

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Modèles paramétriques pour la tomographie sismique bayésienne

Ecole doctorale n°398

Géosciences Ressources Naturelles et Environnement

Spécialité GÉOSCIENCES ET GÉOINGÉNIERIE

Soutenue par Jihane BELHADJ
le 02 décembre 2016

Dirigée par **Jacques RIVOIRARD**

COMPOSITION DU JURY :

M. Guillaume CAUMON
ENSG Nancy, Président du jury

M. Josselin GARNIER
Ecole Polytechnique, Rapporteur

M. Thomas BODIN
CNRS, Examineur

M. Thierry COLEOU
CGG, Examineur

Mme. Mickaële LE RAVALEC
IFPen, Invitée

Mme. Alexandrine GESRET
Mines ParisTech, Examinatrice

M. Mark NOBLE
Mines ParisTech, Examineur

M. Jacques RIVOIRARD
Mines ParisTech, Examineur

M. Thomas ROMARY
Mines ParisTech, Examineur



Résumé :

La tomographie des temps de première arrivée vise à retrouver un modèle de vitesse de propagation des ondes sismiques à partir des temps de première arrivée mesurés. Cette technique nécessite la résolution d'un problème inverse afin d'obtenir un modèle sismique cohérent avec les données observées. Il s'agit d'un problème mal posé pour lequel il n'y a aucune garantie quant à l'unicité de la solution. L'approche bayésienne permet d'estimer la distribution spatiale de la vitesse de propagation des ondes sismiques. Il en résulte une meilleure quantification des incertitudes associées. Cependant l'approche reste relativement coûteuse en temps de calcul, les algorithmes de Monte Carlo par chaînes de Markov (MCMC) classiquement utilisés pour échantillonner la loi a posteriori des paramètres n'étant efficaces que pour un nombre raisonnable de paramètres. Elle demande, de ce fait, une réflexion à la fois sur la paramétrisation du modèle de vitesse afin de réduire la dimension du problème et sur la définition de la loi a priori des paramètres. Le sujet de cette thèse porte essentiellement sur cette problématique.

Le premier modèle que nous considérons est basé sur un modèle de mosaïque aléatoire, le modèle de Johnson-Mehl, dérivé des mosaïques de Voronoï déjà proposées en tomographie bayésienne. Contrairement à la mosaïque de Voronoï, les cellules de Johnson-Mehl ne sont pas forcément convexes et sont bornées par des portions d'hyperboloïdes, offrant ainsi des frontières lisses entre les cellules. Le deuxième modèle est, quant à lui, décrit par une combinaison linéaire de fonctions gaussiennes, centrées sur la réalisation d'un processus ponctuel de Poisson. Ce modèle est exploitable en pratique pour décrire des anomalies locales du champ de vitesse. Pour chaque modèle, nous présentons un exemple de validation sur des champs de vitesse simulés. Nous appliquons ensuite notre méthodologie à un modèle synthétique plus complexe qui sert de benchmark dans l'industrie pétrolière. Les résultats obtenus, montrent la capacité de notre algorithme à reconstruire le champ de vitesse des ondes sismiques en utilisant un nombre raisonnable de paramètres, et ce même pour des modèles de sub-surfaces complexes. Nous proposons enfin, un modèle de vitesse basé sur la théorie du compressive sensing pour reconstruire le champ de vitesse. Ce modèle, encore imparfait, ouvre plusieurs pistes de recherches futures.

Dans ce travail, nous nous intéressons également à un jeu de données réelles acquises dans le contexte de la fracturation hydraulique. Nous développons dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique afin de traiter efficacement la complexité du modèle à couches. Ceci nous permet d'obtenir des résultats cohérents avec un log sismique acquis de manière indépendante sur le même site.

Mots clés : Tomographie sismique, problème inverse, approche bayésienne, paramétrisation, MCMC, trans-dimensionnel, hiérarchique.

Abstract:

First arrival time tomography aims at inferring the seismic wave propagation velocity using experimental first arrival times. This constitutes an inverse problem that is usually ill-posed. In particular, the unicity of the solution is not guaranteed. In our study, we rely on a Bayesian approach to estimate the wave velocity and the associated uncertainties. This approach incorporates the information provided by the data and the prior knowledge of the velocity model. Bayesian tomography allows for a better estimation of wave velocity as well as associated uncertainties. However, this approach remains fairly expensive, and MCMC algorithms that are used to sample the posterior distribution are efficient only as long as the number of parameters remains within reason. Hence, their use requires a careful reflection both on the parameterization of the velocity model, in order to reduce the problem's dimension, and on the definition of the prior distribution of the parameters. In this thesis, we introduce new parsimonious parameterizations enabling to accurately reproduce the wave velocity field with the associated uncertainties.

The first parametric model that we propose uses a random Johnson-Mehl tessellation, a variation of the Voronoï tessellation. Unlike Voronoï tessellation, Johnson-Mehl tessellation contains cells that are not necessarily convex polytopes, therefore providing smooth boundaries between cells. The second one uses Gaussian kernels as basis functions. It is especially adapted to the detection of seismic wave velocity anomalies. Each anomaly is considered to be a linear combination of these basis functions localized at the realization of a Poisson point process. We first illustrate the tomography results with a synthetic velocity model, which contains two small anomalies. We then apply our methodology to a more advanced and more realistic synthetic model that serves as a benchmark in the oil industry. The tomography results reveal the ability of our algorithm to map the velocity heterogeneities with precision using few parameters. Finally, we propose a new parametric model based on the compressed sensing techniques. The first results are encouraging. However, the model still has some weakness related to the uncertainties estimation.

In addition, we analyse real data in the context of induced microseismicity. In this context, we develop a trans-dimensional and hierarchical approach in order to deal with the full complexity of the layered model. The tomography results are consistent with the seismic log acquired independently from a well located on the operating site.

Keywords:

Seismic tomography, inverse problem, Bayesian approach, parameterization, MCMC, transdimensional, hierarchical.

Remerciements :

Cette thèse a été réalisée au sein du centre Géosciences de l'école des Mines de Paris sous la direction de Jacques Rivoirard. Ces trois années de thèse ont été une véritable aventure humaine et scientifique, en très grande partie grâce aux personnes que j'ai pu rencontrer et à qui je souhaite adresser mes sincères remerciements.

Ma profonde reconnaissance va en premier lieu à mes encadrants de thèse Alexandrine Gesret et Thomas Romary, pour leur intérêt et leur soutien, leur grande disponibilité et leur nombreux conseils durant la rédaction de ma thèse.

Je remercie mon directeur de thèse, Jacques Rivoirard, pour la liberté qu'il m'a donnée, pour sa confiance et ses encouragements.

Je tiens à remercier également Mark Noble qui, malgré un emploi du temps souvent très chargé, a toujours été disponible pour soutenir le projet.

Cette thèse doit beaucoup aux idées et conseils de Bruno Figliuzzi du centre Morphologie Mathématique. Bruno, merci pour ta bonne humeur, tes conseils, ton écoute dans les épreuves difficiles et les nombreuses discussions que nous avons eues. Merci pour tout !

J'exprime ma profonde gratitude et mes plus vifs remerciements à Monsieur Guillaume Caumon et Monsieur Josselin Garnier, qui m'ont fait l'honneur d'accepter d'être les rapporteurs de ce travail. Je remercie également Monsieur Thomas Bodin et Monsieur Thierry Coléou, qui m'ont fait l'honneur d'accepter d'être examinateurs de ce travail.

Je tiens à exprimer ma gratitude aux membres de l'équipe Géostatistique pour leur accueil chaleureux. Je remercie en particulier Isabelle Schmitt et Nicolas Dessassis pour leur soutien moral, Fabien Ors pour sa précieuse aide en informatique. Je remercie également Christian Lantuéjoul, Emilie Chautru, Didier Renard, Christian Lajaunie, Gaëlle Le loc'h, Hélène Beucher, Xavier Freulon, Hans Wackernagel et Chantal de Fouquet pour leurs conseils et leur bienveillance. J'aimerais également chaleureusement remercier Nathalie Dietrich, Dominique Vassiliadis et Véronique Lachasse pour leur gentillesse et leur disponibilité. Un grand merci à mes collègues thésards: Clément Barthélemy, Francky Fouedjio, Anna Bubnova, Liên Nguyen, Marine Demangeot, Yuan Chen, Baptiste Sauvaget, Mike Pereira.

Le centre Géosciences à Fontainebleau a été durant ces trois ans un environnement de travail très agréable grâce à de très belles rencontres : Emmanuel, Yves-Marie, Vaïa, Jean-Baptiste, Robin, Sara, Hafsa, Mohamed Mahjoub, Mohamed Amri, Nidhal, Anis, Paule, Dariouche, ...

Enfin, je remercie ma famille, ma sœur, mes frères et mes parents sans qui rien n'aurait été possible. Merci pour votre soutien toutes ces années et vos encouragements.

Table des matières

Introduction générale	11
Chapitre 1. La tomographie sismique	17
1 La paramétrisation du modèle de vitesse :	18
2 Le problème direct:	19
2.1 Equation Eikonal:	19
2.2 Résolution de l'équation Eikonal	22
3 L'inversion :	22
4 Tomographie bayésienne:	25
4.1 Formulation du problème inverse :	25
4.2 Résolution du problème inverse : Estimateurs bayésiens	29
5 Conclusion	29
Chapitre 2. Méthodes MCMC	31
1 Motivation	31
2 Algorithmes de type Metropolis-Hastings	32
2.1 L'algorithme Metropolis-Hastings classique	32
2.2 Delayed Rejection :	34
2.3 Metropolis Hastings adaptatif :	35
2.4 Algorithme de Metropolis-Hastings à saut réversible :	37
3 Distributions tempérées:	38
3.1 Une seule chaîne :	39
3.2 Chaînes en interaction:	41
4 Conclusion :	43
Chapitre 3. La paramétrisation	45
1 État de l'art	45
1.1 Modèle à couches	45
1.2 Voronoï :	47
1.3 Les ondelettes	48
2 Nouvelles paramétrisations	52
2.1 Mosaïques de Johnson Mehl:	52
2.2 Les noyaux gaussiens	59
3 Conclusion:	62
Chapitre 4. Applications	65
1 Analyse de sensibilité:	65
1.1 Mosaïque de Johnson-Mehl	66
1.2 Noyaux gaussiens :	67
1.3 Conclusion	70
2 Données synthétiques 1:	70

3	Données synthétiques 2:	76
4	Conclusion :	83
Chapitre 5. Modèle à couches		85
1	Données synthétiques:	85
1.1	Une seule chaîne :	88
1.2	Chaînes en interaction :	89
2	Données réelles:	94
3	Conclusion :	102
Chapitre 6. Compressed Sensing appliqué à la tomographie		103
1	Compressed sensing :	103
1.1	Matrice de mesure	104
1.2	RIP et matrices aléatoires	107
1.3	Reconstruction :	108
2	Application à la paramétrisation du champ de vitesse :	109
2.1	Intégration dans une approche bayésienne	111
2.2	Résultats et discussion :	112
3	Conclusion et perspectives :	117
Chapitre 7. Conclusion générale		121
Annexes		123
	New Parametric Models For Bayesian Seismic Tomography	125
	First Arrival Travel Time Tomography – Bayesian Approach	151
Bibliographie		157
Liste des figures		163
Liste des tableaux		167

Introduction générale

Parmi les objectifs principaux des études géophysiques figure l'estimation des paramètres physiques correspondant à des caractéristiques du sous-sol, que l'on ne peut pas (ou difficilement) mesurer directement (vitesse, impédance, densité) à partir des données collectées. La résolution du problème inverse permet de remonter au système de paramètres recherchés (Tarantola, 1987).

Le terme « tomographie », composé de deux racines grecques: tomos et graphia signifie littéralement dessin d'une coupe. Cette technique, initialement utilisée dans l'imagerie médicale, a vu le jour grâce à la découverte des rayons X par le physicien allemand Wilhelm Conrad Röntgen en novembre 1895.

Cette technique a été ensuite adoptée par les sismologues dans le but de reconstituer la structure du sous sol. Le principe de la tomographie sismique est analogue à celui de la tomographie médicale. Au lieu des rayons X, les géophysiciens utilisent les ondes sismiques. Ces ondes traversent le sous sol et ralentissent ou accélèrent en profondeur en fonction de la composition physico-chimique du milieu. En exploitant les différents types de données sismiques comme par exemple les temps de propagation, les amplitudes ou les phases (Rawlinson et al., 2010), les géophysiciens peuvent déterminer les variations de vitesses de propagation des ondes sismiques dans le sous-sol. Selon la nature de la source, on distingue deux types de tomographie sismique : La tomographie passive, une technique développée dans le début des années 80 par K.Aki (Aki and Richards, 1980) qui exploite les données fournies par les tremblements de terre dus à des sources naturelles. Et la tomographie dite active qui utilise des sources artificielles (camions vibreurs, dynamite à terre).

Dans ce travail, nous nous intéressons à la tomographie active des temps de première arrivée. Elle a pour objectif de cartographier le modèle de vitesse de propagation des ondes sismiques à partir de temps de première arrivée mesurés. Dans le domaine pétrolier, la tomographie en réfraction est essentiellement utilisée pour déterminer les caractéristiques des couches superficielles et étudier la zone altérée de surface. Cette technique est également très employée en génie civil pour déterminer les propriétés mécaniques des roches par exemple (module d'élasticité, module de Poisson etc) ou encore en hydrogéologie pour déterminer la profondeur de la nappe phréatique et identifier les zones fracturées.

La tomographie des temps de première arrivée est un problème inverse classique, en général mal-posé, le plus souvent résolu avec des approches linéarisées itératives, comme la descente du gradient (Taillandier et al., 2009). Une des difficultés majeures du problème inverse

considéré est qu'il ne présente aucune garantie quant à l'unicité de la solution. Les méthodes linéarisées itératives peuvent donner un modèle de vitesse erroné si le problème est non-linéaire. En effet, ces approches sont facilement piégées dans un minimum local, par conséquent, la solution peut fortement dépendre du champ de vitesse utilisé pour initialiser l'algorithme. De plus, elles ne permettent pas d'estimer de manière fiable les incertitudes associées au modèle de vitesse. Or le modèle de vitesse final issu de la tomographie est utilisé pour des analyses sismologiques et tectoniques. Par exemple, les incertitudes associées à l'estimation du modèle de vitesse peuvent être propagées aux incertitudes de localisation des hypocentres de séismes comme dans (Agostinetti et al., 2015; Gesret et al., 2015). Il est donc indispensable d'avoir un modèle de vitesse fiable avec les incertitudes associées afin d'être en mesure de mener des interprétations fiables.

