propagation. Notons M la projection orthogonale de A sur le rai $R2$. On suppose que B est dans un voisinage E de A , que les temps de propagation sur $R2 \setminus [BM]$ et sur $R1$ sont égaux et que la direction de départ des rais $R1$ et $R2$ est suivant le vecteur lenteur s .



Le calcul de la différence des temps de propagation est effectuée en considérant un temps origine (inconnu) pour chaque séisme,

$$(2.2.18) \quad t_2^{obs} - t_1^{obs} = \int_{\gamma_2} sd\sigma + t_o^2 - \int_{\gamma_1} sd\sigma - t_o^1$$

En utilisant (2.2.17), et en notant $\delta t_o = t_o^2 - t_o^1$, on définit le résidu

$$(2.2.19) \quad r(\delta x, \delta y, \delta z, \delta t_o) = t_2^{obs} - t_1^{obs} - \mathbf{s}^t \mathbf{u} - \delta t_o,$$

avec $\mathbf{u} = (\delta x, \delta y, \delta z)^t$. Ainsi on définit un problème inverse linéaire qui permet de calculer la position relative des évènements entre eux (Fréchet, [13]; Got et coll., [15]). Si on dispose d'une mesure directe d_{21}^{obs} du décalage temporel (par exemple par méthode interspectrale [13]), le résidu peut s'écrire

$$(2.2.20) \quad r(\delta x, \delta y, \delta z, \delta t_o) = d_{12}^{obs} - \mathbf{s}^t \mathbf{u} - \delta t_o,$$

Pour généraliser au cas d'un essaim d'au moins trois séismes, on peut définir un vecteur \mathbf{s} unique par station placé au barycentre supposé de l'essaim. Cette méthode suppose que l'essaim est regroupé dans une zone de faible extension vis à vis de la distance au réseau. Si ce n'est pas le cas, il faut considérer un problème direct non linéaire, où chaque séisme est affecté d'un vecteur lenteur \mathbf{s} . C'est l'objet de la partie suivante.

2.2.2.1. *Définition générale du problème.* La définition d'un problème inverse particulier nécessite, en premier lieu, la modélisation des données à ajuster. Dans cette partie nous considérerons les décalages temporels. Le décalage temporel peut être défini comme la différence de temps de propagation entre deux séismes i et j en une station k ,

$$(2.2.21) \quad d_{ij}^k(\mathbf{m}) = t_j^k(\mathbf{m}) - t_i^k(\mathbf{m})$$

Il s'agit alors de perturber le modèle \mathbf{m} , pour ajuster un ensemble de décalages temporels observés, $\mathbf{d}^{obs} = (d_{ij}^k)_{(i,j,k) \in \mathfrak{S}}$, où \mathfrak{S} est un ensemble d'indice (i,j,k) qui désigne le couple de séismes et la station, associé à chaque décalage temporel. Remarquons d'abord que la notion de décalage temporel a été définie comme une différence de temps de propagation (2.2.21), et qu'il n'est pas possible d'accéder directement à ce type d'information. Il existe néanmoins au moins deux façons de les approximer, soit on effectue des différences de temps pointés manuellement sur des sismogrammes, soit il sont calculés directement en utilisant la méthode interspectrale (cf Annexe 2 et [20] ou [13]).

2.2.2.2. *Localisation en Double Différence.* Le terme "double différence" à été introduit pour caractériser une généralisation du problème de relocalisation relative (Waldhauser et Ellsworth [52]). Ces auteurs considèrent un jeu de données constitué à partir de différences de temps d'arrivées pointés sur des sismogrammes, et éventuellement des décalages temporels calculés par intercorrélation. En perturbant les hypocentres et les temps origine des séismes, dans un modèle de vitesse unidimensionnel, Waldhauser et Ellsworth [52] ajustent ces données au sens des moindres carrés, tout en imposant une condition de conservation de la position du barycentre des essaims relocalisés. La position moyenne de l'essaim est contrôlée par cette contrainte. La méthode de localisation double différence de Waldhauser et Ellsworth [52] est donc, dans la mesure où la contrainte de conservation de la position

du barycentre est appliquée, une méthode de localisation relative au même titre que celle de Fréchet, [13], Got et coll., [15] si ce n'est qu'elle est définie avec un problème direct non linéaire.

Nous allons dans ce chapitre définir une méthode de localisation utilisant les décalages temporels dans laquelle :

- nous n'utiliserons pas de contraintes sur le barycentre des événements², car comme nous le verrons il est parfois possible de déterminer les positions absolues en utilisant des données relatives. Dans le cas particulier où le barycentre ne sera pas contraint par le problème, l'information *a priori* va naturellement stabiliser l'algorithme. Il faut donc utiliser une approche d'optimisation pour déterminer la quantité (optimale) d'information *a priori* à apporter ;
- nous localiserons les séismes dans un modèle de vitesse à trois dimensions représentant au mieux les variations de vitesse.

2.2.2.3. *Paramétrage du problème.* Les données que l'on cherche à ajuster sont les décalages temporels entre séismes mesurés par un ensemble de stations. Le décalage temporel d_{ij}^k entre deux séismes i et j obtenu en une station k est la différence de temps de propagation,

$$(2.2.22) \quad d_{ij}^k = t_j^k - t_i^k.$$

En général, on pointe sur les traces l'instant d'arrivée $t_a^{i,k}$ d'un séisme i en une station k , et par un premier calcul de localisation absolue, on estime les paramètres hypocentaux \mathbf{x}_i du séisme et son temps origine t_o^i . On peut obtenir une première approximation de d_{ij}^k en faisant,

$$(2.2.23) \quad d_{ij}^k \simeq \widetilde{d}_{ij}^k = (t_a^{j,k} - t_o^j) - (t_a^{i,k} - t_o^i).$$

Les temps origine n'étant connus qu'avec une certaine erreur,

$$(2.2.24) \quad d_{ij}^k = (t_a^{j,k} - t_o^j + \delta t_o^j) - (t_a^{i,k} - t_o^i + \delta t_o^i) = \widetilde{d}_{ij}^k + \delta t_o^j - \delta t_o^i,$$

en notant δt_o^i l'erreur sur le temps origine t_o^i et \widetilde{d}_{ij}^k le décalage temporel estimé. Il est alors nécessaire de calculer cette erreur lors du processus d'inversion. Ainsi, les paramètres recherchés sont l'hypocentre et la correction du temps origine de chaque événement. Cela nous conduit à définir les résidus de la façon suivante,

$$(2.2.25) \quad r_k(\mathbf{x}_i, \mathbf{x}_j, \delta t_o^i, \delta t_o^j) = d_{ij}^k - (t^k(\mathbf{x}_j) - t^k(\mathbf{x}_i) + \delta t_o^j - \delta t_o^i).$$

Nous allons commencer par considérer le problème inverse dans le cas gaussien avec une information *a priori* modélisée par la fonction caractéristique d'un ensemble de paramètres (cf partie 1.1.1). Cela est équivalent à ajuster les données au sens des moindres carrés. Il s'agit de minimiser la fonction,

$$(2.2.26) \quad f(\mathbf{x}_i, \mathbf{x}_j, \delta t_o^i, \delta t_o^j) = \sum_k \sum_i \sum_{j>i} \left(\frac{1}{\alpha_{ij}^k} r_k(\mathbf{x}_i, \mathbf{x}_j, \delta t_o^i, \delta t_o^j) \right)^2,$$

²En utilisant l'algorithme LSQR cette contrainte est inutile, en effet si le barycentre n'est pas contraint par les équations, il reste à sa valeur *a priori* car LSQR calcule la solution de norme minimale.

en notant $(\alpha_{ij}^k)_{\substack{1 \leq k \leq nst \\ 1 \leq i < nse, i < j \leq nse}}$ les écart-types des erreurs sur les données. Appelons $\mathbf{m}_e = \left\{ \left(\overline{\mathbf{x}}_i, \overline{\delta t_o^i} \right)_{1 \leq i \leq nse} \right\}$ la solution³ de (2.2.26), et notons δ_l la différence des corrections de temps origine entre les séismes l et $l+1$,

$$(2.2.27) \quad \delta_l = \overline{\delta t_o^{l+1}} - \overline{\delta t_o^l}.$$

Considérons les évènements $\mathbf{m}_\lambda = \left\{ \left(\overline{\mathbf{x}}_i, \lambda - \sum_{l=i}^{l=nse-1} \delta_l \right)_{1 \leq i < nse}, (\overline{\mathbf{x}}_{nse}, \lambda) \right\}$, où λ est un réel quelconque fixé *a priori*. Il est facile de vérifier que

$$(2.2.28) \quad r_k \left(\overline{\mathbf{x}}_j, \overline{\mathbf{x}}_i, \lambda - \sum_{l=j}^{l=nse-1} \delta_l, \lambda - \sum_{l=i}^{l=nse-1} \delta_l \right) = r_k \left(\overline{\mathbf{x}}_j, \overline{\mathbf{x}}_i, \overline{\delta t_o^j}, \overline{\delta t_o^i} \right), \text{ pour tous } j < nse, j < i \leq nse, k \leq nst.$$

Il s'en suit⁴, $f(\mathbf{m}_e) = f(\mathbf{m}_\lambda)$, pour tous λ (noter que lorsque $\lambda = \overline{\delta t_o^{nse}}$ alors $\mathbf{m}_e = \mathbf{m}_\lambda$). A partir d'une solution particulière (en l'occurrence la solution exacte) on vient de construire une famille de solutions paramétrée par le réel λ . Cette famille conserve les mêmes positions absolues mais chaque solution diffère par les corrections sur les temps origines, sans changer le résultat de l'ajustement des données. Il est donc mathématiquement impossible de déterminer la correction du temps origine de chaque évènement, qui n'est connue qu'à une constante additive près. Cette remarque permet de proposer une autre paramétrisation, $\tilde{\mathbf{m}} = \{\mathbf{x}_i, \delta_l \mid i = 1, \dots, nse \text{ et } l = 1, \dots, nse - 1\}$. Dans ce cas, le paramètre λ n'est pas cherché. Les résidus s'écrivent alors

$$(2.2.29) \quad r_k \left(\mathbf{x}_i, \mathbf{x}_j, (\delta_l)_{l=1, \dots, nse-1} \right) = d_{ij}^k - \left(t^k(\mathbf{x}_j) - t^k(\mathbf{x}_i) + \sum_{q=i}^{q=j-1} \delta_q \right),$$

et le problème consiste à minimiser la fonction suivante,

$$(2.2.30) \quad f \left(\mathbf{x}_i, \mathbf{x}_j, (\delta_l)_{l=1, \dots, nse-1} \right) = \sum_k \sum_i \sum_{j>i} \left(\frac{1}{\alpha_{ij}^k} r_k \left(\mathbf{x}_i, \mathbf{x}_j, (\delta_l)_{l=1, \dots, nse-1} \right) \right)^2.$$

Nous allons appliquer l'algorithme de Gauss-Newton au problème (2.2.30). Etant donné l'itéré \mathbf{m}_n , nous calculons l'itéré suivant \mathbf{m}_{n+1} par

$$(2.2.31) \quad \mathbf{m}_{n+1} = \mathbf{m}_n + (\mathbf{G}_n^t \mathbf{C}_d^{-1} \mathbf{G}_n)^{-1} \mathbf{G}_n^t \mathbf{C}_d^{-1} \mathbf{r}(\mathbf{m}_n),$$

où $\mathbf{r}(\mathbf{m}_n)$ désigne le vecteur résidu à l'itération n et \mathbf{G}_n représente la matrice jacobienne, calculée à l'itération n , du problème direct $\mathbf{m} \mapsto \mathbf{d}(\mathbf{m})$ qui modélise le vecteur décalage

³Il s'agit ici de la solution au sens physique.

⁴Il est très facile de voir que les corrections sur les temps origines sont indéterminées, en remarquant que les deux évènements suivants ajustent les données de la même façon,

$$\mathbf{m}_0 = \left\{ (\mathbf{x}_i, \delta t_o^i)_{i=1, \dots, nse} \right\} \text{ et } \mathbf{m}_\lambda = \left\{ (\mathbf{x}_i, \delta t_o^i + \lambda)_{i=1, \dots, nse}, \lambda \in R \right\}.$$

temporel théorique. Plus précisément, la dérivée partielle par rapport à un paramètre ξ s'écrit,

$$(2.2.32) \quad \frac{\partial d_{ij}^k}{\partial \xi}(\mathbf{m}) = \frac{\partial t_j^k}{\partial \xi}(\mathbf{m}) - \frac{\partial t_i^k}{\partial \xi}(\mathbf{m}) + \frac{\partial \sum_{q=i}^{q=j-1} \delta_q}{\partial \xi}.$$

Il suffit alors d'utiliser les relations construites pour la localisation d'un séisme à l'aide des temps de première arrivée pour calculer (2.2.32). La ligne de \mathbf{G}_n appliquée à une perturbation $\delta \mathbf{m} = (\delta \mathbf{x}_i, \delta \mathbf{x}_j, \delta_1, \dots, \delta_{n_{se}-1})$ correspondant au couple de séismes i, j et à la station k , s'écrit,

$$(2.2.33) \quad (\mathbf{G}_n \delta \mathbf{m})_{ijk} = -\mathbf{s}_j^k \cdot \delta \mathbf{x}_j + \mathbf{s}_i^k \cdot \delta \mathbf{x}_i + \sum_{q=i}^{q=j-1} \delta_q.$$

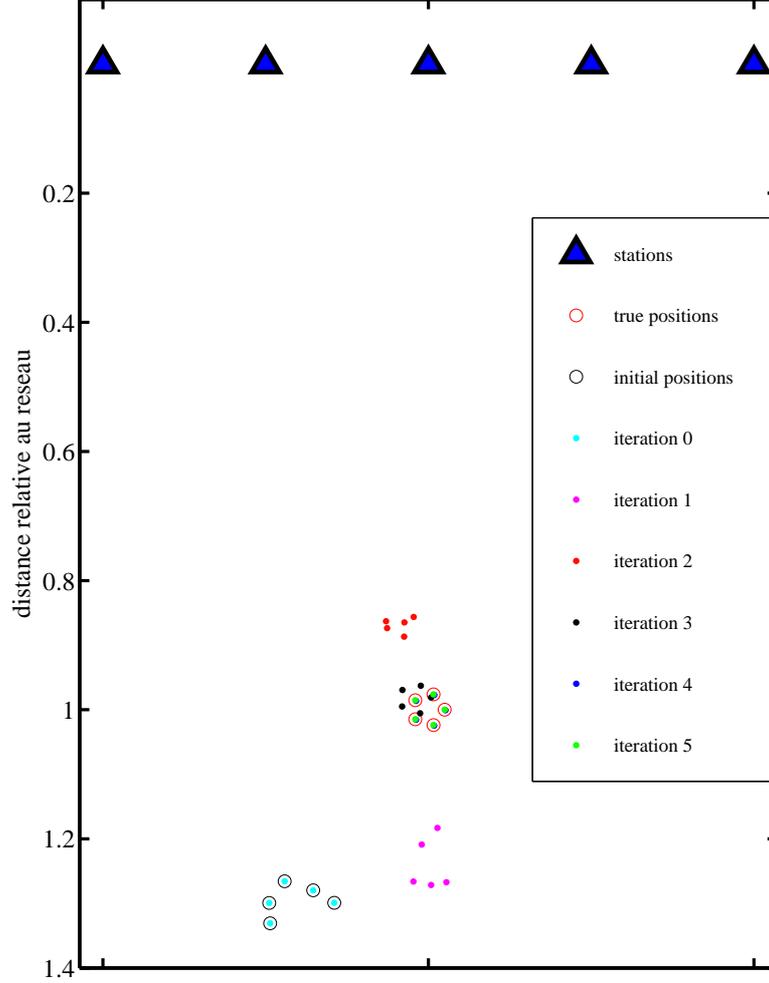
Cette relation (2.2.33) est à comparer avec les équations obtenues par Waldauser et Ellsworth [52] (cf Annexe 3). Ici, le paramétrage est différent. En effet, nous ne cherchons pas le temps origine de chaque séisme puisque il est indéterminé. Ainsi, dans (2.2.33), nous avons un terme $\sum_{q=i}^{q=j-1} \delta_q$ qui diffère par rapport aux équations Doubles Différences de Waldauser et Ellsworth [52].

Montrons qu'il est possible de résoudre le problème (2.2.30) dans un cas numérique synthétique. Considérons un milieu homogène et cinq séismes enregistrés par un réseau de cinq stations situées en surface comme schématisé sur la figure (2.2.7). La figure (2.2.7) nous montre l'évolution des paramètres hypocentaux au cours du processus itératif. En cinq itérations tous les paramètres sont parfaitement retrouvés, ce qui montre que les décalages temporels (2.2.30) peuvent dans certaines conditions contraindre complètement la solution. Il est donc inutile de rajouter à la contrainte apportée par les données de décalages temporels, une contrainte concernant la conservation du barycentre des séismes relocalisés (contrairement à ce qui est fait dans [52]). On atteint donc à ce point la conclusion *a priori* étonnante que des données, **relatives**, de décalages temporels, permettent de déterminer des paramètres de position **absolue** hypocentaux. Seule la correction de temps origine n'est connue que relativement, c'est-à-dire que ce problème ne permet de trouver qu'une chronologie relative des événements. Nous allons examiner ci-dessous dans quelles conditions cette solution peut être obtenue.

2.2.2.4. *Influence de la distance inter-événements.* Reprenons la formulation (2.2.30) du problème de localisation utilisant les décalages temporels. Nous allons montrer à l'aide d'un exemple synthétique qu'il n'est pas toujours possible de déterminer les paramètres hypocentaux. Nous allons aussi voir quelles sont les principales causes qui peuvent amener à une divergence de l'algorithme de Gauss-Newton pour ce problème. La figure (2.2.8) montre une expérience similaire à (2.2.7), dans le cas où les séismes synthétiques sont plus proches les uns des autres : la distance inter-événement est petite devant la distance hypocentrale. La figure (2.2.8) montre que la méthode ne permet plus de résoudre les positions absolues ; seules les positions relatives sont retrouvées. Au cours d'un processus

FIG. 2.2.7. Localisation à l'aide des décalages temporels dans un milieu 2d homogène.

Les données sont générées synthétiquement (sans erreur de mesure) dans un milieu homogène (5 km/s). Les 5 séismes sont régulièrement répartis sur un cercle. Il s'agit de retrouver leur position exacte (en rouge) en minimisant (2.2.30) par la méthode de Gauss-Newton. En cinq itérations on retrouve bien les hypocentres. Le type de problème posé en (2.2.30) résoud, dans ce cas ,tous les paramètres hypocentaux.



itératif convergent nous avons,

$$(2.2.34) \quad \begin{cases} \mathbf{m}_n \rightarrow \bar{\mathbf{m}} \text{ quand } n \rightarrow \infty, \\ \mathbf{G}_n \rightarrow \bar{\mathbf{G}} \text{ quand } n \rightarrow \infty. \end{cases}$$

Lorsque la matrice \mathbf{G}_n se rapproche de $\bar{\mathbf{G}}$, $\text{cond}(\bar{\mathbf{G}})$ tend vers $\text{cond}(\bar{\mathbf{G}})$. Si les positions vraies des séismes sont trop proches, le conditionnement de $\bar{\mathbf{G}}$ est grand (figure(2.2.9)) ; à partir d'une certaine itération le système linéaire associé ne peut plus être résolu correctement numériquement , ce qui entraîne l'échec de la méthode. Le mauvais conditionnement de $\bar{\mathbf{G}}$ provient de la faible distance inter-événements vis à vis de la distance au réseau. Vus

d'une station les vecteurs d'ondes pour deux séismes semblent identiques, si bien que leur différence est négligeable numériquement, $\mathbf{s}=\mathbf{s}_i^k=\mathbf{s}_j^k$, nous aurons alors dans la matrice \mathbf{G}_n ,

$$(2.2.35) \quad (\mathbf{G}_n \delta \mathbf{m})_{ijk} = -\mathbf{s} \cdot \delta \mathbf{x}_j + \mathbf{s} \cdot \delta \mathbf{x}_i + \sum_{q=i}^{q=j-1} \delta q,$$

autrement dit,

$$(2.2.36) \quad (\mathbf{G}_n \delta \mathbf{m})_{ijk} = -\mathbf{s} \cdot (\delta \mathbf{x}_j - \delta \mathbf{x}_i) + \sum_{q=i}^{q=j-1} \delta q.$$

Une telle équation ne permet pas de calculer indépendamment les positions \mathbf{x}_i et \mathbf{x}_j ; elle permet seulement de trouver la position relative entre les deux séismes $\delta \mathbf{x}_{ij}=\mathbf{x}_j-\mathbf{x}_i$. Les inconnues du système ne sont plus linéairement indépendantes ce qui provoque le mauvais conditionnement de la matrice \mathbf{G}_n . On voit donc que les équations (2.2.33) ne permettent de trouver complètement les paramètres hypocentaux (position absolue et relative) que si la distance inter-événements est suffisante. Pour remédier à ce problème, deux possibilités (non exclusives) peuvent être exploitées :

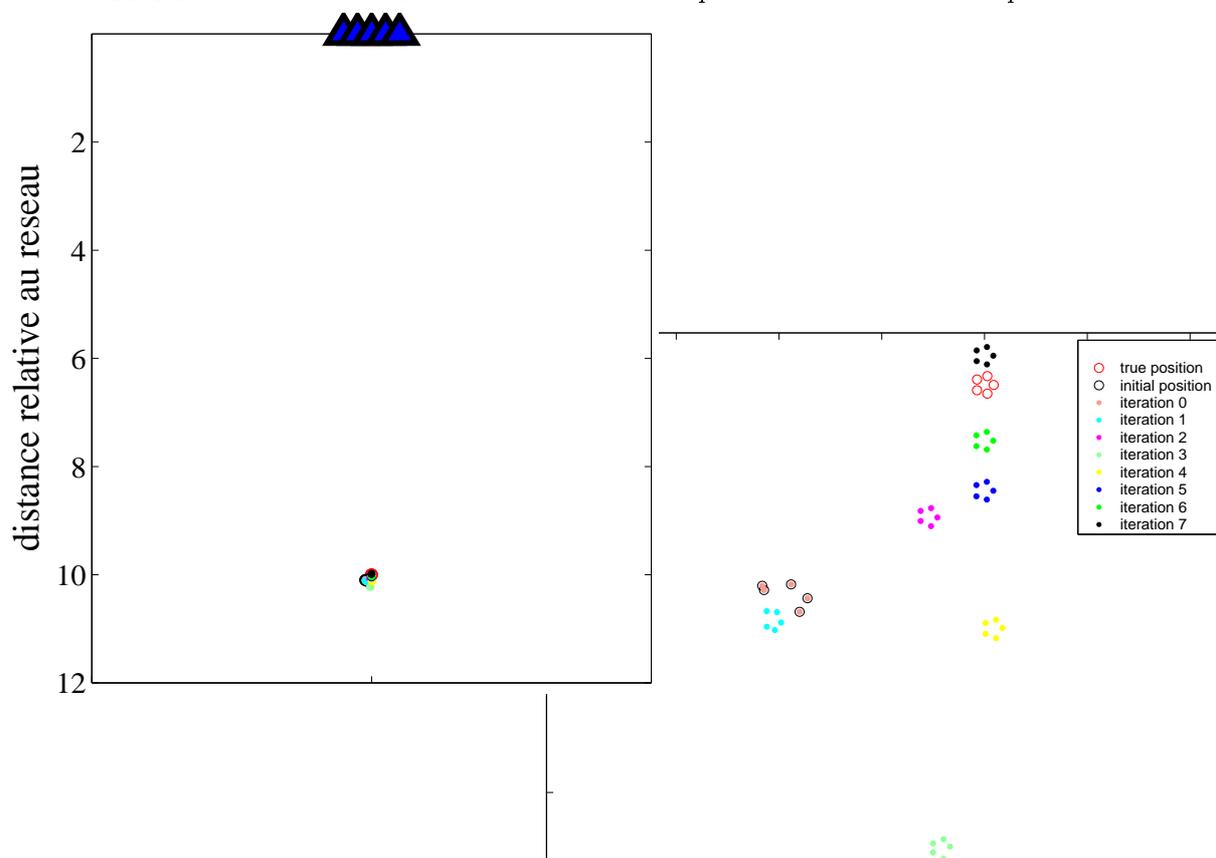
- (1) Introduire de l'information *a priori* pour stabiliser le problème. En effet, avec cette méthode lorsque les équations ne contraignent pas le barycentre, sa valeur restera automatiquement à l'*a priori*. En revanche, lorsque les équations contraignent le barycentre, celui-ci est estimé en ajustant les données. Il est inutile de contraindre le barycentre par des équations supplémentaires comme cela est fait par Waldhauser et Ellsworth [52] (cf Annexe 3).
- (2) Ajouter des séismes supplémentaires situés loin de l'ensemble de séismes trop regroupés.

La figure (2.2.10) montre que si on ajoute un troisième séisme éloigné, le conditionnement de la matrice des dérivées partielles devient acceptable. Nous pouvons alors calculer tous les paramètres hypocentaux des 3 séismes, alors qu'il n'est pas toujours possible de déterminer les positions absolues des séismes 1 et 2 sans utiliser le troisième comme le montre la courbe rouge.

2.2.2.5. *Influence du modèle de vitesse.* Une autre source d'erreur possible sur les localisations provient de la méconnaissance du modèle de vitesse. Les relations entre hypocentres et loi de distribution des vitesses sont complexes. Nous allons essayer d'illustrer certains aspects des corrélations existant entre ces paramètres (corrélations "sources-structure") en utilisant des cas synthétiques. L'étude de la loi gaussienne tangente permet de représenter les erreurs par un ellipsoïde dans l'espace des paramètres. La décomposition en valeurs propres et vecteurs propres de l'inverse de la matrice variance-covariance indique les directions principales de cet ellipsoïde, dont les vecteurs propres associées aux valeurs propres les plus faibles sont les axes les plus allongés. La figure (2.2.11) montre un exemple de spectre de valeurs propres et la base propre dans un cas synthétique de localisation à l'aide des décalages temporels. Le test numérique consiste à générer des décalages temporels entre deux séismes dans un milieu homogène, et à calculer les valeurs propres et vecteurs

FIG. 2.2.8. Condition d'échec de la localisation utilisant les décalages temporels.

Les conditions sont les mêmes qu'en (2.2.7), avec une taille de cluster diminuée. On ne retrouve plus les positions absolues des évènements. La distance inter-évènements étant petite devant la distance au réseau, les vecteurs d'ondes sont quasi-parallèles. Ainsi on ne retrouve que la position relative des séismes. Celle-ci est retrouvée en 2 itérations mais il est impossible de déterminer les positions absolues.

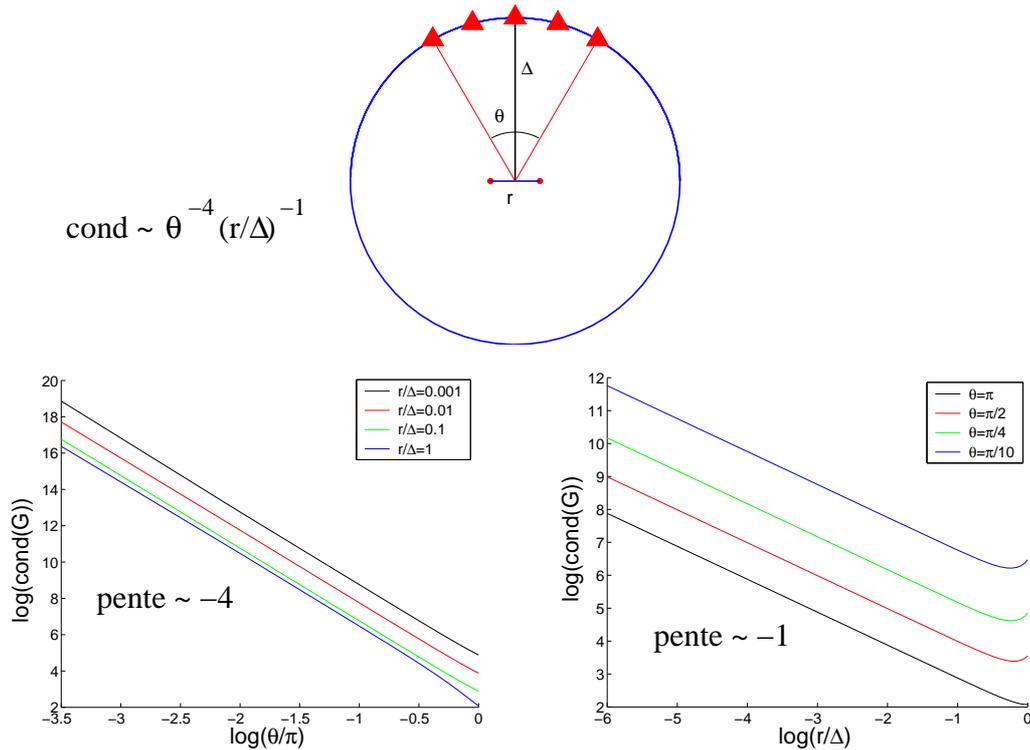


propres issues de la loi gaussienne tangente. Quand les séismes sont loin l'un de l'autre le conditionnement de l'ordre de 1000 représente le rapport entre le grand axe et le petit axe de l'ellipsoïde. On peut lire ces axes sur la matrice V dont les colonnes sont les vecteurs propres. L'ellipsoïde s'allonge au fur et à mesure que les séismes se rapprochent l'un de l'autre. En effet le conditionnement est de l'ordre de 1 000 000 quand le rapport $\frac{r}{\delta}$ est de 0.1 et plus pour $\frac{r}{\delta} = 0.01$. On remarque alors que le 5ième et 6ième vecteur propre correspond aux coordonnées du barycentre des évènements avec des valeurs propres associées de plus en plus basses. Cette quantité est d'autant plus difficile à calculer que la distance inter-évènement est faible. Le dernier vecteur propre fait intervenir la distance entre les séismes, et la vitesse du milieu. On peut l'interpréter comme une corrélation (un "trade-off") entre ces deux quantités. Une erreur positive (resp. négative) sur la vitesse induit une distance plus (resp. moins) importante entre les évènements.

FIG. 2.2.9. Conditionnement en fonction de la taille du cluster.

Dans un milieu homogène, 5 stations équidistantes sont réparties sur un cercle de rayon Δ , qui représente la distance hypocentrale. Deux séismes sont situés à une distance r l'un de l'autre, leur barycentre correspondant au centre du cercle. Notons θ l'angle sous lequel les stations sont vues du centre du cercle. Les deux figures du bas montrent, en échelle logarithmique, le conditionnement en fonction de θ et de r . On en déduit le comportement du conditionnement,

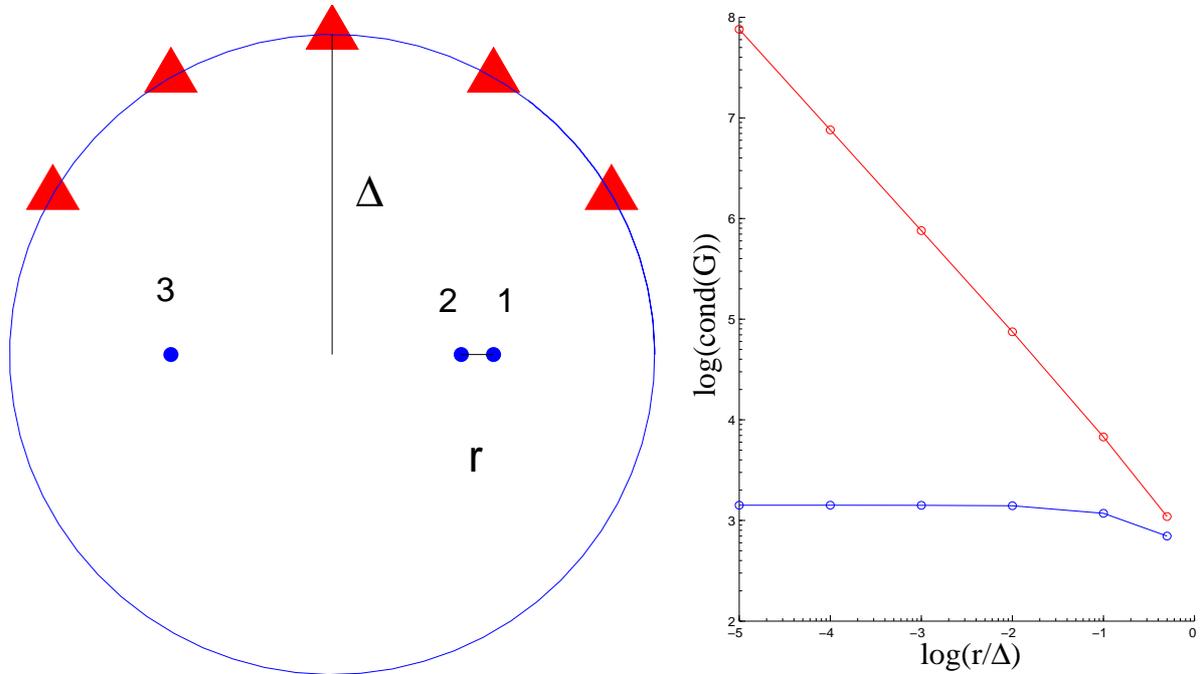
$$\text{cond}(G) \sim \theta^{-4} \left(\frac{r}{\Delta}\right)^{-1}.$$



Une première expérience (figure (2.2.12)) a été menée dans un modèle de vitesse homogène, l'erreur de modélisation étant elle-même homogène. Avec la méthode de localisation à l'aide des temps de première arrivée, l'erreur est essentiellement reportée sur la composante verticale du barycentre des événements, alors que leur distance est peu affectée ; une erreur de vitesse positive applique une translation du barycentre vers le haut, alors qu'une vitesse trop faible place les séismes trop profondément. Le fait que l'erreur de localisation absolue soit surtout verticale est dû à la distribution superficielle des récepteurs. Avec la méthode de localisation à l'aide des décalages temporels (figure (2.2.12)), une erreur homogène sur la vitesse est simplement compensée par l'ajustement de la distance entre les sources. Cette erreur est fonction de l'erreur de modélisation et colinéaire au vecteur position relative. Par contre on constate que le barycentre des séismes n'est pas affecté par cette erreur de modélisation homogène, ce qui s'explique par le fait que les vecteurs d'ondes ne sont pas déviés par l'erreur de modélisation commise. Le barycentre des séismes localisés avec une

FIG. 2.2.10. Conditionnement avec 3 séismes.

La figure de gauche montre la géométrie du problème de localisation. La figure de droite montre le conditionnement de la matrice des dérivées partielles G . La courbe bleue correspond au problème de localisation des 3 séismes et la courbe rouge correspond au problème de localisation des séismes 1 et 2.



méthode utilisant les décalages temporels est donc laissé invariant par une erreur sur la vitesse moyenne du modèle. Une comparaison de l'amplitude de l'erreur sur les hypocentres avec les deux méthodes montre qu'elle est plus faible dans le cas de la localisation à l'aide des décalages temporels que dans le cas de la localisation à l'aide des temps de première arrivées, ce qui tendrait à faire penser intuitivement que les localisations calculées à partir des décalages temporels sont moins affectées par les erreurs de modélisation du milieu que la méthode de localisation utilisant les temps de première arrivées.

Cependant l'erreur de modélisation ne se limite pas à une erreur homogène sur tout le modèle. Pour étudier l'influence d'une erreur de modélisation non homogène sur les deux méthodes de localisation, nous avons choisi d'appliquer différentes perturbations gaussiennes de moyennes, écart-types et amplitudes variés (Figures (2.2.14)-(2.2.15)) à un modèle de référence à gradient. La figure (2.2.13) montre la façon dont nous décomposerons l'erreur de localisation. On distingue une erreur sur le barycentre, une rotation moyenne de la structure géométrique de l'ensemble des séismes et enfin une erreur sur les positions relatives ("erreur relative"). Cette erreur relative est la variance de l'ensemble de la structure par rapport aux localisations exactes, une fois ôtées les contributions dues à l'erreur barycentrique et à la rotation moyenne. Les figures (2.2.14)-(2.2.15) montrent la décomposition de l'erreur de localisation en fonction d'une erreur de modèle de vitesse. Notons

FIG. 2.2.11. Vecteurs propres et spectre.

Le problème considéré est le même que pour la figure (2.2.9) en ajoutant la vitesse comme paramètre. La matrice des dérivées partielles G est décomposée en valeurs singulières,

$$G = USV^t.$$

Les figures à gauche montrent le spectre des valeurs singulières ; à droite la matrice des vecteurs propres V est représentée dans la base des paramètres $(v, x_1, y_1, x_2, y_2, \delta)$.

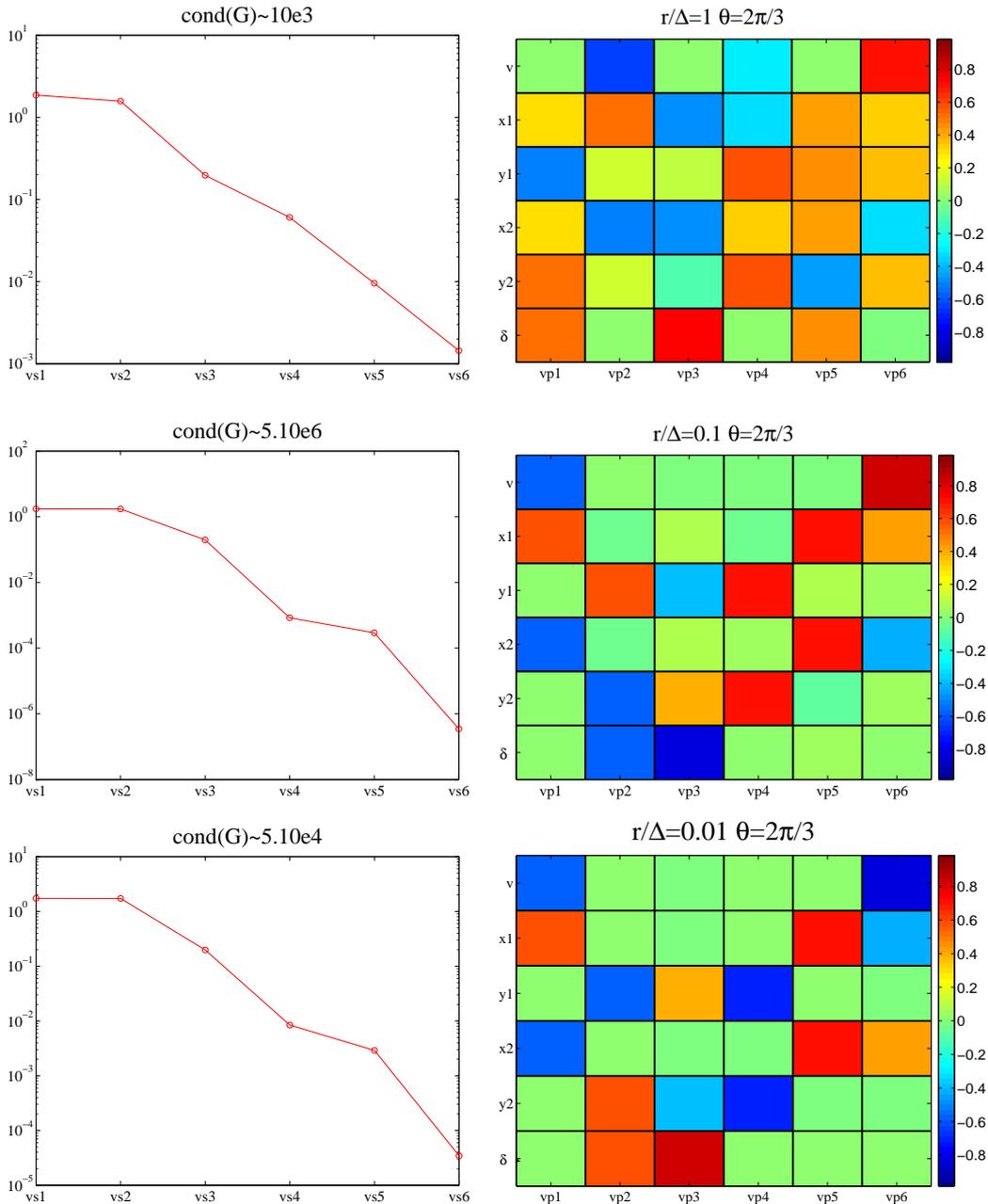
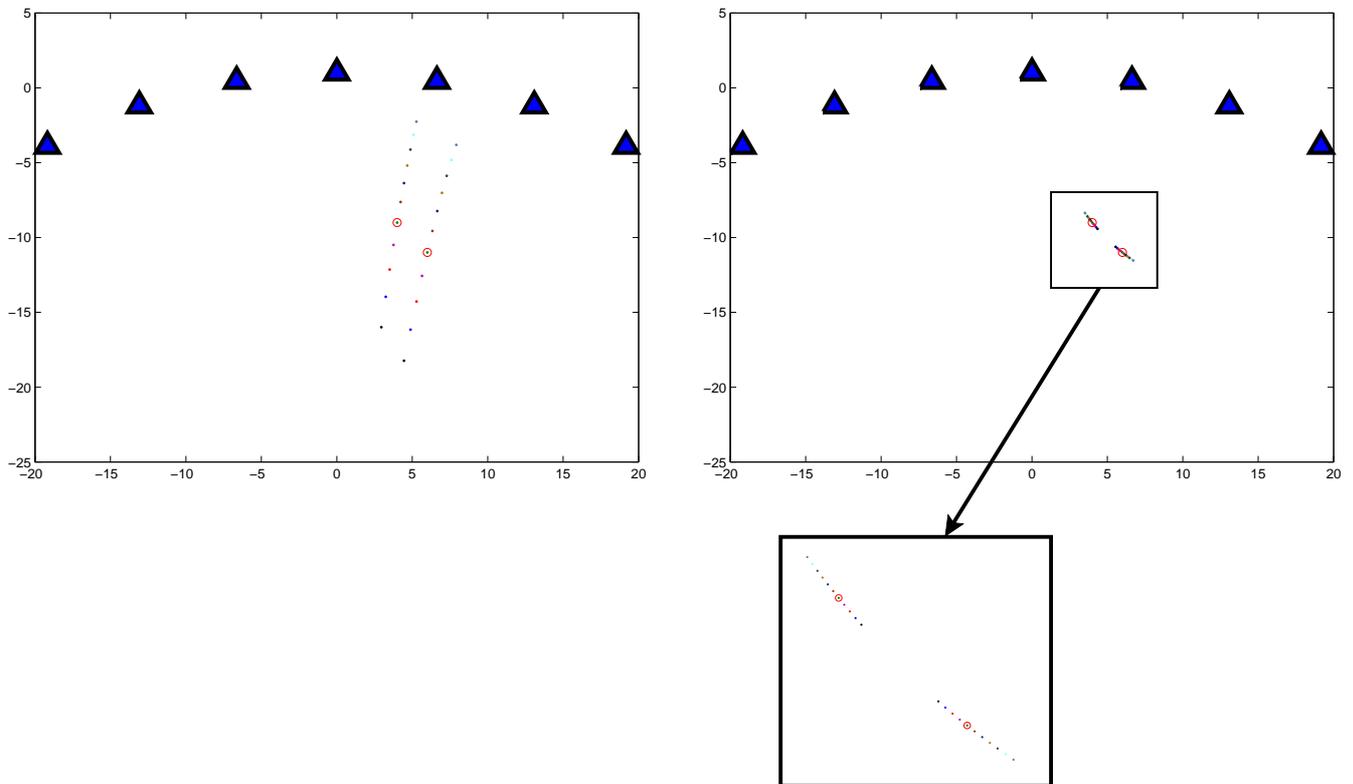


FIG. 2.2.12. Erreur due au modèle de vitesse.