Afin de prendre en compte les incomplétudes et les incertitudes sur les données expérimentales ainsi que la non-linéarité de la relation liant les données aux paramètres, les géophysiciens se sont intéressés aux méthodes probabilistes, en particulier à l'approche bayésienne utilisant des algorithmes de Monte Carlo par Chaînes de Markov (MCMC). Tarantola et Valette (Tarantola and Valette, 1982) ont posé les bases de la résolution des problèmes inverses dans un cadre bayésien. (Pour une revue sur l'histoire de l'utilisation des techniques de Monte Carlo en géophysique, voir (Sambridge and Mosegaard, 2002)).

Cette approche consiste à quantifier l'information apportée par les données sur la connaissance a priori du modèle de vitesse. La tomographie bayésienne permet de mieux estimer la vitesse des ondes tout en quantifiant de manière fiable les incertitudes associées à cette estimation. Cependant, dans le cas de distributions multimodales avec des modes très éloignés, ou de modèles complexes possédant un grand nombre de paramètres, les algorithmes MCMC classiques deviennent inefficaces : la chaîne de Markov générée par ces algorithmes évolue lentement (se mélange mal) et a du mal à se déplacer. Elle risque par conséquent de rester bloquée dans le voisinage d'un mode local. Pour remédier à ce problème et afin d'améliorer l'échantillonnage de l'espace d'état, de nombreuses améliorations ont été apportées aux échantillonneurs classiques, à l'instar du recuit simulé (Kirkpatrick et al., 1983). Plus récemment, la communauté scientifique s'est intéressée aux chaînes de Markov en interaction. Il s'agit d'algorithmes MCMC qui simulent plusieurs chaînes en même temps en leur permettant d'échanger les informations entre elles (Jasra et al., 2007; Romary, 2010; Sambridge, 2013).

Il est important de noter que bien que nous disposions aujourd'hui d'outils informatiques très performants, l'approche bayésienne reste relativement coûteuse en temps de calcul car elle nécessite la résolution du problème direct à chaque itération. Elle demande donc une réflexion sur la paramétrisation du modèle de vitesse afin de réduire la dimension du problème. Le sujet de cette thèse porte essentiellement sur cette problématique. L'objectif est donc de proposer des modèles à la fois parcimonieux et physiquement pertinents du modèle de vitesse des ondes sismiques.

La paramétrisation consiste généralement à exprimer le modèle de vitesse sous forme d'une combinaison linéaire de fonctions de bases. Ces dernières correspondent souvent à des cellules locales, ou à des harmoniques sphériques (Nolet and others, 2008) Elles sont choisies à l'avance et leur nombre est fixé par l'utilisateur. Au cours des dernières décennies, plusieurs améliorations ont été apportées à ces paramétrisations classiques afin de mieux aborder les hétérogénéités spatiales. Il s'agit principalement de grilles adaptatives. Celles-ci sont par exemple construites en utilisant des B-splines comme dans les travaux de (Michelini, 1995) ou en utilisant des algorithmes génétiques voir (Curtis and Snieder, 1997), ou encore à partir de cellules de Voronoï, introduites par (Sambridge et al., 1995) et popularisées en inversion bayésienne par (Bodin and Sambridge, 2009; Bodin et al., 2009, 2012a). (pour une revue (Sambridge and Rawlinson, 2005)).

Dans la présente thèse, nous nous proposons de développer des modèles parcimonieux capables de reconstruire le champ de vitesse des ondes sismiques avec suffisamment de précision. Ces modèles sont particulièrement adaptés à l'approche bayésienne. Dans un premier temps, nous proposons un modèle basé sur un modèle de mosaïque aléatoire, le modèle de Jonhson-Mehl, dérivé des mosaïques de Voronoï déjà proposées en tomographie bayésienne. Le deuxième modèle que nous considérons est un modèle de vitesse décrit par une combinaison linéaire de fonctions gaussiennes, centrées sur la réalisation d'un processus ponctuel de Poisson. Ce modèle est exploitable en pratique pour décrire des anomalies locales du champ de vitesse. Enfin, nous présentons un modèle paramétrique, encore imparfait, basé sur la théorie du compressed sensing. A notre connaissance, l'utilisation du compressed sensing (CS) dans ce contexte est une nouveauté. Le CS est une théorie qui permet de définir des acquisitions linéaires de signaux mathématiques sous une forme directement comprimée. Cette théorie permet donc, dans le cadre d'un problème de tomographie inverse, de proposer une représentation parcimonieuse du champ de vitesse.

Dans cette thèse nous nous intéressons également aux modèles à couches qui sont particulièrement adaptés au cadre de la micro-sismicité induite par l'exploitation des ressources dans les bassins sédimentaires. Le monitoring de la sismicité induite apporte des informations sur la localisation de ruptures dans la structure imagée or la localisation des événements sismiques nécessite de supposer un modèle de vitesse. Il importe alors d'avoir une bonne estimation du modèle de vitesse auquel il faut ajouter une bonne connaissance des incertitudes associées. Nous utilisons dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle où le nombre de couches est inconnu et hiérarchique en ajoutant le bruit à l'ensemble des inconnus du problème inverse. Cette approche nous permet de traiter efficacement la complexité du modèle à couches.

La plupart des problèmes d'inversion traités dans cette thèse considèrent un nombre aléatoire de paramètres. Pour inférer la distribution a posteriori, nous utilisons un algorithme de Metropolis Hastings à saut réversible (Reversible Jump) (Green, 1995). Cet algorithme est

une généralisation de l'algorithme Metropolis Hastings qui permet de simuler la dimension de l'espace des inconnues. Il a été introduit en géophysique, par Malinverno, en 2002, dans le cadre d'une inversion électrique pour un modèle à couches (Malinverno, 2002). L'approche trans-dimensionnelle a reçu un accueil positif de la part des géophysiciens, nombre d'entre eux l'ont adoptée, citons (Agostinetti and Malinverno, 2010; Dettmer et al., 2010, 2012). Quelques années plus tard, son champ d'application s'est élargi aux problèmes en 2D, avec les travaux de (Bodin and Sambridge, 2009; Bodin et al., 2012a). Sa première application en 3D quant à elle est toute récente et a été mise en œuvre par Agostinetti en 2015 (Agostinetti et al., 2015).

Organisation du manuscrit :

Le premier chapitre du présent manuscrit a un caractère introductif. Nous y présentons en premier lieu les éléments de base de la tomographie sismique ainsi que les différentes approches pour sa résolution. Nous nous intéressons plus particulièrement à l'inférence bayésienne que nous détaillons dans un second temps.

Pour résoudre le problème d'inférence bayésienne, nous sommes amenés à échantillonner la loi a posteriori. À cette fin nous utiliserons des méthodes de simulation stochastique. Plus particulièrement, nous nous intéresserons aux méthodes de type Monte Carlo par chaîne de Markov (MCMC) qui seront utilisées avec profit dans le cadre de ce travail de thèse. **Le deuxième chapitre** est entièrement consacré à ces méthodes.

Le troisième chapitre est réservé à la paramétrisation et se décompose en deux parties : Dans la première partie, nous recensons les différentes paramétrisations utilisées dans la littérature dans le cadre bayésien. Dans la seconde partie nous présentons de nouveaux modèles paramétriques adaptés à l'approche bayésienne, nous détaillons dans ce contexte les lois a priori sur les différents paramètres ainsi que la méthode adoptée pour inférer la loi a posteriori.

Le quatrième chapitre présente deux applications des modèles paramétriques présentés au chapitre 3. Le premier modèle est similaire à celui présenté dans (Bodin and Sambridge, 2009; Bodin et al., 2009). Le deuxième modèle est plus complexe et sert de benchmark dans l'industrie pétrolière. Il a été déjà utilisé en tomographie des premières arrivées dans le cadre d'une inversion linéarisée (Taillandier et al., 2009).

Dans **le cinquième chapitre** nous nous intéressons à des données réelles issues de la micro-sismicité induite. Il s'agit des mêmes données présentées dans (Gesret et al., 2015; Bottero et al., 2016). Nous développons dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique afin de traiter efficacement la complexité du modèle à couches.

Enfin, **le sixième chapitre** est dévolu au compressed sensing. Nous commençons par exposer les bases de la théorie du compressed sensing. Nous présentons dans un second temps un algorithme d'inversion bayésienne basé sur le compressed sensing. Enfin, nous concluons en présentant les résultats de l'algorithme sur des modèles synthétiques, et en évoquant plusieurs pistes de recherches futures qu'il nous paraîtrait intéressant d'explorer.

Chapitre 1.

La tomographie sismique

La tomographie sismique est une méthode géophysique de prospection du sous sol se basant sur la propagation d'ondes sismiques. Elle est utilisée depuis de nombreuses années dans le but de retrouver un modèle de vitesse de propagation d'ondes sismiques à partir des temps mesurés entre les sources et diverses stations (Aki and Lee, 1976). Ces vitesses sont reliées aux densités des milieux et permettent ainsi une identification grossière de la lithologie des roches dans des zones inaccessibles par des mesures directes (Improta et al., 2002; Zelt et al., 2006). Cette technique nécessite la résolution d'un problème inverse afin d'obtenir un modèle sismique cohérent avec les données observées.

Nous nous intéressons en particulier à la tomographie en réfraction. Celle-ci se base sur les temps de première arrivée provenant des ondes directes ou réfractées. La tomographie en réfraction utilise les ondes coniques. Celles-ci arrivent à une interface séparant deux couches sous l'angle d'incidence correspondant à la réfraction limite et se propagent le long de l'interface à la vitesse de la couche inférieure. Cette méthode est étudiée en détails dans de nombreux ouvrages et articles scientifiques (Sheriff and Geldart, 1995; Milsom and Eriksen, 2011).

Exemple :

Prenons l'exemple d'un modèle à deux couches de vitesses différentes V_1 et V_2 avec $V_1 < V_2$ (Figure 1). L'onde sismique générée par la source S et enregistrée par les récepteurs peut suivre différents parcours :

- L'onde directe qui voyage de la source aux récepteurs dans le premier milieu à la vitesse de celui-ci (V_1).
- L'onde réfractée à l'interface entre les deux milieux qui se propage à la vitesse V_2 et réémet régulièrement une onde plane le long de l'interface avec le même angle d'incidence critique dans le premier milieu.
- L'onde réfléchi à l'interface, celle-ci ne pouvant jamais arriver avant l'onde directe ou réfractée, n'est donc pas considérée dans la tomographie en réfraction.

Les temps de première arrivée forment des droites. Leurs pentes correspondent à l'inverse de la vitesse dans la couche le long de laquelle l'onde est réfractée. A partir d'une certaine distance x_c , l'onde réfractée, bien qu'elle parcoure un trajet plus long que l'onde directe arrive avant celle-ci. La position x_c est appelée *distance critique* et correspond à la position x pour laquelle l'onde directe et l'onde réfractée arrivent au même temps.

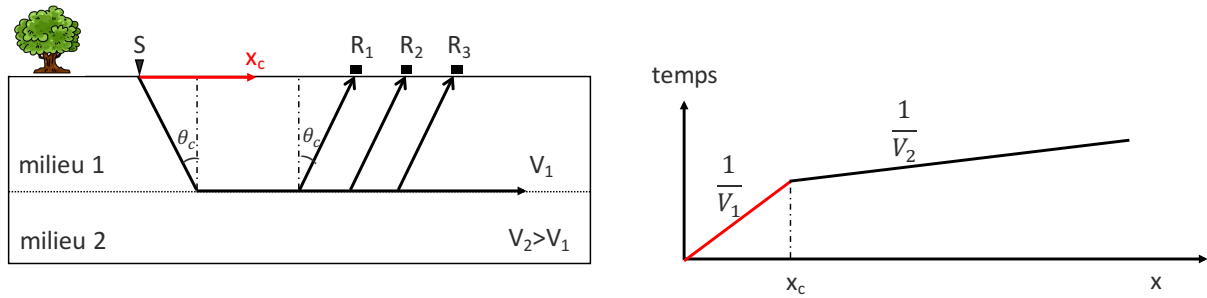


Figure 1: Réfraction pour un terrain à deux couches horizontales.

Eléments de la tomographie sismique :

Comme cité, dans (Tarantola, 2005), les éléments indispensables pour résoudre le problème inverse sont:

- 1- *La paramétrisation du modèle de vitesse* : Elle consiste à donner une expression mathématique à la distribution de la vitesse dans le sous-sol. Cette expression fait intervenir des paramètres qui représentent les inconnues du problème inverse.
- 2- *Le problème direct* : Il s'agit de la génération des temps de propagation des ondes sismiques à partir des paramètres. Le problème direct consiste, dans notre cas, à résoudre l'équation Eikonal, décrite plus loin.
- 3- *L'inversion* : Cette étape consiste à manipuler les valeurs des paramètres du modèle de vitesse afin de minimiser la différence entre les données observées et les données prédites. L'étape a pour but de trouver un ou des modèle(s) qui correspond(ent) aux données observées.

1 La paramétrisation du modèle de vitesse :

En pratique, le modèle de vitesse $V(x)$ que l'on essaie d'inférer est approximé par une combinaison linéaire, $v(x)$, d'un nombre fini de fonctions de base φ_i

$$V(x) \approx v(x) = \sum_i a_i \varphi_i(x) \quad (1)$$

L'objectif de cette étape est la détermination d'un nombre minimal de paramètres, dits paramètres principaux (Tarantola, 1987). Dans toute la suite de ce document, m désignera le vecteur contenant les paramètres qui seront estimés par l'approche inverse.

$$v(x) = f(m) \quad (2)$$

Exemple :

Reprenons, l'exemple de la figure 1. Une manière de paramétrer celui-ci est de considérer les trois paramètres donnés par le vecteur suivant :

$$m = (z, V_1, V_2) \quad (3)$$

Où z correspond à la position de l'interface, V_1 et V_2 sont les vitesses dans les deux couches. La vitesse en chaque point $u(x_u, z_u)$ du modèle s'écrit alors sous la forme suivante :

$$v(u) = f(m) = V_1 1_{[0,z]}(z_u) + V_2 1_{]z,L]}(z_u) \quad (4)$$

La paramétrisation a une influence sur la formulation du problème et sa résolution, le choix des paramètres à estimer est, de ce fait, une étape importante dans la résolution du problème inverse. Plusieurs paramétrisations sont possibles pour le même modèle (Tarantola, 1987). La bonne paramétrisation est celle qui permet de tirer le maximum d'informations à partir des données sans rajouter une complexité injustifiée à la structure du modèle. Le chapitre 3 de ce manuscrit est entièrement dédié à cette problématique.