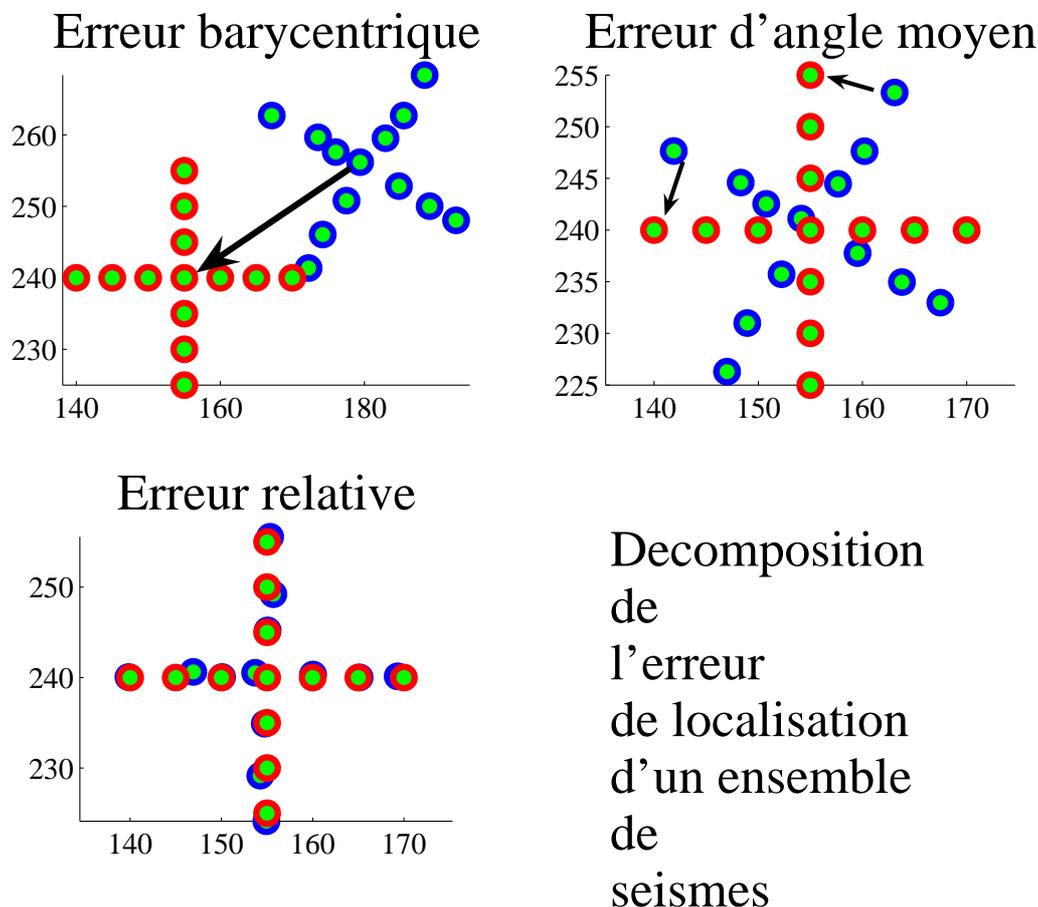
Des données synthétiques (temps de première arrivée et décalages temporels) sont générés dans un milieu homogène de vitesse 5 km/s, les positions vraies des séismes étant représentées par les cercles rouges, et les stations (triangles bleus) ayant une disposition simulant une distribution superficielle. Des inversions de temps de première arrivée (figure de gauche) et des décalages temporels (figures de droite) sont effectuées en considérant des modèles de vitesse faux (de 3 km/s à 8 Km/s avec un pas de 0.5 km/s, l'erreur étant homogène). Les résultats des inversions sont indiqués par un couple de points de même couleur. Notons que les deux méthodes convergent vers la solution exacte avec le modèle de vitesse vrai. L'effet de l'erreur de vitesse est différent selon la méthode utilisée. Pour la localisation à partir des temps de première arrivée, si l'erreur sur la vitesse est positive (resp. négative), les séismes sont localisés plus haut (resp. plus bas) que les positions vraies, la distance entre les deux séismes varie peu. On constate que l'erreur sur le modèle de vitesse provoque essentiellement une erreur sur la composante verticale de la position, pour une distribution superficielle de stations. Pour la localisation utilisant les décalages temporels, on constate que le barycentre est peu affecté par une erreur sur la vitesse moyenne, par contre la distance inter-événements l'est. Si l'erreur sur la vitesse est positive (resp. négative), la distance entre les positions estimées est plus grande (resp. petite) que la distance entre les positions vraies.



d'abord que dans tous les cas, les deux algorithmes d'inversion convergent en un nombre fini d'itérations et que le conditionnement des matrices des dérivées partielles reste dans des plages de valeurs permettant de résoudre correctement les systèmes linéaires en question,

FIG. 2.2.13. **Décompositon de l'erreur de localisation.**

L'erreur de localisation se décompose en trois transformations géométriques : une translation, une rotation et un ajustement résiduel correspondant à l'erreur relative.



les erreurs sur les solutions calculées ne sont donc pas dues à des artefacts numériques. Nous remarquons en premier lieu que les deux méthodes compensent l'erreur de vitesse par une erreur de localisation. Lorsque le modèle de vitesse est exact nous retrouvons les mêmes solutions dans tous les cas : la solution exacte. Si on porte notre attention sur les courbes figurant la rms, on remarque un minimum qui correspond à la connaissance exacte du modèle. Lorsque l'erreur sur le modèle est non nulle, la valeur de la rms est plus importante. Cela veut dire qu'en jouant uniquement sur les hypocentres, on ne peut pas expliquer totalement le résidu. Il reste potentiellement dans ce résidu de l'information portant sur le modèle de vitesse. La méthode de localisation à l'aide des temps de première arrivée perturbe le barycentre quand les erreurs sur le modèle de vitesse sont à grande longueur d'onde alors que la méthode utilisant les décalages temporels est sensible aux

FIG. 2.2.14. Erreur Gaussienne sur le modèle de vitesse : effet de l'amplitude de l'erreur.

Dans un modèle de référence à gradient, on calcule des données synthétiques (temps de première arrivées et décalages temporels) pour les séismes figurés en bleu. On inverse ces données dans un modèle faux, qui est le modèle de référence perturbé par une gaussienne de rayon caractéristique 2 km. Les résultats de l'inversion sont en noir pour la localisation utilisant les temps de première arrivées et en rouge pour la localisation utilisant les décalages temporels. Le résultat est présenté pour une perturbation d'amplitude de 500 m/s (figures de gauche). On teste la sensibilité des localisations en fonction de l'amplitude et la position de l'erreur gaussienne sur le modèle de vitesse utilisé. Les résultats sont reportés sur les différentes courbes.

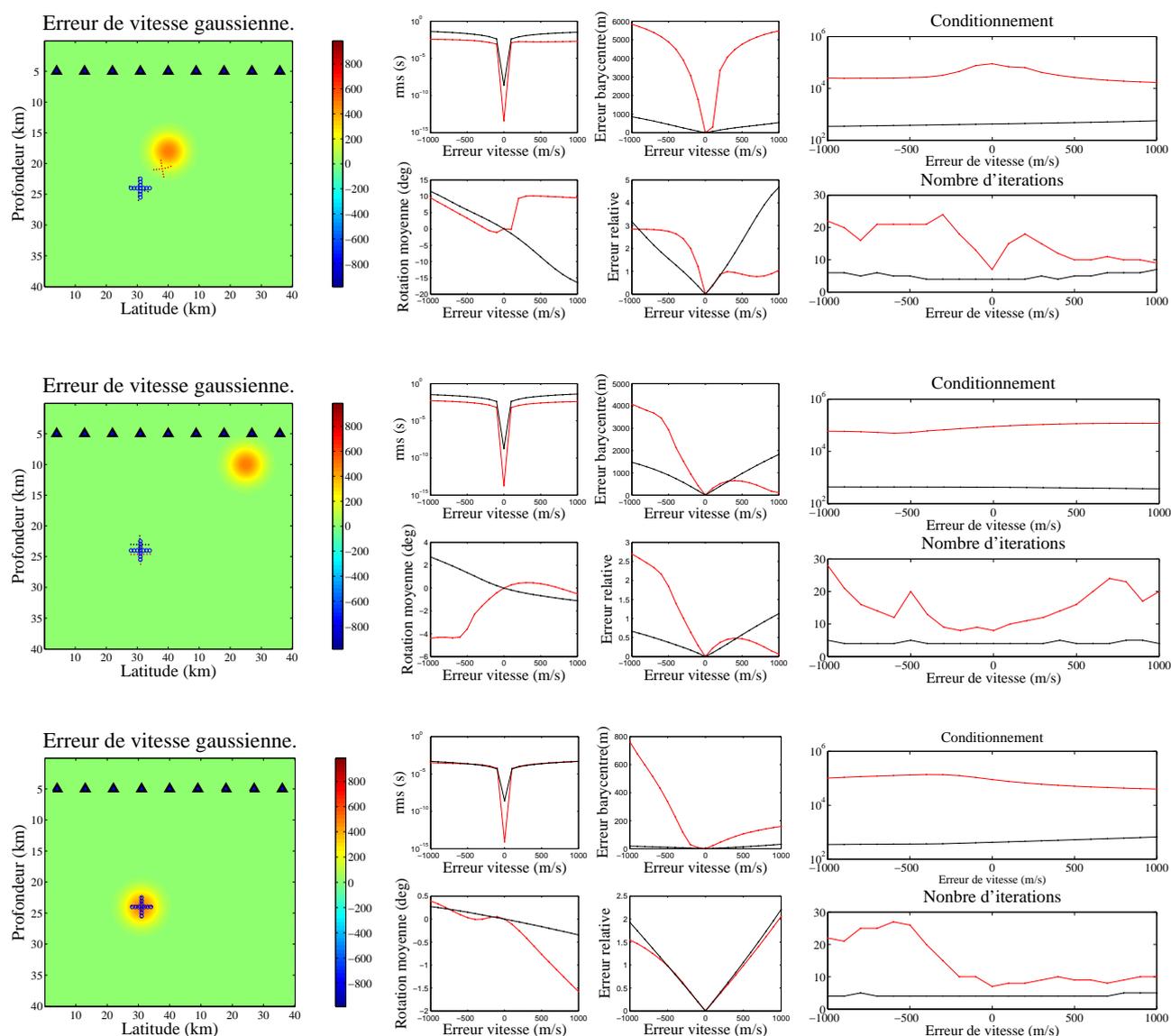
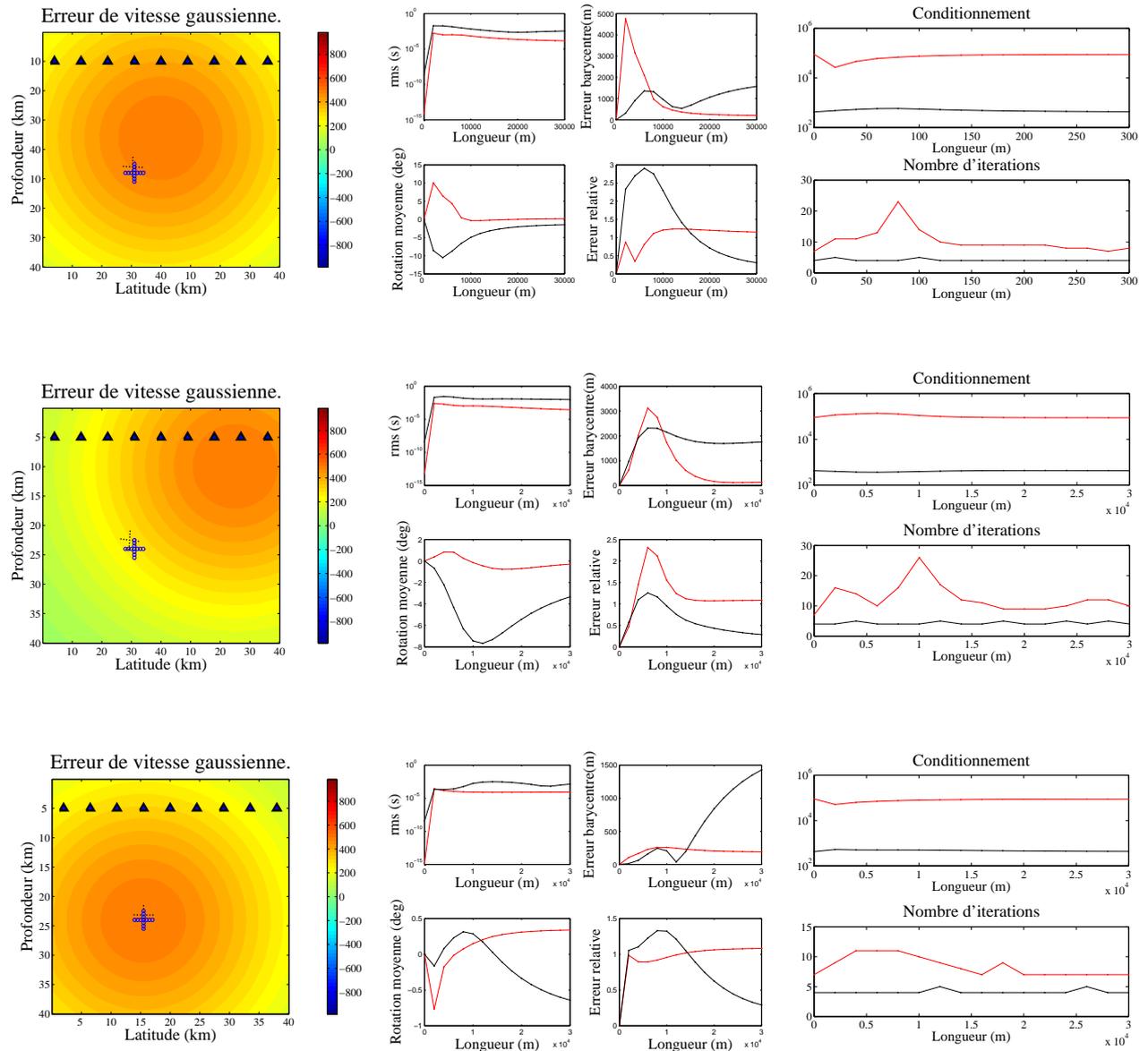


FIG. 2.2.15. Erreur Gaussienne sur le modèle de vitesse : effet de la taille caractéristique de la zone mal modélisée.

Expérience similaire à (2.2.14) ; l'amplitude de l'erreur gaussienne est fixée 500 m/s et le diamètre du disque mal modélisé varie. On reporte les différentes erreurs en fonction de la longueur caractéristique du disque mal modélisé.



longueurs d'ondes plus faibles (2.2.15). Même une perturbation située loin des sources perturbe fortement le barycentre dans cette dernière méthode : faire des différences de temps de propagation entre deux rais n'élimine donc pas nécessairement les erreurs du modèle de vitesse comme on aurait pu le penser intuitivement. Une perturbation du modèle de vitesse

dévie en effet la trajectoire des rais et induit une rotation de l'ensemble de la structure relocalisée, modifiant son orientation, ce qui peut entraîner des conséquences importantes pour les interprétations mécaniques. Elle modifie également le lieu où les rais convergent, c'est-à-dire le barycentre. La méthode de localisation utilisant les décalages temporels est sensible aux variations du champ de vecteur \mathbf{k} , et la structure relocalisée est donc sensible aux erreurs du modèle de vitesse ; cette observation a déjà été faite par Michelini et Lomax [28] dans le cas d'un modèle stratifié horizontal, comportant une erreur de modélisation dans une couche : on observe alors une focalisation plus ou moins forte de la structure relocalisée. Il est donc nécessaire de disposer d'un modèle de vitesse correct si l'on veut pouvoir interpréter les résultats de localisation à partir de décalages temporels. On peut penser qu'un modèle tomographique peut permettre de progresser dans cette direction. On peut également se poser la question de l'inversion simultanée des paramètres hypocentaux et des paramètres de vitesse avec des données de décalages temporels.

2.3. Tomographies.

Dans ce paragraphe nous allons considérer, en plus des séismes, le modèle de vitesse comme inconnu. La valeur de la vitesse \tilde{v} des ondes P sur les noeuds de contrôle du modèle discrétisé de la grille \mathbf{m}_i , sont les paramètres supplémentaire considérés. Comme pour les méthodes de localisations, nous définirons deux méthodes tomographiques, l'une utilisant comme données des temps de première arrivée et l'autre des décalages temporels. D'un point de vue pratique, ces deux méthodes sont proche l'une de l'autre. En effet, dans un algorithme de type Gauss-Newton, le schéma (**gn**) appliqué à la *tomographie des décalages temporels* produit une matrice directement issue (par combinaison linéaire) de la matrice générée lors de la méthode de *tomographie des temps de première arrivée*. Cette remarque nous permet de mener conjointement l'étude des deux méthodes dans la partie suivante.

2.3.0.6. *Tomographie en doubles différences*. Un premier travail concernant l'apport des décalages temporels dans un processus tomographique a été mené par Zhang et Thurber [55]. Ils ont proposé une solution des équations en *doubles différences*, pour la tomographie (cf Annexe 3). Cela consiste à ajuster simultanément des temps de première arrivée et des décalages temporels au sens des moindres carrés. Les auteurs mélangent les données absolues et relatives dans un processus unique, il est de ce fait difficile d'apprécier la contribution de chaque type de données. Ainsi nous avons choisi de conduire deux processus distincts. En premier lieu une *tomographie des temps de première arrivée* nous fournit un premier modèle, qui ensuite sera utilisé comme *a priori* dans une *tomographie des décalages temporels*. Nous pourrons ainsi aisément comparer les deux solutions et mettre en évidence l'apport des décalages temporels dans une méthode tomographique. La suite de cette section décrit les éléments utilisées dans les schémas numérique ((**gn**) ou (**se**)) appliqués au problème de la *tomographie des temps de première arrivée* et de la *tomographie des décalages temporels*.

2.3.1. Matrice Jacobienne dans le cas Gaussien. Pour calculer les dérivées partielles par rapport aux paramètres du modèle de lenteur s , considérons le temps de propagation le long d'un rai (d'après Tarantola [46]),

$$(2.3.1) \quad t(s) = \int_{\gamma(s)} s d\sigma,$$

et le temps de propagation dans un modèle perturbé,

$$(2.3.2) \quad t(s + \delta_s) = \int_{\gamma(s+\delta_s)} (s + \delta_s) d\sigma.$$

Le théorème de Fermat permet d'écrire (Principe de stationnarité du rai γ),

$$(2.3.3) \quad \int_{\gamma(s+\delta_s)} (s + \delta_s) d\sigma = \int_{\gamma(s)} (s + \delta_s) d\sigma + o(\|\delta_s\|^2),$$

ainsi

$$(2.3.4) \quad \int_{\gamma(s+\delta_s)} (s + \delta_s) d\sigma = \int_{\gamma(s)} s d\sigma + \int_{\gamma(s)} \delta_s d\sigma + o(\|\delta_s\|^2),$$

autrement dit,

$$(2.3.5) \quad t(s + \delta_s) - t(s) = \int_{\gamma(s)} \delta_s d\sigma + o(\|\delta_s\|^2).$$

Appliquons ce résultat au modèle de lenteur discrétisé, $\tilde{s}(\mathbf{x}) = \sum_{\mathbf{k}} \lambda_{\mathbf{k}} \varphi_{\mathbf{k}}(\mathbf{x})$, en prenant comme perturbation $\delta_s = \lambda \varphi_{\mathbf{i}}$,

$$(2.3.6) \quad t(\lambda_1, \dots, \lambda_{\mathbf{i}} + \lambda, \dots, \lambda_{\mathbf{n}}) - t(\lambda_1, \dots, \lambda_{\mathbf{i}}, \dots, \lambda_{\mathbf{n}}) = \lambda \int_{\gamma} \varphi_{\mathbf{i}} d\sigma + o(\lambda^2),$$

on en déduit aisément,

$$(2.3.7) \quad \frac{\partial t}{\partial \lambda_{\mathbf{i}}} = \int_{\gamma} \varphi_{\mathbf{i}} d\sigma.$$

Nous pouvons également exprimer cette dérivée par rapport à la vitesse; posons $\nu_{\mathbf{i}} = \frac{1}{\lambda_{\mathbf{i}}}$, (2.3.7) se réécrit

$$(2.3.8) \quad \frac{\partial t}{\partial \nu_{\mathbf{i}}} = -\frac{1}{\nu_{\mathbf{i}}^2} \int_{\gamma} \varphi_{\mathbf{i}} d\sigma.$$

La fonction $\varphi_{\mathbf{i}}$ ayant pour support tous les pavés du maillage contenant le noeud i , pour calculer (2.3.7) il suffit d'intégrer numériquement le long du trajet du rai inclus dans les (au plus 8) pavés concernés. De plus l'expression numérique de la fonction de base change avec les pavés, nous intégrerons sur chacun des pavés avant de faire la somme totale. Pour cela nous repérons les intersections du rai avec les faces des différents pavés par une méthode de dichotomie. Concrètement nous utilisons l'algorithme suivant :

- Le rai est stocké comme une suite de points $(\gamma_k)_{k=1, n_r}$. La valeur de l'intégrale, notée S , est initialisée à zéro.
- **boucle** $k = 1, \dots, n_r - 1$,
- Faire $X = \gamma_k, Y = \gamma_{k+1}$,
- **tant que** $\text{TEST}[X, Y] = \text{vrai}$, faire :
 - $I = \text{DICHOTOMIE}[X, Y]$,
 - $S = S + \text{INTEGRER}[X, I]$,
 - Faire $X = I$
- **fin tant que**
- $S = S + \text{INTEGRER}[X, Y]$
- **fin boucle** ,

en notant $\text{TEST}[X, Y]$ la fonction qui renvoie la variable booléenne "vrai" lorsque le segment $[X, Y]$ traverse la frontière d'au moins une cellule. $\text{DICHOTOMIE}[X, Y]$ fournit le premier point d'intersection du segment $[X, Y]$ avec la première frontière de pavé rencontrée. La fonction $\text{INTEGRER}[X, Y]$ calcule l'intégrale de la fonction de base dans le pavé concerné.

Une fois calculé l'ensemble des dérivées partielles des temps par rapport à chacun des paramètres du modèle, il est trivial de calculer la matrice jacobienne pour le problème de *tomographie des décalages temporels* en utilisant la formule (2.3.8) ; en effet pour un décalage d entre deux rais γ_1 et γ_2 , il suffit de faire une simple différence :

$$(2.3.9) \quad \frac{\partial d}{\partial v_i} = -\frac{1}{v_i^2} \int_{\gamma_2} \varphi_i d\sigma + \frac{1}{v_i^2} \int_{\gamma_1} \varphi_i d\sigma.$$

Chaque ligne de la matrice jacobienne \mathbf{G} est calculée à partir d'un rai pour la *tomographie des temps de première arrivée* et de deux rais pour la *tomographie des décalages temporels*. Les éléments non nuls de chaque ligne correspondent aux paramètres interceptés par le (ou les) rai(s). Or, le rai étant une courbe traversant un milieu à trois dimensions, nous pouvons estimer que le nombre de cellules rencontré par cette courbe est inférieur à $c^{te} n_{max}$, où c^{te} est une constante et $n_{max} = \max(n_x, n_y, n_z)$. Si on note n_d le nombre de données alors la matrice jacobienne \mathbf{G} possède au plus $c^{te} n_{max} n_d$ éléments non nuls sur un nombre total de $n_d \times n_x \times n_y \times n_z$ éléments. Or, en pratique dans les applications tomographiques, nous avons observé que $c^{te} n_{max} \ll n_x \times n_y \times n_z$, autrement dit \mathbf{G} est très creuse. Par ailleurs, la matrice $\mathbf{G}^t \mathbf{G}$ possède $(n_x \times n_y \times n_z)^2$ éléments et n'est pas creuse. De plus, dans les applications tomographiques la taille de la grille d'échantillonnage est telle que : $n_d < n_x \times n_y \times n_z$ (problème sous-déterminé). Ainsi, le nombre d'éléments non nul de \mathbf{G} est bien inférieur aux nombre d'éléments de $\mathbf{G}^t \mathbf{G}$. Il est donc préférable dans les applications tomographiques d'utiliser un schéma numérique utilisant la matrice \mathbf{G} sans avoir à calculer et stocker la matrice $\mathbf{G}^t \mathbf{G}$. C'est ainsi que nous avons choisi d'implémenter les schémas (**gn**) et (**se**). La capacité de stockage ainsi gagné permet d'utiliser un plus grand nombre de données dans le processus d'inversion. De plus, nous avons choisi l'algorithme LSQR qui est particulièrement adapté pour gérer des matrices creuses et de grande taille.

2.3.2. information a priori Gaussienne. Nous supposons que l'information *a priori* sur le modèle de vitesse est représentée par une variable aléatoire Gaussienne de

moyenne \mathbf{m}_0 et de matrice de variance-covariance \mathbf{C}_m . Les schémas numériques ((**se**) et (**gn**)) donnés dans la partie 1.3 utilisent explicitement la matrice $\mathbf{C}_m^{-\frac{1}{2}}$. Cette matrice est facilement calculable en supposant qu'il n'y a pas de corrélations entre les différents paramètres de vitesses, dans ce cas elle est diagonale. Dès que l'on suppose des corrélations entre les paramètres, le calcul de $\mathbf{C}_m^{-\frac{1}{2}}$ n'est plus trivial, on peut par exemple utiliser la méthode de Lanczos pour diagonaliser \mathbf{C}_m et calculer $\mathbf{C}_m^{-\frac{1}{2}}$. Cette méthode demande beaucoup de ressources informatiques et n'est pas directement applicable dans un processus tomographique en raison de la dimension très importante de l'espace des modèles (de l'ordre de 10^6 paramètres pour une tomographie à l'échelle régionale de l'île de Hawaii). Ainsi nous sommes amenés à proposer une méthode d'approximation de $\mathbf{C}_m^{-\frac{1}{2}}$ efficace pour les méthodes tomographiques, utilisant aussi bien les temps de première arrivée que les décalages temporels.

2.3.2.1. *Calcul de $\mathbf{C}_m^{-\frac{1}{2}}$* . Lors du processus d'inversion du modèle de vitesse⁵, nous utiliserons une matrice de variance-covariance *a priori* sur le modèle comme décrite en [46],

$$\mathbf{C}_m = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ & \ddots & \sigma_{i,j} \\ \vdots & \sigma_{i,i} & \vdots \\ & \sigma_{j,i} & \ddots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{pmatrix},$$

avec

$$(2.3.10) \quad \sigma_{i,j} = \sigma(v_i) \sigma(v_j) \exp\left(-\sqrt{\langle \mathbf{C}^{-1}(\mathbf{X}_i - \mathbf{X}_j), \mathbf{X}_i - \mathbf{X}_j \rangle}\right),$$

où $\sigma(v_i)$ est l'écart-type *a priori* sur le paramètre de vitesse v_i et \mathbf{X}_i est le noeud correspondant au paramètre v_i ; pour simplifier nous utiliserons, par la suite, une valeur unique d'écart-type $\sigma_v = \sigma(v_i)$, pour tout i . La matrice \mathbf{C} apparaissant dans (2.3.10) sera prise diagonale,

$$(2.3.11) \quad \mathbf{C} = \begin{pmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & \lambda_z \end{pmatrix},$$

les quantités λ_x , λ_y et λ_z étant les longueurs de corrélation entre paramètres du modèle dans les directions x , y et z du modèle. La matrice \mathbf{C}_m ainsi définie dépend :

- des longueurs de corrélation : λ_x , λ_y et λ_z ,
- de la valeur de l'écart-type *a priori* sur le modèle de vitesse : σ_v ,
- de la taille de la grille de discrétisation : $n = n_x \times n_y \times n_z$.

Dans les schémas numériques d'inversion utilisés dans cette thèse -écrits dans l'espace modèle-, c'est la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ qui apparait explicitement et non pas \mathbf{C}_m . En raison de

⁵Le même raisonnement pourrait être appliqué au modèle de lenteur.

la dimension trop importante de l'espace modèle⁶, le calcul de $\mathbf{C}_m^{-\frac{1}{2}}$ avec les méthodes numériques classiques est impossible. Nous sommes alors dans l'obligation de proposer une méthode d'approximation (cf Annexe 6), numériquement exploitable, afin d'approcher $\mathbf{C}_m^{-\frac{1}{2}}$ par une matrice noté $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$.

La matrice \mathbf{C}_m est diagonalisable :

$$(2.3.12) \quad \mathbf{C}_m = \mathbf{U}\mathbf{S}\mathbf{U}^t,$$

ainsi

$$(2.3.13) \quad \mathbf{C}_m^{-\frac{1}{2}} = \mathbf{U}\mathbf{S}^{-\frac{1}{2}}\mathbf{U}^t.$$

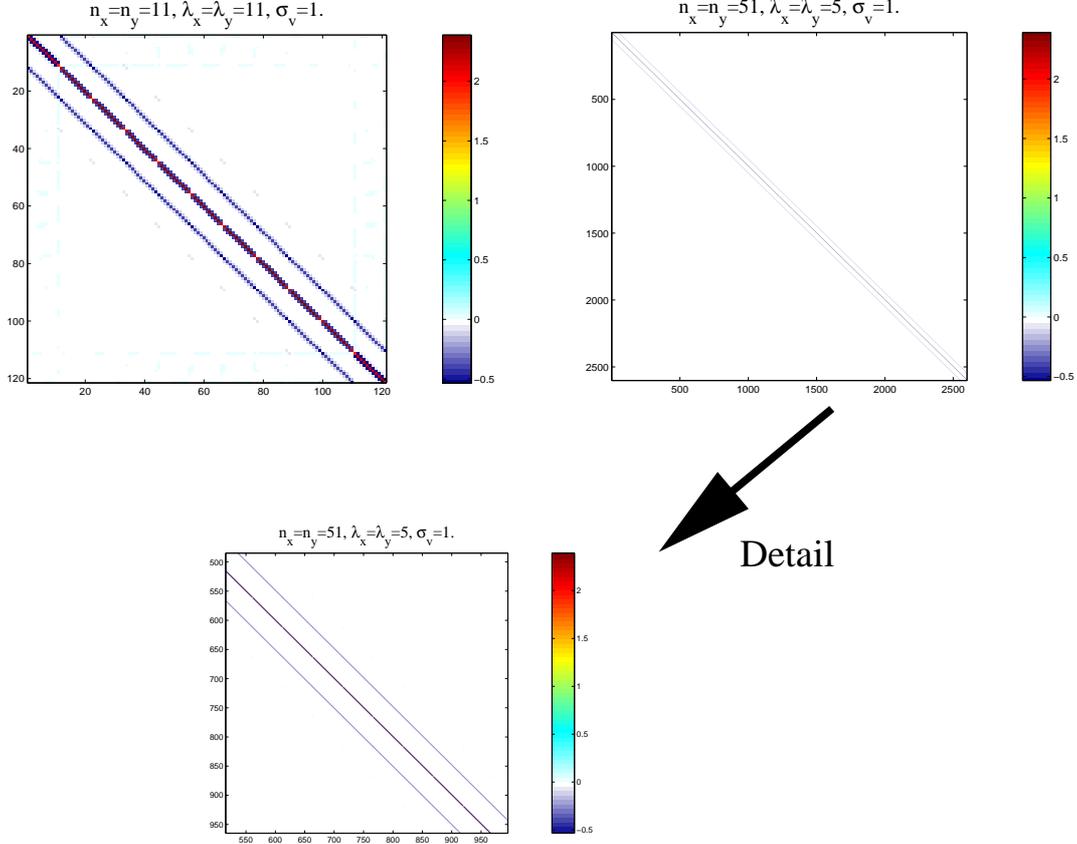
Il est possible de calculer $\mathbf{C}_m^{-\frac{1}{2}}$ dans le cas de modèles à deux dimensions pour des valeurs de n_x et n_y raisonnables (de l'ordre de la centaine). La figure (2.3.1) montre la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ lorsque \mathbf{C} est trois fois l'identité et lorsque σ_v est égal à un. En faisant varier la taille de n_x et n_y , on s'aperçoit que la structure de $\mathbf{C}_m^{-\frac{1}{2}}$ d'une part, que la valeur des coefficients d'autre part, ne "*dépendent pas*" de n_x et n_y , au sens où la différence entre les coefficients reste de l'ordre du bruit numérique lorsque les dimensions du modèle varient. En effet, on peut diviser n_x et n_y par cinq sans changer la valeur des coefficients⁷. La figure (2.3.2) montre, pour n_x et n_y fixés, différentes matrices obtenues en faisant varier les coefficients λ_x et λ_y de la diagonale de la matrice \mathbf{C} . On remarque que seules les valeurs des coefficients de $\mathbf{C}_m^{-\frac{1}{2}}$ changent avec les longueurs de corrélation, alors que la structure de la matrice reste la même. En extrapolant ces propriétés dans le cas de modèles à trois dimensions, on peut proposer une méthode pour calculer une approximation de $\mathbf{C}_m^{-\frac{1}{2}}$ en vue des applications tomographiques réelles. On définit d'abord une grille auxiliaire plus lâche, de taille $n_{0,x} \times n_{0,y} \times n_{0,z}$; la matrice $\mathbf{C}_{0,m}^{-\frac{1}{2}}$ est calculée dans cette grille, et ses coefficients sont redistribués dans une matrice $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$, à taille réelle, en suivant la structure particulière de $\mathbf{C}_{0,m}^{-\frac{1}{2}}$. Il est difficile de donner à trois dimensions des exemples similaires au cas à deux dimensions, car la taille des grilles devient très vite trop importante pour être traitée informatiquement. Dans les applications tomographiques de cette thèse, les grilles des modèles de vitesse ne dépasseront jamais 200x200x50 noeuds. On peut alors penser calculer les coefficients de $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$ en divisant par dix chacune des dimensions de la grille. Le calcul, via la méthode de Lanczos, d'une telle matrice sur une grille de 20x20x5 ne prend que quelques minutes. Il ne reste plus qu'à placer les coefficients obtenus dans la matrice $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$ cherchée. Pour plus de précautions il est possible de vérifier, a posteriori, la validité

⁶La matrice \mathbf{C}_m est de dimension n^2 .

⁷Nous avons effectué le même calcul avec $n_x=101$ et $n_y=101$, le résultat reste identique, nous pouvons diviser par dix sans affecter l'approximation $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$.

FIG. 2.3.1. Matrice de covariance selon les noeuds.

Un exemple de $\mathbf{C}_m^{-\frac{1}{2}}$ calculée avec des valeurs de nombre de noeuds de discrétisation différents. La structure et les coefficients de la matrice sont invariants.



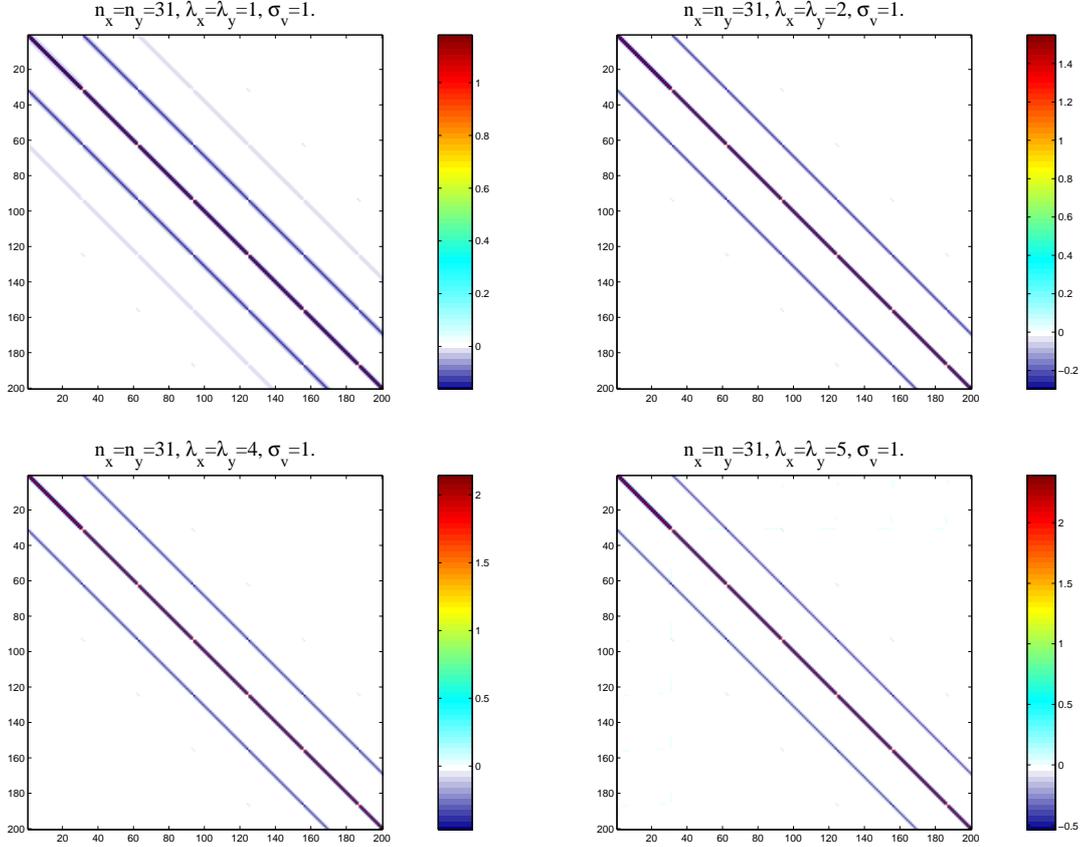
numérique de l'approximation en utilisant la majoration,

$$(2.3.14) \quad \frac{\|\mathbf{C}_m^{-1} - \widetilde{\mathbf{C}}_m^{-1}\|}{\|\mathbf{C}_m^{-1}\|} \leq \|\mathbf{I} - \mathbf{C}_m \widetilde{\mathbf{C}}_m^{-1}\|$$

en notant $\widetilde{\mathbf{C}}_m^{-1}$ le carré de l'approximation de $\mathbf{C}_m^{-\frac{1}{2}}$. Si, par exemple, on utilisait la norme L_∞ sur l'espace des matrices, il faudrait calculer la norme L_1 de chaque ligne de $\mathbf{I} - \mathbf{C}_m \widetilde{\mathbf{C}}_m^{-1}$ et prendre le maximum des valeurs obtenues pour estimer l'erreur relative. En pratique on se bornera à calculer la norme L_1 de certaines lignes pour conclure que l'erreur relative est acceptable ou pas. Si l'approximation n'est pas satisfaisante, il faudra augmenter la dimension de l'espace auxiliaire de calcul et refaire le test et ainsi de suite, jusqu'à l'obtention d'une approximation satisfaisante. Dans la pratique, il n'est pas nécessaire de recalculer la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ à chaque nouvelle inversion : nous pouvons faire un calcul pour des longueurs de corrélation données et stocker le résultat dans un fichier réutilisable pour chaque calcul

FIG. 2.3.2. Matrice de covariance selon les longueurs de corrélations.

Un exemple de $\mathbf{C}_m^{-\frac{1}{2}}$ calculée avec des valeurs de longueurs de corrélations différentes. La structure reste la même, seules les valeurs des coefficients changent.



tomographique, ce qui accélère encore le processus. Nous constituons ainsi une bibliothèque de matrice $\mathbf{C}_m^{-\frac{1}{2}}$ classées en fonction des valeurs $\lambda_x, \lambda_y, \lambda_z$, toutes calculées avec $\sigma_v = 1$. Pour pouvoir faire varier l'écart-type σ_v sur les vitesses, il suffira de diviser chaque coefficient de $\mathbf{C}_m^{-\frac{1}{2}}$ par σ_v . Remarquons enfin que la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ est creuse, son utilisation ne consomme pas beaucoup de ressources informatiques. Ainsi, cette matrice est bien adapté à la mise en oeuvre pratique des schémas (gn) et (se).

2.3.3. Quelques tests synthétiques à 2D. La convergence de l'inversion tomographique va être contrôlée par la quantité d'information *a priori*, et notamment par la longueur de corrélation. Nous présentons ici quelques tests synthétiques afin de comprendre le rôle du modèle *a priori* dans l'inversion et ainsi appréhender une manière de choisir les paramètres les plus adaptés lors d'inversions de données réelles.

2.3.3.1. *Effet de la longueur de corrélation.* Le premier test est effectué dans un cas idéal. Sources et stations sont choisies de façon à obtenir une couverture optimale du milieu

en rais. Nous testons donc les capacités de l'algorithme à retrouver les paramètres dans un cas très bien contraint, et nous faisons varier certains paramètres pour comprendre leur influence sur le résultat. La figure (2.3.3) montre le milieu synthétique, modèle à deux dimensions discrétisé sur une grille 101×101 noeuds, avec une taille de maille de 200m, dans lequel nous calculons un jeu de temps de première arrivée synthétiques. Une douzaine d'inversions tomographiques sont effectuées en faisant varier la longueur de corrélation λ pour l'essentiel par pas de 200m. La figure (2.3.3) montre la variation du coût en fonction de λ : à partir d'une valeur, optimale, de λ le coût est minimal. Le modèle reconstitué pour cette valeur est proche du modèle vrai. **On remarque que cette façon de procéder permet de choisir un modèle suivant un critère objectif, et apporte une réponse au problème classique du choix d'un modèle solution en tomographie.** Les autres paramètres de l'inversion ; les écart-types sur les positions, les temps origine et les vitesses sont fixés à :

$$(2.3.15) \quad \begin{cases} \sigma_v = 100 \text{ km/s} \\ \sigma_p = 100 \text{ km} \\ \sigma_t = 100 \text{ s} \end{cases} .$$

Ces valeurs sont au delà de la variation physiquement admissible des paramètres par rapport au modèle *a priori*. Cela indique que les données contraignent bien les paramètres et qu'il n'est pas nécessaire d'apporter une quantité d'information *a priori* importante pour stabiliser l'inversion. Cela provient de la répartition du réseau et des séismes qui permet de contraindre parfaitement tous les paramètres. En effet, le réseau situé sur les quatre cotés du modèle permet de contraindre l'hypocentre de chaque séisme. Les sources réparties dans tout le milieu et la densité des récepteurs permettent de couvrir en rais, de façon optimale, la zone investiguée.

Une autre tomographie (figure (2.3.4)) est effectuée sur un grille plus fine de 401×401 mailles de 50m de côté. Les résultats montrent que le sur-échantillonnage de la fonction ne gêne en aucun cas le processus d'inversion pourvu que l'on utilise la longueur de corrélation adéquate. On reporte donc le problème du choix du maillage à celui de la longueur de corrélation, la longueur de corrélation en unités physiques étant déterminée par les données. Il ne faut cependant pas utiliser un maillage trop fin, à l'instar des méthodes aux différences finies : si les noeuds sont trop proches, des artefacts numériques (dus à la précision limitée de la machine) pourraient apparaître. La quantité d'information doit également rester traitable en machine. Les données ont donc un rôle important dans le choix du maillage, le pouvoir de résolution étant limité par la densité et de la divergence des rais. Il est donc inutile de sur-échantillonner le modèle de vitesse au point que les données ne contraignent pas le modèle, même si l'utilisation d'une longueur de corrélation adaptée permet la convergence de l'inversion et une estimation correcte, puisque ce sur-échantillonnage n'apportera aucune information supplémentaire sur le milieu tout en consommant plus de ressources informatiques. A contrario, il ne faut surtout pas sous-échantillonner, dans ce cas on n'exploite pas toute l'information contenue dans les données et cela affecte l'image tomographique. Pour résumer il faut choisir la grille la plus fine que l'on puisse traiter et ajuster la longueur de corrélation. La figure (2.3.4) montre qu'une longueur de corrélation trop faible ne

permet pas de collecter l'information cohérente utile. Notons également qu'une longueur de corrélation "trop grande" (au-delà de l'optimum) permet de bien retrouver le modèle notamment de retrouver des variations de faible longueur d'onde pourvu que les données contraignent suffisamment le modèle : la quantité d'information *a priori*, nécessaire pour éventuellement forcer la convergence pour certains paramètres, est toujours faible par rapport à la quantité d'information contenue dans les données. Il faut donc privilégier des grilles fines et rechercher la longueur de corrélation optimale par rapport aux grilles lâches et longueurs de corrélation faibles.

2.3.3.2. *Reconstruction d'une anomalie gaussienne.* La figure (2.3.5) montre un milieu synthétique construit à l'aide d'une perturbation gaussienne de 10 % d'un modèle homogène de 5 km/s. Les sources sont situées en profondeur et les stations en surface. Le modèle de vitesse *a priori* est homogène à 5 km/s. Les sources *a priori* sont générées aléatoirement en prenant un écart-type de 1 km par rapport aux sources exactes. Une localisation des sources est effectuée dans le modèle de vitesse *a priori*, la solution est biaisée par l'erreur de vitesse. Les solutions sont très proches avec les deux méthodes, utilisant soit les temps de première arrivée, soit les décalages temporels. Nous avons vu précédemment que les deux méthodes intègrent de manière différente les erreurs de vitesse sur les hypocentres. Ici l'erreur de vitesse est trop faible pour que cette différence soit significative. On effectue ensuite une tomographie : les modèles sont reconstitués presque identiquement avec les deux méthodes. On ne retrouve pas exactement l'anomalie gaussienne car les rais ne couvrent pas de façon homogène le milieu, ce qui engendre des déformations de l'image tomographique par rapport à l'image vraie. Enfin la dernière expérience consiste à retrouver simultanément les hypocentres et l'anomalie gaussienne : les deux méthodes présentent également des résultats très proches. Nous pouvons conclure, qu'en théorie et avec des données exactes les deux méthodes produisent des résultats identiques lorsque le modèle *a priori* est proche de la solution, c'est-à-dire lorsque l'on doit retrouver une perturbation du modèle autour de 10 %.

2.4. Conclusion.

Dans ce chapitre, on s'est intéressé à un problème inverse particulier : celui de la tomographie sismique - comprenant le problème de la localisation d'un séisme. Ce problème est résolu à partir de deux types de données différentes : les temps de première arrivées \mathbf{t}^{obs} , et les décalages temporels \mathbf{d}^{obs} . Ces deux types de données donnent lieu à des algorithmes d'inversion tomographique très semblables, avec une paramétrisation légèrement différente. Nous avons mis en oeuvre un calcul robuste et précis des temps de parcours théoriques $\mathbf{t}(\mathbf{m})$ et des décalages temporels théoriques $\mathbf{d}(\mathbf{m})$; $\mathbf{d}(\mathbf{m})$ peut se déduire de $\mathbf{t}(\mathbf{m})$ par simple différence, c'est-à-dire par combinaison linéaire. Cette particularité est exploitée pour définir un algorithme unique capable de traiter aussi bien les temps de première arrivée que les décalages temporels. Nous avons ensuite choisi le schéma numérique de Gauss-Newton ((**gn**) ou (**se**)) pour calculer un minimum de la fonction coût,

$$(2.4.1) \quad E(\mathbf{m}) = \frac{1}{2} (\mathbf{m} - \mathbf{m}_0)^t \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \frac{1}{2} (\mathbf{r}(\mathbf{m}))^t \mathbf{C}_D^{-1} (\mathbf{r}(\mathbf{m})),$$

FIG. 2.3.3. Choix de la longueur de corrélation.

Une inversion tomographique des temps de première arrivée est effectuée en utilisant le modèle a priori figuré en haut à droite. Les paramètres de l'inversion sont :

$$\begin{cases} \sigma_v = 100 \text{ km/s}, \\ \sigma_p = 100 \text{ km}, \\ \sigma_t = 100 \text{ s}. \end{cases}$$

Le choix de la longueur de corrélation λ se fait en effectuant des inversions successives pour différents λ . On reporte ensuite le coût en fonction de λ comme on le voit sur la figure en bas à gauche. La valeur du coût se stabilise sur un minimum à partir de $\lambda=0.8 \text{ km}$. Le résultat correspondant est en bas à droite. Comme attendu, le modèle estimé est très proche du modèle exact compte tenu de la répartition optimale des rais.

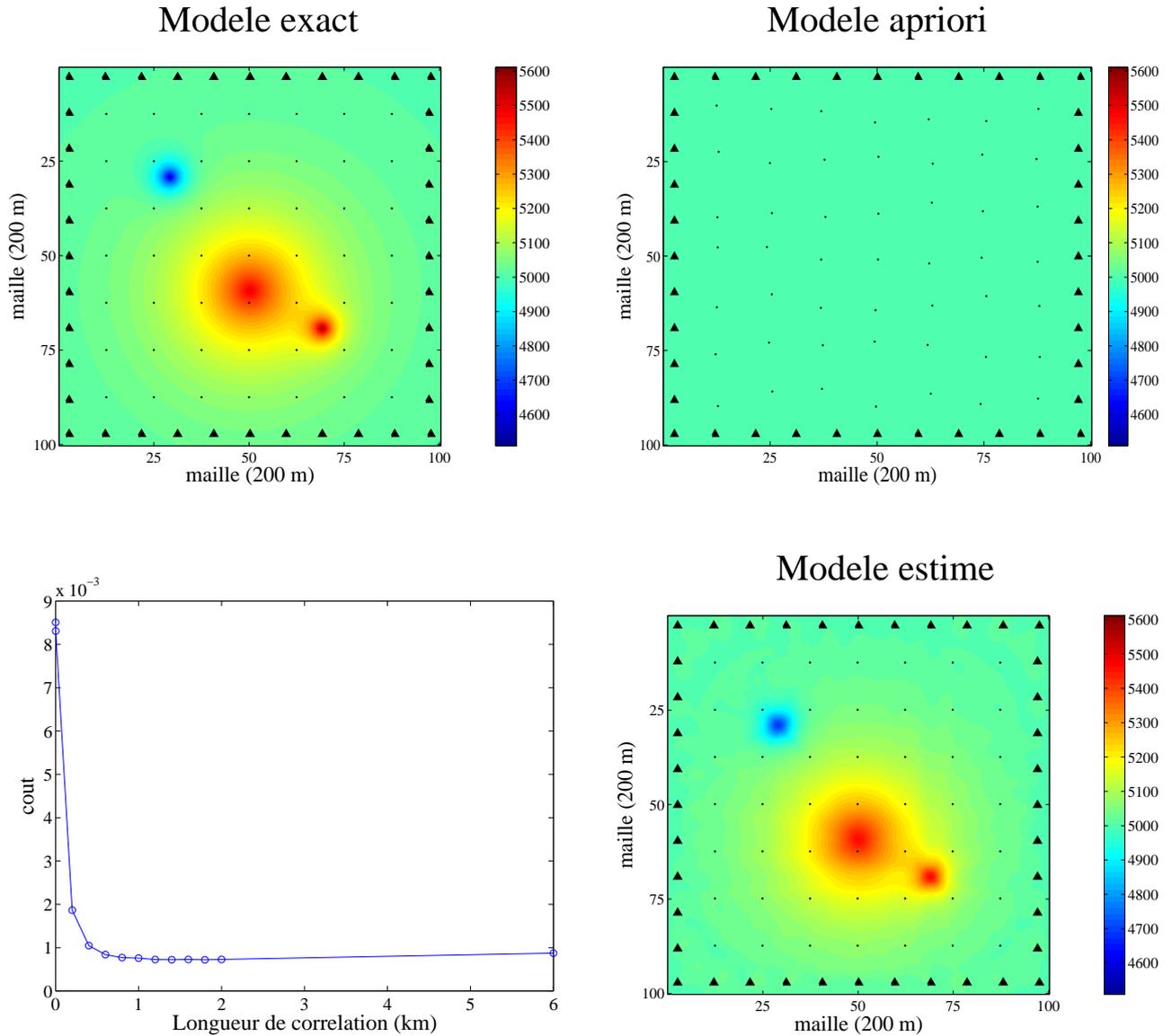
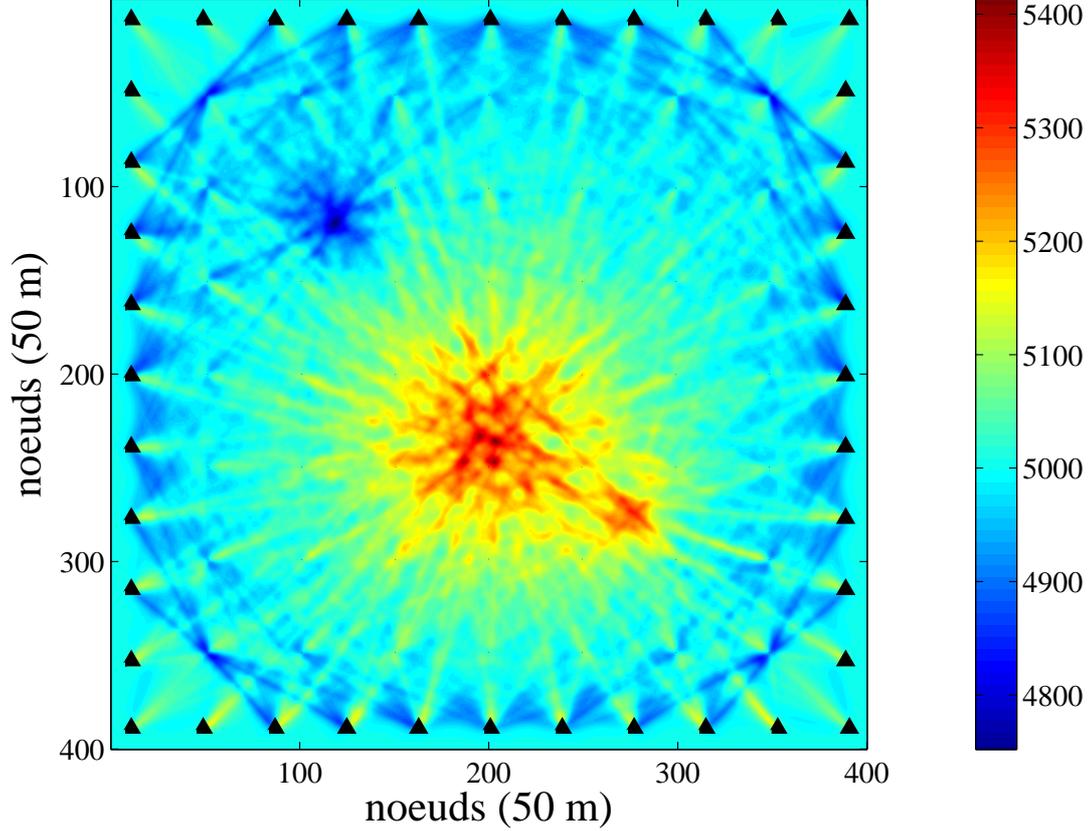


FIG. 2.3.4. Longueur de corrélation insuffisante

Une inversion des temps de première arrivée synthétiques calculés dans le modèle de la figure (2.3.3) avec une longueur de corrélation de 100 m. On voit des “tubes de rais” apparaître, qui sont des indices d’une longueur de corrélation insuffisante : le modèle est reconstruit le long des rais, les noeuds non échantillonnés par les rais ne sont pas perturbés.

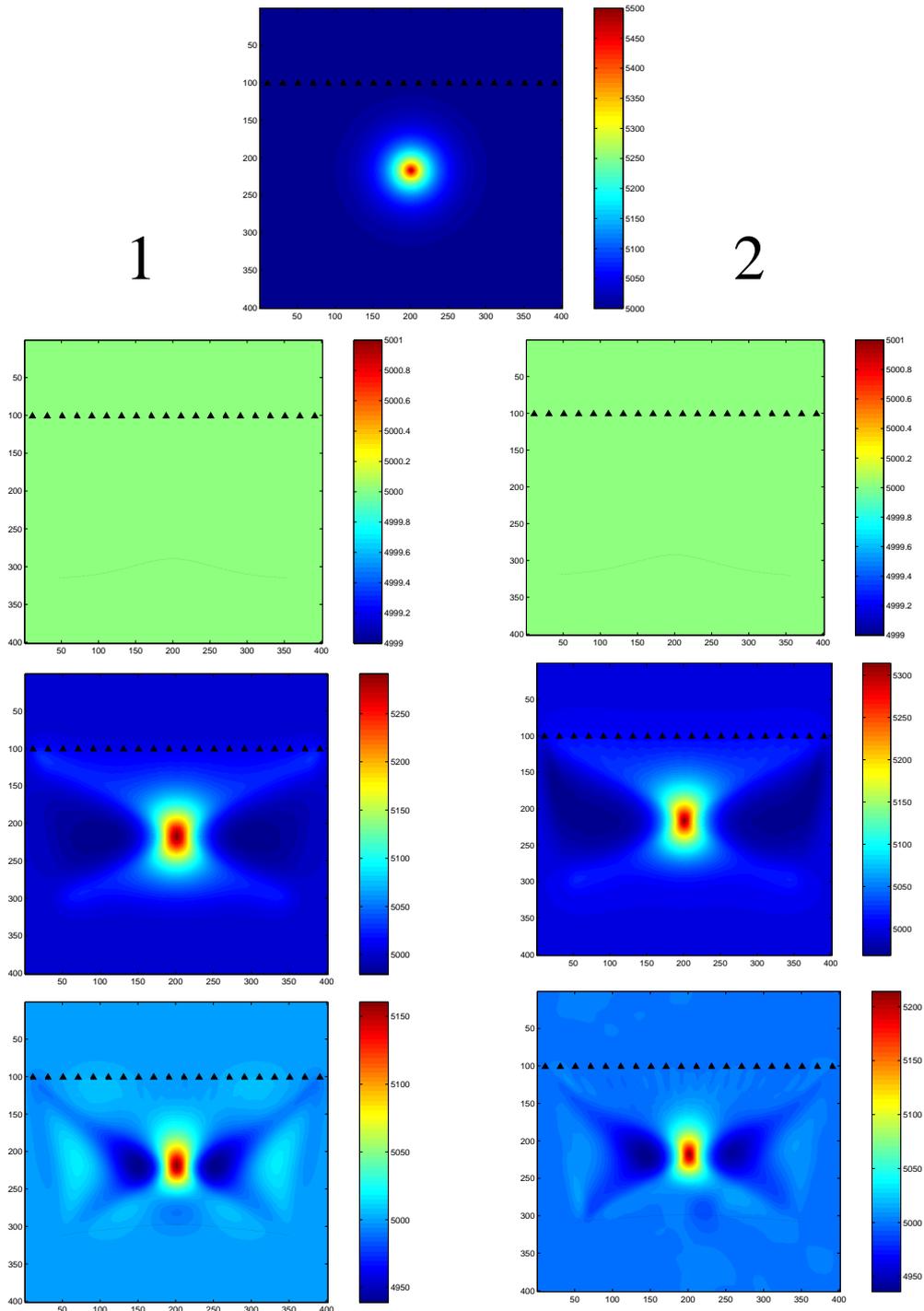


en notant $\mathbf{r}(\mathbf{m})$ le résidu, avec $\mathbf{r}(\mathbf{m}) = \mathbf{t}^{obs} - \mathbf{t}(\mathbf{m})$ pour le problème des temps de première arrivée et $\mathbf{r}(\mathbf{m}) = \mathbf{d}^{obs} - \mathbf{d}(\mathbf{m})$ pour les décalages temporels. L’algorithme d’optimisation de la fonction E débute par le choix de l’itéré initial qui, par simplicité, est pris égal au modèle *a priori*. Les étapes importantes de la méthode sont :

- (1) modèle *a priori* \mathbf{m}_0 donné, faire $\mathbf{m} = \mathbf{m}_0$.
- (2) Calculer les temps de propagation théoriques $\mathbf{t}(\mathbf{m})$, la matrice jacobienne \mathbf{G} et les résidus $\mathbf{r}(\mathbf{m}) = \mathbf{t}^{obs} - \mathbf{t}(\mathbf{m})$.
- (3) Dans le cas de décalages temporels, former les différences à partir de \mathbf{G} et $\mathbf{t}(\mathbf{m})$ pour obtenir la matrice jacobienne encore notée \mathbf{G} et les décalages théoriques $\mathbf{d}(\mathbf{m})$, former ensuite les résidus $\mathbf{r}(\mathbf{m}) = \mathbf{d}^{obs} - \mathbf{d}(\mathbf{m})$.

FIG. 2.3.5. Perturbation gaussienne.

La figure du haut montre le milieu de référence dans lequel sont calculées des données synthétiques ; des temps de première arrivées et des décalages temporels. Les deux colonnes montrent des résultats d'inversion (1) en utilisant les temps et (2) en utilisant les décalages. Le modèle de vitesse a priori est choisi homogène à 5 km/s. Les positions des sources a priori sont les sources exactes perturbées aléatoirement par une loi uniforme. Plusieurs expériences sont effectuées. En premier lieu, nous localisons dans le modèle de vitesse a priori (choisi homogène à 5 km/s), ensuite nous inversons le modèle de vitesse avec les sources exactes, enfin nous effectuons une tomographie complète. Dans les deux cas les résultats sont très proches, en théorie, les deux méthodes semblent produire les mêmes résultats.



(4) Résoudre au sens des moindres carrés,

$$(2.4.2) \quad \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{G} \\ \mathbf{C}_m^{-\frac{1}{2}} \end{bmatrix} \delta \sim \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{r}(\mathbf{m}) \\ \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m}_0 - \mathbf{m}) \end{bmatrix}.$$

(5) Faire $\mathbf{m} = \mathbf{m} + \delta$, si \mathbf{m} est solution alors stop, sinon retourner en 2 et suivre les étapes 2, 3, 4 et 5.

Les deux méthodes diffèrent seulement au niveau du point 3. Les autres étapes sont toujours suivies quelles que soient les données considérées. La réalisation de cet algorithme suppose notamment la connaissance de la matrice $\mathbf{C}_m^{-\frac{1}{2}}$. Lorsque chaque élément de \mathbf{C}_m est représenté par une fonction de corrélation, d'un point de vue pratique il est difficile de calculer exactement $\mathbf{C}_m^{-\frac{1}{2}}$. Ainsi, une approximation de celle-ci a été proposée dans ce chapitre. Cette approximation rend possible l'utilisation du schéma numérique présenté au chapitre précédent, ainsi qu'une démarche d'optimisation, au cours de laquelle la longueur de corrélation et l'amplitude optimale de la fonction de corrélation sont déterminées. Cette démarche permet le choix d'un modèle tomographique unique et stable, de norme minimale et ajustant les données, en suivant des critères objectifs et reproductibles. Elle permet d'obtenir un modèle dont les fluctuations sont contraintes par les données.

CHAPITRE 3

Tomographie de l'île de Hawaii.

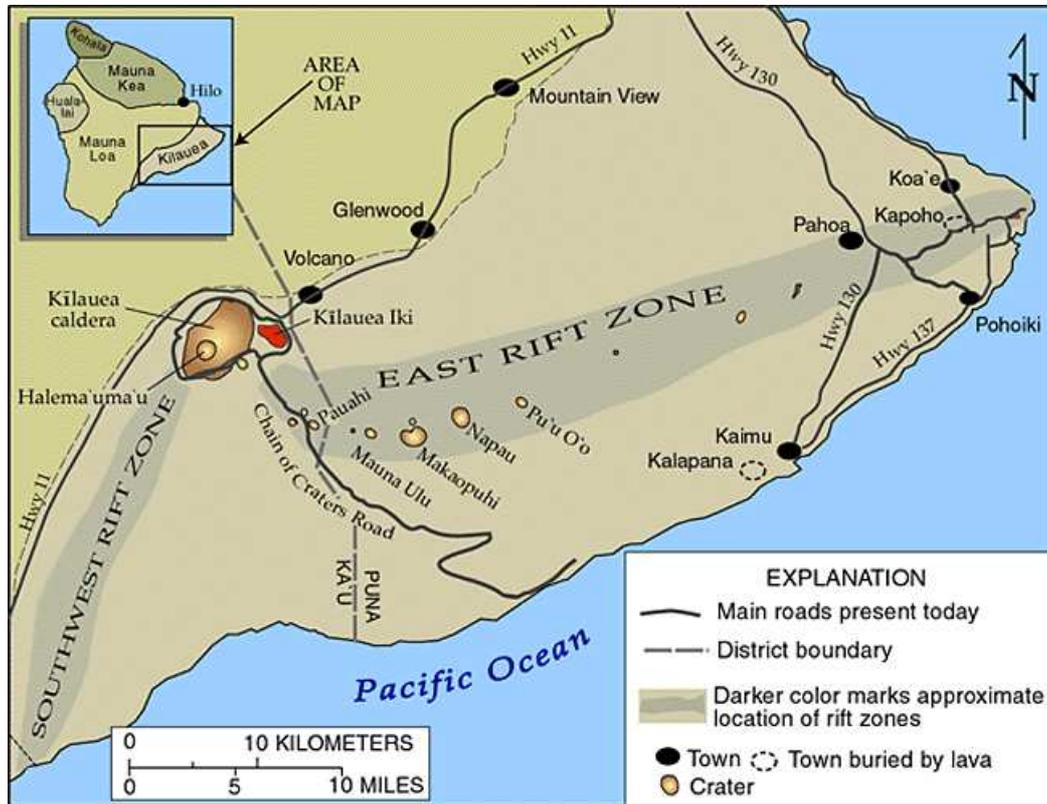


L'archipel Hawaïien est situé dans le Pacifique nord. Différentes îles le composent dont l'île de Hawaii, la plus grande. Elle est formée de plusieurs volcans, notamment au sud le Kilauea qui a la particularité d'être en éruption permanente depuis 1983 (Figure(3.0.1)). Depuis les années soixante-dix, il existe un réseau permanent maintenu par le Hawaiian Volcano Observatory (HVO) qui enregistre la sismicité abondante générée par l'intrusion de magma en force dans les volcans de l'île. Ainsi, ces conditions favorables constituent un terrain d'expérience idéal pour la mise au point d'une nouvelle méthode. En effet, l'abondance et la diversité des enregistrements sismiques donnent accès à des données de bonne qualité; d'autre part cette zone a été très étudiée par de nombreux auteurs, ce qui procure des points de repère essentiels pour l'interprétation des résultats de tomographie.

Afin d'imager la structure de vitesse des ondes P de l'île de Hawaii, nous disposons d'un catalogue de données enregistrées par le réseau permanent du HVO. La sismicité très dense du flanc Sud du Kilauea a permis d'identifier plusieurs milliers de séismes suffisamment semblables pour pouvoir en déduire un très grand nombre de décalages temporels par méthode interspectrale. La précision de ces données est de l'ordre de cinq millièmes de seconde, alors que la précision des temps de première arrivée pointés manuellement est de l'ordre de cinq centièmes de seconde. Peut-on mettre à profit cette précision supérieure

FIG. 3.0.1. L'île de Hawaii, zone d'étude.

La sismicité abondante dans le flanc sud du Kilauea et dans la zone du rift Est permet d'identifier des essaims de séismes semblables qui procurent un jeu de données de grande qualité pour effectuer une tomographie des décalages temporels.



pour obtenir une image tomographique de plus grande qualité ? Peut-on espérer une avancée comparable au domaine des localisations de séismes ? Pour apporter des éléments de réponse, nous allons procéder en plusieurs étapes. La première étape débute par le choix du modèle de vitesse *a priori*. Il s'agit ensuite de choisir des séismes répartis au mieux sur l'île ("déclustérisés") pour effectuer une tomographie des temps de première arrivée. Nous obtiendrons un modèle tomographique régional "classique", à l'échelle de l'île de Hawaii, qui servira de référence. La deuxième étape fait enfin intervenir les décalages temporels. Nous sélectionnerons des essaims de séismes dans le flanc Sud du Kilauea afin d'effectuer une tomographie des décalages temporels dans cette région (échelle "locale"). Le modèle *a priori* sera constitué par la restriction du modèle tomographique calculé lors de la première étape au volume local d'inversion. En comparant le modèle *a priori*, celui obtenu par tomographie "classique" des temps de première d'arrivée dans le volume local et le modèle inversé à partir des décalages temporels, nous aurons des éléments de réponses aux questions posées.

3.1. Tomographie à l'échelle régionale.

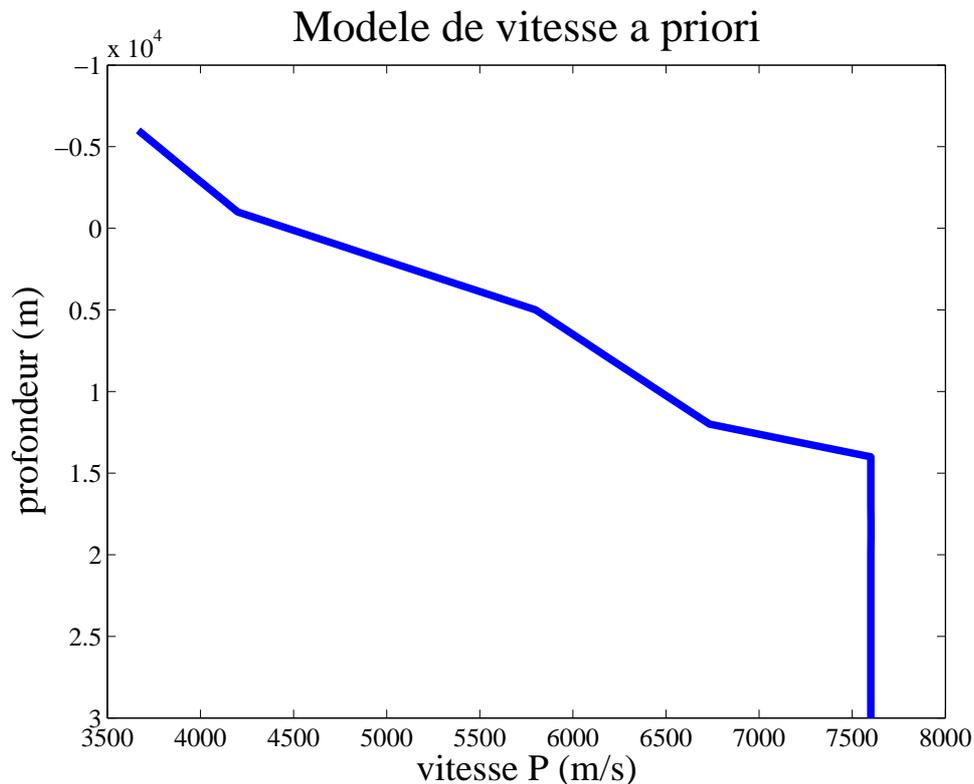
Nous nous intéressons ici à la structure de l'île de Hawaii afin de déterminer un modèle *a priori* pour l'étude tomographique ultérieure du flanc Sud du Kilauea.

3.1.1. Connaissance *a priori*. Le point de départ de la méthode d'inversion consiste à collecter et à modéliser l'information *a priori*.

3.1.1.1. *Modèle a priori.* De nombreux travaux ont été menés sur l'île de Hawaii, ce qui constitue un avantage lors du choix du modèle *a priori*. Kissling [24] a proposé une structure stratifiée du modèle de vitesse à partir d'une inversion de temps d'arrivée d'ondes sismiques. Ce type de structure 1D a également été étudié par Eaton [11], Ryall and Bennett [42], Hill [19], Ward et Gregersen [53], Crosson et Koyanagi [9], Klein [23]. Des modèles à gradient ont été proposés par Klein [23], Okubo et coll. [35] et Benz et coll. [5]. Ce type de modèle (à gradient) semble bien adapté pour fournir un modèle *a priori* convenable sur un volcan. Nous avons choisi d'utiliser le modèle proposé par Okubo et coll. [35] et Benz et coll. [5]. (Figure (3.1.1)). Quant aux localisations *a priori*, elles seront issues d'un premier calcul d'inversion dans le modèle de vitesse choisi comme *a priori*.

FIG. 3.1.1. **Modèle de vitesse *a priori* pour la tomographie à l'échelle régionale.**

Le modèle *a priori* est un modèle à gradient suivant la profondeur, il été choisi d'après les travaux de Okubo et coll. (1997) et Benz et coll. (2003).



3.1.1.2. *Données.* Les données sont issues de ~ 30000 microséismes enregistrés de 1988 à 1999 par l'USGS - Hawaiian Volcano Observatory (HVO) à l'aide d'un réseau permanent de 50 stations. En utilisant les localisations préliminaires de ce catalogue, nous avons choisi 1358 séismes (Figure (3.1.2)) en veillant à ce qu'ils couvrent au mieux la zone d'étude. Les séismes choisis ont été enregistrés par au moins 30 stations, le jeu de données ainsi constitué comporte 41886 pointés de première arrivée d'ondes P de très bonne qualité. Néanmoins, pour un petit nombre de pointés, l'opérateur peut confondre la phase des ondes P avec les ondes S. Cela risque de survenir lorsque la station se trouve dans un plan nodal pour les ondes P. Dans ce cas, sur le sismogramme, il est difficile de reconnaître la première arrivée des ondes P et on pointe par erreur la phase S qui arrive après les ondes P. Ainsi le jeu de données constitué peut contenir des pointés aberrants. Il est alors souhaitable d'utiliser la loi Sech pour inverser de telles données. Nous utiliserons donc cette loi sur les données issues de pointés manuels avec une erreur caractéristique σ que l'on estime à 5 dixièmes de seconde (ordre de grandeur). Pour quantifier¹ l'ajustement des données par un modèle, nous calculerons une *rms pondérée* par la formule suivante,

$$(3.1.1) \quad \text{rms} = \sqrt{\sum_{i,k} \left(\frac{w_i^k}{\sum_{i,k} w_i^k} \frac{1}{\sigma} r_i^k \right)^2}$$

où r_i^k est le résidu du temps d'arrivée du séisme i à la station k . Les coefficients w_i^k sont des poids définies par Mosteller et Tukey [31] de la façon suivante,

$$(3.1.2) \quad w_i^k = \max \left\{ 0, \left(1 - \left(\frac{r_i^k}{\alpha r_{med}} \right)^2 \right)^2 \right\},$$

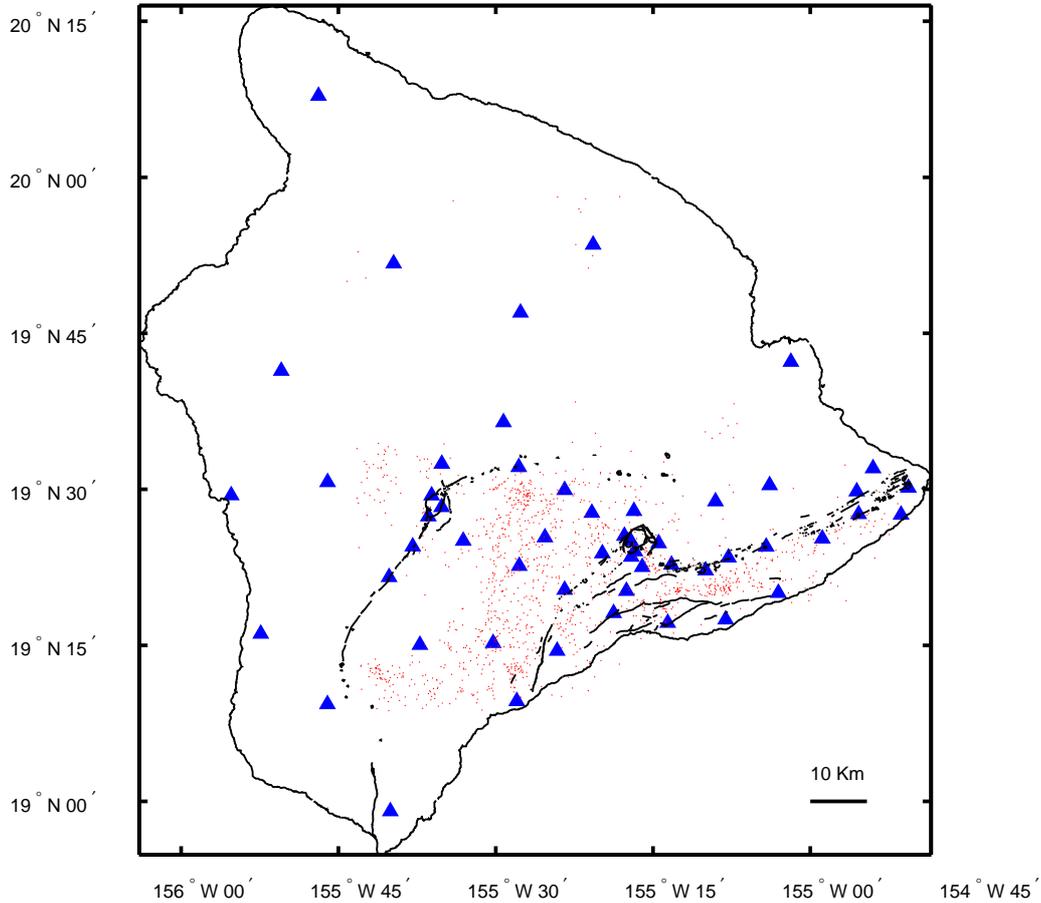
où r_{med} est la médiane des résidus, et α est un coefficient compris entre 4 et 6, nous utiliserons la valeur 5. Cette pondération (3.1.2) permet de produire une valeur de rms, minimisant les effets des données aberrantes.

3.1.1.3. *Localisation préliminaire.* Nous effectuons une localisation des séismes choisis dans le modèle de vitesse à gradient afin d'obtenir un modèle *a priori* (vitesse des ondes P + localisations) cohérent. L'information *a priori* pour cette localisation préliminaire est puisée dans le catalogue des localisations du HVO. Nous utilisons le temps d'arrivée, la localisation et le temps origine consigné dans les données fournies par le HVO. Cette première localisation donne lieu aux choix des paramètres d'inversion, l'écart-type *a priori* sur les temps origines σ_t et l'écart-type *a priori* sur les hypocentres σ_h . Le paramètre σ_t est fixé à valeur physiquement acceptable de 10 secondes, σ_h est choisi en effectuant plusieurs calculs. La figure (3.1.3) montre qu'il existe un seuil en σ_h à partir duquel la solution ne change plus ; cette valeur de σ_h sera retenue pour le calcul des paramètres hypocentraux.

¹Les données auxiliaires définies par le changement de variable dans le schéma (se) ne peuvent pas être interprétés d'un point de vue physique, nous définissons alors une fonction d'ajustement (rms pondérée) exprimée en secondes pour quantifier l'ajustement des données.

FIG. 3.1.2. Sismicité utilisée pour la tomographie à l'échelle régionale.

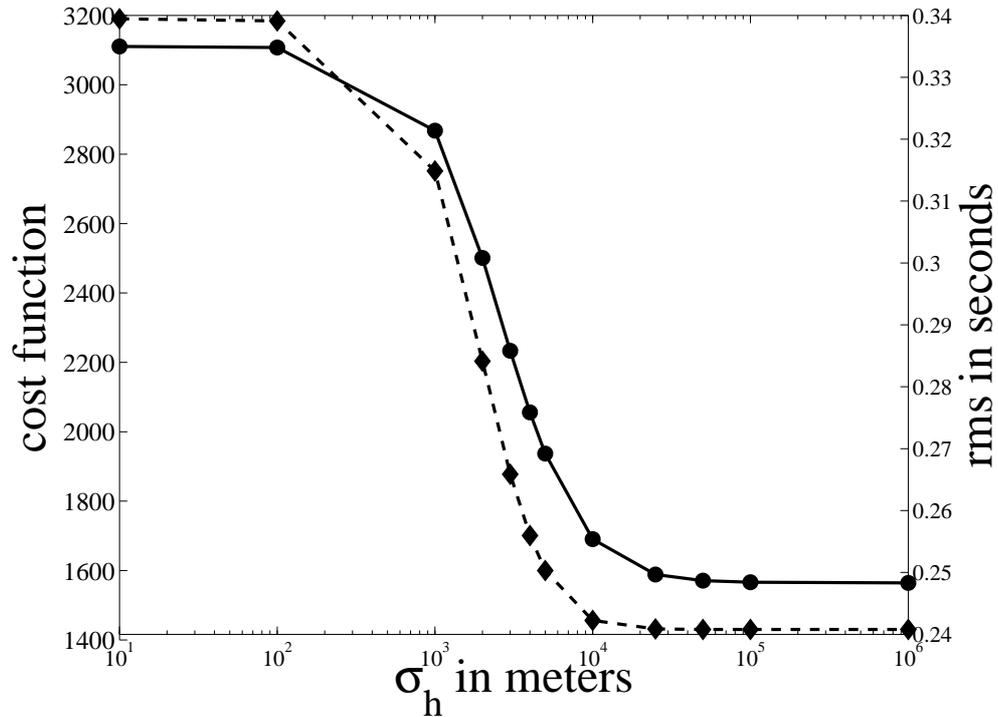
Les points rouges représentent les hypocentres des séismes, les triangles bleus les stations.



3.1.2. Test synthétique. La qualité de l'image tomographique dépend en premier lieu de la répartition des rais dans le volume investigué - une zone peu ou pas couverte ne pourra pas être reconstituée correctement. Il est ensuite très important que les rais traversent les cellules avec des directions différentes pour apporter un maximum d'information contraignant le modèle. La matrice de résolution (cf [45] et [3]) est un outil théorique testant cette propriété. Cette méthode nécessite d'inverser une matrice de très grande taille, un calcul impossible dans notre cas. Nous proposons alors de tester la géométrie des rais par une inversion synthétique consistant à retrouver un damier (Figure (3.1.4)). Une perturbation sinusoïdale d'amplitude 200 m/s et de 12 km de longueur d'onde en carte et 6 km en profondeur est appliquée au modèle à gradient d'Okubo et coll. [35]. Dans ce modèle les temps de propagation synthétiques entre les sources et récepteurs constituent les données à inverser. L'information *a priori* pour ce calcul est : (1) le modèle à gradient,

FIG. 3.1.3. Choix de l'écart-type *a priori* sur les positions.

Les localisations des événements utilisés dans l'inversion sont effectuées à l'aide des temps de propagation, dans le modèle de vitesse *a priori*, en faisant varier l'écart-type σ_h sur les positions *a priori*. On reporte ensuite la rms (tirets) et la fonction coût (trait continu) en fonction de σ_h . Le choix de l'écart-type *a priori* est le point qui débute la partie constante basse de la courbe, en l'occurrence $\sigma_h = 10$ km.



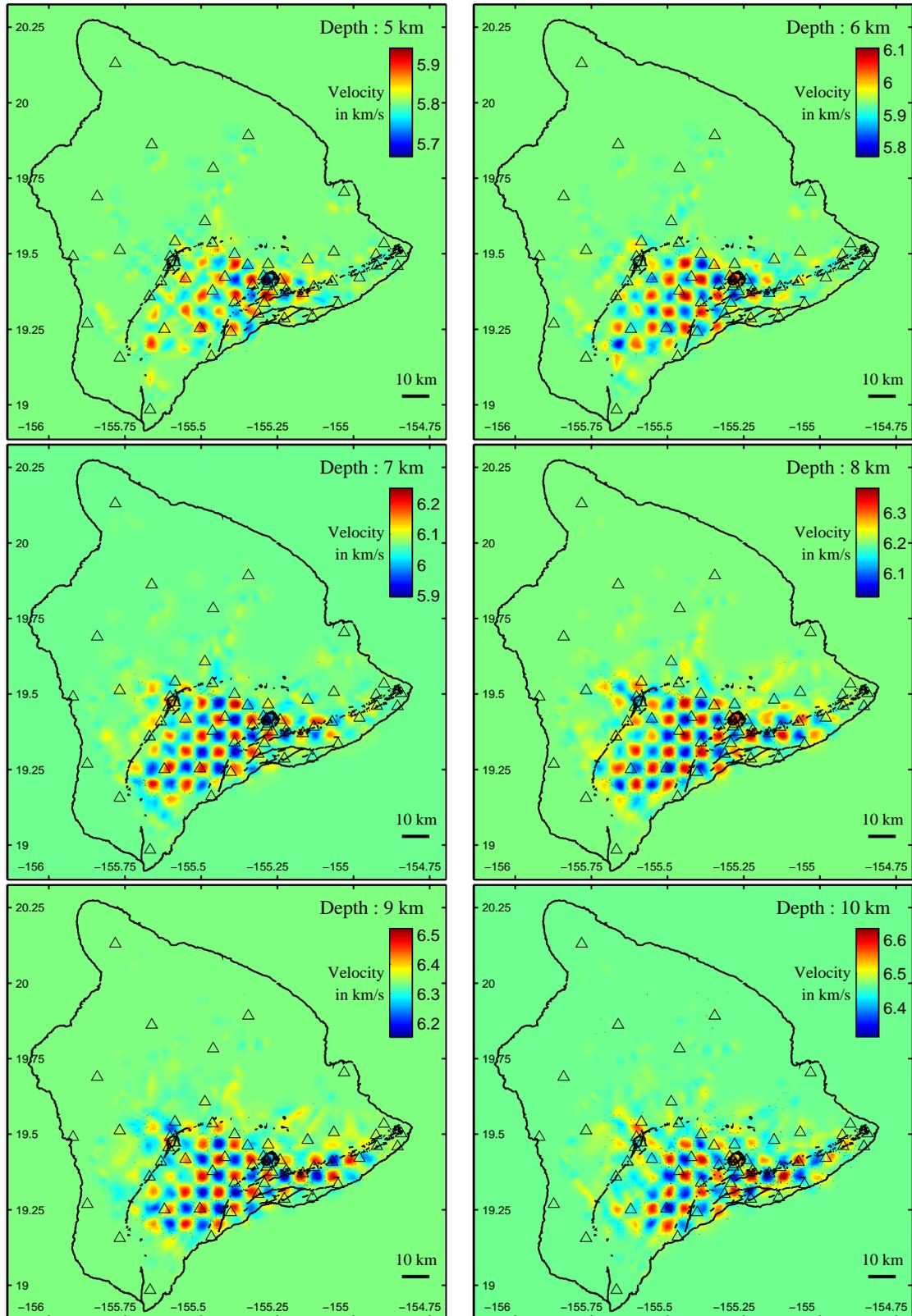
(2) les sources du modèle synthétique. Le meilleur résultat a été obtenu en prenant comme paramètres d'inversion :

- (1) $\lambda = 1.5$ km,
- (2) $\sigma_v = 10$ km/s,
- (3) $\sigma_h = 1$ km.
- (4) $\sigma_t = 10$ s.

La valeur $\sigma_v = 10$ km/s est au delà de la plage de variation admissible, cela signifie que pour stabiliser l'inversion synthétique il n'est pas nécessaire de contraindre fortement le modèle par de l'information *a priori*. La figure (3.1.4) montre le résultat du test, l'algorithme reconstruit le milieu dans une zone située du Mauna Loa au Kilauea, de 5 à 10 km de

FIG. 3.1.4. Damier à l'échelle régionale .

Test de reconstruction d'un damier synthétique dans la géométrie des sources calculées lors de l'inversion tomographique.



profondeur. Ce volume est couvert de façon optimale par les rais, c'est le lieu où le modèle sera le mieux reconstitué. Cette information nous sera utile pour interpréter le résultat tomographique avenir.