2 Le problème direct:

Cette étape a pour objectif de calculer les données à partir des valeurs des paramètres proposées lors de l'inversion sismique. Le résultat de cette étape est ensuite utilisé pour calculer la valeur d'une quantité scalaire appelée fonction coût, représentative de l'écart existant entre les données observées et les données calculées. Cette fonction est décrite plus loin.

Rappelons que dans le cadre de la tomographie sismique que nous étudions, les données sont les temps de première arrivée et le modèle est un modèle de vitesse de propagation des ondes sismiques.

2.1 Equation Eikonal:

Cette partie n'a pas vocation à décrire précisément les lois physiques qui régissent le comportement des ondes sismiques. Le lecteur curieux de ces détails pourra les retrouver dans de nombreux ouvrages (Aki and Richards, 1980; Nolet et al., 2008). Nous nous contenterons de montrer comment l'équation Eikonal se déduit à partir de l'équation de l'élasto-dynamique.

Considérons l'équation de l'élasto-dynamique en milieu isotrope et homogène donnée par (5). Cette équation relie les déformations d'un objet aux forces appliquées à cet objet.

$$\rho \frac{\partial^2 u}{\partial t^2} - (\lambda + \mu) \text{grad}(\text{div}(u)) - \mu \Delta u = 0 \quad (5)$$

$u(x, t)$ est le champ vectoriel de déplacement qui associe la position du point x au repos à sa position actuelle dans la configuration déformée. Les constantes de Lamé du milieu sont λ et μ . Les opérateurs div et Δ représentent respectivement la divergence et le Laplacien.

D'après la décomposition de Helmholtz, pour tout champ vectoriel, en l'occurrence $u(x, t)$, il existe toujours un champ scalaire $\phi(x, t)$ et un champ vectoriel $\psi(x, t)$ tels que :

$$u = \text{grad}(\phi) + \text{rot}(\psi) \quad (6)$$

Le champ de déplacement u peut donc s'écrire sous la forme suivante :

$$u = u_p + u_s \quad (7)$$

Avec :
$$\text{rot}(u_p) = 0 \text{ et } \text{div}(u_s) = 0 \quad (8)$$

Où :

- u_p correspond au champ de déplacement des ondes de **compression P**, appelées aussi ondes longitudinales. Elles sont les plus rapides et se propagent dans tous les milieux en créant successivement des zones de dilatation et des zones de compression (Figure 2a).
- u_s correspond au champ de déplacement des ondes de **cisaillement S**, ou encore ondes transversales. Elles ne se propagent que dans les milieux solides. Les particules oscillent dans un plan perpendiculaire à la direction de propagation de l'onde (Figure 2b).

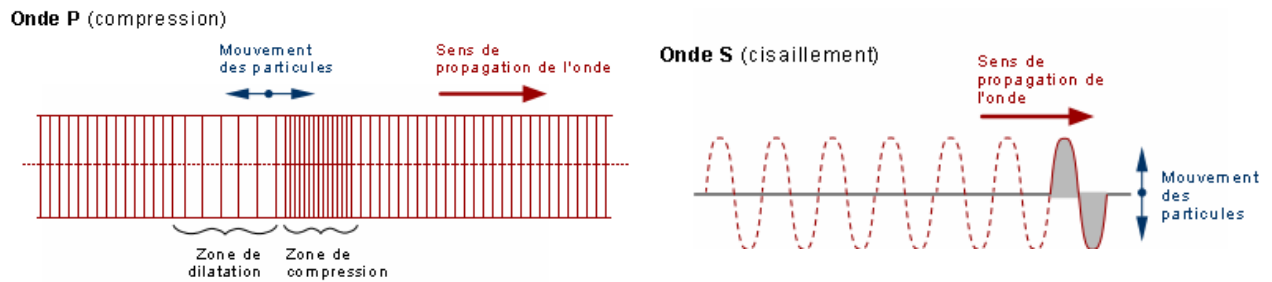


Figure 2: Mouvements des particules d'un milieu dans lequel se propagent les ondes de volume : a) compression b) cisaillement

En remplaçant u par (7), l'équation (5) devient :

$$\rho \frac{\partial^2 u_p}{\partial t^2} + \rho \frac{\partial^2 u_s}{\partial t^2} - (\lambda + \mu) \text{grad}(\text{div}(u_p)) - \mu \Delta u_p - \mu \Delta u_s = 0 \quad (9)$$

sachant que :

$$\text{grad}(\text{div}(u_p)) = \Delta u_p + \text{rot}(u_p) = \Delta u_p \quad (10)$$

L'équation de l'elastodynamique (5) peut finalement s'écrire sous la forme :

$$\rho \frac{\partial^2 u_p}{\partial t^2} + \rho \frac{\partial^2 u_s}{\partial t^2} - (\lambda + 2\mu) \Delta u_p - \mu \Delta u_s = 0 \quad (11)$$

Etant données les propriétés des deux champs u_p et u_s , leurs expressions vectorielles sont nulles séparément :

$$\rho \frac{\partial^2 u_p}{\partial t^2} - (\lambda + 2\mu) \Delta u_p = 0 \quad (12)$$

$$\rho \frac{\partial^2 u_s}{\partial t^2} - \mu \Delta u_s = 0 \quad (13)$$

Nous obtenons ainsi les équations de propagation des ondes P et S :

$$\frac{\partial^2 u_p}{\partial t^2} - \frac{1}{c_p^2} \Delta u_p = 0 \quad (14)$$

$$\frac{\partial^2 u_s}{\partial t^2} - \frac{1}{c_s^2} \Delta u_s = 0 \quad (15)$$

Avec

$$c_p = \sqrt{\frac{\rho}{\lambda+2\mu}} : \text{vitesse des ondes P} \quad \text{et} \quad c_s = \sqrt{\frac{\rho}{\mu}} : \text{vitesse des ondes S}$$

La solution pour chacune de ces équations s'écrit dans le domaine fréquentiel sous la forme d'un produit d'une exponentielle complexe à phase réelle par une amplitude :

$$u(x, \omega) = S(\omega)A(x, \omega)e^{i\omega T(x)} \quad (16)$$

Avec $S(\omega)$ la signature de la source, $A(x, \omega)$ l'amplitude de l'onde et $\omega T(x)$ sa phase.

Du fait de l'approximation des hautes fréquences, l'amplitude, $A(x, \omega)$ peut s'écrire comme une série de puissance inverses de la fréquence.

$$A(x, \omega) = \sum_{n=0}^{+\infty} \frac{A_n(x)}{(i\omega)^n} \quad (17)$$

En pratique, on ne garde pas les termes d'ordre élevé du développement limité, seul le terme d'ordre 0 est utilisé :

$$u(x, \omega) \approx A_0 S e^{i\omega T(x)} \quad (18)$$

Nous obtenons alors à une équation non-linéaire du premier ordre pour les temps de propagation :

$$(\nabla T(x))^2 = s^2(x) \quad (19)$$

s étant la lenteur, l'inverse de la vitesse de propagation de l'onde:

$$s(x) = \frac{1}{c(x)} \quad (20)$$

2.2 Résolution de l'équation Eikonal

(Vidale, 1988) est le premier à introduire un schéma aux différences finies permettant de propager des fronts d'ondes plans ou circulaires. Les gradients des temps de parcours sont discrétisés et le calcul de ces temps est fait à partir d'une grille régulière. Par la suite, (Podvin and Lecomte, 1991) en s'inspirant des travaux de Vidale, ont proposé une méthode basée également sur un schéma aux différences finies. A la différence de la méthode proposée par Vidale, la propagation des ondes prend en considération les phénomènes de diffraction, de transmission. Le calcul des temps de trajet, en chaque point de la grille, se fait en prenant en compte la possibilité d'avoir plusieurs fronts d'ondes localement indépendants mais émanant de la même source, ainsi parmi les solutions candidates (correspondant à chaque front d'onde), seul le temps minimal est retenu. De plus, Podvin et Lecomte ont utilisé une représentation physique du modèle de vitesse qui est constitué de pavés de vitesse constante.

Plus récemment, (Noble et al., 2014) ont mis au point un nouveau solveur Eikonal qui calcule les temps de trajet par différences finies. Ce nouveau solveur utilise conjointement l'équation Eikonal factorisée, qui permet une meilleure approximation des fronts d'ondes au voisinage de la source, et l'approximation des ondes planes plus loin. Ce nouveau solveur utilise la Fast Sweeping Method pour balayer la grille des temps de parcours dans des ordres alternés, permettant ainsi, la mise à jour des temps sur la totalité du modèle. Enfin, (Noble et al., 2014) ont introduit des grilles de temps avec des pas variables permettant de mieux gérer le modèle de vitesse. Une comparaison a été faite dans (Noble et al., 2014) entre, d'une part, les méthodes classiques de (Vidale, 1988) et de (Podvin and Lecomte, 1991) et, d'autre part, le nouveau solveur. Les résultats de cette comparaison montrent que ce dernier permet de calculer de manière plus précise les temps de trajets.

Remarque :

L'approximation des hautes fréquences s'applique à des problèmes de propagation dans lesquels la longueur d'onde est très petite vis-à-vis de la longueur caractéristique du milieu. Elle n'est licite que lorsqu'on observe une séparation d'échelle entre la longueur d'onde et la longueur de variation caractéristique des propriétés physiques du milieu. En d'autre terme, il est nécessaire que les propriétés du milieu (élasticité, densité, etc..) soient constantes en première approximation sur une longueur du même ordre de grandeur que quelques longueurs d'onde.

3 L'inversion :

Notons $d^{obs} \in \mathbb{R}^d$ le vecteur des données observées, $d^{obs} = (d_i^{obs})_{1 \leq i \leq d}$.

L'équation reliant les paramètres qui caractérisent un modèle de vitesse $m \in \mathbb{R}^p$ avec les données observées d^{obs} peut s'écrire sous la forme suivante :

$$d^{obs} = F(m) \quad (21)$$

En réalité, aucun dispositif expérimental n'est complètement affranchi d'incertitude. De plus, dans la majorité des cas le modèle mathématique utilisé ne traduit pas exactement la réalité en raison des différentes hypothèses et approximations utilisées. De ce fait, il est plus réaliste de considérer que le modèle de vitesse et les données mesurées sont reliées par une équation de la forme :

$$d^{obs} = F(m) + \epsilon \quad (22)$$

où ϵ est appelé bruit représentant l'ensemble des erreurs et prenant en compte les erreurs de mesure et de modélisations.

- L'opérateur F est souvent **non-linéaire** ce qui rend le problème inverse difficile à résoudre.

Rappelons tout d'abord, que le but de la tomographie sismique est de trouver les modèles qui expliquent le mieux les données observées. C'est-à-dire les modèles de vitesse qui simulent des données synthétiques les plus proches possibles des données observées. Ces modèles minimisent une certaine quantité qu'on appelle fonction coût. Cette fonction mesure l'écart entre les données prédites pour chaque modèle et les données observées. Il existe plusieurs manières d'exprimer cette fonction. La fonction coût la plus utilisée en tomographie sismique est celle des moindres carrés obtenue sous l'hypothèse gaussienne de l'ensemble des erreurs (modèles et données). Nous allons voir une formulation de cette fonction plus loin dans le cadre de l'inversion bayésienne. Pour l'instant nous allons nous contenter de la noter $C(m)$ (fonction coût pour un modèle m).

Il existe trois approches possibles pour aborder ce problème de non linéarité :

- *l'approche linéaire* : Elle consiste à définir un modèle de référence m^0 et à faire un tracé de rais pour ce modèle initial : Le trajet du rayon est connu a priori, sous l'approximation linéaire, qui néglige la dépendance entre le trajet du rayon et la distribution de la vitesse. Ensuite, à partir de cette hypothèse et en linéarisant la relation reliant la vitesse et le temps de propagation autour du modèle de référence m^0 , une seule correction/ perturbation Δm est effectuée à la vitesse pour obtenir le modèle représentant la solution du problème.

La perturbation Δm est déterminée en faisant appel à des méthodes d'analyse fonctionnelle et d'algèbre linéaire, notamment la décomposition en valeurs singulières (SVD) de l'opérateur choisi (Bertero and Boccacci, 1998; Aster et al., 2005).

– L'approche itérativement non-linéaire : Elle ne considère pas la dépendance des raies par rapport à la vitesse mais elle prend en compte la non-linéarité du problème en faisant une suite de corrections du modèle de vitesse basées sur la linéarisation précédente et un tracé de raies jusqu'à converger à la solution désirée.

Il s'agit de résoudre le problème inverse de manière itérative en linéarisant le problème à chaque itération autour du modèle issu de l'itération précédente. Cette résolution fait appel à des méthodes d'optimisation locale, en partant d'un modèle m^0 supposé localisé dans le bassin d'attraction du minimum global de la fonction coût $C(m)$. A chaque itération k , le minimum de la fonction coût $C(m^k)$ est recherché au voisinage d'un modèle de départ m^{k-1} obtenu à l'itération précédente. Le modèle m^k s'écrit dès lors en fonction du modèle précédent m^{k-1} et d'une certaine perturbation Δm^k .

$$m^k = m^{k-1} + \Delta m^k \quad (23)$$

Pour déterminer la perturbation Δm par une approche itérative, des méthodes de projection et de reconstruction sont utilisées. Nous citons par exemple la *Simultaneous Iterative Reconstruction Technique* (SIRT) qui est la méthode de reconstruction la plus utilisée, et la *Least Squares QR* (LSQR), qui est la méthode de projection dont l'usage est le plus répandu (Nolet, 1985).

– l'approche complètement non-linéaire : Elle propose une solution sans aucune hypothèse de linéarisation. Elle se base seulement sur la résolution de l'équation Eikonal pour trouver le modèle qui correspond le mieux aux données observées.

Les approches précédentes (linéaires et itérativement non-linéaires), n'exploitent que l'information dans des régions limitées de l'espace modèle afin de converger vers une solution. Ces techniques dépendent fortement des modèles initiaux, et ne permettent pas d'estimer de manière fiable les incertitudes associées au modèle de vitesse. Les méthodes complètement non-linéaires s'appuient généralement sur un échantillonnage exhaustif de l'espace des paramètres. Elles sont ainsi en mesure de fournir un ensemble de modèles, qui peuvent être ensuite utilisés pour l'évaluation des incertitudes. L'inconvénient évident de cette approche est que les ressources informatiques nécessaires sont généralement supérieures à celles nécessaires pour un système linéaire ou itérativement non-linéaire.