3.1.3. Tomographie des temps de première arrivée.

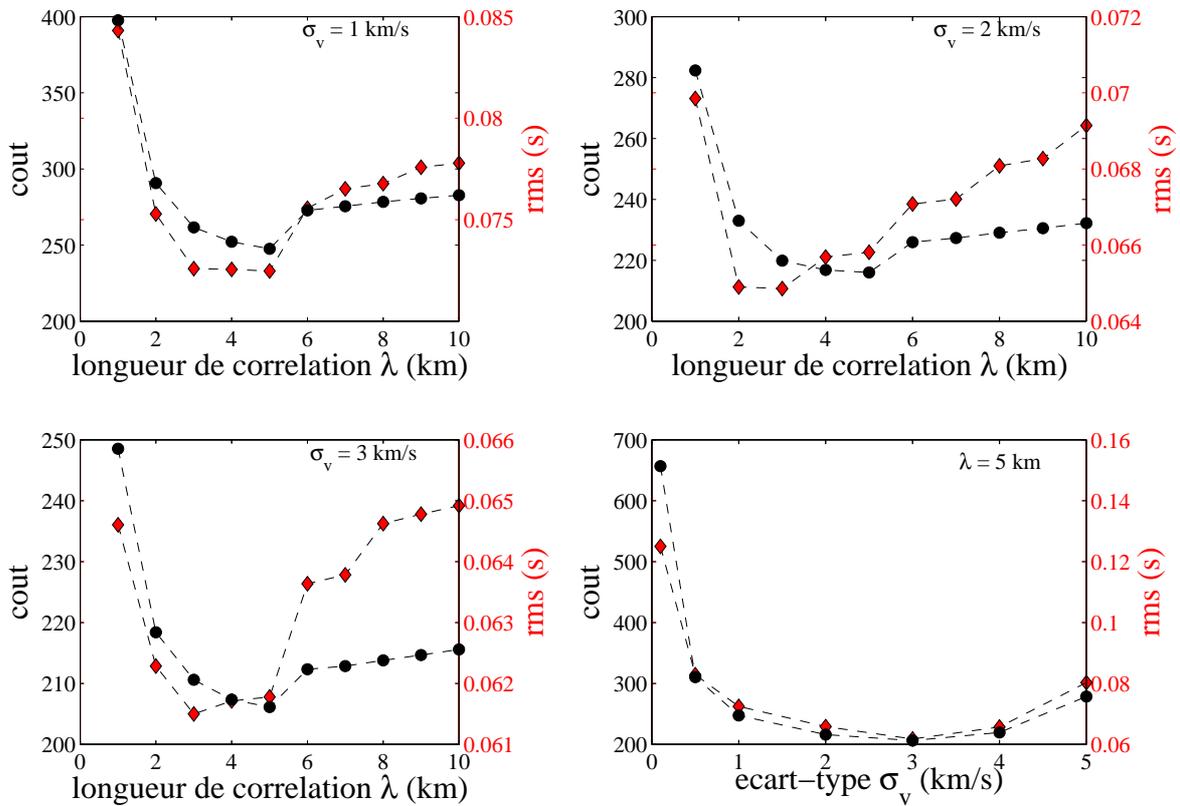
3.1.3.1. *Tomographie.* La zone investiguée est un domaine contenant l'île de Hawaii, échantillonnée par une grille de $160 \times 160 \times 37$ noeuds cubiques de 1 km. Le calcul des temps de parcours par la méthode Podvin-Lecomte [37], est fait sur une grille de cellules cubiques ramenés à 0,5 km de côté. Nous avons également utilisé une grille de 0,25 km de côté, les résultats sont restés comparables avec les deux pas. Pour des raisons de temps de calcul, nous avons préféré utiliser le pas de 0,5 km. Le modèle *a priori* étant déterminé, nous pouvons effectuer la tomographie à l'échelle régionale. Afin de choisir les coefficients d'inversion les mieux adaptés nous effectuons quelques calculs en faisant varier la longueur de corrélation λ et l'écart-type sur les paramètres de vitesses σ_v . La figure (3.1.5) montre la fonction coût et la *rms pondérée* pour différents λ et σ_v . Les valeurs $\lambda=5$ km et $\sigma_v = 3$ km/s minimisent la fonction coût. Les valeurs $\lambda=3$ km et $\sigma_v = 3$ km/s minimisent la rms mais comme on peut le voir sur la figure (3.1.6) les modèles obtenus sont presque identiques. Nous choisirons le modèle minimisant la fonction coût, qui est la fonctionnelle optimisée par l'algorithme, plutôt que la rms. La figure (3.1.7) montre le résultat tomographique sous forme de perturbation par rapport à l'*a priori* en fonction de σ_v . Une valeur trop importante de σ_v ne permet plus de stabiliser numériquement l'algorithme et des artefacts apparaissent sur les images à partir de $\sigma_v = 4$ km/s. Pour les valeurs de σ_v allant de 0.5 à 3 km/s, les images sont proches, mais pour les faibles valeurs le modèle est moins perturbé, la solution est contrainte par l'information *a priori*. La valeur optimale est le meilleur compromis sur σ_v , de manière à ajuster les données, sans que l'*a priori* domine, tout en évitant des artefacts numériques. Le rôle de σ_v peut être considéré comme un poids qui joue sur l'importance relative de l'*a priori* dans la fonction à optimiser. La figure (3.1.8) montre le résultat de l'inversion en faisant varier λ . Pour $\lambda = 1$ km et $\lambda = 2$ km, le modèle possède des fluctuations plus hautes fréquence et des amplitudes plus faibles que pour les autres longueurs de corrélations utilisées. Lorsque l'on allonge λ , les fluctuations sont à plus grande longueur d'onde et de plus forte amplitude tout en ajustant mieux les données. La longueur de corrélation permet d'améliorer le conditionnement de la matrice à inverser, en regroupant l'information pertinente provenant de rais voisins. Quand λ est trop faible, le valeur de conditionnement trop forte ne permet pas d'inverser convenablement la matrice ce qui engendre des fluctuations numériques dans la solution.

3.1.3.2. *Interprétation des résultats.* La figure (3.1.9) montre le modèle de vitesse des ondes P obtenu après inversion avec les paramètres choisis précédemment. Pour les profondeurs de 7 km à 9 km, nous remarquons une zone plus rapide s'étendant du Mauna Loa au Kilauea, le long de ce qui constitue géologiquement les rifts et les calderas de ces deux volcans. Cette anomalie positive se prolonge aux profondeurs 5 km et 6 km, mais elle est découpée par une zone à moindre vitesse qui apparaît nettement entre les deux calderas. Sur toutes les planches, autour de la zone centrale plus rapide, on remarque des régions de moindre vitesse. Ces caractéristiques sont présentes dans les images obtenues par des

FIG. 3.1.5. Choix de σ_v et λ .

Plusieurs tomographies sont effectuées de manière à explorer différents λ et σ_v admissibles, avec $\sigma_h = 10$ km fixé à partir du graphique (3.1.3). Les courbes ci-dessous montrent pour $\sigma_v = 1, 2$ et 3 km/s, la valeur de la fonction coût et de la rms pour différents λ variant de 1 à 10 km. La valeur $\lambda = 5$ km constitue un minimum de la fonction coût sur chaque graphique. Le graphique en bas à droite montre la fonction coût et la rms en fonction de σ_v . Les deux courbes atteignent leur minimum respectif pour le même écart-type, $\sigma_v = 3$ km/s. Les paramètres optimaux pour la tomographie des temps de première arrivée sont :

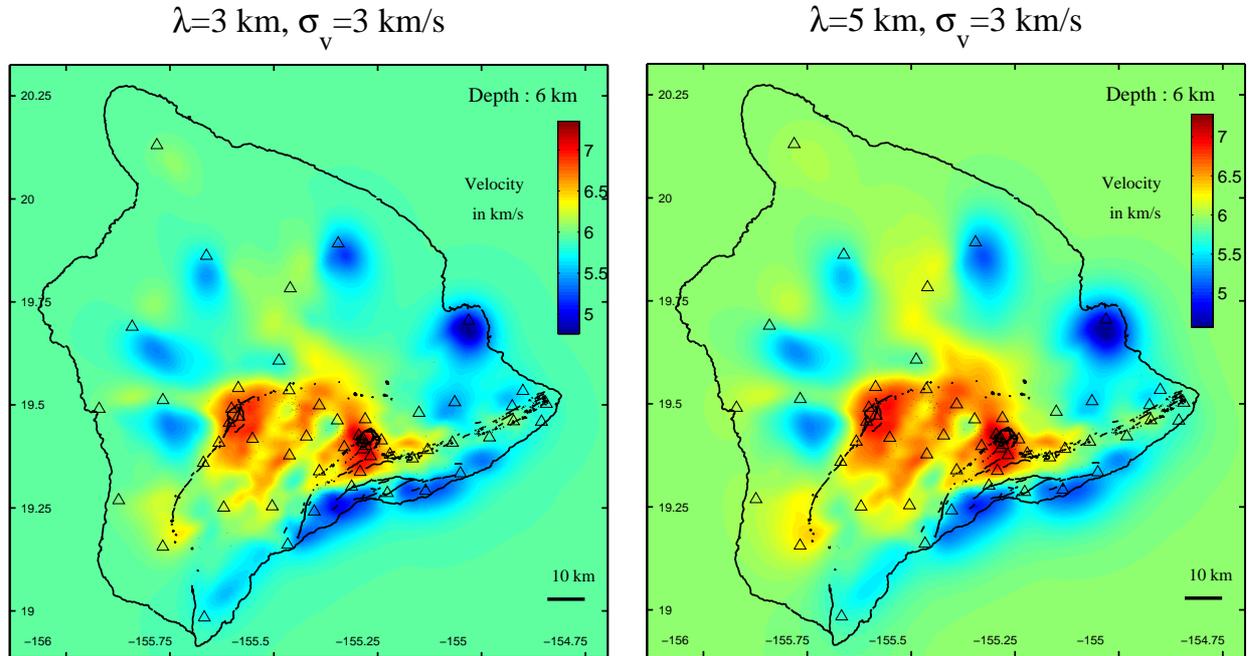
$$\left\{ \begin{array}{l} \lambda = 5 \text{ km} \\ \sigma_v = 3 \text{ km/s} \\ \sigma_h = 10 \text{ km} \\ \sigma_t = 10 \text{ s} \end{array} \right.$$



inversions réalisées avec d'autres longueurs de corrélation et écarts-types. En fait, pour un ensemble de paramètres les résultats sont très proches. Les paramètres ne convenant pas correspondent sur les courbes (3.1.5), aux cas où la fonction coût a une valeur importante,

FIG. 3.1.6. Modèles obtenus avec des coefficients d'inversion différents.

Résultats d'inversion avec des paramètres différents, cartes à 6 km de profondeur.



par exemple lorsque la longueur de corrélation est trop courte, en deçà de 2 km ou dans le cas où σ_v dépasse 4 km/s. Les autres modèles sont équivalents. Ces zones à vitesse plus élevée, entourées de régions à vitesse moins élevée, sont situées dans les volumes bien reconstruits lors de l'inversion, étant donné la géométrie des rais procurés par les données (Figure (3.1.4)). On peut donc extraire une information pertinente des différentes tendances communes observées sur les images obtenues avec les paramètres d'inversion admissibles. Ces images sont par ailleurs comparables à celles obtenues par Okubo et coll. [35] et Benz et coll. [5]. Thurber [48] a également mis en évidence l'existence d'une zone rapide au voisinage de la caldera du Kilauea, et Haslinger et coll. [18], Hansen et coll. [17] ont fait de même pour le rift Est. La zone rapide du centre est l'indice de la présence de matériaux denses, du magma remonté des profondeurs étant resté, sous forme d'intrusion et donc non décomprimé, à l'intérieur de l'édifice. Les zones lentes proviennent probablement du magma décomprimé sorti de l'édifice sous forme de lave, qui s'est répandue à l'extérieur du noyau dense que constitue la zone rapide centrale. La croissance de l'édifice se fait donc selon deux mécanismes : une croissance interne intrusive et une croissance externe effusive. L'existence d'un noyau dense -et plus rigide- à l'intérieur de l'édifice volcanique, entouré par une couverture moins dense et moins rigide -plus plastique-, à une conséquence très importante pour la compréhension des déformations du volcan et leur modélisation. On peut en effet montrer (Monteux et coll., [30]) que cette distribution de la densité et des

FIG. 3.1.7. Exemples de tomographies pour différents σ_v .

Résultats d'inversion exprimés sous forme de perturbation par rapport à l'a priori avec $\lambda = 5 \text{ km}$, $\sigma_h = 10 \text{ km}$, $\sigma_t = 10 \text{ s}$. Différentes cartes à 6 km de profondeur sont représentées en fonction de en fonction σ_v .

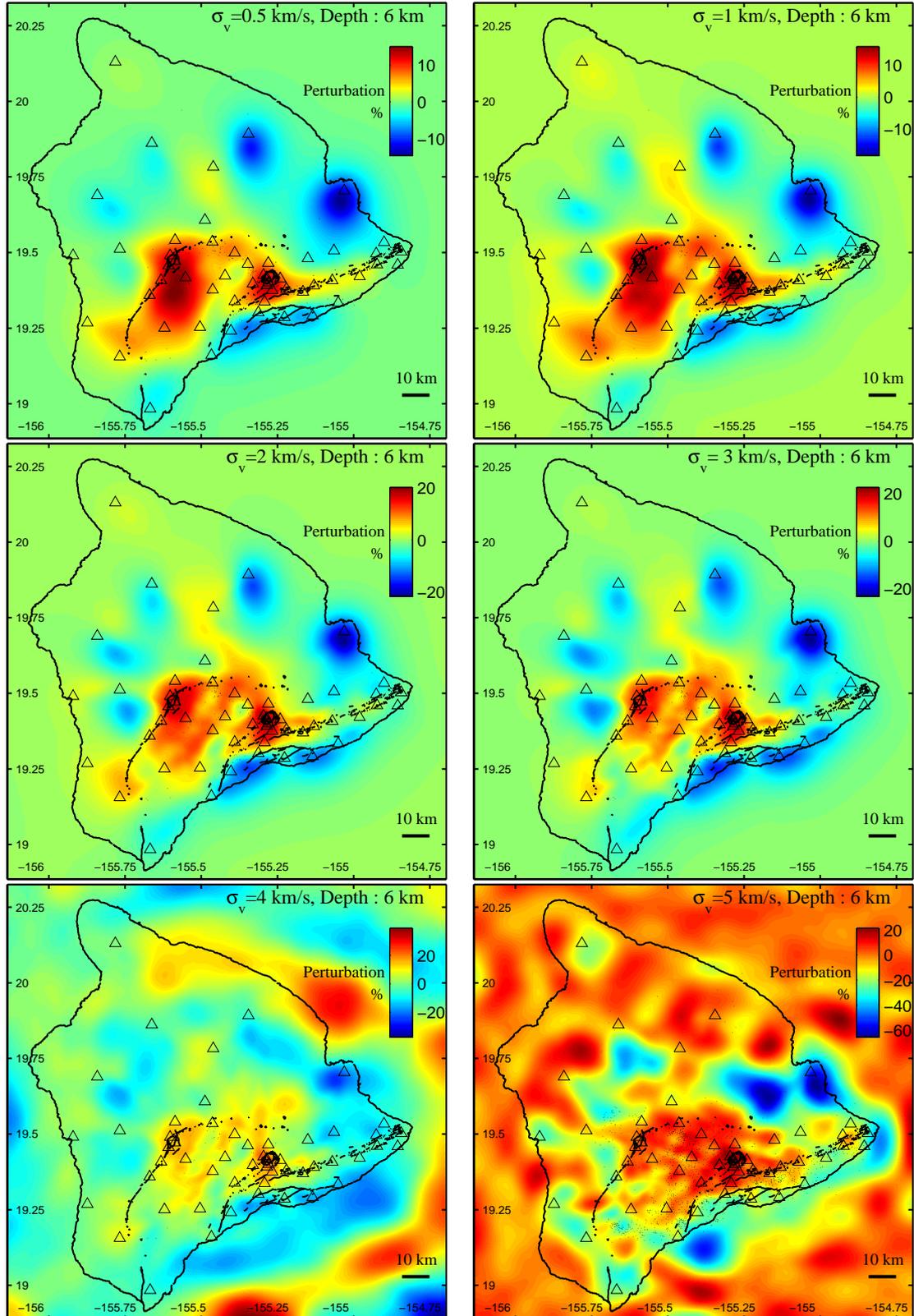


FIG. 3.1.8. Exemples de tomographies en fonction de λ .

Résultats d'inversion exprimés sous forme de perturbation par rapport à l'a priori avec $\sigma_v = 3 \text{ km/s}$, $\sigma_h = 10 \text{ km}$, $\sigma_t = 10 \text{ s}$. Différentes cartes à 6 km de profondeur sont représentées en fonction de en fonction λ .

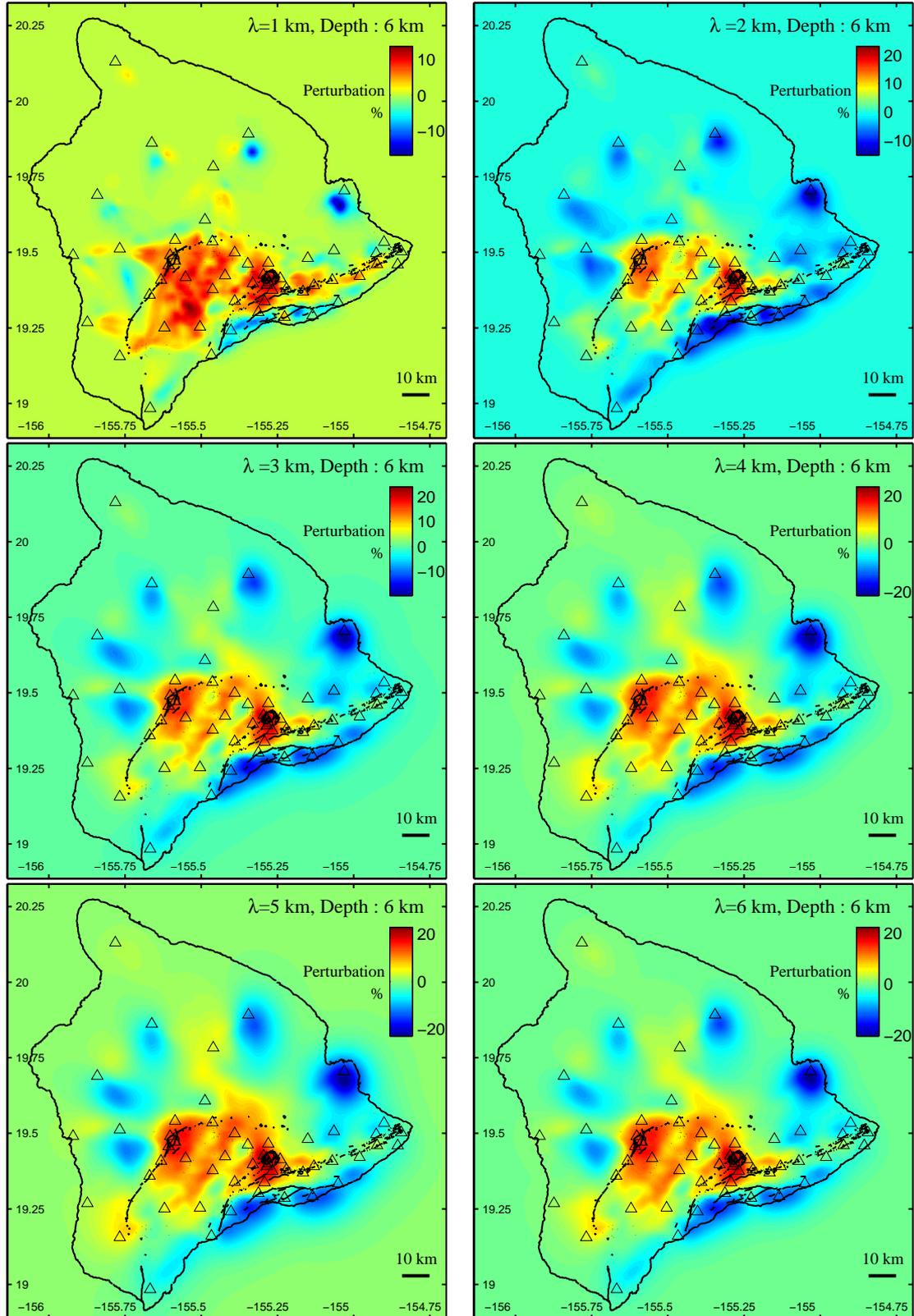
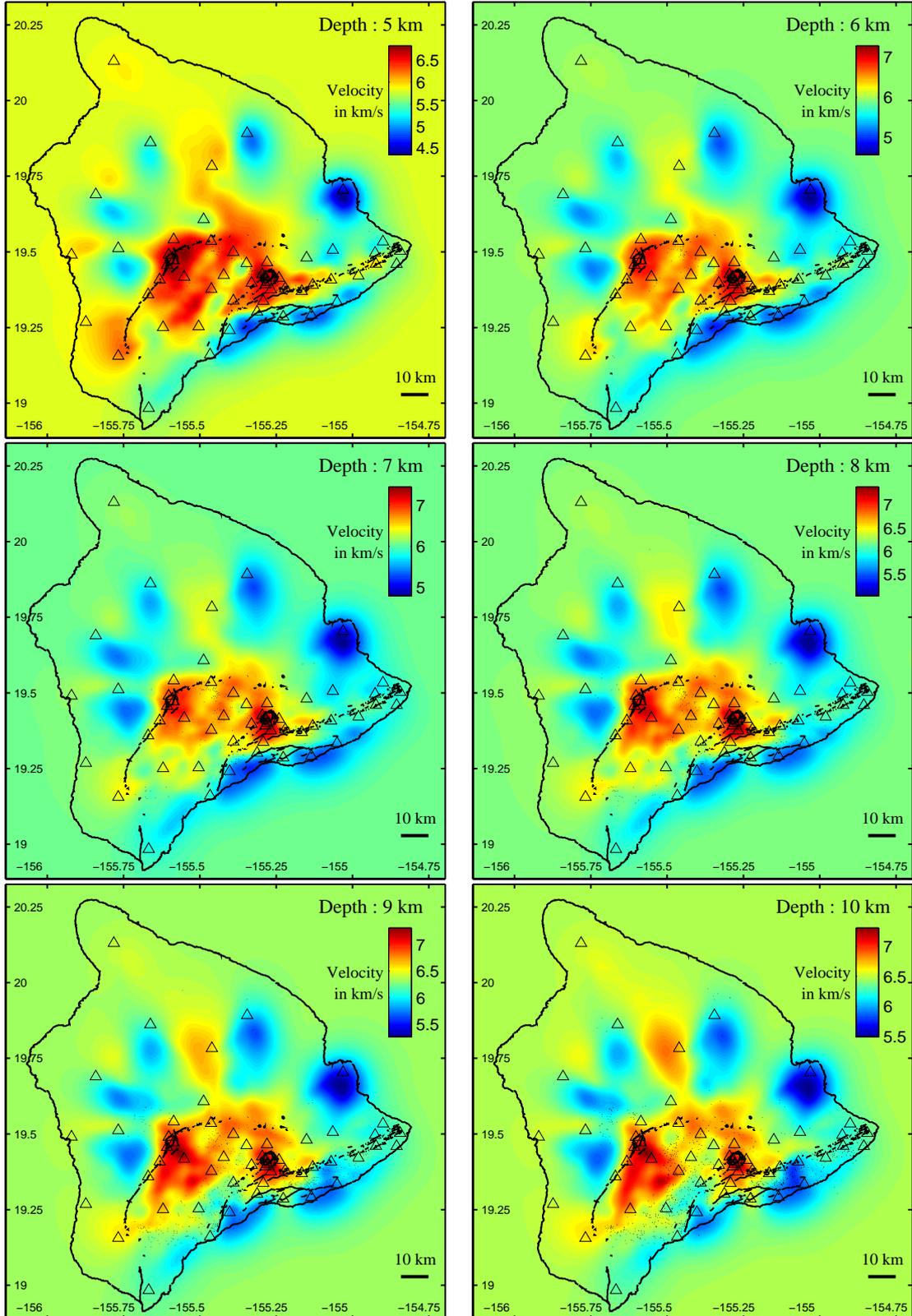


FIG. 3.1.9. Tomographie à l'échelle régionale .

Résultats d'inversion, carte de 5 km à 10 km de profondeur avec,

$$\left\{ \begin{array}{l} \lambda = 5 \text{ km} \\ \sigma_v = 3 \text{ km/s} \\ \sigma_h = 10 \text{ km} \\ \sigma_t = 10 \text{ s} \end{array} \right.$$



paramètres rhéologiques induit la rupture de la croûte océanique sous-jacente et la déformation intense de la couverture plastique, au niveau des flancs des volcans Hawaïiens. Elle permet notamment d'expliquer la distribution confinée de la sismicité et de la déformation intense du flanc Sud du Kilauea.

3.2. Tomographie à l'échelle locale.

L'étape précédente avait pour rôle de fournir une information la plus précise possible à l'échelle régionale de l'île de Hawaïi. Le modèle obtenu va à présent être utilisé comme point de départ pour la tomographie utilisant des décalages temporels interspectraux.

3.2.1. Données, estimation de la résolution, modèle *a priori*.

3.2.1.1. *Données.* Notre objectif est de connaître mieux le modèle de vitesse de la zone comprenant la caldera et la racine du rift Est. Pour cela nous disposons d'un ensemble d'essaims denses de séismes comportant de très nombreux séismes similaires, dans le flanc Sud du Kilauea, ainsi que d'un réseau dense de stations sismiques HVO, autour de la caldera et du rift Est. La sismicité du flanc Sud a la particularité de comporter plusieurs essaims de séismes semblables. Parmi ces événements nous avons sélectionné 416 séismes (figure (3.2.1)) issus de 5 essaims (dont 4 peuvent être intercorrélés figure (3.2.2)) pour lesquels la méthode interspectrale a permis de calculer plus de 500 000 décalages temporels avec une erreur moyenne de 5 millisecondes (Got et Okubo, [16]). Cette valeur dépend de la cohérence entre les signaux. Elle atteint 10 millisecondes pour les couples de séismes les moins cohérents que nous ayons sélectionnés. Nous avons une idée *a priori* de la qualité de chaque décalage temporel, par l'intermédiaire de la fonction de cohérence (cf [13] et Annexe 2) dont les valeurs sont comprises entre 0 et 1 et vaut 1 pour des signaux semblables. Pour décrire la statistique des erreurs de mesure que nous supposons gaussienne, nous utiliserons la formule (4.0.17) (Annexe 2), avec $C_{max}=0.995$ et $\sigma = 5.10^{-4}$ secondes. Nous obtenons ainsi pour chaque donnée un écart-type définissant la matrice \mathbf{C}_D . L'erreur d'ajustement des données sera quantifié par la fonction *rms pondérée*,

$$(3.2.1) \quad \text{rms} = \sqrt{\sum_{i,j,k} \left(\frac{\sum_{i,j,k} \sigma_{ij}^k r_{ij}^k}{\sigma_{ij}^k} \right)^2},$$

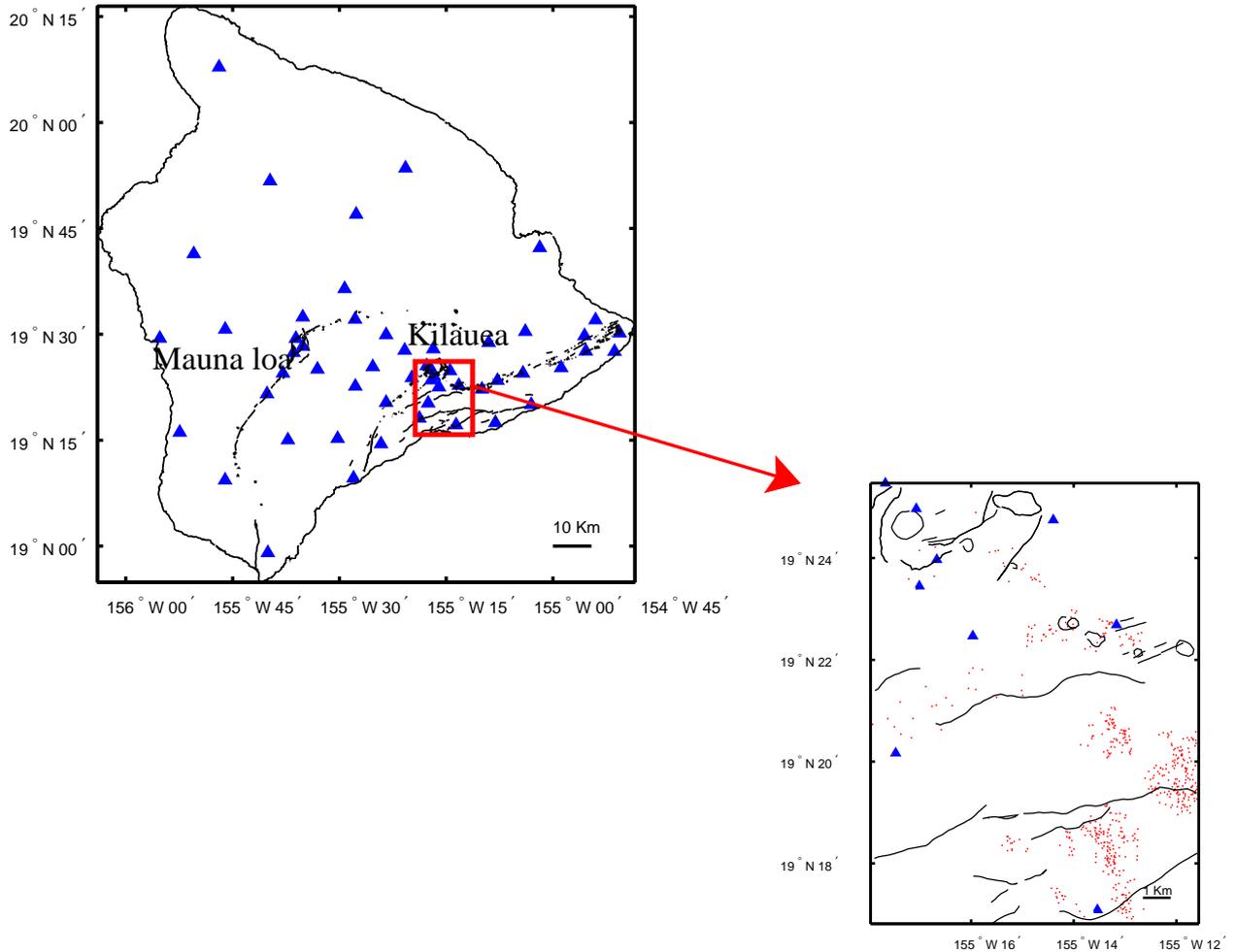
où r_{ij}^k est le résidu du décalage temporel entre les séismes i et j pour la station k et σ_{ij}^k est l'incertitude pour ce même décalage temporel.

L'ensemble des séismes corrélés couvre un volume d'environ $7 \text{ km} \times 7 \text{ km} \times 4 \text{ km}$; la station la plus proche est à une distance de 10 à 15 km de ces séismes, les nombreuses (7) stations de la caldera et des rifts du Kilauea sont situées à une distance de 15 à 25 km. La configuration est donc *a priori* favorable à une détermination complète des paramètres hypocentraux. On peut penser que le problème de la localisation des décalages temporels sera bien conditionné, et donc bien contraint par les données; de même la résolution des équations de tomographie des décalages temporels ne devrait pas être beaucoup moins bien conditionnée que la résolution des équations classiques, les problèmes de conditionnement

du système d'équations provenant plutôt des paramètres de vitesse non échantillonnés par les rais que par les paramètres hypocentraux mal déterminés.

FIG. 3.2.1. Sismicité pour la tomographie des décalages temporels.

Sur la carte de Hawaii, la zone d'étude est représentée par un cadre rouge. Un agrandissement de celle-ci montre les essaims de séismes semblables dans le flanc Sud du Kilauea utilisées pour l'étude tomographique.



3.2.1.2. *Modèle a priori.* La première question concerne le volume à inverser. Etant donné la distribution des sources et des stations, il est probable que la distribution des rais sera favorable dans un volume situé au Sud de la caldera et à la racine du rift Est. Une fois la position moyenne du volume à inverser déterminée, sa dimension sera essentiellement (dans notre cas) contrôlée par les ressources informatiques disponibles. La tomographie des temps de première arrivée nous a fourni un modèle de vitesse régional optimal pour les ondes P. C'est une partie de ce modèle, correspondant au volume à inverser, qui sera retenue comme modèle *a priori* pour la tomographie locale.

FIG. 3.2.2. **Couples de séismes retenus pour la tomographie des décalages temporels.**

Les couples de séismes sélectionnés permettent de calculer des décalages temporels interspectraux de cohérence supérieure à 0,8. Le nombre de stations montrant des signaux semblables par couple de séismes est figuré par différentes couleurs. Nous avons sélectionné les couples de séismes (1) en veillant à la répartition de ceux-ci afin de bien échantillonner le milieu par les rais et (2) en utilisant les couples dont les signaux sont suffisamment similaires afin de calculer des décalages temporels interspectraux les plus précis possibles. On voit que les corrélations entre séismes peuvent être à grande longueur d'onde. Les couples de séismes les plus éloignés atteignent une cohérence suffisante à basse fréquence lorsqu'ils sont enregistrés par des stations lointaines.

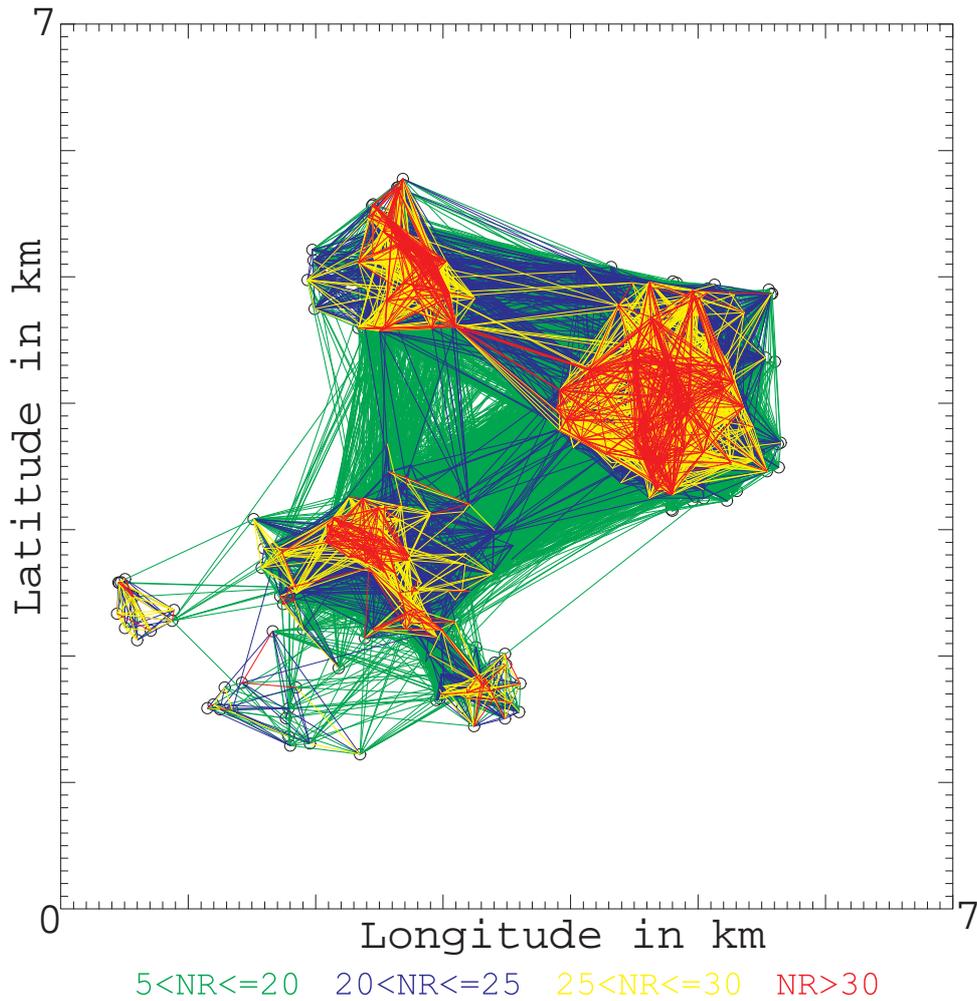
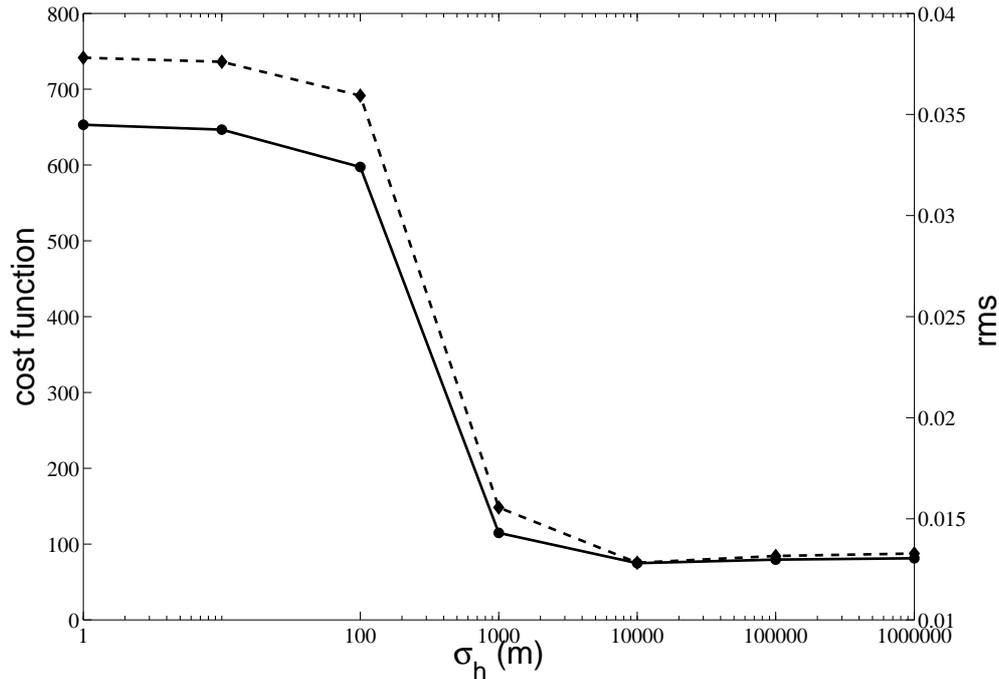


Figure 4

En premier lieu nous effectuons une localisation des séismes à l'aide des décalages temporels dans le modèle choisi afin d'obtenir un *a priori* (séismes + modèle de vitesse) cohérent. Nous devons déterminer l'écart-type sur les hypocentres σ_h adéquat (figure (3.2.3)).

FIG. 3.2.3. **Choix de l'écart-type *a priori* sur les positions.**

Plusieurs localisations avec les décalages temporels sont effectuées afin de choisir l'écart-type σ_h adapté. $\sigma_h = 1\text{km}$ constitue le point à partir duquel les résultats sont indentiques, cette valeur sera utilisée dans tous les calculs.



La valeur $\sigma_h = 1\text{km}$ est le seuil à partir duquel la fonction coût et la rms ne diminuent presque plus et convergent vers la même solution. Pour les paramètres de valeur inférieure, la solution est trop contrôlée par l'*a priori*. De plus, l'erreur estimée sur les positions *a priori* étant de l'ordre de 1 km, il est satisfaisant de voir que cette valeur est admissible pour le processus d'inversion. De la même façon l'écart-type σ_t montre comme σ_h un seuil, situé à une valeur de 10 secondes.

3.2.1.3. *Tests synthétiques.* Le test du damier permet de tester le pouvoir de résolution de la géométrie des rais. Le modèle synthétique est le modèle à gradient de Okubo et coll. [35] et Benz et coll. [5] perturbé par une fonction sinusoïdale de longueur d'onde 4 km en carte et 3 km selon la profondeur avec une amplitude maximale de 200 m/s. L'expérience

consiste à générer des décalages synthétiques dans le damier et ensuite à inverser ces données en prenant le modèle à gradient comme *a priori*. L'objectif de ce test synthétique étant de déterminer la qualité de la couverture en rais du volume investigué, les sources peuvent être contraintes au voisinage de leur position exacte par le biais d'un *a priori* fort. L'idée est de fixer une géométrie de rais proches de celle utilisée dans le processus tomographique des décalages temporels. L'inversion permet de déterminer les zones où la géométrie des rais est optimale, c'est le lieu où le modèle de vitesse sera le mieux reconstitué. Il faut cependant noter que le test du damier est très difficile à satisfaire ; certaines zones où le damier est mal reconstruit peuvent malgré tout contenir certaines anomalies de vitesse à plus grande longueur d'onde estimables avec les mêmes données. Il faudrait en fait produire toute une série de damiers de longueurs d'onde différentes, pour construire littéralement le spectre de la fonction de transfert en chaque cellule. On peut comprendre le résultat d'un test de damier individuel comme indiquant le volume dans lequel on peut reconstruire la longueur d'onde testée. Les figures (3.2.4)-(3.2.5) montrent le modèle résultant du test, avec une longueur d'onde de 4 km horizontalement et 3 km verticalement, selon 12 cartes de 4 km à 9.5 km de profondeur représentées avec un pas vertical de 500 m. Notons que sur chaque figure il existe une zone où de nombreuses cases contigües du damier sont reconstituées à cette longueur d'onde. On peut mettre en évidence la zone la mieux résolue par les rais, qui est un volume débutant à 4 km de profondeur au Nord-Ouest de la carte, traversant le domaine pour finir à 9.5 km de profondeur au Sud-Est de la carte. Au voisinage de ce volume, on distingue une zone où le damier est moins bien reconstitué, mais qui peut éventuellement contenir des informations pertinentes à plus grande longueur d'onde.

3.2.2. Tomographie des décalages temporels.

3.2.2.1. *Tomographie.* La zone investiguée est un domaine situé au sud de la caldera du Kilauea (figure (3.2.1)), échantillonnée par une grille comprenant $33 \times 25 \times 25$ mailles espacées de 0,5 km. Le calcul direct (Podvin-Lecomte [37]) est fait avec un pas d'échantillonnage de 0,25 km. Il s'est avéré inutile de prendre des mailles plus fines, les résultats obtenus alors sont restés comparables. Comme pour la tomographie des temps de première arrivée, nous devons choisir les paramètres d'inversion appropriés afin d'exploiter au mieux les données choisies. Les écart-types choisis sur les positions et les temps origines sont valeurs optimales trouvées lors de la localisation en doubles différences préliminaire. Pour les paramètres de vitesse, les courbes de la figure (3.2.6), nous conduisent à choisir $\lambda = 1\text{km}$ et $\sigma_v = 2\text{km/s}$. Dans ce cas la valeur de la rms passe de $\sim 0.0147\text{ s}$ à $\sim 0.0098\text{ s}$ (figure (3.2.7)) en 4 itérations au cours du processus d'inversion, ce qui représente une diminution de l'ordre de 30 %. Les figures (3.2.8)-(3.2.9) montrent la perturbation par rapport au modèle *a priori*. On observe une variation significative de $\pm 5\%$ des vitesses des ondes P.

3.2.2.2. *Interprétation des résultats.* Le modèle de vitesse estimé par tomographie des décalages temporels est découpé en 12 cartes tous les 500 m de 4 à 9,5 km de profondeur (Figures (3.2.10)-(3.2.11)). Nous distinguons sur chaque carte une zone où la vitesse est plus rapide. Pour la mettre en évidence nous avons tracé la frontière de ce qui constitue cette

FIG. 3.2.4. Test du damier (1).

Résultat du test du damier pour la tomographie des décalges temporels, profondeur de 4 à 6.5 km.

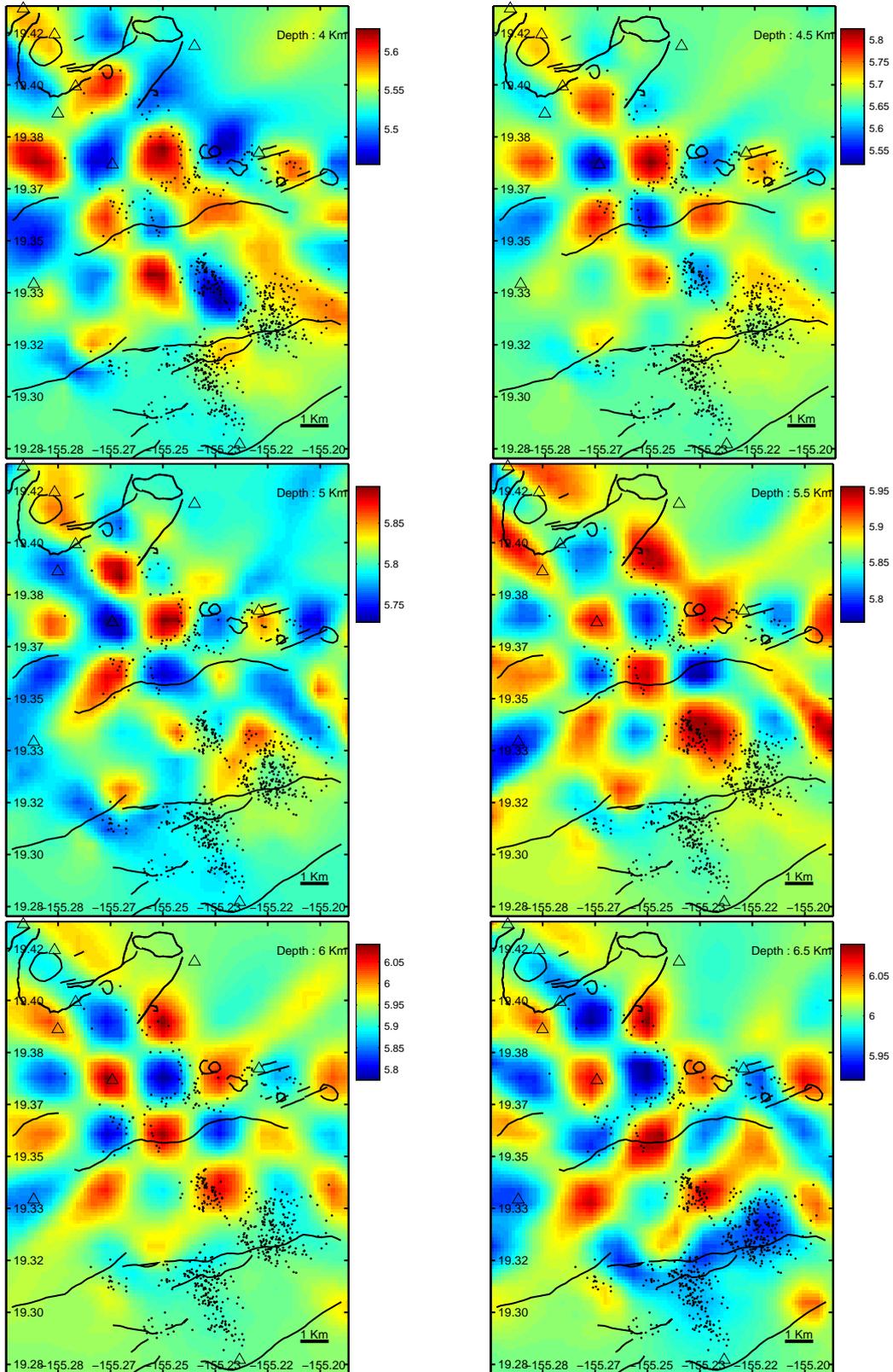


FIG. 3.2.5. Test du damier (2).

Résultat du test du damier pour la tomographie des décalges temporels, profondeur de 7 à 9.5 km.

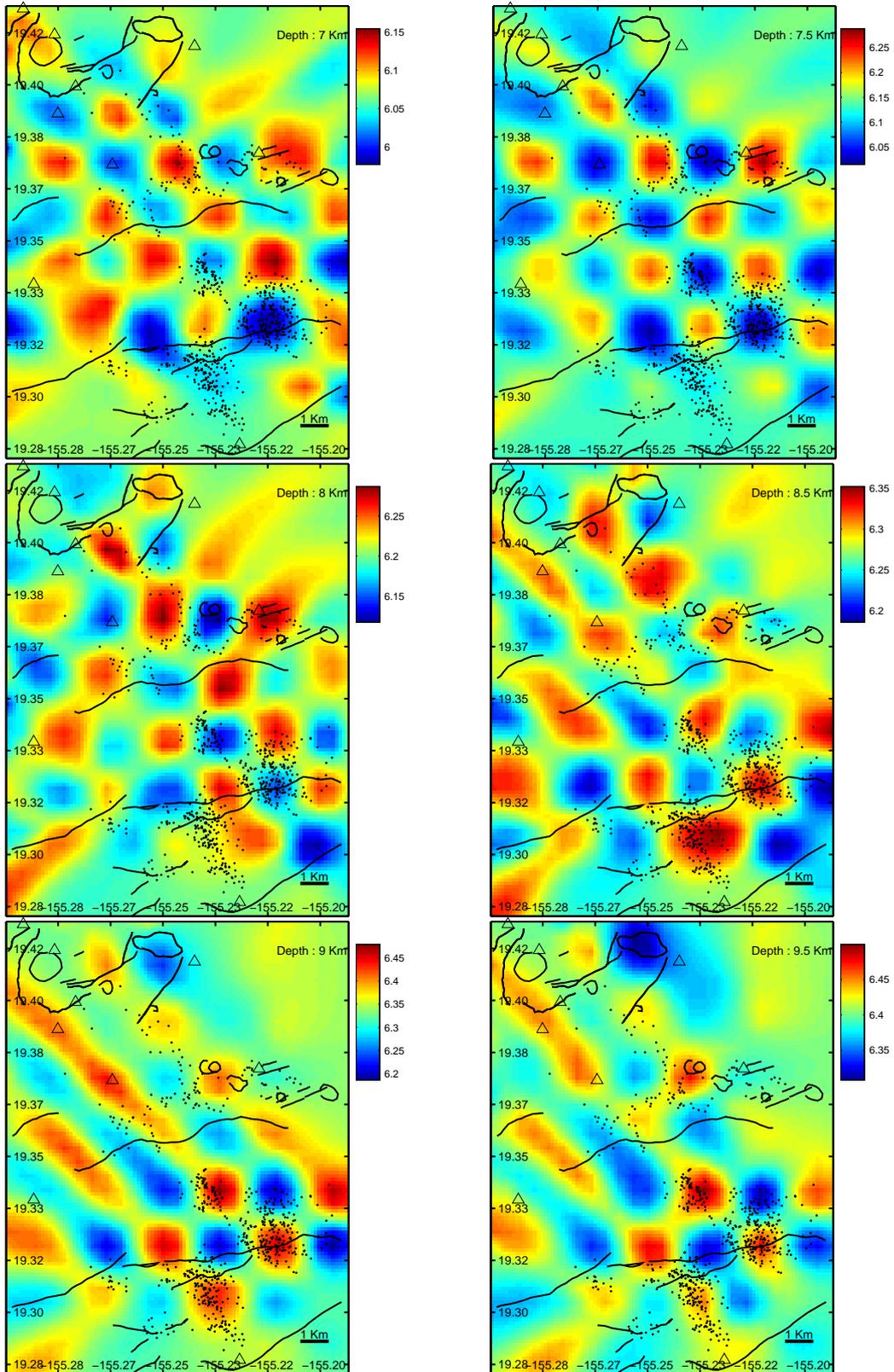
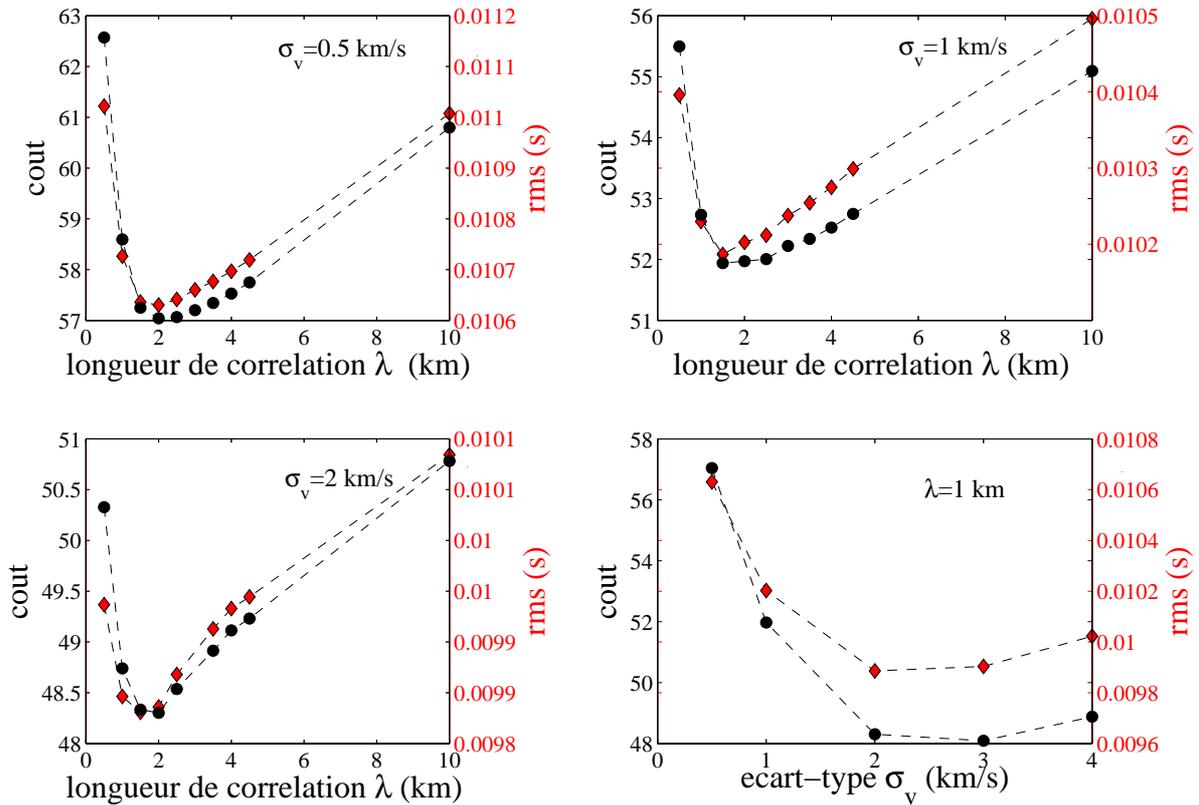


FIG. 3.2.6. Choix des paramètres d'inversion.

Plusieurs tomographies des décalages temporels sont effectuées afin de déterminer les paramètres adaptés, on choisi,

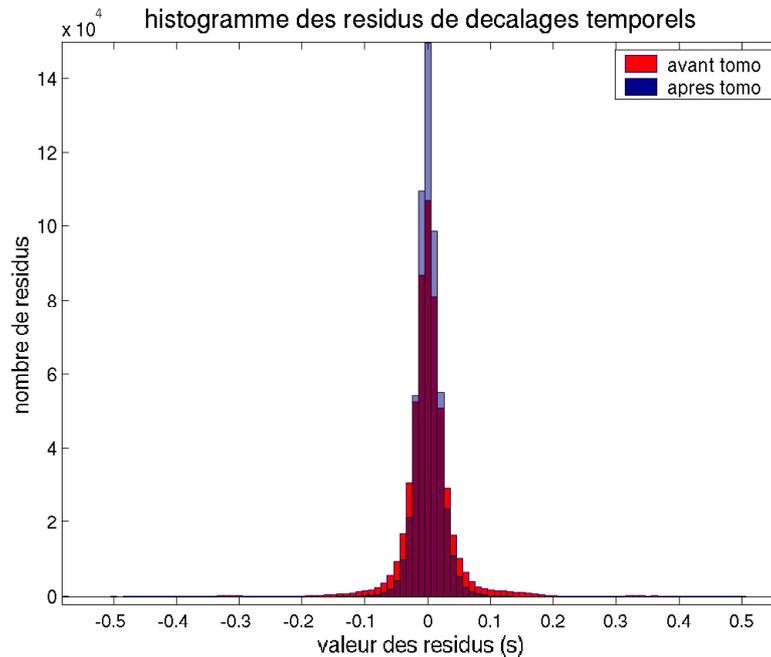
$$\left\{ \begin{array}{l} \lambda = 1 \text{ km} \\ \sigma_v = 2 \text{ km/s} \\ \sigma_h = 1 \text{ km} \\ \sigma_t = 10 \text{ s} \end{array} \right.$$



anomalie positive. Elle est située dans le volume le mieux résolu par les rais ; elle constitue donc l'information principale que contient l'image tomographique. Cet objet est l'indice de la présence d'un corps particulièrement dense sous la caldera et le rift Est du Kilauea. Une reconstitution 3D de sa géométrie (Figure (3.2.12)), montre qu'il a un développement surtout vertical ; il peut être représenté à partir de 8.5 km de profondeur, montre deux conduits qui remontent au Sud de la caldera jusqu'à 5 km de profondeur, puis bifurque vers le rift Est à environ 4.5 - 4 km de profondeur. Ce corps dense doit se prolonger en

FIG. 3.2.7. Résidus pour la tomographie locale.

Histogramme des résidus des décalages temporels après localisation à l'aide des décalages temporels dans le modèle tomographique régional et après tomographie des décalages temporels.



profondeur et en surface, au delà des zones couvertes par la tomographie. On peut penser que ce corps dense donne une indication sur le trajet du magma dans l'édifice volcanique : en effet lorsque le magma remonte, une partie reste prisonnière dans l'édifice et se compacte sous l'effet de la pression autour des conduits pour former une coque dense. Sur le Kilauea, le magma est extrêmement fluide et peut progresser dans l'édifice par des conduits (dykes, sills, tubes) très étroits. Cette caractéristique les rend très difficiles (voire impossible) à détecter par des moyens sismologiques "directs", à moins qu'il existe des volumes dans le volcan comportant de très nombreux conduits voisins simultanément actifs. Or ce n'est pas le cas en surface ; les structures actives pour le transport de magma -dykes et tubes- recoupent la surface suivant des géométries très simples : généralement un seul plan pour les dykes, un point pour les tubes. Les observations géologiques faites dans les calderas ouvertes par l'érosion montrent un chevelu de fractures ouvertes et remplies de magma, mais il est maintenant clair que ces fractures ne fonctionnent pas toutes simultanément. Une raison majeure s'oppose alors à l'identification de structures planaires ou tubulaires fines : plus encore que la résolution limitée de l'inversion (qui pourrait être améliorée en introduisant des données supplémentaires), c'est la limitation haute fréquence du spectre des séismes enregistrés en surface qui ne permet pas de détecter le transport du magma dans un dyke d'épaisseur métrique avec un signal de 500 m de longueur d'onde. Ce que nous observons n'est donc probablement que la signature de ce transport, elle est une bonne indication concernant le système d'approvisionnement magmatique mais elle ne signifie pas

FIG. 3.2.8. Tomographie des décalages temporels, perturbation (1).

Résultat de la tomographie des décalages temporels sous forme de perturbation par rapport au modèle a priori, cartes de 4 à 6.5 km.

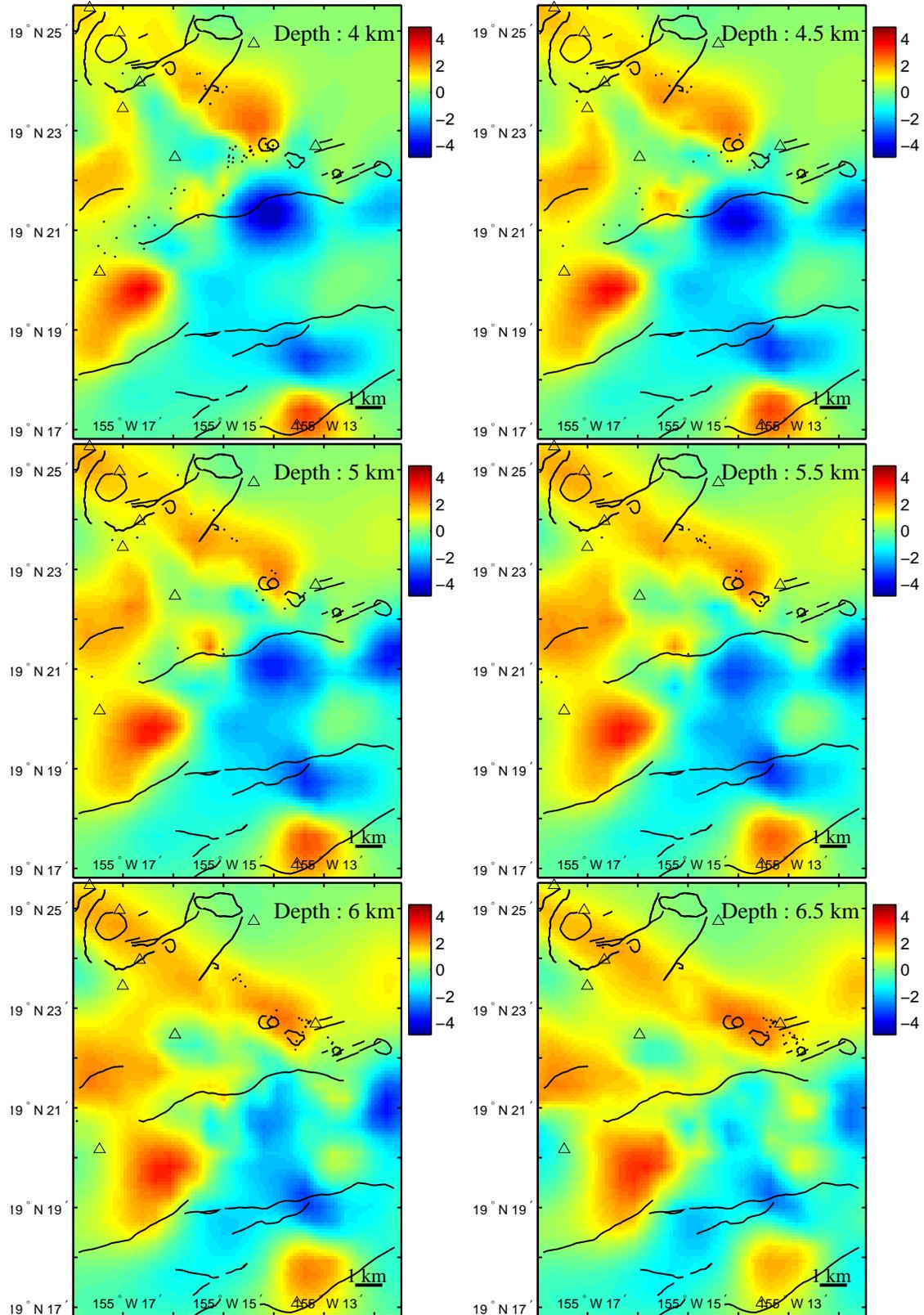


FIG. 3.2.9. Tomographie des décalages temporels, perturbation (2).

Résultat de la tomographie des décalages temporels sous forme de perturbation par rapport au modèle a priori, cartes de 7 à 9.5 km.

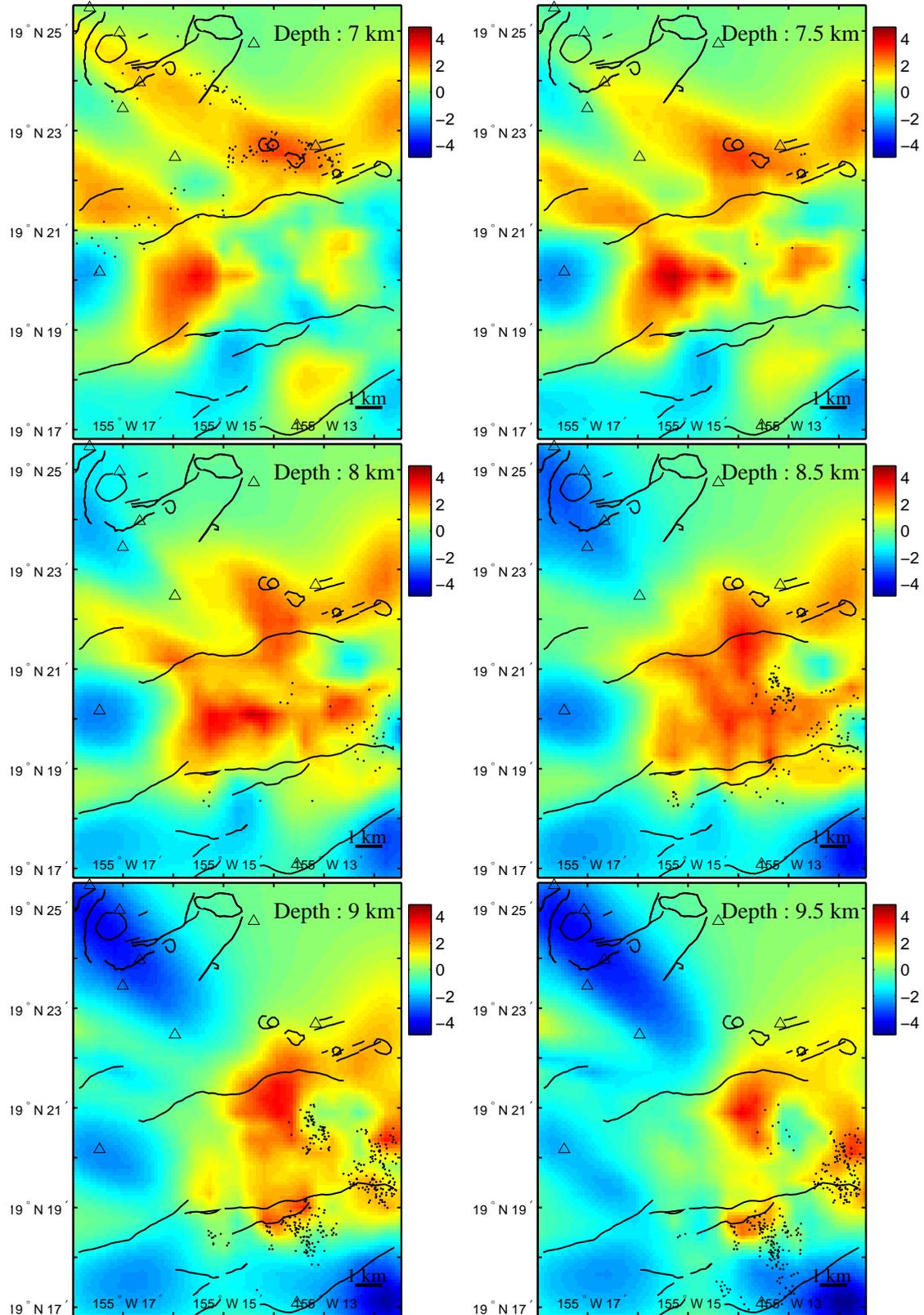


FIG. 3.2.10. **Modèle de vitesse (1).**

Résultat de la tomographie des décalges temporels, profondeur de 4 à 6.5 km. La courbe bleu indique l'anomalie positive qui constitue la principale information du résultat.

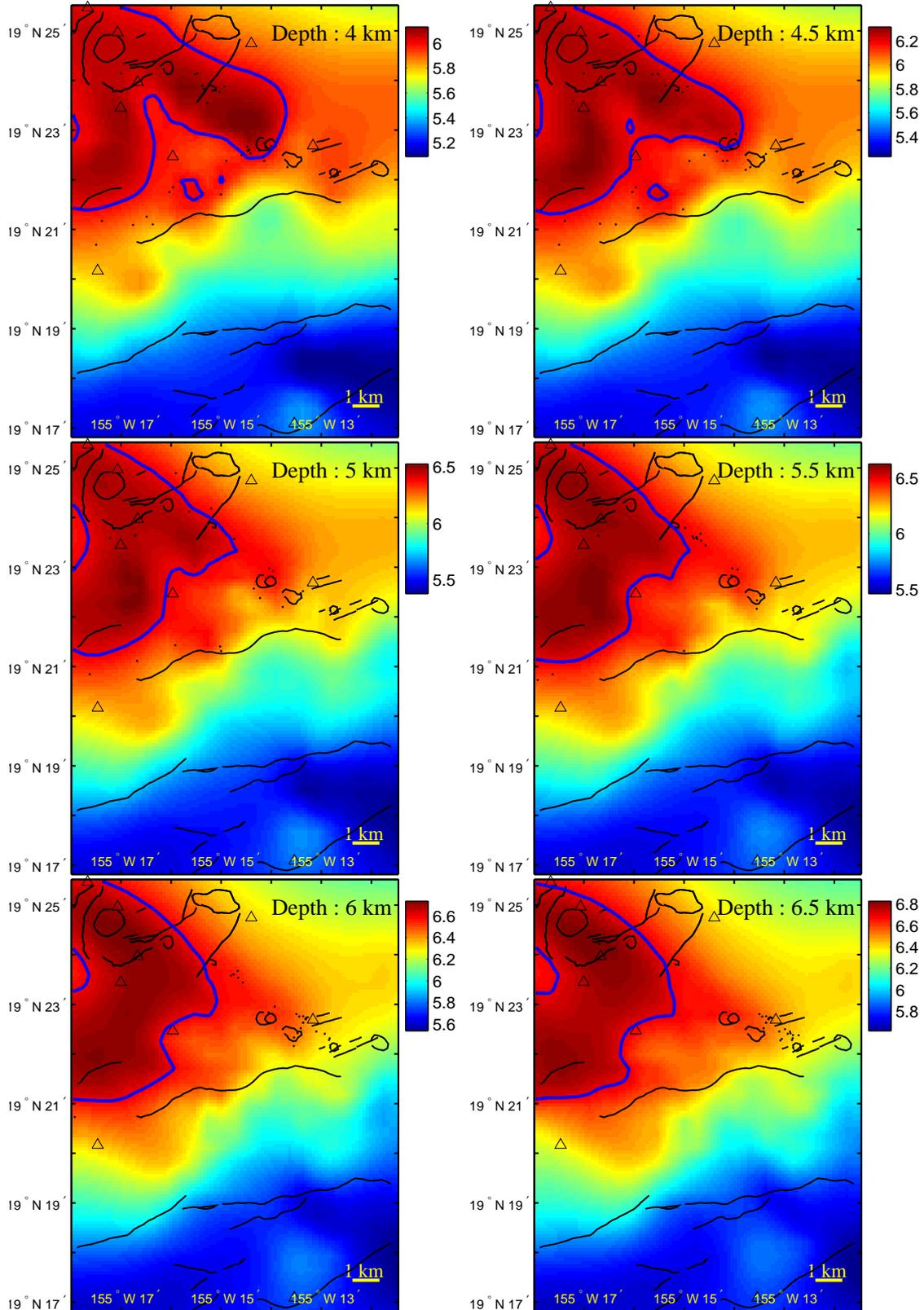


FIG. 3.2.11. **Modèle de vitesse (2).**

Résultat de la tomographie des décalges temporels, profondeur de 7 à 9.5 km. La courbe bleu indique l'anomalie positive qui constitue la principale information du résultat.

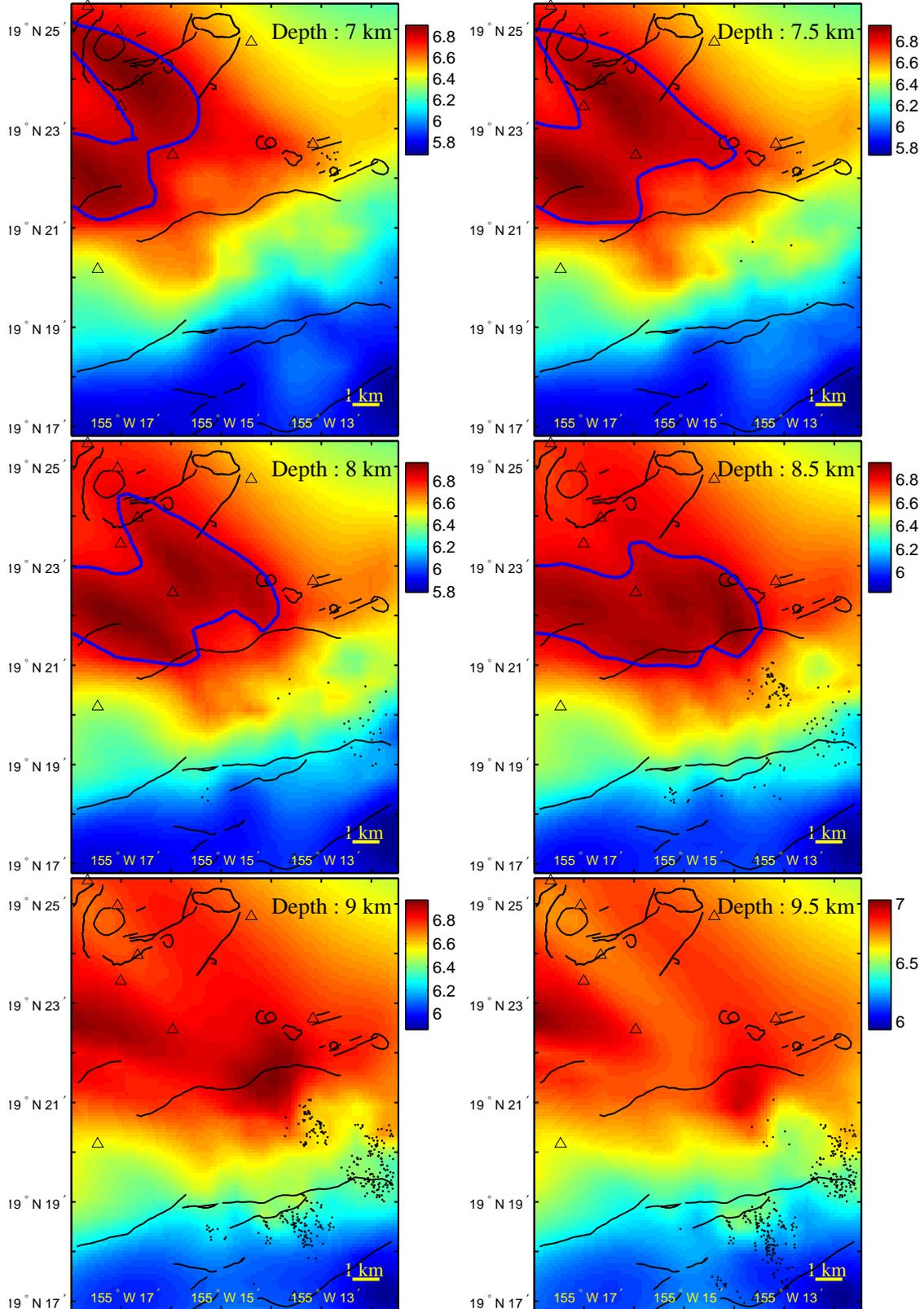
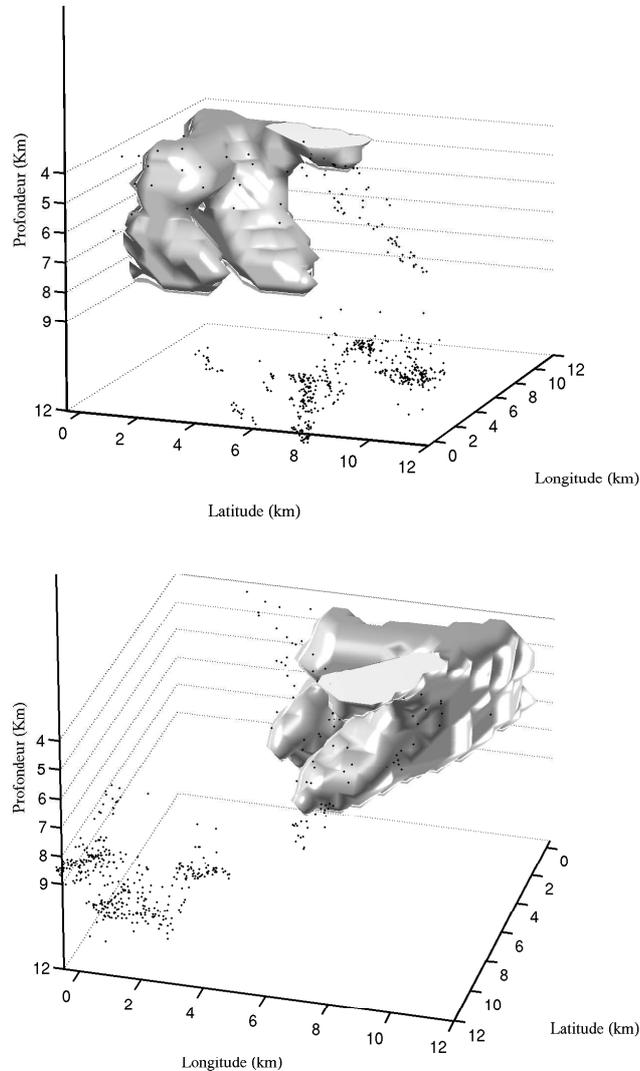


FIG. 3.2.12. **Visualisation 3D de l'anomalie positive.**

Représentation à 3 dimensions de l'anomalie positive mise en évidence par la tomographie des décalages temporels. La volume figuré ici est une isosurface matérialisant la frontière de la zone perturbée à plus de 7 % par rapport au modèle à gradient de Okubo et coll. [35].



que ce système soit actif pendant la durée échantillonnée par les données utilisées dans cette tomographie.

Seuls les résultats de Dawson et coll. [10] semblent pouvoir être comparés avec le modèle de vitesse local obtenu dans le cadre de cette étude. En effet, ces auteurs ont déployé un réseau dense de 67 stations courte période dans la caldera du Kilauea, obtenant un espacement moyen des stations de 650 m. Ils enregistrèrent environ 200 séismes de la caldera ($z < 3$ km) pendant une période de un mois. Le modèle, constitué de cellules cubiques de

500 m de côté a été calculé jusqu'à une profondeur de 2,5 km. Ils trouvent un volume de 27 km³ montrant une vitesse plus lente de 10%, dans le secteur Sud-Est de la caldera, entre la surface et 2,5 km de profondeur. De par leurs caractéristiques géométriques différentes, les deux expériences fournissent des résultats qui ne se chevauchent pas. Les deux approches (réseau de stations dense / essaim de séismes dense) sont complémentaires et ont le potentiel d'atteindre la même résolution. Une autre conclusion que l'on peut tirer de notre étude est que le volume investigué ne contient probablement pas de zone étendue de stockage de magma fluide. En effet, les figures (3.2.4)-(3.2.5) montrent qu'un volume de 27 km³ montrant une anomalie de vitesse négative de 10% ne pourrait passer inaperçue dans les zones les mieux résolues et reconstituées.

3.3. Comparaison avec d'autres méthodes.

Les méthodes tomographiques présentées dans ce travail sont basées sur une théorie probabiliste du problème inverse. Nous avons également investigué les méthodes tomographiques "plus classiques" en sciences de la terre. Nous allons comparer les résultats obtenus avec les méthodes classiques d'une part et les méthodes décrites dans ce document d'autre part.

3.3.1. Méthodes classiques en tomographie sismique. En reprenant les notations de la partie 1, section 3 concernant les schémas numériques, écrivons le schéma typique utilisé en tomographie "classique" (voir à ce sujet Nolet, [33]) : étant donné un modèle initial \mathbf{m}_0 , on définit un processus itératif par la formule de récurrence suivante,

$$(3.3.1) \quad \begin{bmatrix} -\mathbf{C}_D^{-\frac{1}{2}} \mathbf{G}_k \\ \lambda \mathbf{I} \end{bmatrix} \delta_k \sim \begin{bmatrix} \mathbf{C}_D^{-\frac{1}{2}} (\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ 0 \end{bmatrix},$$

l'itéré $k+1$ est calculé en faisant : $\mathbf{m}_{k+1} = \mathbf{m}_k + \delta_k$. La matrice $\mathbf{C}_D^{-\frac{1}{2}}$ est interprétée ici, comme une matrice de poids permettant de favoriser les "bonnes" données qui seront affectées de fortes valeurs au détriment de "mauvaises" données qui seront affectées de valeurs plus faibles. Le terme λ est un facteur d'amortissement permettant de régulariser la résolution de (3.3.1). Dans le cas d'inversion conjointe des paramètres hypocentaux, des temps origine et des paramètres de vitesse, ce terme d'amortissement ne suffit pas à stabiliser le problème. Il faut, en plus, faire un changement de variable en multipliant la matrice \mathbf{G}_k par une matrice diagonale (préconditionnement à droite) \mathbf{P} . La matrice \mathbf{G}_k se présente de la façon suivante,

$$(3.3.2) \quad \mathbf{G}_k = (\mathbf{G}_v, \mathbf{G}_p, \mathbf{G}_t),$$

où \mathbf{G}_v est une sous matrice contenant les dérivées partielles par rapport aux paramètres de vitesse, \mathbf{G}_p contient les dérivées partielles par rapport aux hypocentres et \mathbf{G}_t contient les dérivées partielles par rapport aux temps origine. Notons η_v , η_p et η_t , les valeurs maximales de la norme L2 des colonnes des matrices \mathbf{G}_v , \mathbf{G}_p et \mathbf{G}_t . La matrice \mathbf{P} est définie de la façon

suivante,

$$(3.3.3) \quad \mathbf{P} = \begin{pmatrix} \frac{\alpha_v}{\eta_v} & 0 & \dots & 0 \\ 0 & \ddots & & \\ & & \frac{\alpha_v}{\eta_v} & 0 & \vdots \\ \vdots & 0 & & \frac{\alpha_p}{\eta_p} & \\ 0 & \dots & & 0 & \frac{\alpha_t}{\eta_t} \end{pmatrix},$$

où α_v, α_p et α_t sont des paramètres compris en général entre 0 et 1. La perturbation δ_k se calcule en résolvant,

$$(3.3.4) \quad \left\{ \begin{array}{l} \left[\begin{array}{c} -\mathbf{C}_D^{-\frac{1}{2}} \mathbf{G}_k \mathbf{P} \\ \lambda \mathbf{I} \end{array} \right] \Delta_k \sim \left[\begin{array}{c} \mathbf{C}_D^{-\frac{1}{2}} (\mathbf{g}(\mathbf{m}_k) - \mathbf{d}_o) \\ 0 \end{array} \right], \\ \delta_k = \mathbf{P} \Delta_k. \end{array} \right.$$

En résumé, nous avons une méthode d'inversion contrôlée par quatre paramètres ; λ , α_v, α_p et α_t .

Le schéma numérique (3.3.4) a été utilisé pour implémenter deux méthodes ; la première utilisant comme données des temps de première arrivée et la seconde utilisant des décalages temporels. Les problèmes directs associés sont les mêmes que ceux présentés dans ce document : calcul des temps de parcours par la méthode de Podvin-Lecomte [37] puis tracé du rai plus recalcul du temps de parcours par intégration numérique pour accroître la précision (cf chapitre 2 partie 1). La résolution de (3.3.4) est faite avec l'algorithme LSQR.

3.3.2. Tomographie à l'échelle régionale. L'ensemble des temps de première arrivée utilisé pour la tomographie à l'échelle régionale (cf chapitre 3 partie 1), est utilisé ici pour faire une tomographie basée sur le schéma (3.3.4). Le pas d'échantillonnage du modèle est important dans cette méthode classique. En effet, chaque cellule doit être traversée par un nombre suffisant de rais pour contraindre les paramètres de vitesse. Nous allons ici utiliser une approche multigrille ; des inversions successives sont effectuées en diminuant le pas d'échantillonnage du modèle de vitesse. Cela revient à décomposer la solution en fonction des fréquences des perturbations du modèle de vitesse ; on cherche d'abord les fluctuations basse fréquence en utilisant une maille de grande taille, et on diminue cette maille au cours du processus afin d'ajouter des composantes à plus haute fréquence. On réalise, en utilisant les mailles de grande taille, un préconditionnement du problème. On ne peut pas directement utiliser la maille la plus fine car on ne retrouve pas ainsi les fluctuations basse fréquence. La figure (3.3.2) montre une tomographie basée sur le schéma (3.3.4). Le processus débute avec une cellule de taille 16 km × 16 km × 1 km en choisissant le modèle à gradient utilisé dans la partie (3.1.3) (d'après Okubo et coll. [35], figure (3.1.1)) comme modèle initial. Le processus converge en 5 itérations. Le modèle obtenu est ré-échantillonné avec une cellule de taille 8 km × 8 km × 1 km, et sert de modèle initial pour inverser à nouveau les mêmes données. Cette démarche est appliquée successivement avec une cellule de taille 4 km × 4 km × 1 km et 2 km × 2 km × 1 km. On obtient le modèle final après 25

itérations. Les paramètres d'inversion utilisés dans cet exemple sont :

$$(3.3.5) \quad \begin{cases} \lambda = 0.1, \\ \alpha_v = 0.1, \\ \alpha_p = 1, \\ \alpha_t = 0.5. \end{cases}$$

Ces paramètres ont été choisis en testant plusieurs combinaisons. Nous avons fait varier α_v de 0,1 à 0,5 par pas de 0,1. Les données sont ajustées de la même façon avec ces différents paramètres. Les images tomographiques obtenues se ressemblent mais les variations de vitesse sont plus importantes avec des coefficients α_v forts. Ce comportement se retrouve avec les autres coefficients; la variation des paramètres correspondant à un coefficient α est une fonction croissante de celui-ci. Les coefficients α_v , α_p , et α_t agissent en quelque sorte comme des poids qui guident l'inversion en favorisant certains groupes de paramètres par rapport aux autres. Le facteur d'amortissement λ est choisi une fois les autres coefficients fixés. Si la valeur est trop faible, l'algorithme diverge, si la valeur est trop forte la solution n'ajuste plus les données. En pratique nous essayons $\lambda=0, 1, 10, 10^2, 10^3, \dots$, nous déterminons ainsi deux valeurs λ_0 et λ_1 telles que pour λ_0 la méthode diverge et converge pour λ_1 . Nous choisissons ensuite un pas $\delta\lambda$ et nous essayons les valeurs $\lambda = \lambda_1 - \delta\lambda, \lambda_1 - 2\delta\lambda, \lambda_1 - 3\delta\lambda, \dots, \lambda_1 - n\delta\lambda$, nous obtenons de la sorte d'autres valeurs λ_0 et λ_1 . On peut recommencer la procédure, ou bien choisir $\lambda = \lambda_1$. On choisit ainsi le facteur d'amortissement le plus petit possible (compte-tenu du temps disponible, étant donnée la durée du processus d'inversion).

Les coefficients d'inversion α_v , α_p , et α_t dépendent du modèle initial et le choix de l'ensemble final de coefficients dépend de l'idée *a priori* que nous nous faisons de la solution. En effet, ces critères permettent de définir les poids α_v , α_p , et α_t afin de guider la méthode vers une solution physiquement admissible. On peut donc obtenir une infinité de résultats en faisant varier ces coefficients sans qu'il soit possible de distinguer leur RMS. On peut ainsi privilégier certains paramètres du modèle par rapport à d'autres. Parmi les résultats possibles on peut obtenir des résultats comparables (figure (3.3.2)) avec ceux obtenus par ailleurs avec la méthode probabiliste (figure (3.1.8)). On reconnaît dans les deux cas les zones rapides sous le Mauna Loa, le Kilauea et le rift Est. L'intérêt premier de la méthode probabiliste est de fournir un critère objectif permettant le choix justifié du modèle final (cf 3.1.3). La méthode probabiliste est par ailleurs plus facile à mettre en oeuvre et plus économe en temps de calcul.