- L'opérateur \mathbf{F} est souvent **non bijectif**, le problème inverse est donc mal-posé.

Un problème est dit mal posé au sens de Hadamard s'il satisfait l'une des conditions suivantes : l'inexistence, la non-unicité ou l'instabilité de la solution.

- *L'inexistence* : signifie que le problème n'admet pas de solution.
- *la non-unicité* : signifie que plusieurs combinaisons de valeurs des paramètres du modèle permettent de minimiser la fonction coût.

- *l'instabilité* signifie que de faibles perturbations sur les données peuvent se traduire par une variation arbitrairement grande de la solution associée.

Le caractère mal posé du problème inverse nécessite la régularisation ou l'introduction d'une information a priori. La régularisation Tikhonov (Tikhonov and Arsenin, 1977) est la méthode de régularisation la plus utilisée. La régularisation consiste dans un premier temps à définir un ensemble de solutions admissibles puis de choisir parmi ces solutions celle qui est compatible avec une certaine information a priori $\phi(m)$. La solution obtenue n'est donc pas l'unique solution possible, mais celle qui réalise un compromis entre la fidélité aux données observées et la régularisation. Elle est obtenue en minimisant un critère du type (Tikhonov and Arsenin, 1977).

$$J(m) = d(d^{obs}, F(m)) + \lambda\phi(m) \quad (24)$$

$d(d^{obs}, F(m))$ est la distance entre les observations d^{obs} et les données calculées, en général $d(d^{obs}, F(m)) = \|d^{obs} - F(m)\|_2$ et λ est le paramètre de régularisation.

La difficulté de la tomographie sismique réside dans la non-linéarité du problème direct et le caractère mal posé du problème inverse. L'approche bayésienne permet de combiner naturellement l'information apportée par les observations et l'information a priori. Elle représente un moyen pertinent pour remédier à ces difficultés en prenant en compte les incertitudes expérimentales et théoriques sur les paramètres et les données.

4 Tomographie bayésienne:

4.1 Formulation du problème inverse :

L'approche bayésienne est une approche probabiliste. Dans ce contexte, toutes les informations sont représentées avec des densités de probabilité. Elle consiste à construire une loi appelée *loi a posteriori* représentant la loi de probabilité des paramètres inconnus $m = (m_i)$ conditionnellement aux données observées d^{obs} . Cette loi permet de combiner les informations contenues dans les observations et la connaissance a priori sur les paramètres. Elle est obtenue à partir de la *vraisemblance* contenant les relations théoriques et les propriétés du bruit, et d'une *loi a priori*, représentant les informations sur les inconnues dont on dispose avant de connaître les données. Elle est obtenue par la formule de Bayes (Bayes and Price, 1763):

$$P(m|d^{obs}) = \frac{P(d^{obs}|m)P(m)}{P(d^{obs})} \quad (25)$$

Où $P(a|b)$ désigne la probabilité conditionnelle de a sachant b.

Le terme $P(d^{obs})$ qui apparaît au dénominateur est appelé évidence et représente la loi de probabilité des données. Cette loi est obtenue en intégrant la loi jointe des données par rapport à tous les modèles possibles.

$$P(d^{obs}) = \int P(d^{obs}, m) \cdot dm = \int P(d^{obs}|m) P(m) dm \quad (26)$$

Comme elle est indépendante du modèle m choisi et tout simplement impossible à calculer dans des problèmes de grande dimension, on la considère comme une constante de normalisation.

La loi a posteriori (25) s'écrit alors sous la forme de la relation proportionnelle suivante

$$P(m|d^{obs}) \propto P(d^{obs}|m)P(m) \quad (27)$$

Où \propto signifie proportionnel à

La vraisemblance $P(d^{obs}|m)$:

Il s'agit de la probabilité d'observer les données d^{obs} sachant un modèle m donné. Cette loi fait appel au problème direct reliant les variables inconnues aux données. Elle traduit l'adéquation entre les données prédites $F(m)$ et les observations et dépend de la modélisation du bruit. Une forme usuelle de la vraisemblance est la vraisemblance gaussienne (Mosegaard and Tarantola, 1995).

$$P(d^{obs}|m) = \frac{\exp\left(-\frac{1}{2} \left(d^{obs} - F(m)\right)^t C_d^{-1} \left(d^{obs} - F(m)\right)\right)}{(2\pi)^{\frac{NP}{2}} |C_d|^{\frac{1}{2}}} \quad (28)$$

- C_d représente la matrice de covariance des données.
- $F(m)$ les données prédites à partir du modèle m .
- $|C_d|$ le déterminant de la matrice.

Outre sa manipulation relativement aisée, l'utilisation de la loi gaussienne pour modéliser le bruit est justifiée par le théorème de la limite centrale qui implique qu'un bruit résultant de l'action indépendante d'un grand nombre de différents facteurs physiques de même moyenne et même variance obéit à la loi normale.

Très souvent, la matrice de covariance est proportionnelle à l'identité

$$C_d = \sigma_d^2 I_{Na} \quad (29)$$

La complexité de la solution dépend fortement de l'écart-type des données σ_d . Si sa valeur est incertaine, il est judicieux de le considérer comme un paramètre inconnu (hyper-paramètre) du problème inverse. Le modèle bayésien hiérarchique (Raudenbush and Bryk, 2002;

Goldstein, 2011) introduit en géophysique par (Malinverno and Briggs, 2004; Malinverno and Parker, 2006) permet de traiter ce problème.

La loi a priori $P(m)$:

Le choix de la loi a priori $P(m)$ est un choix subjectif. Il est l'aspect le plus critiqué de l'approche bayésienne (Robert, 2006). La loi a priori est généralement choisie pour transcrire fidèlement les connaissances sur le modèle dont on dispose avant l'observation. Toute information sur les paramètres qui est disponible indépendamment des données peut /doit être utilisée comme une information a priori et incluse dans la formulation de $P(m)$. Celle-ci peut venir par exemple de la connaissance géologique du milieu. En pratique, elle est choisie au cas par cas et dépend de la paramétrisation.

En tomographie sismique, la loi normale est souvent utilisée. La connaissance a priori sur les paramètres consiste dans ce cas en un ordre de grandeur de leur valeur associé à un intervalle d'incertitude.

$$P(m) = \frac{\exp\left(-\frac{1}{2}(m - m_{prior})^t C_p^{-1}(m - m_{prior})\right)}{(2\pi)^{\frac{N^p}{2}} |C_p|^{\frac{1}{2}}} \quad (30)$$

Où N^p est le nombre de paramètres, m_{prior} , le vecteur de paramètres contenant les valeurs supposées et C_p la covariance associée.

La loi normale appartient à la famille dite conjuguée pour la vraisemblance gaussienne. On dit qu'une loi a priori est conjuguée pour la vraisemblance si la loi a posteriori est de la même classe que la loi a priori. L'intérêt majeur de ce type de lois est de faciliter les calculs.

Il peut arriver toutefois qu'aucune information pertinente sur les paramètres ne soit disponible a priori. Dans de telles situations, il convient d'utiliser des lois non-informatives ne privilégiant aucune valeur des paramètres.

(Laplace, 1921), propose d'utiliser une loi uniforme pour décrire le manque d'information.

$$P(m_i) = \frac{1}{\Delta m_i} \text{ si } m_i \in [m_{i_{min}}, m_{i_{max}}] \quad (31)$$

m_i étant la i ème composante i du vecteur des paramètres m . Toutes les valeurs sont équiprobables a priori.

(Jeffreys, 1946) quant à lui propose d'adopter une distribution uniforme entre $-\infty$ et ∞ :

$$P(m_i) \propto 1 \quad (32)$$

La densité est impropre étant donné que l'intégrale $\int_{-\infty}^{+\infty} P(x) dx$ n'est pas définie.

Prenant l'exemple de l'écart-type σ_d . Jeffreys propose d'utiliser une loi uniforme pour le logarithme de σ_d entre $-\infty$ et ∞ :

$$P(\log(\sigma_d)) \propto 1 \Leftrightarrow P(\sigma) \propto \frac{1}{\sigma} \quad (33)$$

La loi a priori sur l'écart-type (33) est non-informative car il est impossible de dire s'il est plus probable que σ soit inférieur ou supérieur à une certaine valeur σ_m étant donné que les intégrales $\int_0^{\sigma_m} P(\sigma)d\sigma$ et $\int_{\sigma_m}^{+\infty} P(\sigma)d\sigma$ ne sont pas définies.

Par ailleurs, cette loi est invariante par une transformation de type puissance : Si la loi a priori de l'écart-type σ est non informative, cela est tout aussi vrai pour la variance (σ^2). En effet, si $P(\log(\sigma)) \propto 1$ alors $P(\log(\sigma^a)) \propto 1$ (avec $a > 0$) car $\log \sigma^a = a \log \sigma$.

Plus tard, ce résultat d'invariance fût généralisé par (Jeffreys, 1961) qui proposa une loi non informative plus générale fondée sur l'information de Fisher donnée par :

$$I(m_i, m_j) = \mathbb{E} \left[\left(\frac{\partial \log(P(d^{obs}|m))}{\partial m_i} \right) \left(\frac{\partial \log(P(d^{obs}|m))}{\partial m_j} \right) \right] \quad (34)$$

La matrice d'information de Fisher permet de quantifier l'information apportée par les données sur le vecteur des paramètres m (voir (Fisher, 1956)). La loi non-informative de Jeffreys est définie par :

$$P(m) = |I(m)|^{\frac{1}{2}} \quad (35)$$

Ce type de loi favorise les valeurs de m pour lesquelles $I(m)$ est grande et permet donc de minimiser l'influence de la loi a priori. D'autres approches permettant de représenter l'ignorance sont présentées dans (Robert, 2006).

La loi a posteriori (25) peut finalement s'écrire sous la forme de la relation proportionnelle suivante :

$$P(m|d^{obs}) \propto \frac{1}{(\sigma_d)^{N_d}} \cdot \exp \left(-\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2 \right) \cdot P(m) \quad (36)$$

Lien avec la régularisation classique :

Considérons le cas d'une loi a priori sous forme d'exponentielle d'une certaine fonction potentiel $\phi(m)$ ie $P(m) = \exp(\lambda\phi(m))$.

$$P(m|d^{obs}) \propto \exp \left(-\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2 + \lambda \phi(m) \right) \quad (37)$$

Choisir un estimateur au sens du maximum a posteriori est équivalent à trouver le minimum de la quantité suivante :

$$-\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2 + \lambda \phi(m) \quad (38)$$

Ce qui correspond à la formulation de Tikhonov. Le formalisme bayésien a plusieurs avantages par rapport à l'analyse classique. Notamment le choix de l'estimateur ne se cantonne pas au maximum a posteriori (MAP). De plus, l'approche bayésienne permet de définir non seulement une estimation sur les inconnues mais également un intervalle de confiance associé à cette estimation.

4.2 Résolution du problème inverse : Estimateurs bayésiens

La solution bayésienne pour la tomographie sismique est la densité de probabilité a posteriori (25). Celle-ci contient toute l'information a posteriori disponible sur le vecteur des paramètres m et par conséquent sur le modèle de vitesse. Cette information est souvent résumée dans des quantités comme le maximum a posteriori, la moyenne et variance a posteriori ou d'autres quantiles.

La plupart de ces quantités peuvent être représentées par une intégrale de la forme suivante :

$$E[h(m)|d^{obs}] = \int h(m)P(m|d^{obs})dm \quad (39)$$

comme par exemple $h(m) = m$ pour la moyenne a posteriori $E(m)$, ou encore $h(m) = (m - E(m))^2$ pour la variance.

Pour la plupart des problèmes inverses que nous traitons dans cette thèse, nous utilisons l'estimateur de la moyenne a posteriori. Cet estimateur a la propriété d'être le meilleur au sens de l'erreur quadratique moyenne c'est à dire de minimiser le risque bayésien associé au coût quadratique (Robert, 2006).

$$C(m, \hat{m}) = \|m - \hat{m}\|_2 \quad (40)$$

Cette fonction (40) représente le coût de se tromper lorsque la vraie mesure est \hat{m} .

5 Conclusion

Dans ce chapitre nous avons posé certaines généralités concernant la tomographie sismique, plus particulièrement la tomographie en réfraction en présentant brièvement les étapes requises pour imager le sous sol. Nous avons également abordé les différentes approches pour la résolution de la tomographie des temps de première arrivée et nous avons vu l'intérêt de formuler le problème inverse dans un cadre bayésien. Toutefois, la mise en

pratique de l'approche bayésienne peut se révéler problématique car il est souvent impossible de trouver une expression analytique de la loi a posteriori. La complexité de la loi a posteriori ne permet pas d'obtenir une expression simple de l'estimateur bayésien. Par conséquent l'intégrale (39) peut être difficile à évaluer. Il importe ainsi d'utiliser des méthodes de simulation appropriées. Les méthodes de Monte-Carlo par chaînes de Markov sont les mieux adaptées. Elles permettent de fournir des échantillons distribués approximativement suivant la loi a posteriori. Ces échantillons sont ensuite utilisés pour calculer l'estimateur bayésien. Ces méthodes sont présentées dans le chapitre suivant.

Chapitre 2.

Méthodes MCMC

Dans cette partie, nous nous intéressons aux méthodes de Monte Carlo par chaînes de Markov (MCMC). Nous exposons brièvement les propriétés élémentaires des chaînes de Markov et rappelons les principaux résultats de convergence. Pour de plus amples détails sur le sujet nous renvoyons le lecteur vers les références (Robert and Casella, 2004; Meyn and Tweedie, 2012). Nous présentons également quelques algorithmes classiques des méthodes de Monte Carlo par chaînes de Markov.

1 Motivation

Les méthodes de Monte Carlo par chaînes de Markov sont des méthodes de simulation qui permettent d'approcher des quantités associées à une certaine loi de probabilité. Elles consistent à générer des variables aléatoires suivant une loi cible connue à une constante près.

L'objectif principal des méthodes de Monte-Carlo est d'approcher des intégrales du type :

$$E[h(m)|d^{obs}] = \int h(m)P(m|d^{obs})dm \quad (41)$$

En règle générale, le calcul de cette expression est très complexe, car la forme des fonctions $P(\cdot|d^{obs})$ et h ne permet pas de manipuler facilement l'intégrale. Le but est alors d'approcher l'espérance par une moyenne empirique de $h(x)$:

$$E[h(m)] \approx \frac{1}{N} \cdot \sum_{n=1}^N h(m^n) \quad (42)$$

où $\{m^n, n \in [1, N]\}$ est un ensemble de réalisations de la variable m générées *indépendamment* suivant la loi a posteriori $P(m|d^{obs})$.