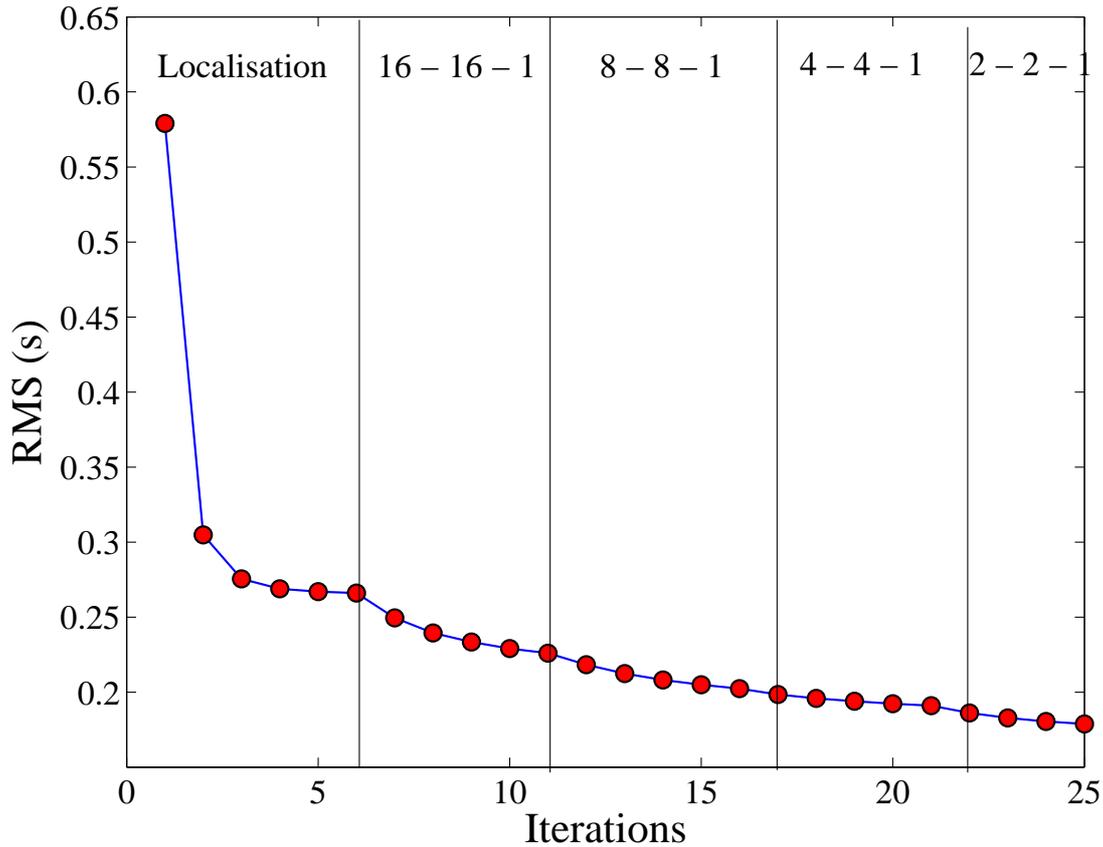
3.3.3. Tomographie à l'échelle locale. A partir du schéma numérique (3.3.4), nous avons développé une méthode tomographique utilisant des décalages temporels. Cette méthode a été utilisée avec les données du flanc sud du Kilauea. Les paramètres d'inversion sont choisis en procédant de la même façon que précédemment, lors de la tomographie à l'échelle régionale. Les figures (3.3.3)-(3.3.4) montrent le résultat obtenu avec cette méthode avec les coefficients :

$$(3.3.6) \quad \begin{cases} \lambda = 0.01, \\ \alpha_v = 0.1, \\ \alpha_p = 1, \\ \alpha_t = 1. \end{cases}$$

FIG. 3.3.1. RMS en fonction des itérations pour la méthode multigrille.

Le processus commence par une localisation préliminaire des séismes dans le modèle à gradient. Ensuite débute la tomographie multigrille avec des tailles de cellules : 1 : $16 \times 16 \times 1$; 2 : $8 \times 8 \times 1$; 3 : $4 \times 4 \times 1$; 4 : $2 \times 2 \times 1$ kilomètres. Les paramètres de l'inversion sont :

$$\begin{cases} \lambda = 0.1 \\ \alpha_v = 0.1 \\ \alpha_p = 1 \\ \alpha_t = 0.5 \end{cases}$$



Ces résultats sont à comparer avec les figures (3.2.10)-(3.2.11). Contrairement à la tomographie à l'échelle régionale, nous n'avons pas d'idée a priori du résultat final ; l'utilisation des coefficients donnés conduit à un résultat montrant de nombreuses fluctuations haute fréquence dont on ne peut apprécier la signification physique, et qui sont donc ininterprétables. Il n'y a donc pas plus d'information dans ce modèle que dans le modèle initial. Une solution serait d'introduire une contrainte de lissage à l'intérieur du système 3.3.4 (voir par exemple Nolet, [34]). Mais quelle support de lissage choisir ? pour répondre à cette question correctement, il semble opportun d'adopter une approche probabiliste.

FIG. 3.3.2. Modèle tomographique avec la méthode multigrille.

En partant du haut à gauche, les quatre premières figures montrent une carte de l'île de Hawaï à la profondeur 6 km au cours du processus tomographique multigrille. La taille des mailles sont respectivement : 1 : $16 \times 16 \times 1$; 2 : $8 \times 8 \times 1$; 3 : $4 \times 4 \times 1$; 4 : $2 \times 2 \times 1$ kilomètres. Les deux dernières figures représentent le modèle final aux profondeurs 7 et 8 km. Les paramètres de l'inversion sont :

$$\begin{cases} \lambda = 0.1 \\ \alpha_v = 0.1 \\ \alpha_p = 1 \\ \alpha_t = 0.5 \end{cases}$$

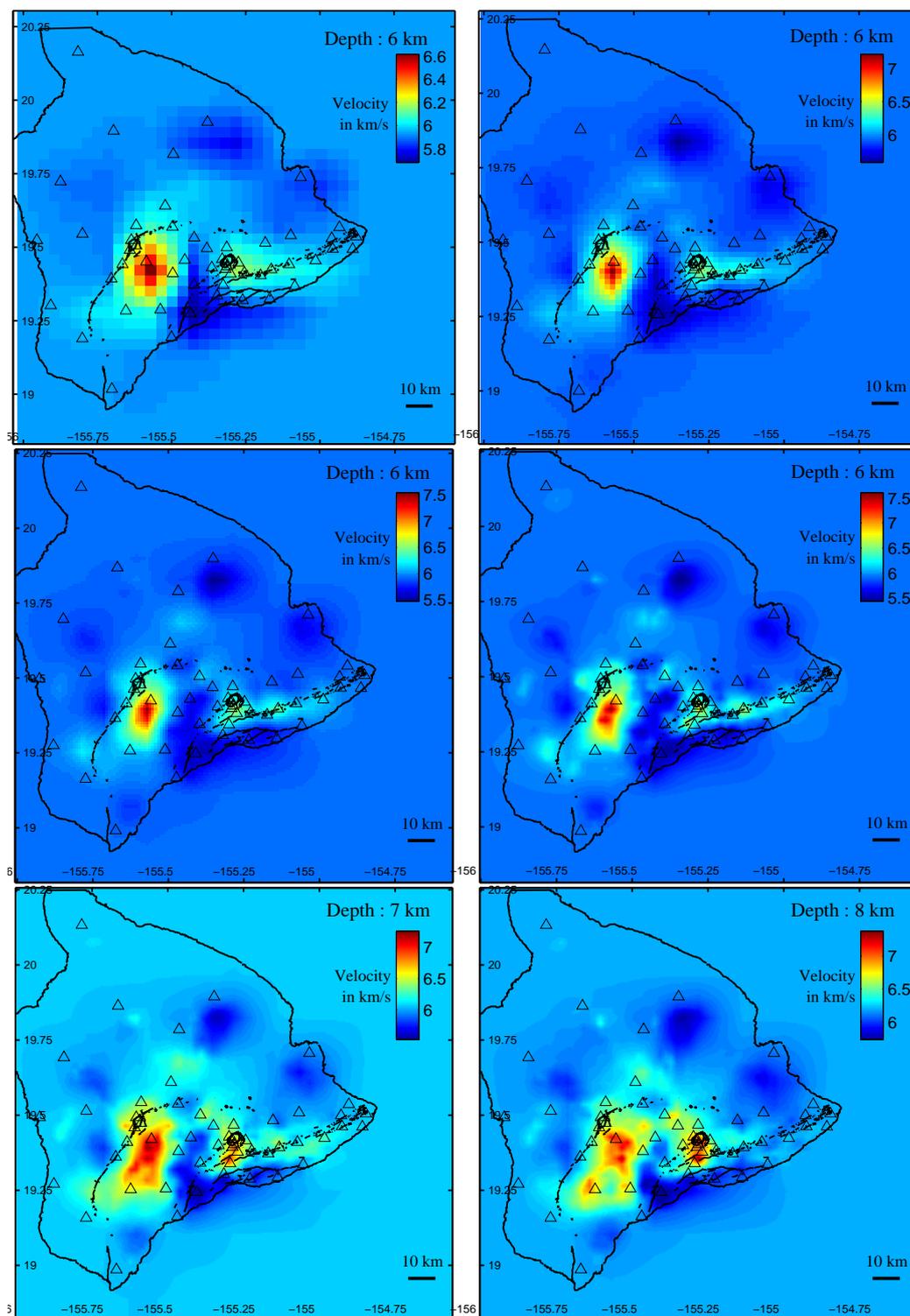


FIG. 3.3.3. Tomographie à l'échelle locale (1).

Modèle du vitesse dans le flanc sud du Kilauea, carte de 4 à 6.5 km de profondeur. Les paramètres de l'inversion sont :

$$\begin{cases} \lambda = 0.1 \\ \alpha_v = 0.1 \\ \alpha_p = 1 \\ \alpha_t = 0.5 \end{cases}$$

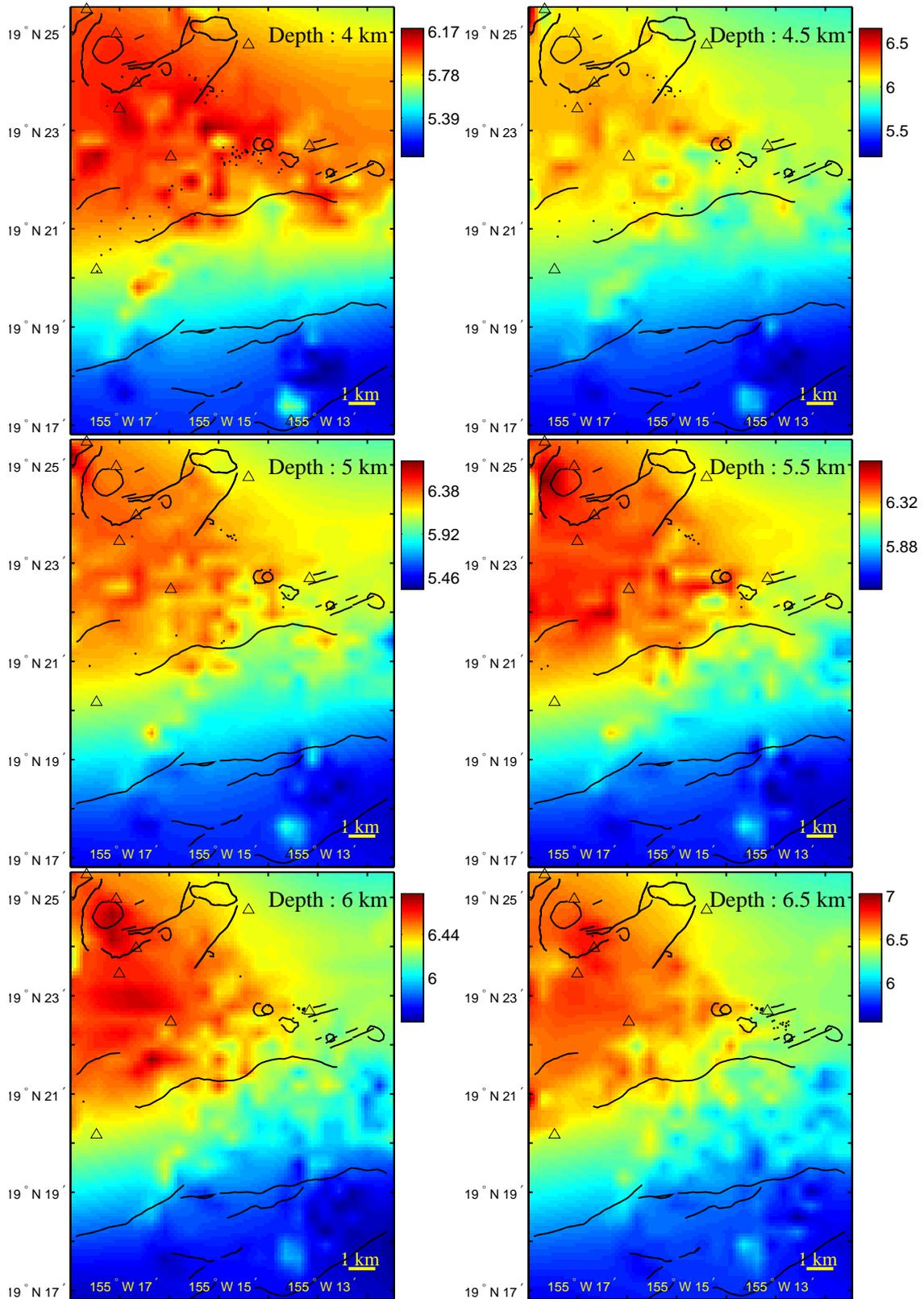
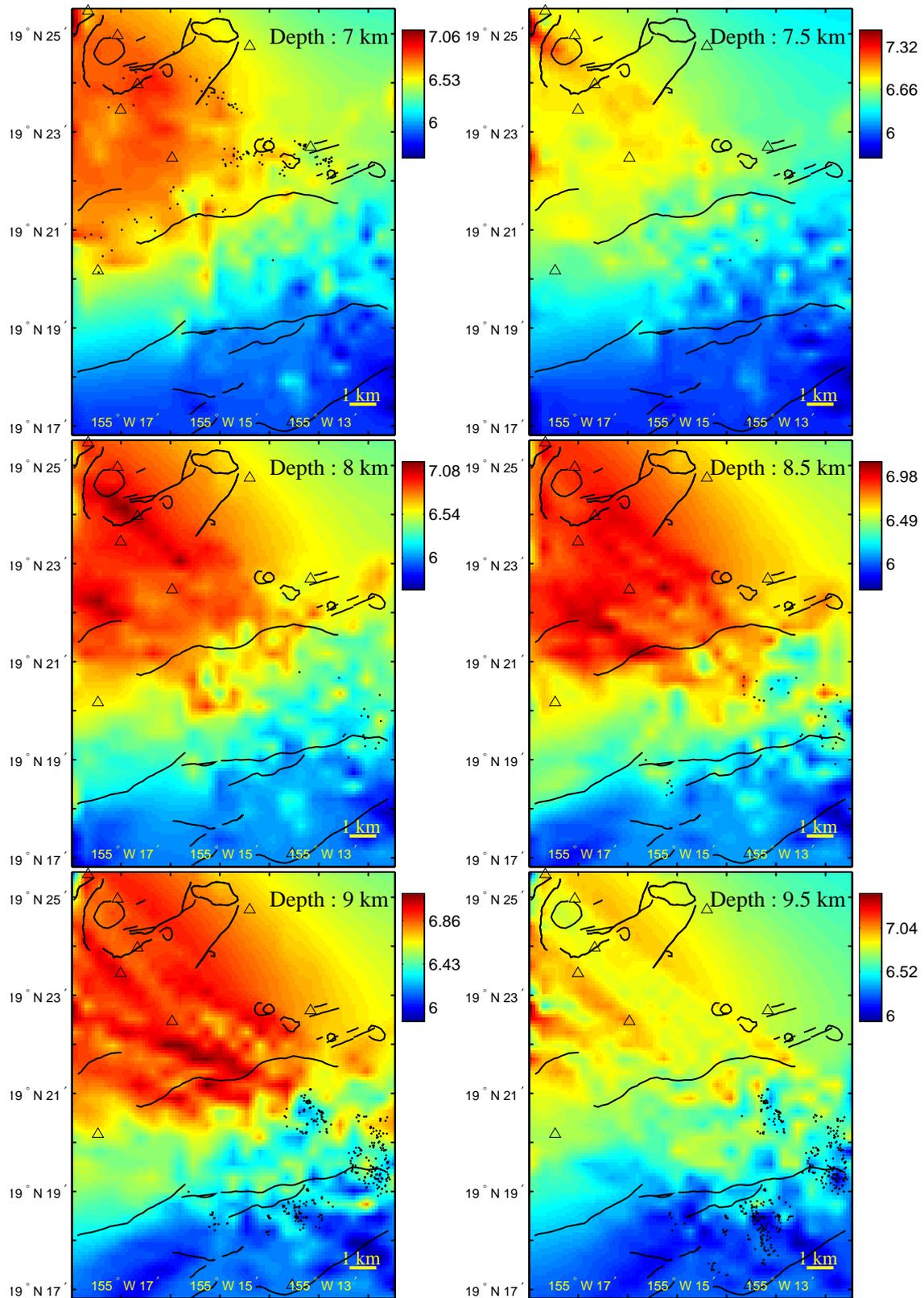


FIG. 3.3.4. Tomographie à l'échelle locale (2).

Modèle du vitesse dans le flanc sud du Kilauea, carte de 7 à 9.5 km de profondeur. Les paramètres de l'inversion sont :

$$\begin{cases} \lambda = 0.1 \\ \alpha_v = 0.1 \\ \alpha_p = 1 \\ \alpha_t = 0.5 \end{cases}$$



3.3.4. Tomographie des temps de première arrivée à l'échelle locale. Dans la partie 3.2, nous avons sélectionné un ensemble de séismes dans le flanc Sud du Kilauea qui a permis de calculer les décalages temporels inversés par la méthode probabiliste. Cet ensemble de séismes fournit aussi des temps de première arrivée que l'on peut également inverser. Les figures (3.3.5)-(3.3.6) montrent les résultats obtenus par une tomographie probabiliste des temps de première arrivée à l'échelle locale du flanc Sud du Kilauea. Les paramètres d'inversion utilisés sont :

$$(3.3.7) \quad \left\{ \begin{array}{l} \lambda = 1 \text{ km} \\ \sigma_v = 1 \text{ km/s} \\ \sigma_p = 0.02 \text{ km} \\ \sigma_t = 0.01 \text{ s.} \end{array} \right.$$

Pour obtenir des fluctuations du modèle de vitesse nous avons été obligé de contraindre les sources vers les positions *a priori* en utilisant des valeurs σ_p et σ_t faibles. Ces résultats sont à comparer avec les figures (3.2.10)-(3.2.11). Les solutions sont différentes :

- (1) La sismicité est mieux localisé avec les décalages temporels.
- (2) La zone dense n'est pas retrouvée avec les temps de première arrivée.

La différence entre les deux résultats peut s'expliquer par la différence de qualité des données. En effet, l'estimation des décalages temporels par méthode intespectrale est environ dix fois plus précise (lorsque les signaux sont semblables) que la lecture directe des temps de première arrivée pointés manuellement. Nous essayons ici d'expliquer des données (pointés manuels des temps de première arrivée) par des perturbations du modèle qui vont générer des fluctuations du résidu se confondant avec le bruit sur les données. Dans ce cas, l'inversion ne peut pas produire un modèle pertinent quelle que soit la valeur des paramètres d'inversion. Dans le cas où la valeur de σ_p est proche de l'*a priori* physique (1 km), la valeur des paramètres de vitesse reste inchangée. On n'obtient une perturbation des paramètres de vitesse que si les positions des séismes sont fortement contraintes (3.3.7) et ne sont donc elles-mêmes pas perturbées. Les données ne contiennent pas d'information qui permette de perturber le modèle de manière cohérente, c'est donc l'information *a priori* qui guide l'inversion. Nous ne pouvons donc pas interpréter les résultats des figures (3.3.5)-(3.3.6). Cela montre la nécessité de disposer, à cette échelle de données suffisamment précises comme le sont les décalages temporels utilisés dans le paragraphe 3.2.

3.3.5. Conclusion. On peut mettre en avant plusieurs différences essentielles entre la conception "classique" et le point de vue probabiliste :

- (1) Les méthodes classiques ne cherchent pas un modèle de norme minimale (cad qui minimise à la fois les fluctuations et le nombre de paramètres indépendants du modèle), qui ajuste les données.
- (2) Elles ne permettent pas de choisir un modèle sur un critère objectif d'optimisation.
- (3) Elles nécessitent la construction d'un ensemble de coefficients de préconditionnement (paramètres d'inversion), ce qui est un processus long et dépendant du modèle initial.

FIG. 3.3.5. Tomographie des temps de première arrivée à l'échelle locale (1).

Modèle du vitesse dans le flanc sud du Kilauea, carte de 4 à 6.5 km de profondeur.

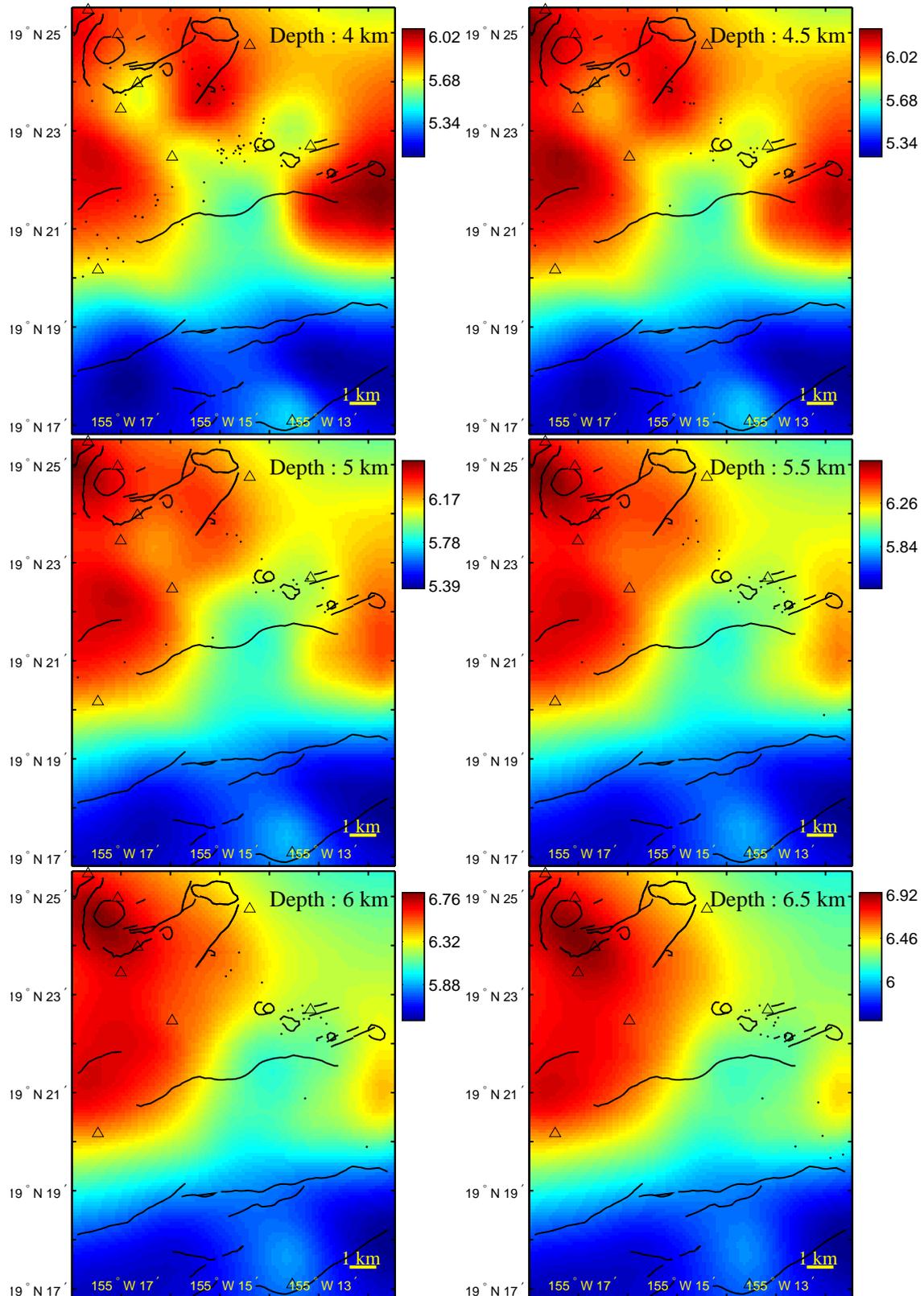
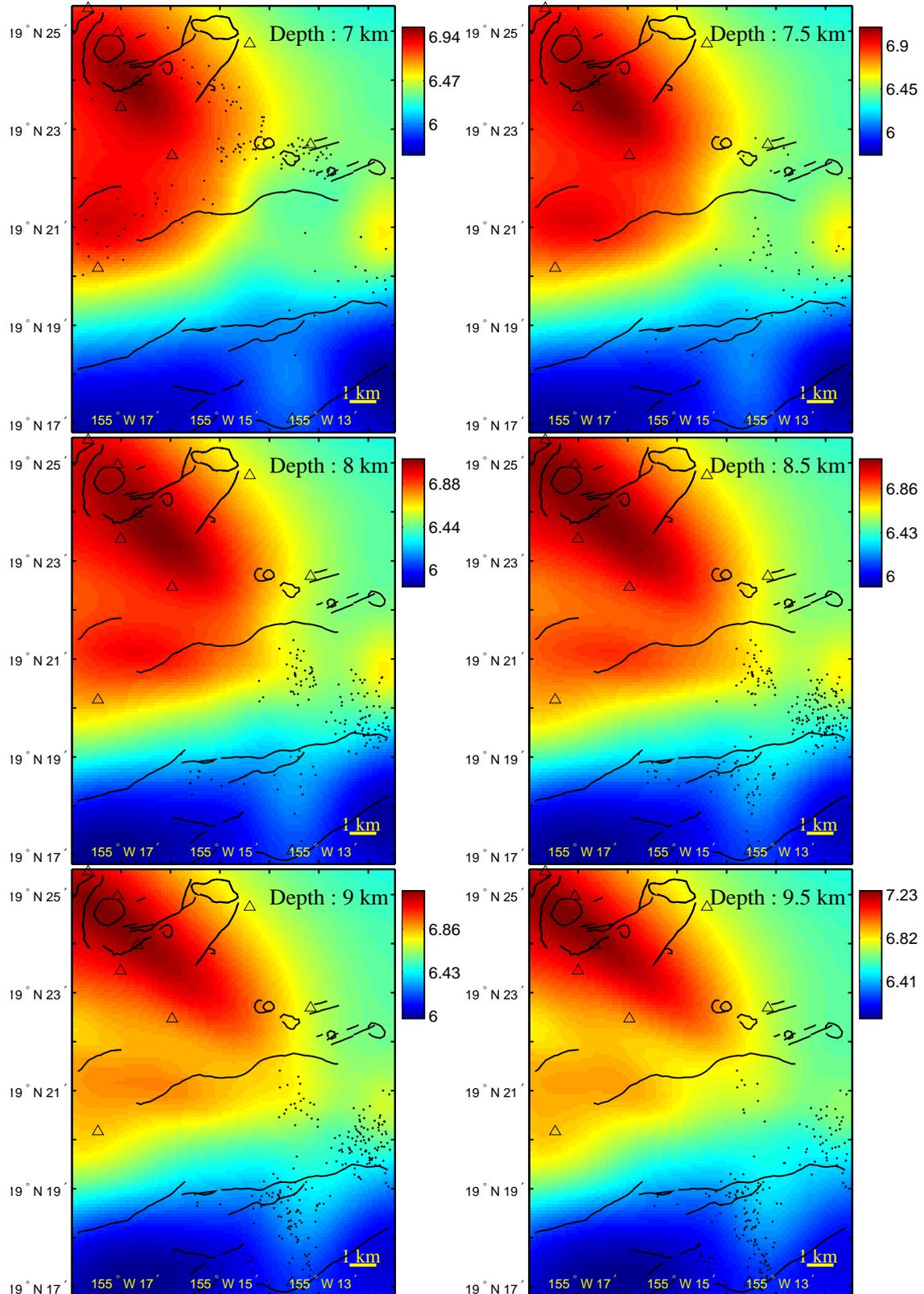


FIG. 3.3.6. Tomographie des temps de première arrivée à l'échelle locale (2).
 Modèle du vitesse dans le flanc sud du Kilauea, carte de 7 à 9.5 km de profondeur.



- (4) Elles ne permettent pas d'utiliser une loi statistique spécifique des erreurs de mesures (par ex. la loi Sech).
- (5) Dans les méthodes classiques, le choix du pas d'échantillonnage est important et rend nécessaire une approche multigrille, ce qui n'est pas le cas pour la méthode probabiliste.

Si les deux méthodes peuvent produire des solutions comparables à l'échelle régionale, ce n'est plus le cas à l'échelle locale. La solution calculée par la méthode classique n'est pas interprétable, même si elle ajuste bien les données. Le schéma numérique (3.3.4) ne convient donc pas pour une tomographie précise utilisant les décalages temporels. Ce constat nous a conduit à explorer d'autres schémas numériques d'inversions et à utiliser finalement une approche probabiliste. Par ailleurs, on constate que la précision des décalages temporels interspectraux est nécessaire pour préciser de façon cohérente les modèles de vitesse à l'échelle locale.

3.4. Conclusion.

Cette thèse avait pour objectif de mettre au point un algorithme de tomographie des décalages temporels (ou de tomographie "en doubles différences") dans un milieu fortement hétérogène comme peut l'être un volcan. Les données utilisées par cet algorithme sont des décalages temporels interspectraux mesurés avec une précision moyenne de l'ordre de 5 ms. Il est donc clair que chaque élément de l'algorithme doit atteindre une précision bien inférieure à 5 ms pour pouvoir utiliser pleinement l'information contenue dans les données. Nous avons d'abord exploré systématiquement les algorithmes des problèmes directs et inverses afin de les optimiser. Le problème direct (calcul des temps de première arrivée) a été résolu de façon robuste dans un milieu hétérogène en calculant le champ des temps de parcours par résolution de l'équation de l'Eikonale en différences finies (Podvin et Lecomte, [37]). La précision du calcul est améliorée d'un ordre de grandeur par retracé du rai et recalcul du temps de parcours le long du rai. Une précision de 1 ms est atteinte pour une discrétisation de 500 m.

Trouver une solution stable au problème inverse a été l'un des objectifs majeurs de ce travail : s'agissant d'améliorer les modèles existants, il faut s'assurer que le modèle résultant contient le moins possible de fluctuations numériques non résolues par les données. Pour cette raison, il a été choisi de suivre l'approche de Tarantola et Valette [47] et de minimiser non seulement l'écart aux données mais aussi la norme du modèle. On cherche ainsi le modèle le plus lisse, le plus simple possible qui ajuste le mieux possible les données. Cette stratégie permet de ne pas introduire de hautes fréquences non contraintes par les données - s'il existe des hautes fréquences dans le modèle, c'est qu'elles sont contraintes par les données. Pour cela, la matrice de covariance a priori du modèle a été construite à l'aide d'une fonction de corrélation dont on peut faire varier l'amplitude et la longueur de corrélation. Une approximation correcte de la racine carré de l'inverse de cette matrice de covariance a été trouvée. Une approche d'optimisation a été utilisée, qui permet de déterminer les longueurs de corrélation et amplitudes optimales de la fonction de corrélation.

Cet algorithme a été utilisé à deux échelles :

- une échelle régionale, avec des données classiques de temps d'arrivée ; le modèle trouvé montre les grands traits de l'appareil magmatique en profondeur, sous les calderas et les rifts, sous la forme de corps plus rapides (donc plus denses et plus rigides) entourés d'un remplissage lent (moins denses et moins rigides). Cette distribution de la densité et des caractéristiques rhéologiques permet d'expliquer les déformations de la croûte océanique sous-jacente, le confinement de la sismicité et les déformations intenses dans le flanc Sud du Kilauea. Ce modèle fournit un a priori pour une inversion tomographique à l'échelle locale ;

- une échelle locale, avec des données de décalages temporels interspectraux ; le modèle trouvé montre un détail de l'appareil magmatique du Kilauea, sous la forme d'un conduit magmatique vertical qui s'oriente suivant le rift Est vers 4 km de profondeur. On obtient ainsi une rms inférieure à un centième de seconde.

Dans les deux cas, la résolution de l'inversion est estimée à l'aide de tests synthétiques (damiers) ; ils permettent de délimiter les volumes où les résultats peuvent être interprétés.

CHAPITRE 4

Annexes

Annexe 1

Modélisation du milieu géophysique.

Dans le cadre de cette thèse, nous supposons que les sources sismiques engendrent des déformations faibles du milieu vis à vis de la taille du domaine considéré, de sorte que l'on puisse supposer que le sous-sol obéit aux lois de l'élasticité linéaire. Une source sismique sera alors vue comme un point géométrique du milieu en lequel une petite déformation engendre une propagation d'ondes élastiques dans le milieu.

Équation des ondes élastiques. Le milieu géophysique peut être représenté par Ω , ouvert de \mathbb{R}^3 , et par deux fonctions v_P et v_S de Ω dans \mathbb{R}^+ . v_P est la vitesse des ondes P et v_S est la vitesse des ondes S. L'équation des ondes peut alors s'écrire :

$$(4.0.1) \quad \frac{1}{v^2} \frac{\partial^2 \mathbf{u}}{\partial t^2} - \Delta \mathbf{u} = 0 \quad \text{sur } Q$$

$$(4.0.2) \quad \mathbf{u}(x, 0) = u_0 \delta_{x_0}(x) \quad \text{sur } \Omega$$

$$(4.0.3) \quad \frac{\partial \mathbf{u}}{\partial t}(x, 0) = 0 \quad \text{sur } \Omega$$

où $Q = \Omega \times [0, \infty[$, v ($=v_P$ ou $=v_S$) est la vitesse de propagation des ondes, \mathbf{u} est une fonction de Q dans \mathbb{R}^+ dans le cas des ondes P, de Q dans \mathbb{R}^3 dans le cas des ondes S (figure(4.0.1)). La source ponctuelle en x_0 est représentée par la fonction de Dirac $u_0 \delta_{x_0}$, où u_0 est une constante donnée. La résolution de ces équations peut être longue et délicate, c'est pourquoi il est courant dans les applications tomographiques de faire des hypothèses simplificatrices dans le but de définir des modèles numériquement exploitables. Ces hypothèses portent généralement sur les longueurs d'ondes et sur le milieu de manière à pouvoir négliger les phénomènes de diffraction.

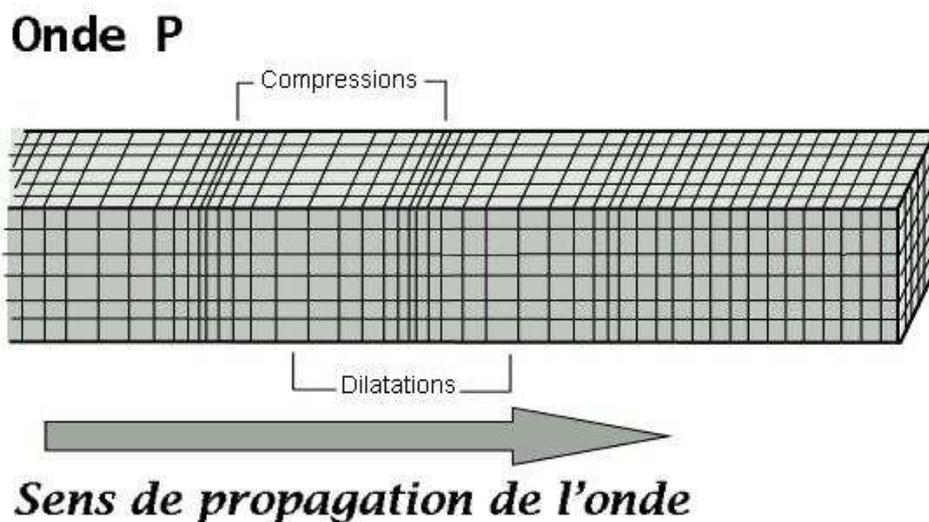
Notion de rai sismique. Soit une courbe $\gamma : \zeta \in]a, b[\mapsto \gamma(\zeta) \in \Omega$, on appelle *temps de propagation* le long de γ l'intégrale :

$$(4.0.4) \quad \tau(\gamma) = \int_{\gamma} \frac{d\sigma}{v},$$

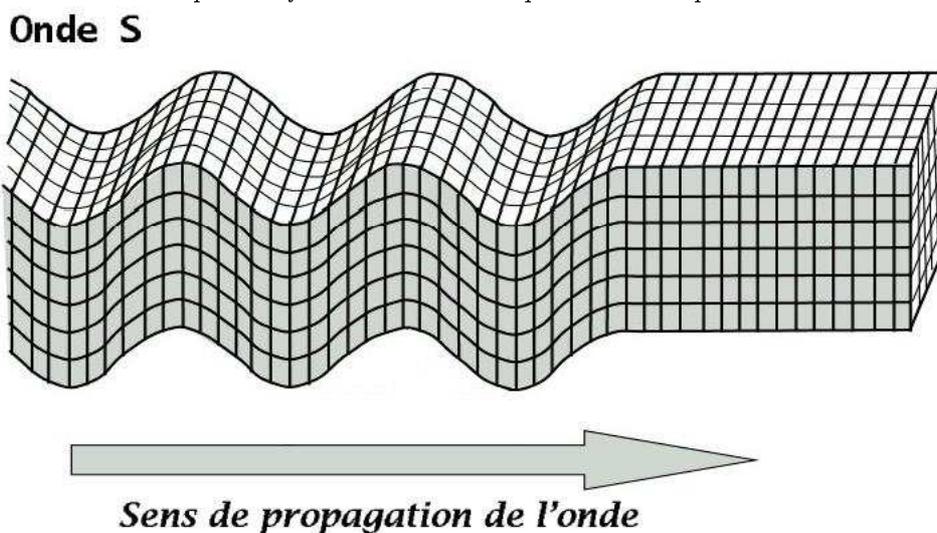
où σ est la longueur de la courbe. σ est l'abscisse curviligne définie par $\sigma := \int_a^b \|\dot{\gamma}(\zeta)\| d\zeta$. Soit $A = \gamma(a)$ et $B = \gamma(b)$, notons C_{AB} l'ensemble des courbes joignant le point A au point B. Le principe de Fermat permet de définir une courbe particulière de C_{AB} appelé *rai sismique* : c'est la courbe $\gamma \in C_{AB}$ telle que l'intégrale définie en (4.0.4) soit minimale

FIG. 4.0.1. Propagation des ondes élastiques.

Les ondes P sont modélisées par une fonction réelle qui donne la valeur de la pression en tout point du domaine.



Les ondes S sont modélisées par une fonction vectorielle qui donne le déplacement du milieu en tout point.



sur C_{AB} . En utilisant le calcul variationnel, on démontre que le rai satisfait les équations d'Euler (cf : [2] vol 1 p 94) :

$$(4.0.5) \quad \frac{d}{d\zeta} \left(\frac{1}{v(\gamma)} \frac{\dot{\gamma}}{\|\dot{\gamma}\|} \right) = \|\dot{\gamma}\| \nabla \left(\frac{1}{v(\gamma)} \right),$$

avec les conditions aux limites $\gamma(a) = A$ et $\gamma(b) = B$.

On peut donner une autre définition des rais en utilisant l'équation des ondes (4.0.1)-(4.0.3). Si on suppose que la vitesse des ondes varie faiblement localement, la longueur

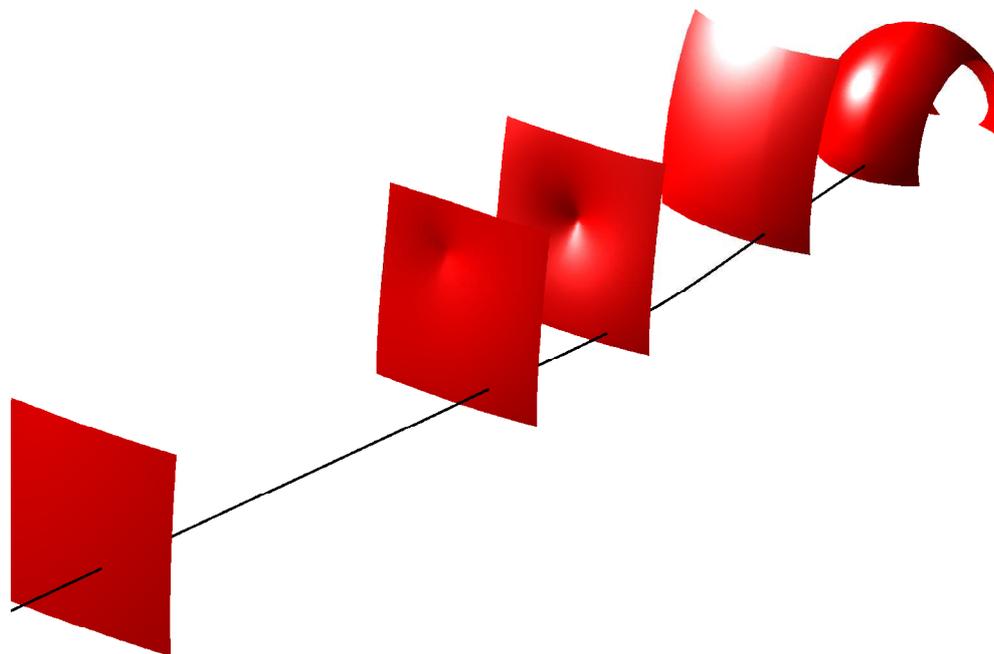
d'onde est faible en regard des hétérogénéités de vitesse et les ondes sont localement planes. On peut alors simplifier les équations (4.0.1)-(4.0.3), pour obtenir l'équation de l'eikonale ;

$$(4.0.6) \quad \nabla T^2 = \frac{1}{v^2}.$$

où \mathbf{T} est une fonction de Ω dans \mathbb{R}_+ qui représente le temps de première arrivée de l'onde sismique. On appelle *fronts d'ondes* les surfaces $T^{-1}(t)$: c'est la frontière qui, à l'instant t , sépare la partie du milieu qui a été affectée par l'onde de celle qui ne l'est pas encore. On démontre (théorème de Malus) que les rais sismiques sont perpendiculaires aux fronts d'ondes (figure(4.0.2)). Autrement dit en tout point, le rai est tangent au gradient de la

FIG. 4.0.2. **Fronts d'ondes et rai.**

Les surfaces rouges représentent les fronts d'ondes, produits par une source ponctuelle, à différents temps. La courbe noire est un rai sismique issue de la source.



fonction T (figure (4.0.3)). Le rai doit donc vérifier le système d'équation différentielle ordinaire (cf : [2] vol 1 p 91),

$$(4.0.7) \quad \dot{\gamma}(\zeta) = g(\gamma(\zeta)) \nabla T(\gamma(\zeta)),$$

avec pour condition initiale, $\gamma(\zeta_0) = \gamma_0$, où g est une fonction réelle qui est déterminée par le choix du paramétrage ζ . Par exemple si on choisit de paramétrer le rai par son abscisse

curviligne σ , le système (4.0.7) devient,

$$(4.0.8) \quad \frac{d\gamma}{d\sigma} = -v(\gamma(\sigma)) \nabla T(\gamma(\sigma)),$$

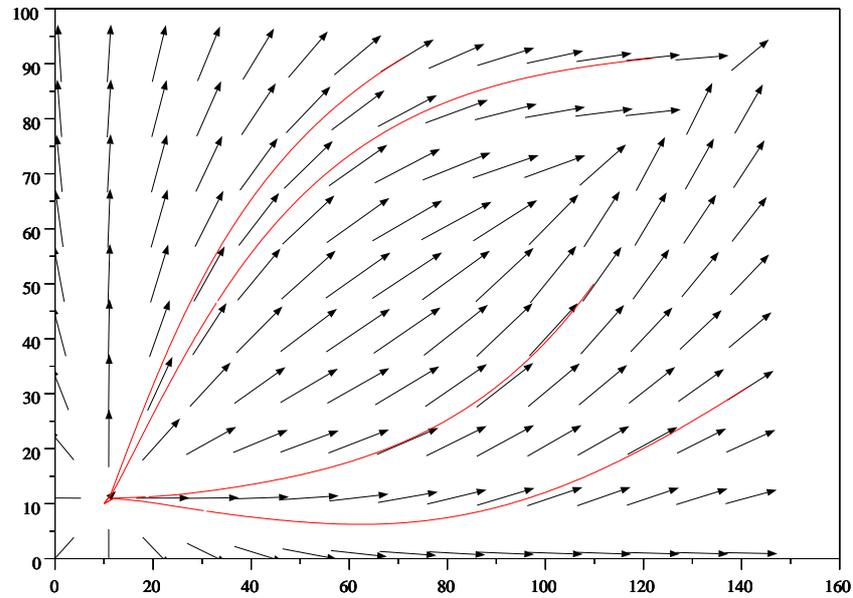
ou bien,

$$(4.0.9) \quad \frac{d\gamma}{d\sigma} = v(\gamma(\sigma)) \nabla T(\gamma(\sigma)),$$

selon le sens de parcours choisi le long du rai.

FIG. 4.0.3. Champs de vecteur et rais.

Les flèches noires représentent le champ de vecteur $g\nabla T$, (avec les notations de la relation 4.0.7). Les courbes rouges sont les trajectoires de ce champ, en sismologie elles sont appelées rais sismiques.



Lorsque on ne connaît pas le gradient de T en tout point du domaine, on peut introduire une nouvelle inconnue : le *paramètre du rai*,

$$p(\sigma) = \nabla T(\gamma(\sigma)).$$

En dérivant cette expression et en utilisant l'équation de l'eikonale on obtient,

$$\frac{dp}{d\sigma} = \nabla s,$$

où s est l'inverse de la vitesse, encore appelée *lenteur*. En reprenant (4.0.9) le rai doit vérifier le système d'équations différentielles du premier ordre (cf : [7]),

$$(4.0.10) \quad \begin{aligned} \dot{\gamma}(\sigma) &= v(\gamma(\sigma)) p(\sigma), \\ \dot{p}(\sigma) &= -\nabla s(\sigma), \\ \gamma(0) &= x_0, p(0) = p_0. \end{aligned}$$

Comme on l'a défini plus haut (4.0.4), le temps de parcours de l'onde le long d'un rai γ s'écrit,

$$\tau(\gamma) = \int_{\gamma} \frac{d\sigma}{v},$$

donc d'après la définition de σ ,

$$(4.0.11) \quad \tau(\gamma) = \int_a^b \frac{\|\dot{\gamma}(t)\|}{v(\gamma(t))} dt.$$

Méthodes numériques. Le paragraphe précédent a introduit la notion de rai sismique. Ces rais ont été caractérisés par 3 définitions (4.0.5), (4.0.7) et (4.0.10). Ces formulations sont à l'origine des différentes méthodes de tracé de rais utilisées en sismologie (figure(4.0.4)).

Tout d'abord la méthode du *bending* résout numériquement le problème aux limites d'Euler (4.0.5). C'est une méthode d'optimisation numérique ; on part d'un rai initial γ_0 , on définit une suite de rais $(\gamma_n)_{n \in \mathbb{N}}$ qui converge vers le rai γ qui minimise (4.0.4). Le calcul de la perturbation $\delta\gamma_k$ à apporter à l'itéré k se fait par l'intermédiaire des équations d'Euler linéarisées. Des problèmes de convergence peuvent apparaître lorsque le milieu de propagation est hétérogène.

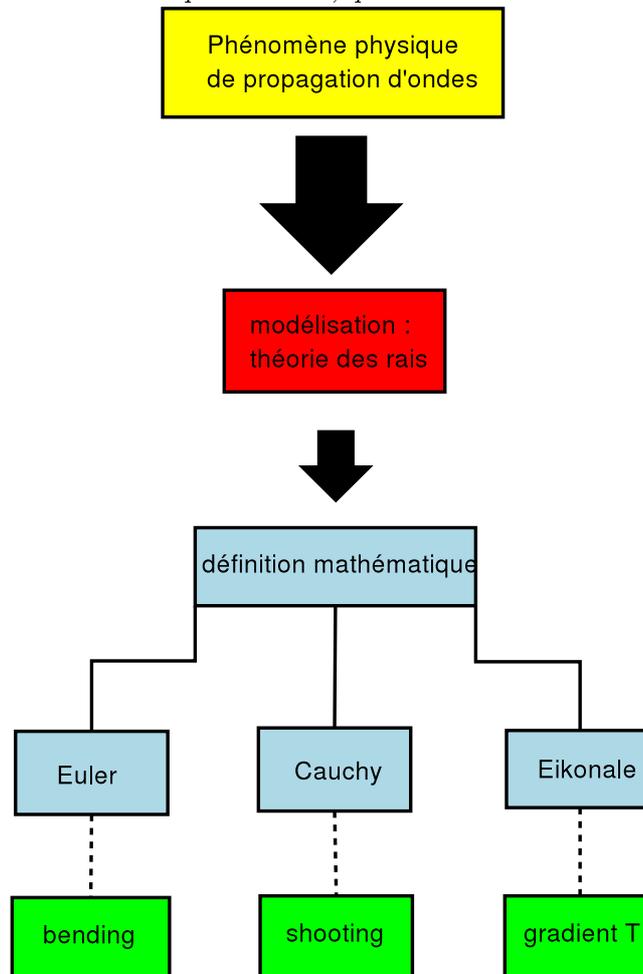
La méthode du *shooting* consiste à résoudre numériquement le problème de Cauchy (4.0.10) à partir de conditions initiales données. Cette formulation n'est pas très pratique dans les applications sismologiques, car en général on cherche le rai joignant deux points du domaine. On connaît donc la position initiale, mais on ne connaît pas le paramètre du rai initial qui va assurer que le rai rejoindra le deuxième point du domaine. On résout les équations avec plusieurs paramètres initiaux différents jusqu'à ce que le rai tombe suffisamment près du deuxième point. On peut ne pas trouver de paramètre de rai satisfaisant, et dans ce cas on ne peut pas tracer le rai correspondant.

La méthode de tracé de rais choisie pour traiter le problème direct dans cette thèse utilise la formulation (4.0.7). Pour cela il faut résoudre l'équation de l'Eikonale (4.0.6). Cette résolution suit l'algorithme dû à P.Podvin et I.Lecomte ([37]) qui résout numériquement, en différences finies, l'équation de l'Eikonale en propageant localement le front d'onde, suivant le principe de Huygens. Cet algorithme permet donc de trouver le champ des temps de parcours du premier train d'onde émis par une source ponctuelle donnée ; il

est très stable, et une fois la solution trouvée, il est facile et toujours possible de tracer le rai joignant la source à un point quelconque du domaine¹.

FIG. 4.0.4. Diagramme récapitulant différentes façons de concevoir le problème direct.

Le phénomène physique de propagation des ondes est simplifié, il est modélisé par théorie des rais. Il existe plusieurs définitions mathématiques d'un rai, qui donnent toutes lieu à une méthode numérique.



¹La méthode est détaillée au chapitre 2, paragraphe 1.

Annexe 2

Décalages temporels interspectraux.

Soient deux signaux x_1 et x_2 , notons \widehat{x}_1 la transformée de Fourier de x_1 et \widehat{x}_2 la transformée de Fourier de x_2 ; on peut écrire

$$\widehat{x}_1(\omega) = a_1(\omega) e^{-i(\omega t + \varphi_1)},$$

et

$$\widehat{x}_2(\omega) = a_2(\omega) e^{-i(\omega t + \varphi_2)}.$$

donc

$$\widehat{x}_2(\omega) = a(\omega) \widehat{x}_1(\omega) e^{-i(\varphi_2 - \varphi_1)},$$

où $a(\omega) e^{-i(\varphi_2 - \varphi_1)}$ est la fonction de transfert qui permet de passer de \widehat{x}_1 à \widehat{x}_2 . La quantité $\varphi = \varphi_2 - \varphi_1$ est le déphasage entre x_2 et x_1 ; c'est aussi la phase de l'interspectre $\widehat{x}_1 \widehat{x}_2^*$. Le modèle du filtre de Wiener (qui s'applique lorsque les signaux x_1 et x_2 sont très semblables), qui est un filtre linéaire, implique que la phase est une fonction linéaire de la fréquence, $\varphi = \omega\theta = 2\pi f\theta$. Pratiquement, le calcul du décalage temporel θ s'effectue par ajustement linéaire au sens des moindres carrés de la phase φ de l'interspectre. On déduit θ de la pente p de cette phase : $\theta = \frac{p}{2\pi}$.

La *méthode interspectrale* appliquée aux signaux sismologiques met en oeuvre le principe exposé ci-dessus pour deux signaux semblables enregistrés par une station. Considérons deux séismes, notons t_a^1 et t_a^2 les temps d'arrivée des séismes 1 et 2, t_o^1 et t_o^2 leur temps origine. Nous devons positionner une fenêtre d'étude définie pour chaque signal. Notons t_f^1 et t_f^2 le début de chaque fenêtre. Le décalage temporel θ_{12} entre les séismes 1 et 2 se décompose (figure (4.0.5)),

$$(4.0.12) \quad \theta_{12} = (t_a^2 - t_a^1) - (t_f^2 - t_f^1).$$

En introduisant le temps origine,

$$(4.0.13) \quad \theta_{12} = (t_a^2 - t_o^2 - t_a^1 + t_o^1) - (t_f^2 - t_o^2 - t_f^1 + t_o^1).$$

En notant t_p^i le temps de parcours de l'onde du séisme i ,

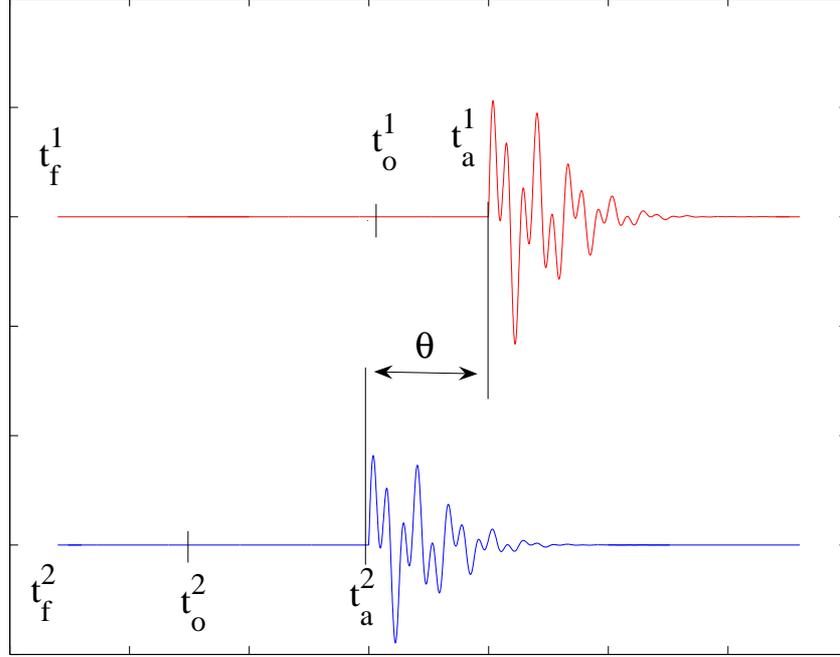
$$(4.0.14) \quad \theta_{12} = (t_p^2 - t_p^1) - (t_f^2 - t_o^2 - (t_f^1 - t_o^1)).$$

Nous remarquons que si l'origine des fenêtres coïncide avec le temps origine, le décalage temporel θ_{12} est égal à la différence de temps de propagation $\theta_{12}^p = (t_p^2 - t_p^1)$. En pratique, dans le cas d'un essaim de séismes, le positionnement des fenêtres se fait en utilisant un temps origine estimé \widetilde{t}_o^i pour chaque séisme. On calcule ensuite un temps de propagation

FIG. 4.0.5. Application de la méthode interspectrale aux signaux sismologiques.

La figure montre deux signaux semblables enregistrés par une station. Une origine est choisie pour chaque signal, notée t_f^i ; le temps de déclenchement du séisme (inconnu) est noté t_o^i et le temps d'arrivée est noté t_a^i . La méthode interspectrale mesure le décalage temporel θ entre deux signaux semblables, qui est la quantité,

$$\theta = (t_a^2 - t_f^2) - (t_a^1 - t_f^1).$$



moyen $t_p^c = \frac{1}{n} \sum_{i=1}^{i=n} \tilde{t}_p^i$, où \tilde{t}_p^i est le temps de parcours estimé du séisme i . Le début de la fenêtre se positionne à l'instant $t_f^i = \tilde{t}_o^i + t_p^c$. Nous devons alors considérer le décalage temporel comme la somme,

$$(4.0.15) \quad \theta_{12} = \theta_{12}^p + \delta\tau_{12},$$

où $\delta\tau_{12} = (t_f^2 - t_f^1) - (t_o^2 - t_o^1) = (\tilde{t}_o^2 - \tilde{t}_o^1) - (t_o^2 - t_o^1)$ est un terme correctif dû aux erreurs de positionnement des fenêtres, qui sera considéré comme un paramètre à chercher dans un processus d'inversion. Ce terme représente l'erreur sur la différence de temps origine.

On définit la fonction de cohérence des signaux x_1 et x_2 ,

$$(4.0.16) \quad C(\omega) = \frac{\overline{\hat{x}_1(\omega) \hat{x}_2^*(\omega)}}{\sqrt{\overline{\hat{x}_1(\omega) \hat{x}_1^*(\omega)}} \sqrt{\overline{\hat{x}_2(\omega) \hat{x}_2^*(\omega)}}},$$

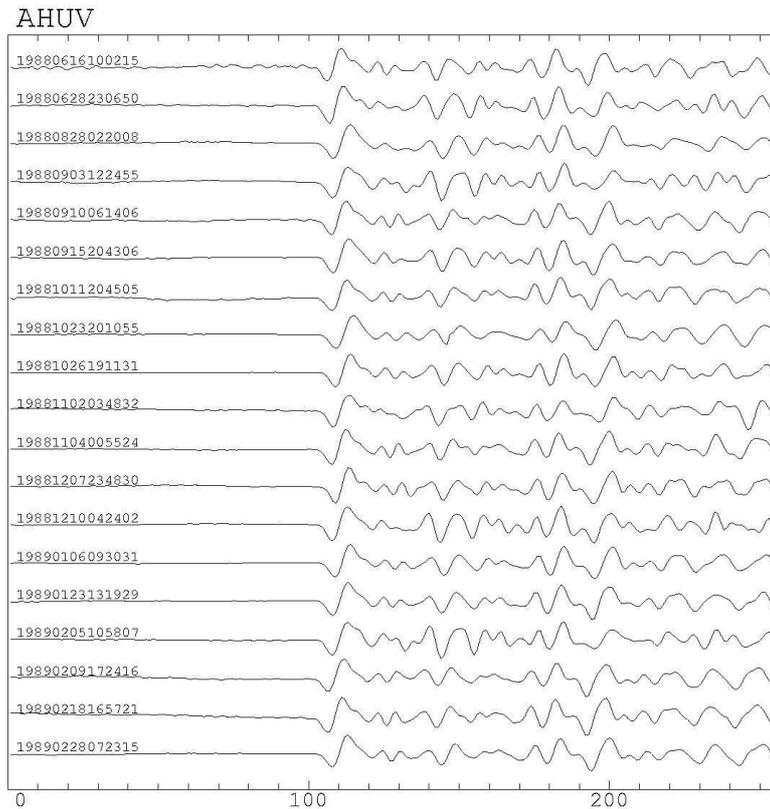
où $\overline{\hat{x}_1(\omega) \hat{x}_2^*(\omega)}$ est l'interspectre lissé de \hat{x}_1 et \hat{x}_2 , et $\sqrt{\overline{\hat{x}_1(\omega) \hat{x}_1^*(\omega)}}$ l'autospectre lissé de \hat{x}_1 . On calcule la moyenne \overline{C} de cette fonction de cohérence sur un support contenant le maximum d'énergie. Nous définissons ainsi pour chaque décalage temporel θ_{12} une valeur de

cohérence qui est comprise entre 0 et 1 ; cette quantité vaut 1 pour des signaux semblables et 0 pour des signaux totalement décorrélés. Cette valeur de cohérence est utilisée pour estimer une incertitude σ_{12} sur le décalage temporel par la formule suivante,

$$(4.0.17) \quad \sigma_{12} = \begin{cases} \sigma \frac{\bar{C}^2}{1-\bar{C}^2} \frac{1-C_{max}^2}{C_{max}^2} \text{ si } \bar{C} < C_{max} \\ \sigma \text{ si } \bar{C} \geq C_{max} \end{cases} ,$$

où C_{max} est une valeur de cohérence prise proche de 1 ($C_{max} = 0.995$ dans nos exemples sur l'île de Hawaii) et σ ($\sigma = 5.10^{-4}$ secondes dans nos exemples sur l'île de Hawaii (figure (4.0.6))) est l'erreur estimé sur les décalages temporels présentant une cohérence de valeur C_{max} . Cette valeur de σ indique que l'erreur ne peut être nulle, même pour une valeur unité de \bar{C} , car cette quantité est elle-même estimée avec une incertitude (c'est-à-dire qu'une valeur unité de \bar{C} ne peut être que fortuite pour des signaux réels entachés de bruit, et il n'y pas de raison de leur donner un poids infini). La valeur σ sert de référence pour établir la statistique des erreurs de mesure d'un ensemble de décalages temporels dont les incertitudes seront calculées par la formule (4.0.17).

FIG. 4.0.6. **Signaux semblables.**
Signaux semblables enregistrés par une station du réseau du HVO.



Annexe 3

Equations en doubles différences.

Ces dernières années, les méthodes de localisation qualifiées de *doubles différences* (e.g. Jordan and Sverdrup [22], Poupinet et coll. [38], Got et coll. [15], Shearer [43], Waldhauser et Ellsworth [52], voir Wolfe [54] pour une revue.) ont permis dans certains cas d'améliorer la connaissance de la position des séismes. Le terme *doubles différences* est utilisé en référence au système d'équations linéaires résolu à chaque itération de l'algorithme. Nous allons expliciter ces équations. Posons d'abord le problème de localisation classique. Il s'agit de calculer les hypocentres de n_s séismes, enregistrées par n_t stations. Le séisme i est représenté par ses coordonnées hypocentriques (x_1^i, x_2^i, x_3^i) et son temps origine τ^i . L'ajustement des données se fait aux sens des moindres carrés. On utilise un algorithme itératif, qui calcule la perturbation $(\Delta x_1^i, \Delta x_2^i, \Delta x_3^i, \Delta \tau^i)$ à appliquer sur chaque séisme i en résolvant les équations,

$$(4.0.18) \quad r_k^i = \sum_{l=1}^{l=3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_l^i + \Delta \tau^i, \quad i = 1, \dots, n_s; k = 1, \dots, n_t,$$

en notant T_k^i le temps de parcours de l'onde émise par la source i et enregistrée par station k . Le résidu r_k^i est la différence entre la donnée observée et le temps théorique,

$$(4.0.19) \quad r_k^i = (T_k^i)^{obs} - T_k^i.$$

Pour une station k , il est possible de former la différence entre les équations (4.0.18) correspondantes aux séismes i et j ,

$$(4.0.20) \quad r_k^i - r_k^j = \sum_{l=1}^{l=3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_l^i + \Delta \tau^i - \sum_{l=1}^{l=3} \frac{\partial T_k^j}{\partial x_l^j} \Delta x_l^j - \Delta \tau^j.$$

Le terme de gauche de l'équation précédente (4.0.20) est qualifié de *double différence*, puisque on peut le réécrire comme une "différence de différences",

$$(4.0.21) \quad r_k^i - r_k^j = (T_k^i - T_k^j)^{obs} - (T_k^i - T_k^j).$$

Waldhauser et Ellsworth [52] définissent à partir de l'équation (4.0.20) un système d'équations en *doubles différences*,

$$(4.0.22) \quad \begin{cases} r_k^i - r_k^j = \sum_{l=1}^{l=3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_l^i + \Delta \tau^i - \sum_{l=1}^{l=3} \frac{\partial T_k^j}{\partial x_l^j} \Delta x_l^j - \Delta \tau^j, \\ \text{avec } (i, j, k) \in \mathfrak{S} \subset \{1, \dots, n_s\} \times \{1, \dots, n_s\} \times \{1, \dots, n_t\}. \end{cases}$$

qu'ils résolvent en ajoutant à (4.0.22) quatre lignes supplémentaires définissant une contrainte de conservation barycentrique,

$$(4.0.23) \quad \left\{ \begin{array}{l} \sum_{i=1}^{i=n_s} \Delta x_1^i = 0 \\ \sum_{i=1}^{i=n_s} \Delta x_2^i = 0 \\ \sum_{i=1}^{i=n_s} \Delta x_3^i = 0 \\ \sum_{i=1}^{i=n_s} \Delta \tau^i = 0 \end{array} \right. ,$$

en suivant un algorithme itératif.

Le principe de faire des différences est appliqué par Zhang et Thurber [56] pour définir une tomographie en *doubles différences*. Supposons que le modèle de vitesse soit divisé en n_v blocs de vitesse constantes (v_1, \dots, v_{n_v}) . Le système d'équations de tomographie classique s'écrit alors,

$$(4.0.24) \quad r_k^i = \sum_{l=1}^{l=3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_i + \Delta \tau^i + \sum_{m=1}^{m=n_v} \frac{\partial T_k^i}{\partial v_m}, i = 1, \dots, n_s; k = 1, \dots, n_t.$$

En formant des différences, on obtient,

$$(4.0.25) \quad \left\{ \begin{array}{l} r_k^i - r_k^j = \sum_{l=1}^{l=3} \frac{\partial T_k^i}{\partial x_l^i} \Delta x_i + \Delta \tau^i - \sum_{l=1}^{l=3} \frac{\partial T_k^j}{\partial x_l^j} \Delta x_j - \Delta \tau^j + \sum_{m=1}^{m=n_v} \left(\frac{\partial T_k^i}{\partial v_m} - \frac{\partial T_k^j}{\partial v_m} \right), \\ \text{avec } (i, j, k) \in \mathfrak{S} \subset \{1, \dots, n_s\} \times \{1, \dots, n_s\} \times \{1, \dots, n_t\}. \end{array} \right.$$

La solution proposée par Zhang et Thurber [55] combine les systèmes d'équations (4.0.24) et (4.0.25) pour construire un algorithme itératif ajustant les temps et différences de temps de propagation.

Annexe 4

Eléments d'algèbre linéaire.

Notations. Nous noterons $\mathcal{M}_{m,n}(\mathbb{R})$ l'ensemble des matrices à coefficients réels ayant m lignes et n colonnes. L'image d'une matrice $\mathbf{A} \in \mathcal{M}_{m,n}(\mathbb{R})$ est un sous espace vectoriel de \mathbb{R}^m , noté $\text{im}(\mathbf{A})$, est l'ensemble

$$(4.0.26) \quad \text{im}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m, \text{ tel que } \exists \mathbf{x} \in \mathbb{R}^n \text{ avec } \mathbf{y} = \mathbf{A}\mathbf{x}\}.$$

Le noyau de la matrice \mathbf{A} est un sous espace vectoriel de \mathbb{R}^n , noté $\text{ker}(\mathbf{A})$, est l'ensemble

$$(4.0.27) \quad \text{ker}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n, \text{ tel que, } \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

Moindres carrés. Soit une matrice \mathbf{A} dans $\mathcal{M}_{m,n}(\mathbb{R})$ et un vecteur \mathbf{b} dans \mathbb{R}^m . Le système linéaire

$$(4.0.28) \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

possède une solution au sens classique lorsque \mathbf{b} est dans l'image de \mathbf{A} , $\text{im}(\mathbf{A})$. Dans le cas contraire, si \mathbf{b} n'est pas dans $\text{im}(\mathbf{A})$, il n'existe pas de vecteur \mathbf{x} dans \mathbb{R}^n , tel que (4.0.28) soit vérifiée. L'idée est alors de chercher les vecteurs \mathbf{x} tels que $\mathbf{A}\mathbf{x}$ soit le plus proche possible de \mathbf{b} . En utilisant la norme L2, nous définissons ainsi le problème de moindres carrés,

$$(4.0.29) \quad \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|,$$

que l'on notera aussi,

$$(4.0.30) \quad \mathbf{A}\mathbf{x} \sim \mathbf{b}.$$

Les solutions de (4.0.29) et (4.0.30) vérifient les équations normales

$$(4.0.31) \quad \mathbf{A}^t \mathbf{A}\mathbf{x} = \mathbf{A}^t \mathbf{b}.$$

Si $\text{ker}(\mathbf{A}) = \{0\}$, le problème possède une solution unique, \mathbf{x}_s . Si $\text{ker}(\mathbf{A}) \neq \{0\}$, il existe une infinité de solutions $\mathbf{x}_s + \text{ker}(\mathbf{A}) = \{\mathbf{x}_s + \mathbf{v}, \mathbf{v} \in \text{ker}(\mathbf{A})\}$, les paramètres de $\text{ker}(\mathbf{A})$ ne sont pas contraints par les équations.

Décomposition SVD.

THEOREM. Soit \mathbf{A} une matrice de $\mathcal{M}_{m,n}(\mathbb{C})$ ($m > n$) ayant r valeurs singulières strictement positives. Il existe deux matrices unitaires $\mathbf{U} \in \mathcal{M}_{n,n}(\mathbb{C})$ et $\mathbf{V} \in \mathcal{M}_{m,m}(\mathbb{C})$ et une matrice diagonale $\tilde{\mathbf{S}} \in \mathcal{M}_{m,n}(\mathbb{R})$ telles que

$$\mathbf{A} = \mathbf{V}\tilde{\mathbf{S}}\mathbf{U}^*, \quad \tilde{\mathbf{S}} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

où $\mathbf{S} \in \mathcal{M}_{r,r}(\mathbb{R})$ est une matrice diagonale, formée des r valeurs singulières de A strictement positives. Par convention, ces valeurs singulières seront classées ainsi :

$$\mu_1 \geq \mu_2 \geq \dots \geq \mu_r > 0.$$

On a $\mathbf{A}^*\mathbf{A} = \mathbf{U}\tilde{\mathbf{S}}^t\tilde{\mathbf{S}}\mathbf{U}^*$. Les colonnes de la matrice \mathbf{U} sont donc les vecteurs propres de la matrice hermitienne $\mathbf{A}^*\mathbf{A}$ et les éléments diagonaux de la matrice carrée $\tilde{\mathbf{S}}^t\tilde{\mathbf{S}} \in \mathcal{M}_{n,n}(\mathbb{R})$ sont les valeurs propres de $\mathbf{A}^*\mathbf{A}$, c'est à dire les carrés des valeurs singulières de la matrice \mathbf{A} . D'autre part $\mathbf{A}\mathbf{A}^* = \mathbf{V}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^t\mathbf{V}^*$. Les colonnes de la matrice \mathbf{V} sont donc les valeurs propres de la matrice hermitienne $\mathbf{A}\mathbf{A}^*$ et les éléments diagonaux de la matrice carrée $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^t \in \mathcal{M}_{m,m}(\mathbb{R})$ sont les valeurs propres de $\mathbf{A}\mathbf{A}^*$. Comme on a supposé $m > n$, les m valeurs propres de $\mathbf{A}\mathbf{A}^*$ sont les n valeurs μ_i^2 et $(m - n)$ fois la valeur propre nulle.

Le problème de moindres carrés (4.0.29) se réécrit en utilisant la décomposition SVD,

$$(4.0.32) \quad \min_{\mathbf{x} \in \mathbb{R}^n} \left\| \mathbf{V}\tilde{\mathbf{S}}\mathbf{U}^*\mathbf{x} - \mathbf{b} \right\|.$$

Or, nous savons que la matrice \mathbf{V} est unitaire ($\mathbf{V}^*\mathbf{V} = \mathbf{V}\mathbf{V}^* = \mathbf{I}$) ainsi,

$$(4.0.33) \quad \left\| \mathbf{V}\tilde{\mathbf{S}}\mathbf{U}^*\mathbf{x} - \mathbf{b} \right\| = \left\| \tilde{\mathbf{S}}\mathbf{U}^*\mathbf{x} - \mathbf{V}^*\mathbf{b} \right\|.$$

En posant $\mathbf{y} = \mathbf{U}^*\mathbf{x}$ et $\mathbf{d} = \mathbf{V}^*\mathbf{b}$, le problème (4.0.29) se ramène à trouver les solutions de

$$(4.0.34) \quad \min_{\mathbf{y} \in \mathbb{R}^n} \left\| \tilde{\mathbf{S}}\mathbf{y} - \mathbf{d} \right\|,$$

et faire le changement de variable $\mathbf{x} = \mathbf{U}\mathbf{y}$. Si nous décomposons le vecteur $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$, où

\mathbf{y}_1 représente les r premières coordonnées. De la même façon, on décompose $\mathbf{d} = \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{pmatrix}$.

Avec ces notations,

$$(4.0.35) \quad \left\| \tilde{\mathbf{S}}\mathbf{y} - \mathbf{d} \right\| = \left\| \begin{pmatrix} \mathbf{S}\mathbf{y}_1 \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{pmatrix} \right\|,$$

le minimum de l'expression précédente est atteint lorsque $\mathbf{S}\mathbf{y}_1 = \mathbf{d}_1$. Nous remarquons de plus que le problème ne contraint pas \mathbf{y}_2 . Ainsi, les solutions du problème (4.0.34) sont de la forme $\begin{pmatrix} \mathbf{S}^{-1}\mathbf{d}_1 \\ \mathbf{y}_2 \end{pmatrix}$, où \mathbf{y}_2 est un élément arbitraire. Les solutions de (4.0.29) sont

donc de la forme, $\begin{pmatrix} \mathbf{US}^{-1}\mathbf{d}_1 \\ \mathbf{y}_2 \end{pmatrix}$. On appelle *solution de norme minimale*, la solution de composante nulle sur le noyau de \mathbf{A} ,

$$(4.0.36) \quad \mathbf{x}_s = \begin{pmatrix} \mathbf{US}^{-1}\mathbf{d}_1 \\ 0 \end{pmatrix}.$$

Pseudo-inverse. On appelle matrice pseudo-inverse de \mathbf{A} , la matrice $\mathbf{A}^\dagger \in \mathcal{M}_{n,m}(\mathbb{C})$ définie par $\mathbf{A}^\dagger = \mathbf{US}^\dagger \mathbf{V}^*$, avec

$$\tilde{\mathbf{S}}^\dagger = \begin{pmatrix} \mathbf{S}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}(\mathbb{R}).$$

On vérifie aisément que la solution de norme minimale (4.0.36) s'exprime grâce au pseudo-inverse,

$$(4.0.37) \quad \mathbf{x}_s = \mathbf{A}^\dagger \mathbf{b}.$$

Dans certaines applications, la matrice \mathbf{A} est connue avec une précision finie, cela pose des problèmes pour les valeurs singulières les plus faibles. En effet, le bruit numérique de la matrice \mathbf{A} ne permet pas de calculer les plus petites valeurs singulières avec suffisamment de précision. Il faut alors fixer ces valeurs singulières à zéro. Cela revient en quelque sorte à définir un noyau "numérique" destiné à stabiliser la solution. Pour calculer la solution de norme minimale, nous définissons alors un pseudo-inverse, $\mathbf{A}^{\dagger,k} = \mathbf{US}_k^\dagger \mathbf{V}^*$, avec

$$\tilde{\mathbf{S}}_k^\dagger = \begin{pmatrix} \mathbf{S}_k^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{n,m}(\mathbb{R}),$$

où $\mathbf{S}_k^{-1} \in \mathcal{M}_{k,k}(\mathbb{R})$ est une matrice diagonale constituée des inverses des k premières valeurs singulières. La solution de norme minimale s'écrit alors,

$$\mathbf{x}_s^k = \mathbf{A}^{\dagger,k} \mathbf{b}.$$

En effet si la valeur singulière μ_{k+1} est très faible et mal connue, la solution \mathbf{x}_s^{k+1} peut alors être erronée car la valeur $\frac{1}{\mu_{k+1}}$ est grande et fautive, si bien qu'elle peut générer de fortes fluctuations dans le calcul de la solution

$$\mathbf{x}_s^{k+1} = \mathbf{A}^{\dagger,k+1} \mathbf{b}.$$

Annexe 5

Eléments de Calcul différentiel.

Le problème inverse défini dans cette thèse consiste à minimiser une fonction aux moindres carrés non linéaire, qui s'écrit, en notant les résidus $\mathbf{r}(\mathbf{m}) = (r_1(\mathbf{m}), \dots, r_i(\mathbf{m}), \dots, r_n(\mathbf{m}))^t$,

$$f(\mathbf{m}) = \frac{1}{2} \|\mathbf{r}(\mathbf{m})\|^2 = \sum_{i=1}^m \frac{1}{2} r_i^2(\mathbf{m}),$$

avec $\mathbf{m} = (m_1, \dots, m_l, \dots, m_n)^t$, le vecteur représentant le modèle. Les algorithmes d'optimisation à directions de descente nécessitent de connaître le gradient, ∇f et le hessien, $\nabla^2 f$. Nous allons expliciter ici ces quantités.

Gradient. Le gradient de f calculé en \mathbf{m} se décompose en fonction des gradient des résidus,

$$(4.0.38) \quad \nabla f(\mathbf{m}) = \sum_{i=1}^m \nabla \left\{ \frac{1}{2} r_i^2(\mathbf{m}) \right\},$$

or,

$$(4.0.39) \quad \nabla \left\{ \frac{1}{2} r_i^2(\mathbf{m}) \right\} = r_i(\mathbf{m}) \nabla r_i(\mathbf{m}),$$

ainsi,

$$(4.0.40) \quad \nabla f(\mathbf{m}) = \sum_{i=1}^m r_i(\mathbf{m}) \nabla r_i(\mathbf{m}).$$

Sous forme matricielle, en notant \mathbf{G} la matrice jacobienne de $\mathbf{m} \mapsto \mathbf{r}(\mathbf{m})$, le gradient de f s'exprime,

$$(4.0.41) \quad \nabla f(\mathbf{m}) = \mathbf{G}^t \mathbf{r}(\mathbf{m}).$$

Hessien. Le hessien de f calculé en \mathbf{m} se décompose en fonction des hessiens des résidus,

$$(4.0.42) \quad \nabla^2 f(\mathbf{m}) = \sum_{i=1}^m \nabla^2 \left\{ \frac{1}{2} r_i^2(\mathbf{m}) \right\},$$

or,

$$(4.0.43) \quad \frac{\partial^2 \left(\frac{1}{2} r_i^2(\mathbf{m}) \right)}{\partial m_k \partial m_l} = \frac{\partial}{\partial m_k} \left(\frac{\partial \left(\frac{1}{2} r_i^2(\mathbf{m}) \right)}{\partial m_l} \right) = \frac{\partial r_i(\mathbf{m})}{\partial m_k} \frac{\partial r_i(\mathbf{m})}{\partial m_l} + \frac{\partial^2 r_i(\mathbf{m})}{\partial m_k \partial m_l},$$

anisi,

$$(4.0.44) \quad \nabla^2 \left\{ \frac{1}{2} r_i^2(\mathbf{m}) \right\} = \nabla r_i(\mathbf{m}) (\nabla r_i(\mathbf{m}))^t + r_i(\mathbf{m}) \nabla^2 r_i(\mathbf{m}).$$

Sous forme matricielle le hessien de f s'écrit,

$$(4.0.45) \quad \nabla^2 f(\mathbf{m}) = \mathbf{G}^t \mathbf{G} + \sum_{i=1}^{i=m} r_i(\mathbf{m}) \nabla^2 r_i(\mathbf{m}).$$

Application. La méthode de Newton pour optimiser la fonction f , consiste à calculer une perturbation δ de l'itéré courant \mathbf{m} en résolvant le système suivant,

$$(4.0.46) \quad \nabla^2 f(\mathbf{m}) \delta = -\nabla f(\mathbf{m}),$$

c'est à dire,

$$(4.0.47) \quad \left(\mathbf{G}^t \mathbf{G} + \sum_{i=1}^{i=m} r_i(\mathbf{m}) \nabla^2 r_i(\mathbf{m}) \right) \delta = -\mathbf{G}^t \mathbf{r}(\mathbf{m}).$$

L'algorithme de Gauss-Newton est une variante qui nous évite de calculer les termes $\nabla^2 r_i(\mathbf{m})$, en effet la perturbation δ vérifie,

$$(4.0.48) \quad (\mathbf{G}^t \mathbf{G}) \delta = -\mathbf{G}^t \mathbf{r}(\mathbf{m}).$$

Dans cette thèse la fonction utilisée contient en plus un terme incluant un modèle *a priori* \mathbf{m}_0 ,

$$(4.0.49) \quad E(\mathbf{m}) = f(\mathbf{m}) + \frac{1}{2} \left\| \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m} - \mathbf{m}_0) \right\|^2 = \frac{1}{2} \left\| \mathbf{C}_d^{-\frac{1}{2}} \mathbf{r}(\mathbf{m}) \right\|^2 + \frac{1}{2} \left\| \mathbf{C}_m^{-\frac{1}{2}} (\mathbf{m} - \mathbf{m}_0) \right\|^2,$$

où \mathbf{C}_d est la matrice de variance-covariance (supposée diagonale) sur les données et \mathbf{C}_m est la matrice de variance-covariance sur le modèle *a priori* \mathbf{m}_0 . D'après la formule (4.0.41),

$$(4.0.50) \quad \nabla E(\mathbf{m}) = \mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{r}(\mathbf{m}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0).$$

De la même façon, le hessien se calcule en utilisant la formule (4.0.47),

$$(4.0.51) \quad \nabla^2 E(\mathbf{m}) = \mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{G} + \sum_{i=1}^{i=m} \frac{1}{\sigma_i} r_i(\mathbf{m}) \nabla^2 r_i(\mathbf{m}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0),$$

où σ_i sont les éléments diagonaux de la matrice \mathbf{C}_d . L'algorithme de Gauss-Newton appliqué à la fonction E , nous amène donc à résoudre,

$$(4.0.52) \quad (\mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{G} + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0)) \delta = \mathbf{G}^t \mathbf{C}_d^{-1} \mathbf{r}(\mathbf{m}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0).$$

Annexe 6

Matrice de covariance.

Dans cette thèse, le schéma d'inversion tomographique utilise explicitement une matrice $\mathbf{C}_m^{-\frac{1}{2}}$ qui est la racine carrée de l'inverse de la matrice de covariance *a priori* \mathbf{C}_m définie au paragraphe (2.5.4). Elle s'écrit de la façon suivante,

$$(4.0.53) \quad \mathbf{C}_m = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,n} \\ & \ddots & \sigma_{i,j} \\ \vdots & \sigma_{i,i} & \vdots \\ & \sigma_{j,i} & \ddots \\ \sigma_{n,1} & \cdots & \sigma_{n,n} \end{pmatrix}.$$

La covariance $\sigma_{i,j}$ entre les noeuds \mathbf{i} et \mathbf{j} de coordonnées \mathbf{X}_i et \mathbf{X}_j s'écrit,

$$(4.0.54) \quad \sigma_{i,j} = \sigma(v_i) \sigma(v_j) \exp\left(-\sqrt{\langle \mathbf{C}^{-1}(\mathbf{X}_i - \mathbf{X}_j), \mathbf{X}_i - \mathbf{X}_j \rangle}\right),$$

où \mathbf{C} est une matrice diagonale,

$$(4.0.55) \quad \mathbf{C} = \begin{pmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & \lambda_z \end{pmatrix}$$

Pour des raisons de ressources informatiques, il n'est pas possible de calculer $\mathbf{C}_m^{-\frac{1}{2}}$ par des méthodes classiques d'analyse numérique (par exemple, Lanczos,...). En effet, le modèle de vitesse utilisé pour la tomographie de l'île de Hawaii, comprend $161 \times 161 \times 37 = 959077$ mailles, la matrice \mathbf{C}_m est donc composée de l'ordre 10^{12} éléments (\mathbf{C}_m occuperait donc 4 To). A l'heure actuelle on ne peut pas traiter informatiquement de telles matrices pour calculer $\mathbf{C}_m^{-\frac{1}{2}}$. Néanmoins en exploitant quelques propriétés de la matrice \mathbf{C}_m , on peut proposer un algorithme qui calcule une approximation de $\mathbf{C}_m^{-\frac{1}{2}}$. Nous allons voir comment procéder dans le cas de modèles à une dimension d'espace et nous extrapolerons au cas à trois dimensions d'espace.

1) Structure de la matrice $\mathbf{C}_m^{-\frac{1}{2}}$. Considérons un milieu à une dimension discrétisé avec n mailles. Les coefficients de \mathbf{C}_m s'écrivent,

$$(4.0.56) \quad \sigma_{i,j} = \sigma(v_i) \sigma(v_j) \exp\left(-\frac{|x_i - x_j|}{\lambda}\right).$$

Nous supposons que pour tous les noeuds i et j ,

$$(4.0.57) \quad \sigma(v_i) = \sigma(v_j) = \sigma,$$

ainsi (4.0.56) devient,

$$(4.0.58) \quad \sigma_{i,j} = \sigma^2 \exp\left(-\frac{|x_i - x_j|}{\lambda}\right).$$

Pour des valeurs de n inférieures à 1000 et dans le cas unidimensionnel, il est possible de calculer $\mathbf{C}_m^{-\frac{1}{2}}$ par la méthode de Lanczos. On écrit la décomposition en valeurs propres et vecteurs propres,

$$(4.0.59) \quad \mathbf{C}_m = \mathbf{U}\mathbf{S}\mathbf{U}^t,$$

ainsi,

$$(4.0.60) \quad \mathbf{C}_m^{-\frac{1}{2}} = \mathbf{U}\mathbf{S}^{-\frac{1}{2}}\mathbf{U}^t.$$

Cette méthode a été appliquée sur de nombreux exemples, cela nous a permis de tirer de manière empirique des propriétés remarquables de $\mathbf{C}_m^{-\frac{1}{2}}$. Sur la figure (4.0.7), nous voyons que la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ possède des éléments de grande valeur absolue près de la diagonale; en s'éloignant de celle-ci, les coefficients tendent vers zéro. D'un point de vue numérique, nous approximerons les éléments éloignés de la diagonale par des zéros, ainsi nous obtenons une matrice creuse. Cette propriété est favorable puisque nous n'aurons à stocker qu'un petit nombre de coefficients. De plus, nous remarquons (figure (4.0.7)) que la structure de la matrice ne dépend pas de λ ainsi nous savons *a priori* où se trouvent les coefficients non nuls. Il s'agit maintenant de définir les valeurs que prennent les coefficients de $\mathbf{C}_m^{-\frac{1}{2}}$.

2) Coefficients de $\mathbf{C}_m^{-\frac{1}{2}}$.

Quelques notations. Nous noterons la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ de la façon suivante,

$$(4.0.61) \quad \mathbf{C}_m^{-\frac{1}{2}} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_{1,n} \\ & \ddots & \alpha_{i,j} \\ \vdots & & \alpha_i & \vdots \\ & \alpha_{j,i} & \ddots & \\ \alpha_{n,1} & \cdots & & \alpha_n \end{pmatrix}.$$

Nous appellerons diagonale supérieure n_d , les éléments de $\mathbf{C}_m^{-\frac{1}{2}}$ vérifiant,

$$(4.0.62) \quad \{\alpha_{i,j} \text{ tel que } i \in \{1, \dots, n - n_d\} \text{ et } j = i + n_d\}.$$

Nous examinerons la matrice $\mathbf{C}_m^{-\frac{1}{2}}$ selon ses diagonales supérieures, la matrice étant symétrique ce point de vue suffit à son étude.

Structures des diagonales supérieures. La figure (4.0.8) montre les coefficients de la diagonale, nous pouvons distinguer cinq valeurs différentes :

FIG. 4.0.7. Structure de $\mathbf{C}_m^{-\frac{1}{2}}$.

Quelques matrices \mathbf{C}_m (à gauche) et $\mathbf{C}_m^{-\frac{1}{2}}$ (à droite) en fonction de λ . Nous remarquons que $\mathbf{C}_m^{-\frac{1}{2}}$ est une matrice creuse dont la structure ne dépend pas de λ . Seules les valeurs des coefficients varient avec λ .