La convergence de la moyenne empirique vers la moyenne théorique est assurée grâce à la loi des grands nombres.

C'est souvent un problème non trivial que de simuler un échantillon suivant la loi $P(m|d^{obs})$. En effet, les méthodes de simulation classiques, comme les méthodes de rejet et d'inversion, atteignent rapidement leurs limites quand la dimension du problème augmente.

Une solution est alors de construire une chaîne de Markov $(m^n)_{n \in \mathbb{N}}$ qui converge en loi vers la densité $P(m|d^{obs})$. En d'autres termes, on cherche à construire une chaîne de Markov telle qu'à partir d'une certaine itération M (burn-in), le noyau de transition itéré K^n ait approximativement la densité $P(m|d^{obs})$:

$$K(m^{n+1}|m^n) \approx P(m^{n+1}|d^{obs}) \forall n \geq M \quad (43)$$

Une fois la convergence établie, on peut obtenir des échantillons aléatoires suivant la loi $P(m|d^{obs})$ et ainsi approcher l'espérance a posteriori de notre fonction d'intérêt $h(m)$, $E[h(m)]$, par :

$$E[h(m)] \approx \frac{1}{N} \cdot \sum_{n=M+1}^{N+M} h(m^n) \quad (44)$$

Convergence de chaînes de Markov homogènes

Nous n'allons pas détailler la théorie des chaînes de Markov, pour plus de détails le lecteur peut se référer au chapitre 6 de (Robert and Casella, 2004). Nous nous contenterons de rappeler brièvement les conditions qui assurent la convergence en loi de la chaîne de Markov de noyau de transition K vers la loi cible $P(m|d^{obs})$. Ces conditions sont regroupées sous la notion d'ergodicité qui permet de valider un algorithme MCMC.

- **apériodicité** : la chaîne de Markov n'adopte pas un comportement périodique lors de ses transitions.
- **irréductibilité** : la chaîne de Markov visite tous les ensembles de mesure $P(m|d^{obs})$ non nuls. Elle ne peut pas, théoriquement, rester bloquée dans une sous-partie de l'espace d'état.
- **invariance** : le noyau de transition K doit avoir $P(.|d^{obs})$ comme point fixe.

Les algorithmes de Monte Carlo par chaînes de Markov (MCMC) permettent de construire une chaîne de Markov ergodique de loi stationnaire la distribution a posteriori $P(m|d^{obs})$ (25). Dans la littérature on distingue deux familles de méthodes MCMC classique : l'échantillonneur de Gibbs¹ (Geman and Geman, 1984), et l'algorithme de Metropolis Hastings (Metropolis et al., 1953; Hastings, 1970). Dans ce qui suit, notre attention se portera essentiellement sur l'algorithme Metropolis et sur quelques-unes de ses variantes.

2 Algorithmes de type Metropolis-Hastings

2.1 L'algorithme Metropolis-Hastings classique

Il s'agit de l'une des méthodes de Monte Carlo les plus générales. Il repose sur l'utilisation d'une densité de probabilité conditionnelle appelée probabilité propositionnelle /instrumentale $q(y|m)$. L'algorithme Metropolis-Hastings simule une chaîne de Markov en proposant un candidat m^* tiré selon la loi q pour succéder à l'échantillon courant. Ce candidat est accepté ou rejeté selon une procédure d'acceptation/rejet.

¹ Gibbs peut être vu comme un Metropolis-Hastings particulier.

L'algorithme produit alors une chaîne de Markov (m^n) qui converge vers la loi a posteriori $P(m|d^{obs})$, quel que soit l'état initial de celle-ci.

Algorithme Metropolis-Hasting (MH)

Étant donnée une initialisation m^0 de la chaîne.

À l'itération ($n + 1$),

1. tirer $m^* \sim q(m^*|m^n)$

2. $\left\{ \begin{array}{l} \text{prendre } m^{n+1} = m^* \text{ avec une probabilité } \alpha(m^n, m^*) \text{ définie par :} \\ \alpha(m^n, m^*) = \min \left\{ \frac{P(m^*|d^{obs}) \cdot q(m^n|m^*)}{P(m^n|d^{obs}) \cdot q(m^*|m^n)}, 1 \right\} \\ \text{prendre } m^{n+1} = m^n \text{ avec une probabilité } 1 - \alpha(m^n, m^*) \end{array} \right. \quad (45)$

Remarque : La loi a posteriori $P(m|d^{obs})$ n'apparaissant que dans le ratio, l'algorithme Metropolis Hastings permet de simuler la loi a posteriori sans avoir à connaître la constante de normalisation.

Le noyau de transition K de la chaîne de Markov associé à l'algorithme est donné par:

$$K(x, A) = \int_A \alpha(x, y)q(x, y)dy + 1_A(x) \int_{\mathcal{M}} (1 - \alpha(x, y))q(x, y)dx \quad (46)$$

Où \mathcal{M} représente l'espace des paramètres.

Ce noyau vérifie la condition d'équilibre ci-dessous dite de la balance détaillée (47), la chaîne admet alors la loi a posteriori comme mesure invariante. De plus quand q vérifie $q(y|x) > 0 \forall x, y$ la chaîne est irréductible. La chaîne est également apériodique car l'événement $m^{n+1} = m^n$ est possible en tout temps.

$$K(x, y)P(x|d^{obs}) = K(y, x)P(y|d^{obs}) \quad (47)$$

Selon le choix de la loi q , on distingue différentes classes d'algorithme MH : Quand la loi q ne dépend pas de l'état courant de la chaîne, $q(m^n|m^*) = q(m^*)$, l'algorithme est appelé échantillonneur indépendant (Metropolis-Hastings indépendant IMH). Dans le cadre de l'inversion bayésienne, q est souvent identique à la loi a priori sur les paramètres m . Il s'agit alors d'un échantillonneur global qui permet une large exploration de l'espace des paramètres. Néanmoins, il peut parfois être incapable d'identifier les différents modes de la loi a posteriori $P(m|d^{obs})$.

Quand q est symétrique, l'algorithme se réduit à l'algorithme Metropolis (Metropolis et al., 1953). Parmi les algorithmes Metropolis-Hastings symétriques les plus utilisés, figure l'algorithme de Metropolis-Hastings à marche aléatoire à accroissements symétriques SIMH, où l'on se sert de la variable précédemment simulée pour tirer la valeur suivante selon une loi symétrique. La probabilité de transition dépend de la distance entre les états, $q(y|m^{(n)}) = q(|y - m^{(n)}|)$.

Il s'agit ici d'un échantillonneur local, qui permet une exploration plutôt fine et précise de l'espace des paramètres. Cependant, il ne permet pas d'avoir un aperçu global de la loi a posteriori quand la dimension de celle-ci est grande, comme c'est le cas en tomographie sismique.

Un indicateur de la qualité de l'exploration de l'espace d'état par un échantillonneur SIMH est le taux d'acceptation moyen de la chaîne. Ce dernier ne doit être ni trop grand, ni trop petit. En effet, un taux d'acceptation trop bas signifie que les candidats sont souvent rejetés et par conséquent que les échantillons successifs sont fortement corrélés. Il en résulte une mauvaise exploration : c'est le cas quand le pas de la marche aléatoire est trop grand. Par ailleurs, un taux d'acceptation trop élevé peut signifier que les sauts sont souvent petits : c'est le cas rencontré quand le pas de la marche est trop petit. La marche aléatoire dans ce cas, évoluera trop lentement et les échantillons résultants risquent d'être fortement corrélés. Une valeur optimale pour le taux d'acceptation est de 0.234 dans le cas de l'algorithme de Metropolis-Hastings à marche aléatoire gaussienne (Voir (Roberts et al., 1997)). Le choix des pas de la marche aléatoire a une influence directe sur les performances de l'échantillonneur. Plusieurs algorithmes ont été développés pour pallier à cette problématique.

2.2 Delayed Rejection :

(Green and Mira, 2001) proposent de retarder le rejet du candidat proposé (Delayed Rejection). L'idée principale est la suivante : lorsqu'un premier candidat est rejeté, cela peut signifier que ce dernier se trouve dans une région où la loi a posteriori (25) est faible ; un second candidat est alors proposé généralement avec une loi instrumentale moins étendue. Les candidats sont proposés consécutivement à l'intérieur d'une même itération, jusqu'à l'acceptation d'un candidat ou jusqu'au nombre d'essais permis fixé par l'utilisateur.

Algorithme Delayed Rejection

Étant donnée une initialisation m^0 de la chaîne.

À l'itération $(n + 1)$,

1. tirer $m_1^* \sim q_1(m_1^*|m^n)$ et prendre $m^{n+1} = m_1^*$ avec une probabilité $\alpha_1(m^n, m_1^*)$ définie par :

$$\alpha_1(m^n, m_1^*) = \min \left\{ \frac{P(m_1^*|d^{obs}) \cdot q_1(m^n|m_1^*)}{P(m^n|d^{obs}) \cdot q_1(m_1^*|m^n)}, 1 \right\} \quad (48)$$

2. Si m_1^* est rejeté, tirer $m_2^* \sim q_2(m_2^*|m^n, m_1^*)$ et prendre $m^{n+1} = m_2^*$ avec une probabilité $\alpha_2(m^n, m_2^*)$ définie par :

$$\alpha_2(m^n, m_2^*, m_1^*) = \min \left\{ \frac{P(m_2^*|d^{obs}) \cdot q_1(m_1^*|m_2^*) \cdot [1 - \alpha_1(m_2^*, m_1^*)] \cdot q_2(m^n|m_2^*, m_1^*)}{P(m^{(n)}|d^{obs}) \cdot q_1(m_1^*|m^{(n)}) \cdot [1 - \alpha_1(m^{(n)}, m_1^*)] \cdot q_2(m_2^*|m^n, m_1^*)}, 1 \right\} \quad (49)$$

...

p. Si m_{p-1}^* est rejeté, tirer $m_p^* \sim q_p(m_p^*|m^n, m_{p-1}^*, \dots, m_1^*)$ et prendre $m^{n+1} = m_p^*$ avec une probabilité $\alpha_p(m^n, m_1^*, \dots, m_p^*)$ définie par :

$$\alpha_p(m^n, m_1^*, \dots, m_p^*) = \min \left\{ \frac{P(m_p^*|d^{obs}) \cdot q_1(m_{p-1}^*|m_p^*) \cdot q_2(m_{p-2}|m_p^*, m_{p-1}^*) \dots q_p(m^n|m_p^*, \dots, m_1^*)}{P(m^{(n)}|d^{obs}) \cdot q_1(m_1^*|m^{(n)}) \cdot q_2(m_2^*|m^n, m_1^*) \dots q_p(m_p^*|m^n, \dots, m_1^*)}, 1 \right\} \quad (50)$$

Notons que la probabilité d'acceptation à chaque étape est calculée de manière à préserver la réversibilité de la chaîne de Markov par rapport à la loi a posteriori $P(m|d^{obs})$ (25). L'ergodicité de la chaîne est ainsi assurée.

Si on considère le cas où les lois instrumentales sont gaussiennes, un moyen pratique est alors de proposer des candidats indépendants à chaque étape i tel que : $q_i(m_i^*|m^n, \dots, m_1^*) = q_i(m_i^*|m^n) = \mathcal{N}(m^n, \sigma_i)$ en prenant $\sigma_1 > \sigma_2 > \sigma_3 > \dots > \sigma_p$, p étant le nombre d'essais permis.

2.3 Metropolis Hastings adaptatif :

Dans la littérature, on trouve également des algorithmes MCMC dits adaptatifs. L'idée principale est d'exploiter le passé de la chaîne pour construire une loi de proposition efficace. On cite par exemple l'algorithme de Metropolis Hastings adaptatif (AM) de (Haario et al., 1999, 2001): le pas de la marche aléatoire est modifié (adapté) au cours des itérations en fonction du passé. Ces méthodes reposent sur une famille paramétrée $\{K_\theta, \theta \in \Theta\}$ de noyaux de transition, (Θ, Γ) étant un espace mesurable de paramètres. Le paramètre θ est choisi dans le but de rendre l'algorithme plus efficace. Dans le cas de l'algorithme de Metropolis Hastings adaptatif (AM), le paramètre θ_n est le suivant $\theta_n = (\mu_n, \Gamma_n)$. La mise à jour de ce paramètre s'effectue selon une procédure d'approximation donnée par

$$\begin{aligned}\mu_{n+1} &= \frac{1}{n+1} \sum_{i=1}^{n+1} m_i \\ \Gamma_{n+1} &= \frac{1}{n+1} \sum_{i=1}^{n+1} (m_i - \mu_i)(m_i - \mu_i)^T\end{aligned}\tag{51}$$

Algorithme Metropolis-Hasting Adaptatif

Étant donnée une initialisation \mathbf{m}^0 de la chaîne.

À l'itération $(n+1)$,

1. tirer $\mathbf{m}^* \sim \mathcal{N}(\mathbf{m}^n, \Gamma_n)$

2. $\left\{ \begin{array}{l} \text{prendre } \mathbf{m}^{n+1} = \mathbf{m}^* \text{ avec une probabilité } \alpha(\mathbf{m}^n, \mathbf{m}^*) \text{ définie par :} \\ \alpha(\mathbf{m}^n, \mathbf{m}^*) = \min \left\{ \frac{P(\mathbf{m}^* | d^{obs})}{P(\mathbf{m}^n | d^{obs})}, 1 \right\} \\ \text{prendre } \mathbf{m}^{n+1} = \mathbf{m}^n \text{ avec une probabilité } 1 - \alpha(\mathbf{m}^n, \mathbf{m}^*) \end{array} \right.$ (52)

3. $\mu_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} m_i$

4. $\Gamma_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} (m_i - \mu_i)(m_i - \mu_i)^T$

(Andrieu and Thoms, 2008) ont posé le cadre général de l'algorithme en se basant sur la procédure d'approximation stochastique de (Robbins and Monro, 1951):

A chaque itération $n+1$, un nouveau candidat est proposé selon la loi $\mathcal{N}(m_n, \lambda_n \Gamma_n)$, λ_n étant des constantes strictement positives, la covariance Γ_n est donnée par la formule suivante qui utilise une famille de pas $(\gamma_n)_{n \in \mathbb{N}}$ avec $\sum_{i=1}^{\infty} \gamma_n = \infty$ et $\sum_{i=1}^{\infty} \gamma_n^{\nu+1} < \infty$ pour un certain $\nu > 0$:

$$\begin{aligned}\mu_n &= \mu_{n-1} + \gamma_n (m_n - \mu_{n-1}) \\ \Gamma_n &= \Gamma_{n-1} + \gamma_n ((m_n - \mu_{n-1})(m_n - \mu_{n-1})^T - \Gamma_{n-1})\end{aligned}\tag{53}$$

Le paramètre λ_n est à son tour mis à jour à chaque itération d'une manière à ce que le taux d'acceptation s'approche au mieux de sa valeur optimale.

$$\log(\lambda_n) = \log(\lambda_{n-1}) + \frac{1}{\gamma_n} (\alpha(m^{n-1}, m_n^*) - \alpha^*)\tag{54}$$

m_n^* correspond au candidat proposé à l'itération n , $\alpha(m^{n-1}, m_n^*)$, la probabilité d'acceptation, enfin α^* est le taux d'acceptation optimal fixé par l'utilisateur, en l'occurrence 0.234.