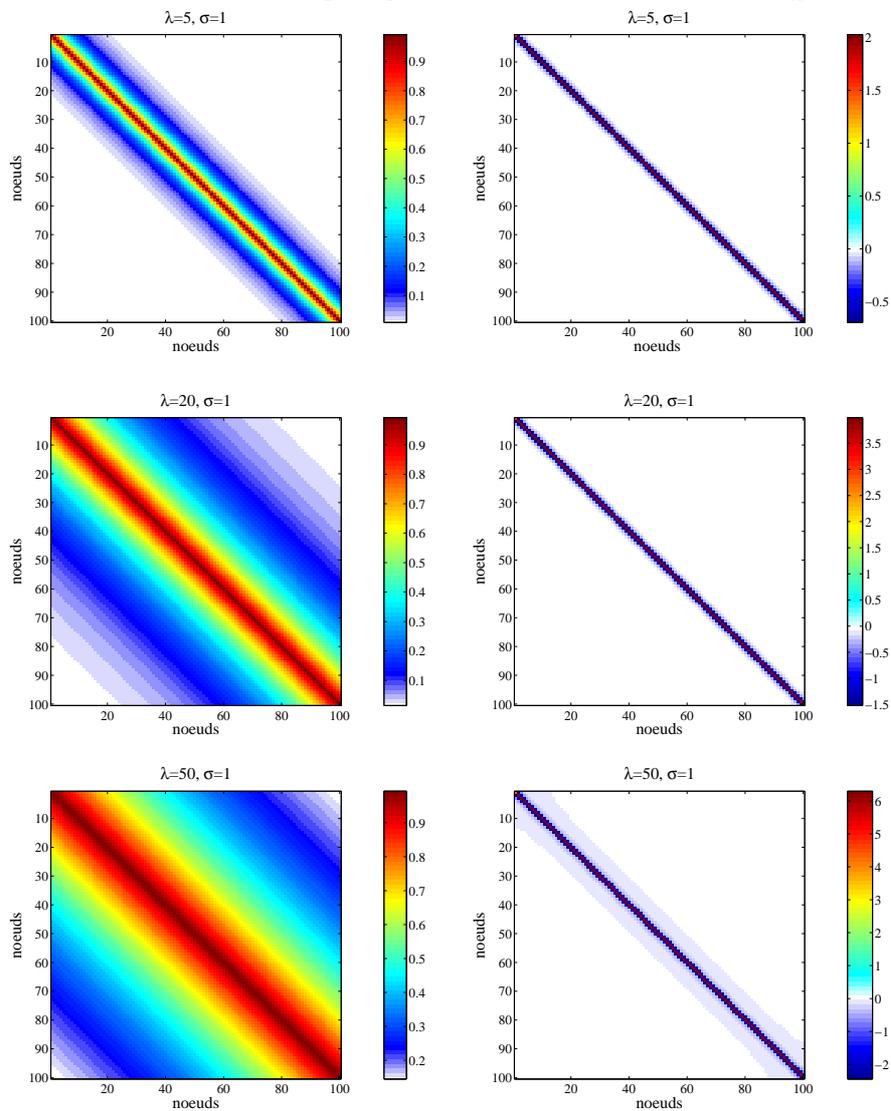
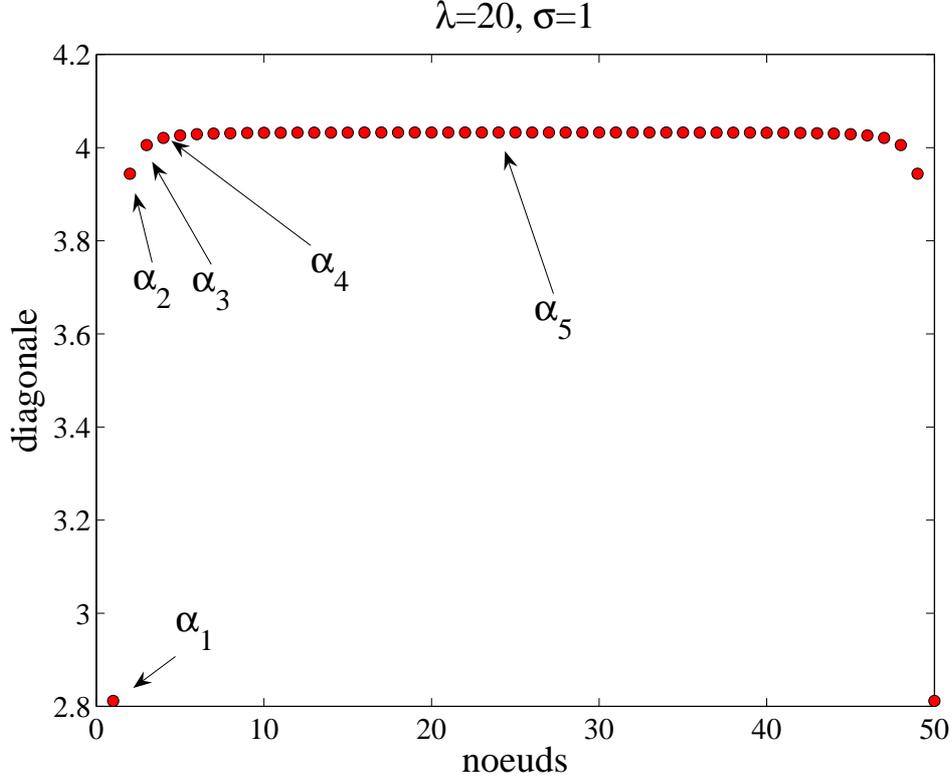


FIG. 4.0.8. Diagonale de $C_m^{-\frac{1}{2}}$.

Diagonale de la matrice $C_m^{-\frac{1}{2}}$ pour $\lambda = 20$, $\sigma = 1$ et $n = 50$. La valeur des coefficients est représentée en ordonnée, en abscisse on lit la ligne (égale à la colonne) correspondant au coefficient. Chaque ligne de la matrice correspond à un noeud du modèle.



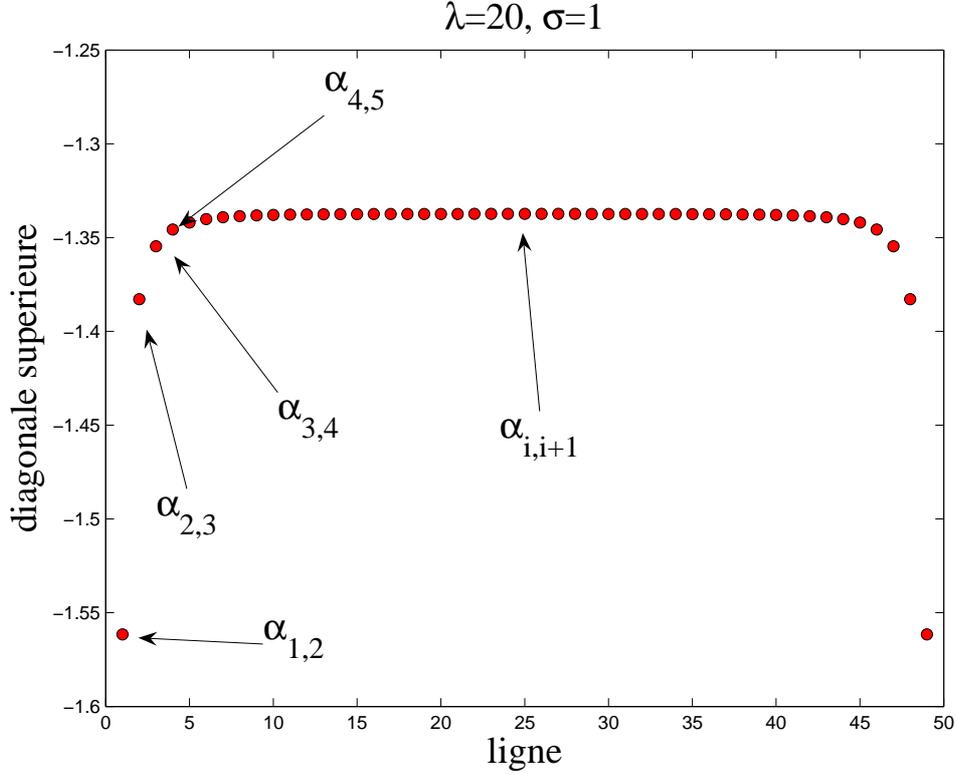
$$(4.0.63) \quad \begin{cases} \alpha_1 & \text{pour } i = 1, i = n, \\ \alpha_2 & \text{pour } i = 2, i = n - 1, \\ \alpha_3 & \text{pour } i = 3, i = n - 2, \\ \alpha_4 & \text{pour } i = 4, i = n - 3, \\ \alpha_5 & \text{pour } i \in \{5, \dots, n - 4\}. \end{cases}$$

La figure (4.0.9) montre la diagonale supérieure $n_d = 1$, on distingue encore cinq coefficients différents,

$$(4.0.64) \quad \begin{cases} \alpha_{1,2} & \text{pour } i = 1, i = n - 1, \text{ et } j = i + 1 \\ \alpha_{2,3} & \text{pour } i = 2, i = n - 2, \text{ et } j = i + 1 \\ \alpha_{3,4} & \text{pour } i = 3, i = n - 3, \text{ et } j = i + 1 \\ \alpha_{4,5} & \text{pour } i = 4, i = n - 4, \text{ et } j = i + 1 \\ \alpha_{5,6} & \text{pour } i \in \{5, \dots, n - 5\}, \text{ et } j = i + 1. \end{cases}$$

FIG. 4.0.9. Diagonale supérieure $n_d = 1$ de $C_m^{-\frac{1}{2}}$.

Diagonale supérieure $n_d = 1$ de la matrice $C_m^{-\frac{1}{2}}$ pour $\lambda = 20$, $\sigma = 1$ et $n = 50$. La valeur des coefficients est représentée en ordonnée, en abscisse on lit la ligne (égale à la colonne-1) correspondant au coefficient. Chaque ligne de la matrice correspond à un noeud du modèle.



En continuant de la sorte, la diagonale supérieure n_d a aussi la même structure, cinq coefficients différents ordonnés de la même manière :

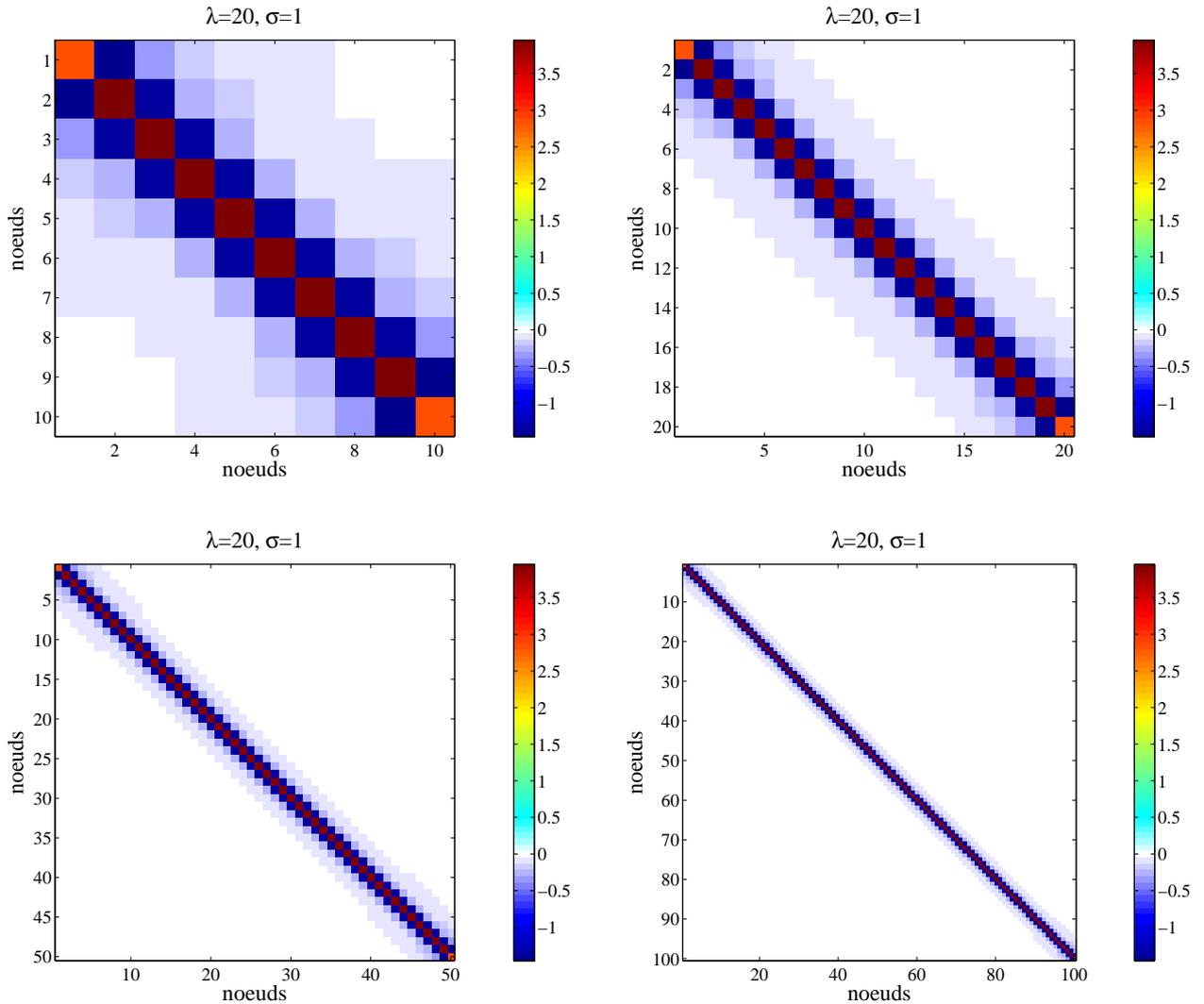
$$(4.0.65) \quad \begin{cases} \alpha_{1,1+n_d} & \text{pour } i = 1, i = n - n_d, \text{ et } j = i + n_d \\ \alpha_{2,2+n_d} & \text{pour } i = 2, i = n - n_d - 1, \text{ et } j = i + n_d \\ \alpha_{3,3+n_d} & \text{pour } i = 3, i = n - n_d - 2, \text{ et } j = i + n_d \\ \alpha_{4,4+n_d} & \text{pour } i = 4, i = n - n_d - 3, \text{ et } j = i + n_d \\ \alpha_{5,5+n_d} & \text{pour } i \in \{5, \dots, n - n_d - 4\}, \text{ et } j = i + n_d. \end{cases}$$

De nombreux autres calculs effectués avec diverses valeurs de n et λ , nous ont montré des configurations similaires à l'exemple considéré ici. En général, sur chaque diagonale supérieure, il ya entre 5 et 10 valeurs différentes. Au centre toutes les valeurs sont identiques, α_c . Une valeur particulière α_e est définie sur les extrémités. On observe un effet de bord matérialisé par des coefficients $\{\alpha_b^1, \dots, \alpha_b^8\}$ de valeurs différentes tendant vers la valeur centrale α_c en s'éloignant de l'extrémité. Selon les cas nous avons observé entre 3 et 8 valeurs dûes aux effets de bords.

3) Propriétés de $C_m^{-\frac{1}{2}}$. Une autre observation importante a été faite. La figure (4.0.10) montre quatre matrices $C_m^{-\frac{1}{2}}$ calculées numériquement avec $\lambda = 20$ et $\sigma = 1$, et $n = 10, 20, 50$ et 100 . On remarque que, dans une plage de valeur assez large (voir 5)), la valeur des coefficients ne dépend pas de n : il existe une plage de valeurs de n dans laquelle, les coefficients observé ne dépendent que de λ et σ . En résumé, nous supposons vraies les

FIG. 4.0.10. Coefficients de $C_m^{-\frac{1}{2}}$.

Matrices et $C_m^{-\frac{1}{2}}$ en fonction du nombre n de noeuds du modèle. Les calculs sont effectués pour $\lambda = 20, \sigma = 1$ et $n=10, 20, 50$ et 100 . On remarque que la valeur des coefficients ne dépend pas de n .



propriétés² suivantes : il existe deux entiers n_{min} et n_{max} tels que pour $n_{min} \leq n \leq n_{max}$, $\mathbf{C}_m^{-\frac{1}{2}}$ vérifie ,

- (1) la structure de $\mathbf{C}_m^{-\frac{1}{2}}$ est diagonale par bandes, indépendante de n ;
- (2) la valeur des coefficients non nuls ne dépend que de λ et σ ;
- (3) sur la diagonale et sur les diagonales supérieures et inférieures le nombre de coefficients différents est de l'ordre de 5 à 10 selon les cas.

4) Principe de l'algorithme. Le but de l'algorithme est de calculer une approximation de $\mathbf{C}_m^{-\frac{1}{2}}$, pour λ , σ et n donnés. Supposons que n soit trop grand pour être traité informatiquement. Nous allons donc considérer une matrice $\mathbf{C}_{0,m}^{-\frac{1}{2}}$ pour une valeur de $n_0 \ll n$ de manière à pouvoir la calculer numériquement par l'algorithme de Lanczos. Les coefficients de $\mathbf{C}_{0,m}^{-\frac{1}{2}}$ sont *a priori* les mêmes que ceux de $\mathbf{C}_m^{-\frac{1}{2}}$ (Propriété 2) . La propriété (1) indique où se trouvent les coefficients non nuls de $\mathbf{C}_m^{-\frac{1}{2}}$. Nous allons donc extraire de chaque diagonale supérieure de $\mathbf{C}_{0,m}^{-\frac{1}{2}}$, les coefficients (Propriété 3) et nous les redistribuerons dans une matrice $\widetilde{\mathbf{C}}_m^{-\frac{1}{2}}$, approximation de $\mathbf{C}_m^{-\frac{1}{2}}$. Ainsi, le calcul des coefficients est rapide ($n_0 \ll n$), ainsi que la distribution des coefficients dans $\widetilde{\mathbf{C}}_m^{-\frac{1}{2}}$ puisque cette matrice est creuse et que nous savons où se trouvent les coefficients non nuls.

5) Validité de l'approximation. Notons $\widetilde{\mathbf{C}}_m = \left(\widetilde{\mathbf{C}}_m^{-\frac{1}{2}} \widetilde{\mathbf{C}}_m^{-\frac{1}{2}} \right)^{-1}$; à partir de cette matrice, nous pouvons extraire la fonction de corrélation utilisée. La figure (4.0.11) montre cette fonction de corrélation comparée avec la fonction de référence (4.0.58). Pour des valeurs de n_0 supérieures à 25, les deux fonctions peuvent être considérées comme égales. Pour n_0 inférieur, la différence entre les deux fonctions est trop importante. Il existe donc un seuil n_0^{min} à partir duquel la matrice $\widetilde{\mathbf{C}}_m^{-\frac{1}{2}}$ peut être considérée comme une approximation valide de $\mathbf{C}_m^{-\frac{1}{2}}$. De nous remarquons que l'approximation devient valide lorsque l'erreur relative, $e_r = \frac{\| \mathbf{C}_m^{-\frac{1}{2}} - \widetilde{\mathbf{C}}_m^{-\frac{1}{2}} \|}{\| \mathbf{C}_m^{-\frac{1}{2}} \|}$ est inférieure à 10^{-2} . Nous pouvons définir un critère ;

l'approximation $\widetilde{\mathbf{C}}_m^{-\frac{1}{2}}$ est valide lorsque l'erreur est inférieure à 10^{-2} . Ce critère est utilisé sur la figure (4.0.11) pour tracer une courbe n_0^{min} en fonction de λ . Nous remarquons que n_0^{min} croît en fonction de λ . Cette courbe sera utilisée dans les applications tomographiques pour choisir n_0 (figure (4.0.12)) afin d'obtenir une matrice $\widetilde{\mathbf{C}}_m^{-\frac{1}{2}}$ approximant correctement $\mathbf{C}_m^{-\frac{1}{2}}$.

²Ce sont des propriétés numériques, mises en évidence empiriquement

FIG. 4.0.11. **Choix de n_0 .**

Plusieurs calculs de $\widetilde{\mathbf{C}}_{\mathbf{m}}^{-\frac{1}{2}}$ sont effectués avec $\lambda = 10$, $\sigma = 1$ et $n_0 = 15, 20, 25$ et 30 . La matrice $\mathbf{C}_{\mathbf{m}}^{-\frac{1}{2}}$ de référence est calculée avec $n=100$. Pour décider si l'approximation $\widetilde{\mathbf{C}}_{\mathbf{m}}^{-\frac{1}{2}}$ est valide, on forme

$\widetilde{\mathbf{C}}_{\mathbf{m}} = \left(\widetilde{\mathbf{C}}_{\mathbf{m}}^{-\frac{1}{2}} \widetilde{\mathbf{C}}_{\mathbf{m}}^{-\frac{1}{2}} \right)^{-1}$, on en extrait la fonction de corrélation (en rouge) et on la compare à la fonction de corrélation de référence (en bleu). On peut considérer que l'approximation est valide quand

n_0 est supérieur à 25, c'est à dire à dire pour une erreur $e_r = \frac{\left\| \widetilde{\mathbf{C}}_{\mathbf{m}}^{-\frac{1}{2}} - \mathbf{C}_{\mathbf{m}}^{-\frac{1}{2}} \right\|_{\infty}}{\left\| \mathbf{C}_{\mathbf{m}}^{-\frac{1}{2}} \right\|_{\infty}}$ inférieure à 10^{-2} . On définit ainsi un critère pour choisir une approximation valide.

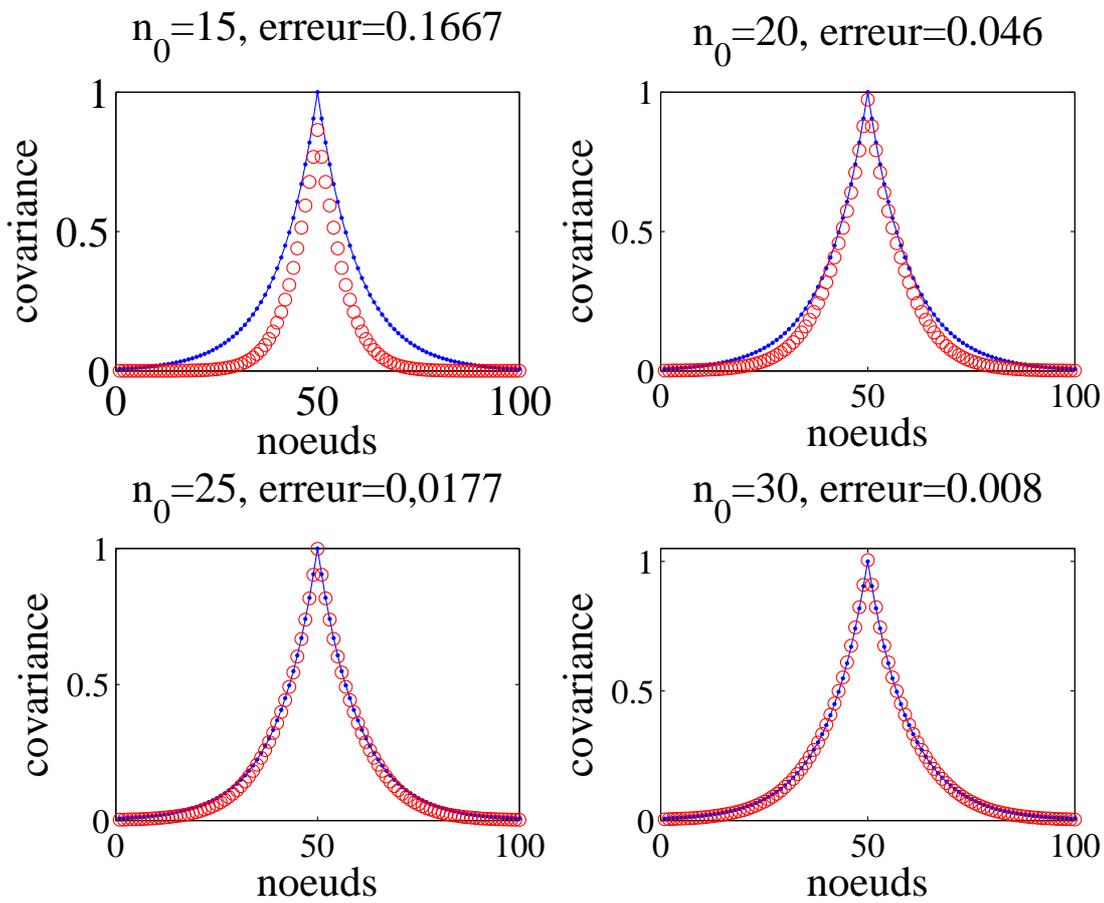
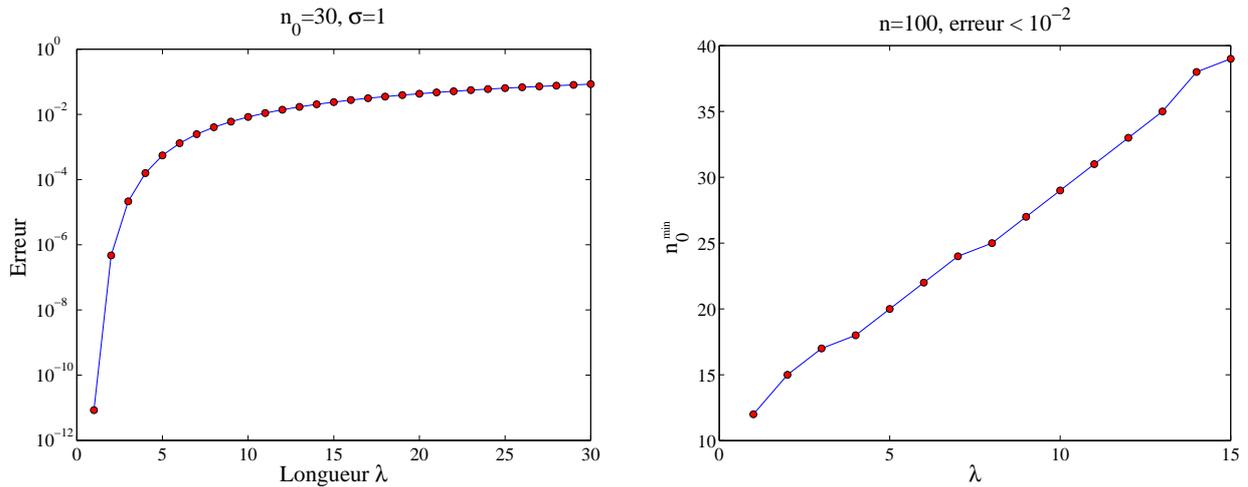


FIG. 4.0.12. Erreur d'approximation de $\mathbf{C}_m^{-\frac{1}{2}}$.

Erreur d'approximation $\frac{\|\mathbf{C}_m^{-\frac{1}{2}} - \widetilde{\mathbf{C}_m^{-\frac{1}{2}}}\|_\infty}{\|\mathbf{C}_m^{-\frac{1}{2}}\|_\infty}$ en fonction de λ , pour $n=100$ et $n_0 = 30$. A gauche est tracée l'erreur en fonction de λ . A droite on peut lire directement n_0^{min} en fonction de λ , une fonctionnalité utile en tomographie.



6) **Cas à 2D et 3D.** Nous supposons que nous pouvons extrapoler les propriétés définies en 3) au cas à deux et trois dimensions d'espace. Néanmoins il faut tenir compte des effets de bords spécifiques à ces cas. Ces effets de bord apparaissent chaque fois qu'un sommet ou une arête font partie du support de la fonction de corrélation. La figure (4.0.13) montre la diagonale principale d'une matrice $\mathbf{C}_m^{-\frac{1}{2}}$ calculée avec $n_x = 5$, $n_y = 5$, $\lambda = 1$, et $\sigma = 1$. Nous voyons que les coefficients prennent différentes valeurs. On distingue, sur la figure 4.0.13 les valeurs des coefficients correspondant :

- aux 4 sommets (noeuds en noir sur la grille de la figure (4.0.13))
- aux 4 arêtes (noeuds en bleu, gris, vert et bleu clair sur la figure (4.0.13))
- aux autres points situés à l'intérieur du milieu (en rouge sur la figure (4.0.13)).

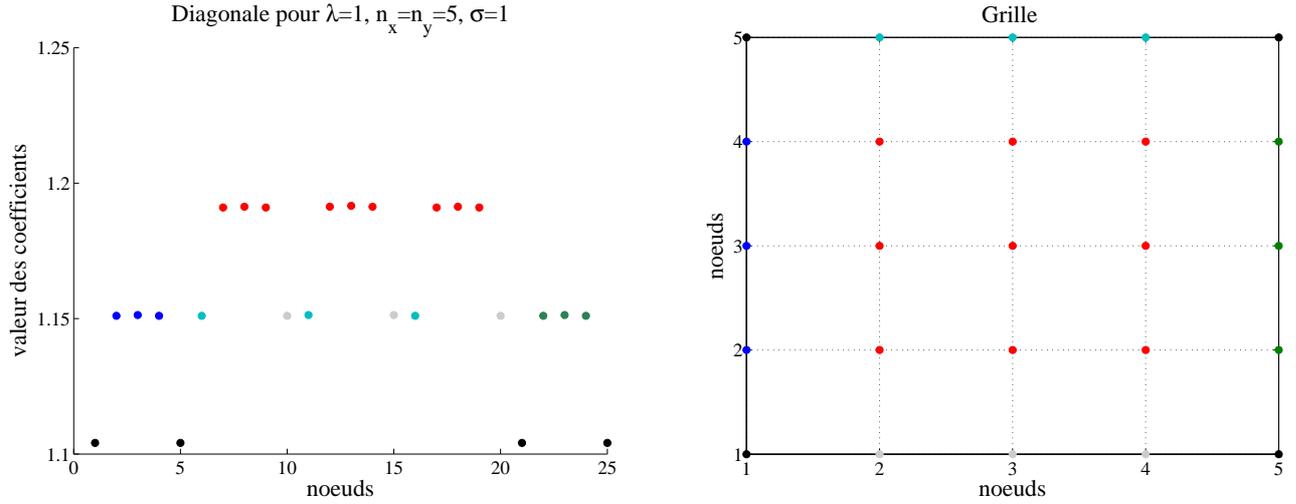
Ces effets de bords existent sur chaque diagonale supérieure. Dans le cas à deux dimensions d'espace, l'algorithme d'approximation de $\mathbf{C}_m^{-\frac{1}{2}}$ est implémenté suivant un schéma identique à une dimension d'espace, mais en tenant compte des effets de bords spécifiques. Pour le cas à trois dimensions d'espace, il faut également considérer l'ensemble des effets de bords ; on distingue alors les valeurs des coefficients correspondant :

- aux 8 sommets,
- aux 12 arêtes,
- aux 6 faces,
- aux autres points situés à l'intérieur du milieu.

Cela nous permet d'implémenter un algorithme d'approximation de $\mathbf{C}_m^{-\frac{1}{2}}$ tenant compte des effets de bords spécifiques au cas à trois dimensions d'espace.

FIG. 4.0.13. **Diagonale à 2D.**

Diagonale de $\mathbf{C}_m^{-\frac{1}{2}}$ dans un milieu à deux dimensions d'espace (à gauche). Le calcul est fait pour $\lambda_x = \lambda_y = 1$, $n_x = n_y = 5$ et $\sigma = 1$. Le code de couleurs permet de relier la valeur des coefficients lue sur la figure de gauche, avec le noeud correspondant sur la grille représentée sur la figure de droite. Les effets de bords spécifique aux deux dimensions d'espace, induisent une structure différente de la diagonale par rapport au cas à une dimension. Les sommets sont représentés en noir, les noeuds du centre en rouge, les autres couleurs correspondent aux 4 arêtes.



7) Utilisation pour la tomographie. Dans les applications tomographiques, nous aurons à calculer une matrice $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$ approximant $\mathbf{C}_m^{-\frac{1}{2}}$, de longueur de corrélation λ_x , λ_y et λ_z km et σ km/s, dans un milieu à trois dimensions discrétisé par n_x , n_y et n_z noeuds, avec des pas de h_x , h_y et h_z km. Il s'agit de choisir une grille $n_{0,x}$, $n_{0,y}$ et $n_{0,z}$ et des pas $h_{0,x}$, $h_{0,y}$ et $h_{0,z}$ pour calculer les coefficients de $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$. Nous allons normaliser en utilisant la valeur des pas ; ainsi nous poserons ; $\lambda_{0,x} = \frac{\lambda_x}{h_x}$, $\lambda_{0,y} = \frac{\lambda_y}{h_y}$ et $\lambda_{0,z} = \frac{\lambda_z}{h_z}$. Nous avons alors à calculer les coefficients de $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$ avec $n_{0,x}$, $n_{0,y}$, $n_{0,z}$, $\lambda_{0,x}$, $\lambda_{0,y}$, $\lambda_{0,z}$ et $\sigma_0 = 1$. Pour choisir les paramètres $n_{0,i=1,2,3}$ nous pouvons alors nous référer à la courbe (4.0.12) qui donne en fonction de λ le paramètre n_0 minimal pour obtenir une approximation valide. On utilise ce critère pour choisir $\lambda_{0,x}$, $\lambda_{0,y}$ et $\lambda_{0,z}$ en fonction de $n_{0,x}$, $n_{0,y}$ et $n_{0,z}$. Les coefficients calculés sont finalement divisés par σ , pour revenir en unité physique puisque le calcul a été fait avec $\sigma_0 = 1$. Cette façon de faire permet de calculer les coefficients, de les stocker dans un fichier et ainsi d'éviter de recalculer la matrice à chaque nouvelle tomographie. On constitue ainsi une bibliothèque de fichiers contenant les coefficients, rangée en fonction de $\lambda_{0,x}$, $\lambda_{0,y}$ et $\lambda_{0,z}$. Par exemple dans la tomographie des décalages temporels nous avons fait

une inversion pour $\lambda_x = \lambda_y = \lambda_z = 1.25$ km et $\sigma = 2$ km/s. Nous avons discrétisé le milieu par $n_x = 25$, $n_y = 33$ et $n_z = 25$ noeuds et $h_x = h_y = h_z = 0.5$ km. Pour calculer l'approximation nous avons $\lambda_{0,x} = \lambda_{0,y} = \lambda_{0,z} = 2.5$, la courbe (4.0.12) nous indique qu'il faut choisir $n_{0,x} = n_{0,y} = n_{0,z} = 16$ pour calculer les coefficients de $\widetilde{\mathbf{C}_m^{-\frac{1}{2}}}$. Dans cet exemple, le calcul se fait par une décomposition de Lanczos sur une matrice de taille de l'ordre de 4000×4000 , que l'on peut traiter informatiquement, alors que le vrai calcul exige une matrice de taille 20000×20000 .

Bibliographie

- [1] Aki, K., and W. H. K. Lee, Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes : 1. A homogeneous initial model, *J. Geophys. Res.*, 81, 4381-4399, 1976.
- [2] Aki, K. and P.G. Richards, *Quantitative Seismology Theory and Methods*. W. H. FREEMAN AND COMPANY, NEW YORK, 1980.
- [3] Backus, G. E. and F. Gilbert, The resolving power of gross earth data, *Geophys. J. Roy. Astron. Soc.*, 16, 169-205, 1968.
- [4] Benz, H. M., Chouet, B., Dawson, P. B., Lahr, J. C., Page, R. A. and J. A. Hole, Three-dimensional P and S wave velocity structure of Redoubt Volcano, Alaska, *J. Geophys. Res.*, 101, 8111 à 8128, 1996.
- [5] Benz, H. M., P. Okubo, and A. Villasenor, Three-dimensional crustal P-wave imaging of Mauna Loa and Kilauea volcanoes, Hawaii, *International Handbook of Earthquake Engineering and Seismology, International Geophysics Series, 81A*, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger ed., 407-421, 2003.
- [6] Bonnans, J. F., J. C. Gilbert, C. Lemaréchal, C. Sagatizabal, *Optimisation numérique Aspects théoriques et pratiques*, Springer, 1997.
- [7] Cerveny, V. Ray tracing algorithms in three-dimensional laterally varying layered structures, in *Seismic Tomography*, (ed. G. nolet), Reidel, Hingham, MA, pp. 99-133, 1987
- [8] Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, Robust elastic nonlinear waveform inversion : application to real data, *Geophysics*, 55, 527-538, 1990.
- [9] Crosson, R., and R. Koyanagi, Seismic velocity structure below the island of Hawaii from local earthquake data, *J. Geophys. Res.*, 84, 2331-2342, 1979.
- [10] Dawson, P.B., B.A. Chouet, P.G. Okubo, A. Villasenor, and H.M. Benz, Three-dimensional velocity structure of the Kilauea caldera, Hawaii, *Geophys. Res. Lett.*, 26, 2805-2808, 1999.
- [11] Eaton, J. P., Crustal structure and volcanism in Hawaii, *Am. Geophys. Soc. Mon.*, 6, 13-29, 1962.
- [12] Evjen, H. M., The place of the vertical gradient in gravitational interpretation, *Geophysics*, 1, 127-136, 1936.
- [13] Fréchet, J. sismogénèse et doublets sismiques, Thèse d'état, 207 pp., Univ. Sci. Technol. Médoc. , Grenoble, 1985.
- [14] Geiger, L., Herdbestimmung bei Erdbeben aus den Ankunftszeiten, *K. Ges. Wiss. Gött.*, 4, 331-349, 1910.
- [15] Got, J.-L., J. Fréchet, and F.W. Klein, Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.*, 99, 15375-15386, 1994.
- [16] Got, J.-L., and P. G. Okubo, New insights into Kilaueas volcano dynamics brought by large-scale relative relocation of microearthquakes, *J. Geophys. Res.*, 108, 2337, doi :10.1029/2002JB002060, 2003.

- [17] Hansen, S., C.H. Thurber, M. Mandernach, F. Haslinger, and C. Doran, Seismic velocity and attenuation of the East Rift Zone and South Flank of Kilauea Volcano, Hawaii, *Bull. Seismol. Soc. Am.*, 94, 1430-1440, 2004.
- [18] Haslinger, F., C.H. Thurber, M. Mandernach, and P.G. Okubo, Tomographic imaging of the P-velocity structure beneath Kilauea's East Rift Zone and South Flank : Seismic evidence for the presence of a deep magma body, *Geophys. Res. Lett.*, 28, 375-378, 2001.
- [19] Hill, D., Crustal structure of the island of Hawaii, from seismic refraction measurements, *Bull. Seismol. Soc. Am.*, 59, 101-130, 1969. I
- [20] Jenkins, G. M., Watts, D. G., 1968, *Spectral Analysis and its Applications*, Holden-Day, San Francisco.
- [21] Ito, A., High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *J. Phys. Earth.*, 33, 279-294, 1985.
- [22] Jordan, T. H., and K.A. Sverdrup, Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seismol. Soc. Am.*, 71, 1105-1130, 1981.
- [23] Klein, F.W., A linear gradient crustal model for south Hawaii, *Bull. Seismol. Soc. Am.*, 71, 1503-1510, 1981.
- [24] Kissling, E., W.L. Ellsworth, D. Eberhart-Phillips, and U. Kradolfer, Initial reference models in local earthquake tomography, *J. Geophys. Res.*, 99, 19635-19646, 1994.
- [25] Latorre, D., J. Virieux, T. Monfret, V. Monteiller, T. Vanorio, J.-L. Got, and H. Lyon-Caen, A new seismic tomography of Aigion area (Gulf of Corinth à Greece) from a 1991 dataset, accepted to *Geophys. J. Int.*, 2005.
- [26] Le Meur, H., J. Virieux, and P. Podvin, Seismic tomography of the Gulf of Corinth : a comparison of methods, *Ann. Geophys.*, XL, 1-25, 1997.
- [27] Menke, W., and D. Schaff, Absolute earthquake locations with differential data, *Bull. Seismol. Soc. Am.*, 94, 2254-2264, 2004.
- [28] Michelini, A., and A. Lomax, The effect of velocity structure errors on double-difference earthquake location, *Geophys. Res. Lett.*, 31, doi :10.1029/2004GL019682, 2004.
- [29] Monteiller, V., J.-L. Got, J. Virieux, and P.G. Okubo, Double-difference tomography of Kilauea Volcano, Hawaii, Proceedings of the EGS - AGU - EUG Joint Assembly, Nice, France, April 2003.
- [30] Monteux, J., J.-L. Got, R. Hassani, Modeling kilauea volcano's topography from sismological constraints. soumis au JGR.
- [31] Mosteller, F., and J. W. Tukey, *Data analysis and regression*, Addison-Wesley, Reading, Mass., 1977.
- [32] Necessian, A., A. Hirn, A. Tarantola, Three-dimensional seismic transmission prospecting of the Mont Dore volcano, France, *Geophys. J. R. astr. Soc.* 76, 307-315, 1984.
- [33] Nolet, G., Seismic waves propagation and seismic tomography, *Seismic tomography : with applications in global seismology and exploration geophysics*, Kluwer Academic Publishers, 1987.
- [34] Nolet, G., Solving large linearized tomographic problems, *Seismic tomography theorie and practice*, Chapman & Hall, 1993.
- [35] Okubo, P., H.M. Benz, and B.A. Chouet, Imaging the crustal magma sources beneath Mauna Loa and Kilauea volcanoes, Hawaii, *Geology*, 25, 867-870, 1997.
- [36] Paige, C.C., and M. A. Saunders, LSQR : sparse linear equations and least squares problems, *ACM Trans. Math. Softw.* 8, 195-209, 1982.
- [37] Podvin, P., and I. Lecomte, Finite difference computation of traveltimes in very contrasted velocity models : a massively parallel approach and its associated tools, *Geophys. J. Int.*, 105, 271-295, 1991.

- [38] Poupinet, G., W.L. Ellsworth, and J. Fréchet, Monitoring velocity variations in the crust using earthquake doublets : An application to the Calaveras fault, California, *J. Geophys. Res.*, 89, 5719-5731, 1984.
- [39] Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran 77*, Second Edition, Cambridge University Press, 933 p.
- [40] Rowan, L.R., and R.W. Clayton, The three-dimensional structure of Kilauea Volcano, Hawaii, from travel time tomography, *J. Geophys. Res.*, 98, 4355-4375, 1993.
- [41] Rubin, A.M., D. Gillard, and J.-L. Got, Streaks of microearthquakes along creeping faults, *Nature*, 400, 635-641, 1999.
- [42] Ryall, A., and D. L. Bennett, Crustal structure of southern Hawaii related to volcanic processes in the upper mantle, *J. Geophys. Res.*, 73, 4561-4582, 1968.
- [43] Shearer, P.M., Improving local earthquake locations using the L1 norm and waveform cross-correlation : application to the Whittier Narrows, California, aftershock sequence, *J. Geophys. Res.*, 102, 8269-8283, 1997.
- [44] Slunga, R., S. Rögnvaldsson, and R. Bödvarsson, Absolute and relative locations of similar events with application to microearthquakes in southern Iceland, *Geophys. J. Int.*, 123, 409-419, 1995.
- [45] Tarantola, A., *Inverse Problem Theory*, Elsevier, New York, 613 pp, 1987.
- [46] Tarantola, A., A. Nercessian, Three-dimensional inversion without blocks, *Geophys. J. R. astr. Soc* 76, 299-306, 1984
- [47] Tarantola, A., and B. Valette, Generalized nonlinear inverse problems solved using the least-squares criterion, *Rev. Geophys. Space Phys.*, 20, 219-232, 1982.
- [48] Thurber, C.H., Seismic detection of the summit magma complex of Kilauea Volcano, Hawaii, *Science*, 223, 165-167, 1984.
- [49] Thurber, C.H., S. Roecker, H. Zhang, S. Baher, and W. Ellsworth, Fine-scale structure of the San-Andreas fault and location of the SAFOD target earthquakes, *Geophys. Res. Lett.*, 31, doi :10.1029/2003GL019398, 2004.
- [50] Valette, B. and P. Lesage, Inferring mean earth mechanical models from normal modes, mass and inertia : 1. Theoretical developments, submitted to *Geophys. J. Int.*, 2005.
- [51] Vidale, J.E., Finite-difference calculation of traveltimes, *Bull. Seismol. Soc. Am.*, 78, 2062-2076, 1988.
- [52] Waldhauser, F., and W. L. Ellsworth, A double difference earthquake location algorithm : method and application to the northern Hayward fault, CA., *Bull. Seismol. Soc. Am.*, 90, 1353-1368, 2000.
- [53] Ward, P. L., and S. Gregersen, Comparison of earthquake locations determined with data from a network of stations and small tripartite arrays on Kilauea volcano, Hawaii, *Bull. Seismol. Soc. Am.*, 63, 679-711, 1973.
- [54] Wolfe, C.J., On the mathematics of using difference operators to relocate earthquakes, *Bull. Seismol. Soc. Am.*, 92, 2879-2892, 2002.
- [55] Zhang, H., and C. H. Thurber, Double difference tomography : the method and its application to the Hayward fault, California, *Bull. Seismol. Soc. Am.*, 93, 1875-1889, 2003.
- [56] Zhang, H., C.H. Thurber, D. Shelly, S. Ide, G.C. Beroza, and A. Hasegawa, High-resolution subducting slab structure beneath northern Honshu, Japan, revealed by double-difference tomography, *Geology*, 32, 361-364, doi :10.1130/G20261, 2004.