D'autres variantes de l'algorithme AM peuvent être trouvées dans (Andrieu and Thoms, 2008).

2.4 Algorithme de Metropolis-Hastings à saut réversible :

La méthode de Monte Carlo par Chaînes de Markov à sauts réversibles (Green, 1995) (Reversible Jump Monte Carlo Markov Chain RjMCMC) appelée aussi Metropolis-Hastings-Green est une généralisation des méthodes MCMC classiques qui permet de simuler la dimension de l'espace des inconnues.

Le mécanisme de cet algorithme est similaire à celui de Metropolis-Hastings présenté précédemment.

Supposons que la chaîne se trouve dans l'état m de dimension d . Le passage de m à un nouvel état m^* se fait en choisissant tout d'abord un mouvement de type i avec une probabilité $j_i(m)$. On définit ensuite deux variables auxiliaires u et u^* de dimension r et r^* respectivement, de telle manière que la transformation $(m, u) \rightarrow (m^*, u^*)$ soit un difféomorphisme. Les deux variables sont tirées selon les lois de proposition q et q' et permettent de compenser les écarts de dimension entre m et m^* : $d + r = d^* + r^*$

Enfin, le nouvel état m^* est obtenu à partir d'une fonction déterministe inversible h

$$h(m, u) = m^* \quad (55)$$

Le rapport de Green est défini de manière à assurer la réversibilité de la chaîne :

$$r = \frac{P(m^* | d^{obs}) \cdot j_i(m^*) \cdot q'(u^*)}{P(m | d^{obs}) \cdot j_i(m) \cdot q(u)} \left| \frac{\partial(m^*, u^*)}{\partial(m, u)} \right| \quad (56)$$

Où : $\left| \frac{\partial(m^*, u^*)}{\partial(m, u)} \right|$ est le jacobien associé à la transformation.

La dynamique de l'algorithme est résumée ci-dessous.

Algorithme Metropolis-Hasting à saut reversible (MH)

Étant donnée une initialisation m^0 de la chaîne.

À l'itération $(n + 1)$,

1. tirer un mouvement $i \sim j_i(m^{(n)})$
2. simuler u selon $q(\cdot | m^n, i)$
3. Calculer m^* avec $h(m, u) = m^*$

4. $\left\{ \begin{array}{l} \text{prendre } m^{n+1} = m^* \text{ avec une probabilité } \alpha(m^n, m^*) \text{ définie par :} \\ \alpha(m^{(n)}, m^*) = \min\{1, r\} \\ \text{prendre } m^{n+1} = m^n \text{ avec une probabilité } 1 - \alpha(m^n, m^*) \end{array} \right. \quad (57)$

3 Distributions tempérées:

Les méthodes MCMC classiques, comme l'algorithme de Metropolis-Hastings présenté précédemment, génèrent une seule chaîne de Markov. Pour pouvoir simuler la loi a posteriori $P(m|d^{obs})$ (25) à partir de cette chaîne, elle doit avoir convergé et avoir visité l'ensemble de l'espace d'état. Ces méthodes classiques deviennent inefficaces dans le cas où l'on a affaire à des modèles possédant un grand nombre de paramètres et/ou à des distributions multimodales. En effet, la chaîne de Markov évoluera lentement et aura du mal à visiter tous les modes de la distribution cible. Plus les modes sont éloignés et séparés par de larges régions peu probables et plus la chaîne aura du mal à se déplacer. Elle peut par conséquent rester bloquée dans un mode local de l'espace d'état.

Pour remédier à ce problème, de nombreuses améliorations ont été apportées aux algorithmes classiques. Nous nous intéressons particulièrement aux chaînes en interaction : l'idée principale est de faire interagir l'algorithme avec d'autres algorithmes MCMC échantillonnant une version tempérée de la loi cible (Geyer, 1991). L'idée fondamentale de ces algorithmes est qu'un changement d'échelle (Figure 3), appelée température, permet de parcourir plus rapidement l'espace d'états dans le but d'empêcher la chaîne de rester bloquée dans un mode local. Avant de présenter les chaînes en interaction, nous passons tout d'abord en revue des méthodes utilisant des distributions tempérées.

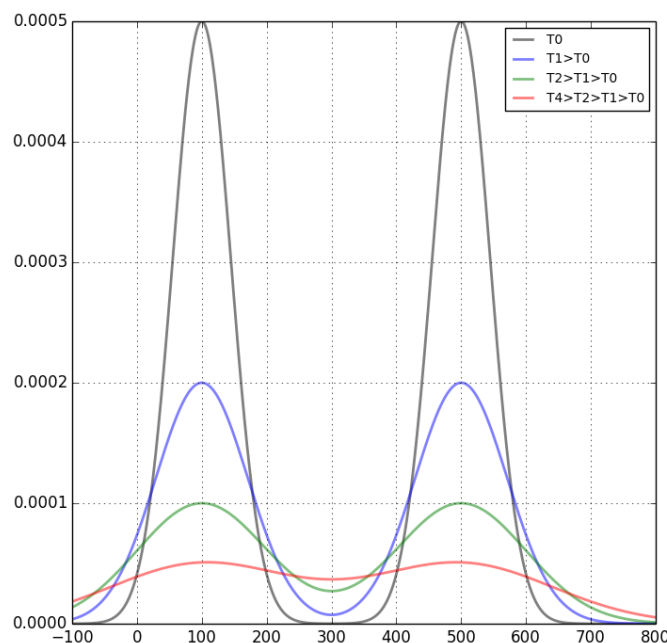


Figure 3: Distributions tempérées.

3.1 Une seule chaîne :

3.1.1 Recuit simulé: Simulated Annealing.

Cette méthode a été proposée par (Metropolis et al., 1953). Sa convergence a été démontrée dans (Kirkpatrick et al., 1983). Le processus du recuit en métallurgie consiste, dans un premier temps, à chauffer un solide à une température très élevée jusqu'à sa fusion -- les particules du système sont donc mélangées et peuvent se déplacer plus facilement-- puis, à diminuer graduellement la température jusqu'à refroidissement. Les particules vont ainsi s'arranger jusqu'à la cristallisation, on obtient alors un solide dans un état de faible énergie. Ce processus physique a été adapté à un algorithme de Metropolis-Hastings, dans le but de résoudre des problèmes d'optimisation.

Soit une distribution d'intérêt que nous cherchons à maximiser, en l'occurrence la loi a posteriori $P(m|d^{obs})$ (25).

On définit par analogie l'énergie du système comme suit :

$$E(m) = -\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2 + \log(P(m)) \quad (58)$$

L'évaluation de l'équation (25) est équivalente à l'évaluation de la quantité suivante :

$$P(m|d^{obs}) \propto e^{-E(m)} \quad (59)$$

Soit une suite de température décroissante $(T_n)_{n \in \mathbb{N}}$ avec $\lim_{n \rightarrow \infty} T_n = 0$ quand il s'agit des problèmes d'optimisation. Quand l'objectif est d'échantillonner suivant (25), la température minimale est fixée à 1, soit $T_N = 1$.

La méthode consiste à lancer successivement des échantillonneurs de distributions stationnaires $P_i(m|d^{obs}) = e^{-\frac{E(m)}{T_i}}$:

Il s'agit de lancer l'algorithme au départ avec des températures élevées, les chaînes auront des distributions plates faciles à échantillonner puis de diminuer graduellement la température pour se diriger vers des distributions "serrées" plus difficiles à échantillonner afin d'atteindre le maximum de la distribution cible $P(m|d^{obs})$ (25). Comme l'état initial de l'étape i est le dernier état de la chaîne à l'étape $(i-1)$, l'intérêt majeur de cette méthode est alors de pouvoir choisir des états initiaux pertinents à chaque étape de l'algorithme pour pouvoir accélérer la convergence de la chaîne.

En admettant qu'on utilise un échantillonneur symétrique, i.e $q(m^*|m^n) = q(m^n|m^*)$, la probabilité d'acceptation α s'écrit comme suit :

$$\alpha(m^*, m^n) = \min \left(1, e^{\frac{E(m^*) - E(m^n)}{T}} \right) \quad (60)$$

Il est important de noter qu'en théorie les températures doivent décroître à une vitesse logarithmique pour assurer la convergence :

$$T \propto \log\left(\frac{1}{n}\right)$$

Tandis qu'en pratique, une distribution géométrique est généralement adoptée.

Notons au passage que l'algorithme du recuit simulé est mieux (plus) adapté au problème d'optimisation. Cependant pour un problème d'échantillonnage, cette technique peut atteindre ses limites dès que la distribution d'intérêt est multimodale.

Algorithme Simulated Annealing

1. Déterminer la température $T_k = T_{\max}$.
2. Choisir un état initial m^k .
3. lancer un nombre d'itérations à partir de l'état m^k avec un échantillonneur MCMC ayant π_k comme distribution stationnaire.
4. Pour i allant de $K-1$ à 0
 Prendre comme état initial le dernier état de la chaîne voisine ayant π_{i+1} comme distribution stationnaire avec une probabilité α .
 lancer N_i itérations d'un échantillonneur MCMC ayant la loi π_i comme distribution stationnaire.

3.1.2 Trempe simulée : Simulated Tempering

Cet algorithme a été introduit d'une façon indépendante par (Marinari and Parisi, 1992) et (Geyer and Thompson, 1995) pour l'appliquer à des problèmes d'échantillonnage. Tout comme le Simulated Annealing (SA), le Simulated Tempering (ST) utilise des distributions tempérées. Cependant, contrairement au SA où la température décroît systématiquement, ici elle peut croître ou décroître permettant ainsi à la chaîne d'échapper à l'attraction d'un mode local. L'indice de la distribution tempérée est choisi de façon aléatoire. A chaque itération un indice j est proposé en utilisant une marche aléatoire avec une probabilité d'acceptation :

$$\alpha_{ij} = \min\left(1, \frac{c_j P_j(m|d^{obs}) q_{j,i}}{c_i P_i(m|d^{obs}) q_{i,j}}\right) \quad (61)$$

Les c_i sont des constantes qui doivent être ajustées par l'utilisateur. (Geyer, 1991) propose une procédure appelée « reverse logistic regression ».

$q_{i,j}$, $q_{j,i}$ sont respectivement la probabilité de proposer de transition de la température d'indice i vers celle d'indice j et de la température T_j vers T_i . Pour plus de détails voir [Geyer et Thompson, 1995].

Pour que l'algorithme ST soit efficace, Le nombre de températures et la répartition de celles-ci doivent être choisis avec discernement. En effet, le nombre de températures ne doit pas être très grand pour permettre à la chaîne de parcourir l'ensemble des distributions en un temps

raisonnable. En revanche, si le nombre de températures est trop faible, les transitions de températures risquent d'être souvent rejetées car les distributions de températures différentes, dans ce cas, sont trop éloignées. Un inconvénient de cet algorithme réside dans la nécessité de déterminer les constantes c_i à l'avance, pour permettre à la chaîne de visiter chacune des distributions P_i . En pratique la détermination de ces constantes peut s'avérer difficile.

Algorithme Simulated Tempering

Initialiser la chaîne $(\mathbf{m}_{init}, i_{init})$. Chaque itération actualise l'état (\mathbf{m}, i) de la chaîne

A l'itération N

1. Actualiser \mathbf{m} avec un échantillonneur MCMC ayant la loi $P_i(\cdot | \mathbf{d}^{obs})$ comme distribution stationnaire. T est fixée à l'indice i.
2. Actualiser l'indice de la température, m étant fixé :
Proposer un indice j de température suivant les probabilités de proposer de transition.
Accepter j avec la probabilité d'acceptation α_{ij}

3.2 Chaînes en interaction:

L'idée des chaînes en interaction a été introduite pour la première fois par (Geyer, 1991) sous le nom de Metropolis Coupled Markov Chain Monte Carlo (**MC**)³. On retrouve cet algorithme dans la littérature sous différentes appellations, telle que : « Exchange Monte Carlo » [Hukushima & Nemoto, 1996]. Pour une revue voir (Iba, 2001; Andrieu and Thoms, 2008).

Il s'agit de générer K+1 chaînes de Markov en parallèle, chacune à une température différente (énergie différente) au sens de la trempe simulée, et de leur permettre d'échanger leur configuration, c'est-à-dire leur état courant ou passé.

Le but est de permettre aux chaînes à basse température notamment la première chaîne ($T_0=1$) qui représente notre distribution d'intérêt, de ne pas rester bloquées dans des modes locaux en assurant le passage entre les différents modes. En effet, les chaînes à haute température sont capables de mieux explorer l'espace d'état. Les chaînes à basse température peuvent alors, moyennant les échanges d'états, accéder à des états qu'elles ne pourraient pas atteindre si ces échanges n'avaient pas lieu.

Depuis peu, la communauté scientifique s'est intéressée aux chaînes de Markov en interaction. Selon la façon d'échanger les états entre les chaînes, on distingue trois algorithmes; le Parallel Tempering, l'Equi-Energy Sampler, le Population Importance Resampling. (Romary, 2009) recommande d'utiliser ce dernier dans le cadre de l'inversion bayésienne. (Pour plus de détails concernant la comparaison faite entre les différents algorithmes, voir (Romary, 2009, 2010).

Nous rappelons que le but est de simuler la loi à posteriori (25).

Soit une séquence de température croissante $T_0 = 1 < T_1 < \dots < T_K = T_{max}$.

La distribution de chaque chaîne, comme pour le SA et la TS peut s'écrire comme étant une loi de Gibbs.

Nous avons donc pour chaque chaîne l de $[0, K]$:

$$\pi_l = e^{-\frac{E(m)}{T_l}} \quad (62)$$

Notons m l'état courant de la chaîne.

On définit la fonction d'importance (pour plus de détails sur la fonction d'importance voir [Robert & Casella, 2004]) donnée par:

$$r^{(l)}(m) = \frac{P_l(m|d^{obs})}{P_{l+1}(m|d^{obs})} = e^{-E(m)\left(\frac{1}{T_l} - \frac{1}{T_{l+1}}\right)} \quad (63)$$

A chaque itération de l'algorithme, un mouvement d'échange entre deux chaînes adjacentes l et $l + 1$ avec $l \in [0, K - 1]$ est proposé suivant une certaine probabilité θ :

- Si la proposition d'échange est refusée, une transition MH classique est réalisée.
- Sinon, un état $m_j^{(l+1)}$ avec $j < n$ est proposé en ré-échantillonnant à partir de tous les états générés par la chaîne $l + 1$ dans le passé $\{m_1^{(l+1)}, m_2^{(l+1)}, \dots, m_{n-1}^{(l+1)}\}$ pondérés par la valeur de la fonction d'importance : $\{r(m_1^{(l+1)}), r(m_2^{(l+1)}), \dots, r(m_{n-1}^{(l+1)})\}$.

Dans (Atchadé and Perron, 2007) l'état $m_j^{(l+1)}$ est ensuite utilisé pour proposer un nouvel état suivant une nouvelle transition. Étant donné que cette nouvelle transition risque d'être rejetée, (Romary, 2009) suggère d'annuler cette étape pour augmenter la chance d'accepter l'échange. Cet échange est accepté avec une probabilité β définie comme suit :

$$\beta(m^{(i,n)}, m^{(i+1,j)}) = \min\left(1, \frac{r^{(l)}(m^{(i+1,j)})}{r^{(l)}(m^{(i,n)})}\right) \quad (64)$$

Le candidat $m_j^{(l+1)}$ est choisi de telle manière à favoriser le candidat avec l'énergie la plus basse. Pour plus de détails, voir (Bottero et al., 2016).

Finalement, dans (Kou et al., 2006) il est fortement conseillé d'utiliser l'ensemble des candidats générés par toutes les chaînes pour mieux estimer l'intégrale (39)

$$E[h(m)] \approx \sum_{l=0}^K \frac{\sum_{i=1}^n \eta^{(l)}(m_i^{(l)}) h(m_i^{(l)})}{\sum_{i=1}^n \eta^{(l)}(m_i^{(l)})} \quad (65)$$

avec

$$\eta^{(l)}(m_i^{(l)}) = e^{-E(m)(1-\frac{1}{T_l})} \quad (66)$$

L'efficacité de l'algorithme dépend du type d'échantillonneur utilisé pour chaque chaîne lors des transitions MH classiques. En effet, il est évident que plus la température est élevée et plus l'échantillonneur associé peut se déplacer facilement, car la distribution cible devient aplatie sous l'effet de la température et par conséquent facile à parcourir. Un échantillonneur global (IMH par exemple) est de ce fait, le mieux adapté car il permet d'explorer facilement l'espace d'état. Tandis que pour les chaînes à basse température, il est intéressant d'utiliser un échantillonneur local (SIMH par exemple). En effet, ce comportement local permet d'échantillonner plus précisément les régions pertinentes qui ont été découvertes par les chaînes à haute température lors des itérations précédentes. Nous verrons dans le chapitre 6, une application de cet algorithme dans le cadre d'une inversion à couches.

4 Conclusion :

Ce chapitre a été consacré aux méthodes de Monte Carlo par Chaînes de Markov. Ces méthodes sont généralement utilisées lorsque la loi a posteriori (25) n'est pas simulable directement. Nous avons présenté plusieurs algorithmes MCMC qui sont des variantes de l'algorithme Metropolis Hastings classique. Ces algorithmes visent à traiter de manière plus performante certains cas difficiles tels que l'échantillonnage de lois cibles multimodales avec des modes distants. Nous nous intéressons en particulier aux chaînes en interaction que nous utiliserons dans le chapitre 6. Nous avons présenté également l'algorithme RJ-MCMC qui permet de simuler la dimension de l'espace des inconnues. Nous utiliserons cet algorithme pour simuler les processus ponctuels (chapitre 4 et 5) et pour inférer la loi a posteriori pour les modèles à couches trans-dimensionnels (chapitre 6).

Chapitre 3.

La paramétrisation

La paramétrisation du modèle de vitesse est une étape clé en tomographie sismique en particulier dans un cadre bayésien. Il est nécessaire d'utiliser des paramétrisations permettant de réduire la dimension du problème tout en restant suffisamment riches pour permettre une bonne qualité de reconstruction du champ des vitesses des ondes sismiques. La principale difficulté est alors de trouver un compromis entre des modèles à la fois parcimonieux et physiquement pertinents pour les vitesses des ondes sismiques. Dans un premier temps, nous dressons un bref état de l'art de la littérature relative aux paramétrisations existantes. Ensuite, nous introduisons de nouvelles paramétrisations qui sont à même de reconstruire le modèle de vitesse des ondes sismiques. Après une présentation des modèles dans leur forme la plus générale, une attention particulière est portée aux lois a priori des paramètres ainsi qu'à l'algorithme adopté pour inférer la distribution a posteriori pour chaque modèle.

1 État de l'art

1.1 Modèle à couches

L'hypothèse d'un modèle à couches est bien adaptée pour décrire les milieux dans lesquels les hétérogénéités sont gouvernées par la stratigraphie (autrement dit, les milieux sédimentaires), tel que rencontrés dans les réservoirs pétroliers ou de gaz de schiste. Cette hypothèse n'est pas forcément adaptée pour les milieux fracturés et faillés comme les réservoirs géothermiques.

Il s'agit de modéliser le sous-sol par un nombre fini $N \in \mathbb{N}$ de couches superposées (Figure 4) : Chaque couche i est caractérisée par sa profondeur z_i , son module de vitesse v_i , et en présence de pendage, par les angles (α_i, β_i) .

La paramétrisation à couches classique usuellement utilisée considère le nombre de couches comme étant connu (Gesret et al., 2015); il est choisi et fixé à l'avance. Le vecteur de paramètres \mathbf{m} est dans ce cas :

$$\mathbf{m} = (Z, V, A, B) \tag{67}$$

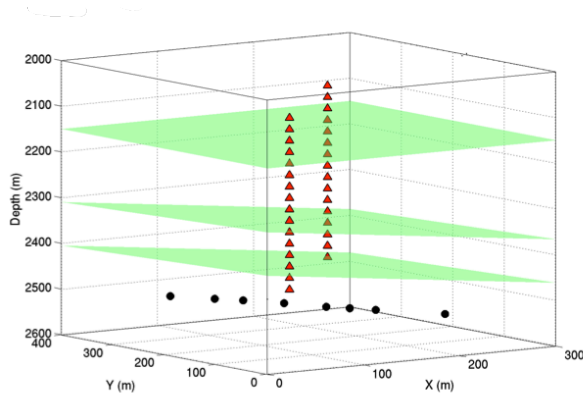


Figure 4: Exemple papier (Gesret et al, 2015).

Z : le vecteur des z_i .

V : le vecteur des vitesses des ondes sismiques.

A : le vecteur des angles α_i et B : le vecteur des angles β_i . Le choix des valeurs des angles doit assurer le non recoupement des couches.

Depuis peu, une paramétrisation « dynamique » a été proposée, où le nombre de couches fait partie des inconnues à estimer. Elle a été proposée pour la première fois par [Malinverno, 2002], dans le cadre d'une inversion électrique, elle a été depuis appliquée dans plusieurs contextes nous citons le cas de l'inversion geoacoustique avec (Dettmer et al., 2010, 2012) et celui des fonctions récepteur (Receiver functions) qui ont fait l'objet des travaux de (Agostinetti and Malinverno, 2010). Le nombre de couches est déterminé par les données observées limitant ainsi l'intervention de l'utilisateur. A notre connaissance, ce type de paramétrisation n'a été appliqué qu'au cas où le sous-sol est modélisé par un empilement de couches homogènes séparées par des interfaces horizontales.

Le vecteur m s'écrit :

$$m = (k, Z, V) \quad (68)$$

Avec k le nombre de couches.

Concernant la loi a priori sur les paramètres, deux lois sont souvent utilisées dans la littérature, la loi uniforme pour le nombre de couches ainsi que pour la profondeur des interfaces et la loi normale pour les vitesses des ondes sismiques. L'inférence de la loi a posteriori quant à elle se fait en utilisant l'algorithme RJ-MCMC. Pour plus de détails concernant l'algorithme utilisé voir (Malinverno, 2002; Agostinetti and Malinverno, 2010). Nous verrons dans le chapitre 5, une application de cet algorithme dans le cadre de la micro-sismicité induite.

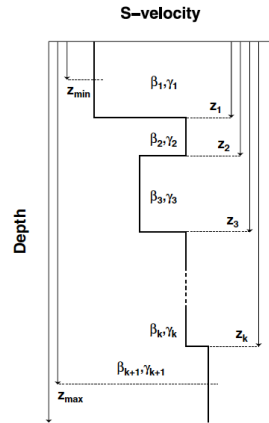


Figure 5: Exemple papier (Agostinetti and Malinverno, 2010)

1.2 Voronoï :

(Sambridge et al., 1995) ont été les premiers à introduire les cellules de Voronoï pour la résolution des problèmes en tomographie utilisant l'algorithme de voisinage (Neighbourhood Algorithm NA). Plus récemment, Bodin (Bodin and Sambridge, 2009) les a utilisées, dans le cadre de l'inversion sismique utilisant l'algorithme MCMC avec un échantillonneur à sauts réversibles (RJMCMC).

Définition :

Etant donné un ensemble C de n points $C = \{c_1, \dots, c_n\}$ avec c_i appartient à \mathbb{R}^d . Le diagramme de Voronoï de l'ensemble C de n points de \mathbb{R}^d est une partition de l'espace en n cellules représentant les zones d'influence des points de C : la cellule de Voronoï d'un point c_i de C est constituée de l'ensemble des points plus proches de c_i que de tout autre point de C . les c_i sont appelés noyaux de Voronoï.

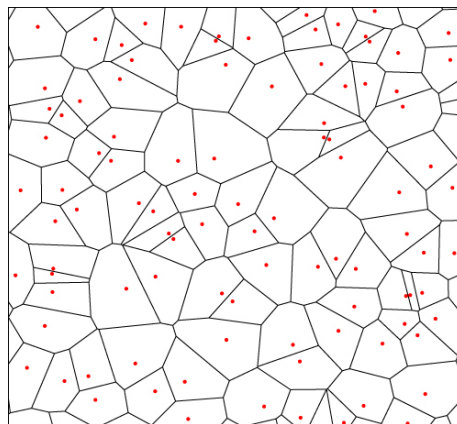


Figure 6 Cellules de Voronoï

Paramétrisation :

Le champ de vitesse est discrétisé en un ensemble de cellules de *Voronoi*. Chaque cellule est par la suite caractérisée par les coordonnées de son noyau c_i et le module de vitesse des ondes sismiques dans la cellule.

Notons C le vecteur contenant la position des noyaux de Voronoï et V celui contenant les modules de vitesse attribués à chaque partition.

On a deux possibilités pour traiter le problème :

- Soit seules la position des noyaux et la valeur de vitesse dans chaque cellule sont variable, le nombre de noyaux quant à lui est maintenu fixe. Ce cas est traité dans l'article (Bodin et al., 2009)

Le vecteur des paramètres m à estimer, s'exprime comme suit :

$$m = (C, V), \quad (69)$$

- Soit le nombre des noyaux est inconnu et variable (Bodin and Sambridge, 2009). Le vecteur m , devient :

$$m = (n, C, V), \quad (70)$$

Avec n le nombre des noyaux. Dans ses articles (Bodin and Sambridge, 2009; Bodin et al., 2012a), Bodin utilise une loi a priori uniforme sur ce paramètre. La loi uniforme est aussi utilisée pour les autres paramètres.

Lien avec le modèle à couches.

(Bodin et al., 2012b) propose dans le cadre de l'inversion des fonctions récepteur, une alternative plus souple au modèle à couches classiques, en utilisant les cellules de Voronoï en une seule dimension. En effet, l'épaisseur de la couche dans ce cas correspond à la taille de la cellule de Voronoï.

1.3 Les ondelettes

1.3.1 Généralités sur les ondelettes:

Depuis 1980, l'analyse par ondelettes est devenue un outil important dans de nombreux domaines scientifiques comme la sismologie, le traitement d'image etc. (Pour plus de détails sur l'origine et l'évolution des ondelettes voir (Graps, 1995)). La transformée en ondelettes discrète consiste à caractériser une fonction de carré intégrable par une famille dénombrable de fonctions de carré intégrable appelées *ondelettes*. De plus, les ondelettes sont d'intégrale nulle et se déduisent toutes par dilatation et translation d'une fonction appelée *ondelette-mère*.

(Pour plus de détails voir par exemple (Mallat, 1999)). On considère la famille : $\{\psi^{m,n} : x$

$$\mapsto \frac{1}{\sqrt{2^m}} \psi\left(\frac{x-2^m n}{2^m}\right)\}_{(m,n) \in \mathbb{Z}^2}$$
 qui forme une base hilbertienne de $L^2(\mathbb{R})$.

Toute fonction $h \in L^2(\mathbb{R})$ s'exprime de façon unique dans cette base :

$$h = \sum h_{m,n} \psi^{m,n} \quad (71)$$

avec $h_{mn} = \langle h | \psi^{m,n} \rangle$. h_{mn} sont appelés les coefficients d'ondelettes associés à h .

Approximation non-linéaire de h :

L'approximation non-linéaire est utilisée pour tronquer des décompositions de signaux sur des bases d'ondelettes. Pour plus de détails voir (Mallat, 1999) p, 389-394.

Généralement, l'approximation non-linéaire d'une fonction h sur M termes dans une base consiste à tronquer le développement de h dans cette base à M termes choisis de telle façon que l'erreur ε_M (73) soit la plus petite possible.

Posons h_M la projection de h sur M vecteurs.

$$h_M = \sum_{(m,n) \in I_M} h_{m,n} \psi^{m,n} \quad (72)$$

Où I_M : l'ensemble des indices des M vecteurs choisis.

L'erreur ε_M s'écrit sous la forme suivante :

$$\begin{aligned} \varepsilon_M = \|h - h_M\| &= \left\| h - \sum_{(m,n) \in I_M} h_{m,n} \psi^{m,n} \right\| \\ &= \left\| \sum_{(m,n) \notin I_M} h_{m,n} \psi^{m,n} \right\| \end{aligned} \quad (73)$$

Etant donné que la base d'ondelettes est une base hilbertienne. Les M termes à sélectionner qui minimisent ε_M , correspondent aux **M plus grands coefficients** de la décomposition de h sur la base c'est-à-dire les plus grands produits scalaires $h_{mn} = \langle h | \psi^{m,n} \rangle$. Les termes de h_M peuvent s'interpréter comme étant les structures principales de h .

La fonction h est alors approximée par h_M

$$h \approx h_M = \sum_{(m,n) \in I_M} h_{m,n} \psi^{m,n} \quad (74)$$

Exemples d'ondelettes :

Il existe une infinité de fonctions d'ondelettes, pour chaque type d'ondelette. Nous présenterons ici les ondelettes dont l'usage est le plus répandu. Il s'agit des ondelettes de Haar et de Daubechies. D'autres exemples d'ondelettes peuvent être trouvés par exemple dans (Mallat, 1999).

- Les ondelettes de Haar :

Elles sont les plus simples et premières à être utilisées dans l'analyse du signal. L'ondelette-mère de cette famille est la fonction définie comme suit :

$$\varphi(x) = \begin{cases} 1 & \text{si } x \in \left[0, \frac{1}{2}\right[\\ -1 & \text{si } x \in \left[-\frac{1}{2}, 1\right[\\ 0 & \text{sinon} \end{cases} \quad (75)$$

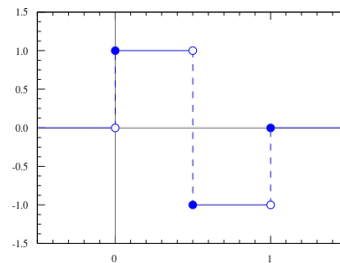


Figure 7: Exemple ondelette de Haar

- Les ondelettes de Daubechies :

Les ondelettes utilisées usuellement dans le cadre du traitement du signal discret sont les ondelettes de Daubechies.

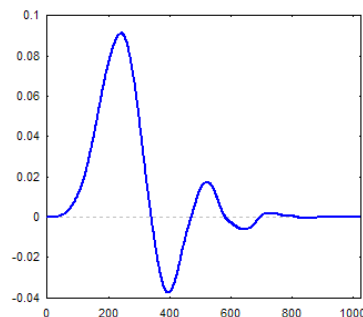


Figure 8: Exemple ondelette de Daubechies

Les ondelettes de Daubechies, développées par Ingrid Daubechies (voir (Daubechies and others, 1992)), concilient trois contraintes à savoir :

- L'orthogonalité de la base d'ondelettes

- La compacité du support de l'ondelette mère (ce qui implique que toute ondelette de la base est à support compact).
- le nombre fixe de moments nuls.

Rappel :

Une fonction est dite à support compact si elle s'annule en-dehors d'un compact de \mathbb{R} .

Une fonction f a n moments nuls si elle est orthogonale à l'espace $\mathbb{R}^n[X]$ des polynômes de degré n ou moins ie si :

$$\forall k \in \{0, \dots, n\} \langle f | t^k \rangle = \int_{-\infty}^{+\infty} t^k f(t) dt = 0 \quad (76)$$

1.3.2 Logs sismiques et décomposition en ondelettes:

Cette paramétrisation peut être utile dans le contexte de la fracturation hydraulique où des logs de vitesse sont généralement mesurés dans des puits. Elle a été suggérée par (Bottero et al., 2016) comme une alternative aux modèles à couches pour les milieux sédimentaires afin de pouvoir exploiter l'information a priori fournie par ces logs. Le but de l'utilisation de la décomposition en ondelette discrète des logs soniques est de pouvoir utiliser les coefficients de la décomposition (74) comme paramètres de l'inversion. À cette fin, il faut déterminer l'ondelette-mère ainsi que les coefficients à conserver. L'approche proposée par (Bottero et al., 2016) consiste dans un premier temps, à étendre horizontalement le log de vitesse, ce qui permet d'obtenir un modèle de vitesse 3D. Ce modèle 3D est considéré par la suite comme étant le modèle réel de vitesse. Il est utilisé pour générer les premiers temps d'arrivée synthétiques d_0 . Ensuite pour différentes décompositions en ondelettes, des temps d'arrivées d_M sont générés et une fonction de coût $\Delta_M = \left\| d_0 - d_M \right\|_2$ est calculée. Cette fonction est ensuite représentée pour chaque type d'ondelettes en fonction de nombre de coefficients gardés dans chaque filtrage, ce qui permet à la fin de choisir le couple, ondelette mère et nombre de coefficients qui minimise la fonction Δ_M .

1.3.3 La structure d'arbre et les ondelettes:

(Hawkins and Sambridge, 2015) présentent une approche Bayésienne pour la tomographie inverse qui s'appuie sur les espaces multi-résolutions. Plutôt que de chercher à déterminer la valeur de la vitesse en chaque point d'une grille cartésienne, l'approche consiste à décomposer le champ de vitesse de plus en plus finement en partant d'une approximation grossière. Hawkins et Sambrige s'appuient pour ce faire sur une structure d'arbre, où chaque noeud correspond à une approximation du champ de vitesse. En descendant dans l'arbre, on obtient des approximations de plus en plus fines. La racine de l'arbre correspond ainsi à une approximation, grossière, où le champ de vitesse est supposé homogène en valeur sur tout le

domaine d'intérêt. Les noeuds suivants, au nombre de quatre pour un arbre quaternaire, viennent diviser le domaine étudié en quatre régions. La vitesse est alors affinée localement dans chacune de ces quatre régions, cette fois-ci à une résolution plus fine correspondant au quart de l'image. La procédure est répétée itérativement, en décomposant le domaine en régions de plus en plus fines.

Afin de contrôler la décomposition, Hawkins et Sambrige proposent un algorithme MCMC qui, à chaque noeud de l'arbre, accepte ou rejette la décomposition à une échelle plus fine et fixe la valeur des coefficients associés aux fonctions de base qui permettent de représenter le signal, en général des ondelettes. Le nombre de noeuds de l'arbre est donné par une loi de probabilité a priori de Jeffrey (voir chapitre 1), ce qui permet de favoriser les décompositions sur un petit nombre de coefficients et de conserver une solution parcimonieuse. L'intérêt principal de la méthode est qu'elle permet d'obtenir une base de décomposition particulièrement bien adaptée au champ de vitesse.

2 Nouvelles paramétrisations

2.1 Mosaïques de Johnson Mehl:

Le premier modèle proposé dans ce manuscrit est une structure spatiale de type mosaïque. Les mosaïques aléatoires sont une décomposition de l'espace en polytopes ou encore cellules. Elles trouvent des applications dans de nombreux domaines, tels que la cristallographie, la métallurgie ou plus généralement la modélisations de géométries de microstructures aléatoires. La mosaïque de Voronoï est sans doute l'exemple le plus classique de mosaïque aléatoire. Pour plus de détails sur les mosaïques aléatoires en particulier et la géométrie stochastique en général, le lecteur peut se référer aux livres de (Matheron, 1975; Schneider and Weil, 2008; Chiu et al., 2013).

Dans le cadre de cette thèse, nous nous intéressons plus précisément aux cellules de Johnson-Mehl. Ce modèle a été introduit par Johnson et Mehl (Johnson and Mehl, 1939) dans le but de représenter le processus de cristallisation dans des microstructures métalliques générées par des germes en croissances. Les cellules de Johnson-Mehl peuvent être considérées comme une généralisation des cellules de Voronoï : Au cours du temps, des germes apparaissent en différents endroits et grossissent de manière isotrope à vitesse constante. Les nouveaux germes n'apparaissent que dans les endroits qui ne sont pas encore occupés par les cellules en cours de croissance. De ce fait, tout germe apparaissant dans une zone déjà occupée est automatiquement supprimé. La croissance des frontières des cellules stoppe dès que deux frontières se rencontrent. La mosaïque finale est obtenue lorsque tout l'espace est occupé par les germes. La figure ci-après (Figure 9) montre les différentes étapes de la construction de la mosaïque de Johnson-Mehl.

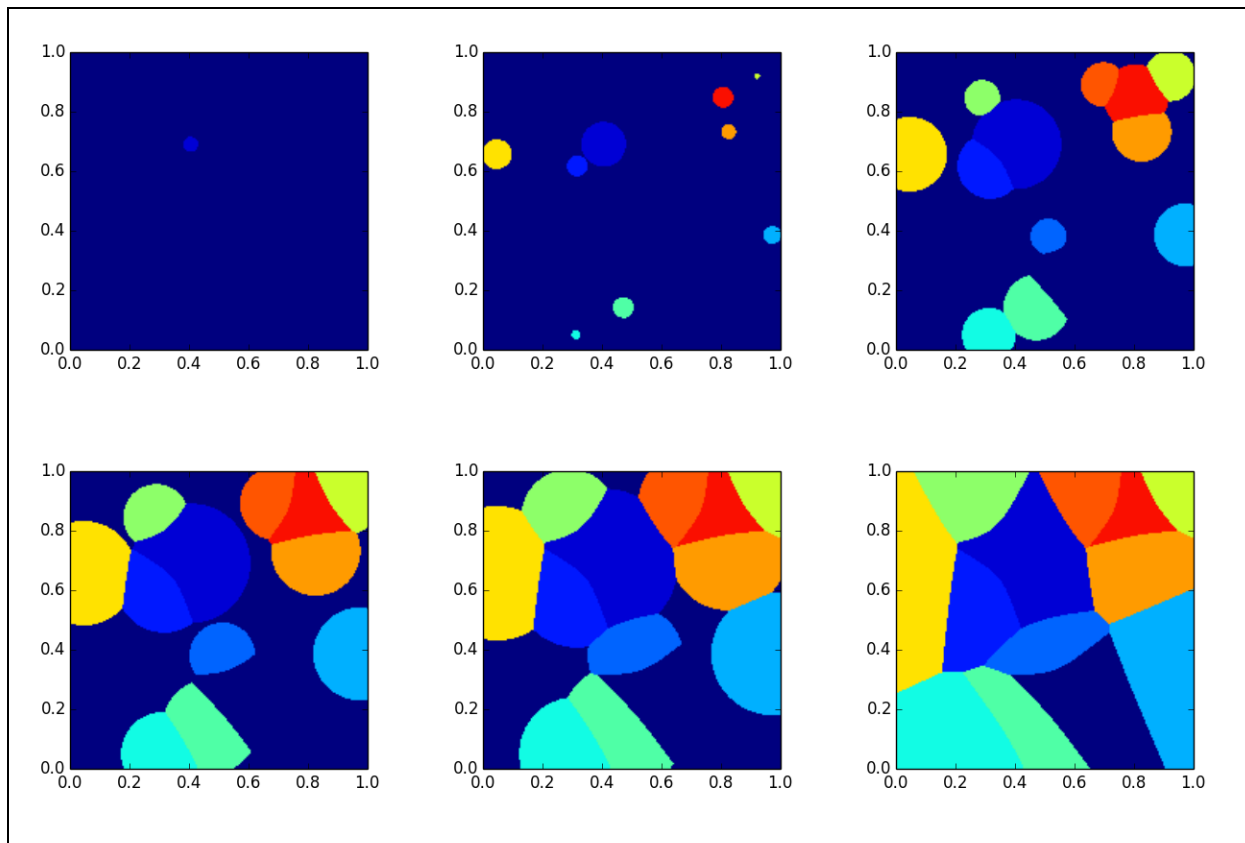


Figure 9: étapes de construction de la mosaïque de Johnson-Mehl

Contrairement à la mosaïque de Voronoï, les cellules de Johnson-Mehl ne sont pas forcément convexes et sont bornées par des portions d'hyperboloïdes, offrant ainsi des frontières lisses entre les cellules. L'intérêt essentiel du modèle de Johnson-Mehl, notamment en comparaison du modèle de Voronoï, est qu'il permet, d'une part l'obtention de formes de cellules plus générales, et d'autre part des frontières plus lisses entre cellules .

Les propriétés mathématiques de la mosaïque de Johnson-Mehl ont été étudiées par (Gilbert, 1962; Møller, 1992). Le lecteur pourra trouver dans ce dernier article un exposé unifié de ce modèle.

D'un point de vue mathématique, la mosaïque de Johnson-Mehl est construite à partir d'un processus ponctuel de Poisson. Dans notre cas, ce processus ponctuel est $\Phi = \{(u_1, t_1), (u_2, t_2), \dots\} \subset \mathbb{R}^d \times \mathbb{R}^+$ de mesure d'intensité $\nu(du)\Lambda(dt)$, où ν désigne la mesure de Lebesgue sur \mathbb{R}^d et Λ est une mesure localement finie sur \mathbb{R}^+ qui satisfait les conditions suivantes :

$$\Lambda([0, \infty)) > 0 \tag{77}$$

$$\lambda = \int p(t)\Lambda(dt) < +\infty \tag{78}$$

Dans cette relation, la quantité p est définie par

$$p(t) = \exp \left(-\omega_d \int_0^t (t-s)^d \Lambda(ds) \right)$$

et correspond à la probabilité qu'un point quelconque dans \mathbb{R}^d ne soit atteint par aucun germe à l'instant t . La quantité ω_d correspond au volume de la boule unité de dimension d :

$$\omega_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2} + 1\right)}$$

les grains u_i sont implantés additivement à l'instant t_i . Les cellules C_i associées aux points u_i sont définies par

$$C_i = \left\{ u \mid \forall j, j \neq i, t_i + \frac{\|u - u_i\|_2}{v_c} \leq t_j + \frac{\|u - u_j\|_2}{v_c} \right\} \quad (79)$$

$t_i + \frac{\|u - u_i\|_2}{v_c}$ correspond à l'instant auquel le germe atteint le point u .

Remarques:

- La première condition (77) est équivalente à $\Phi \neq \emptyset$.
- La deuxième condition (78) signifie que l'espérance du nombre des cellules non-vide est finie autrement dit tout borélien borné ne rencontre qu'un nombre fini de cellules.
- Lorsque tous les germes apparaissent en même temps (Λ est une mesure de Dirac), on retrouve la mosaïque de Voronoï.

Dans ce travail, nous utilisons les indicatrices des cellules Johnson-Mehl comme fonctions de base :

$$v(u) = \sum_{1 \leq i \leq n} v_i \phi_i(u) \begin{cases} \phi_i(u) = 1 & u \in C_i \\ \phi_i(u) = 0 & \text{sinon} \end{cases} \quad (80)$$

Le modèle de vitesse est complètement décrit par le nombre de cellules, la position des germes, les temps d'implantation et les modules de vitesse assignés à chaque cellule. Le vecteur des paramètres s'écrit sous la forme suivante :

$$m = (U, V, T, N)$$

où N est le nombre de cellules qui constitue la mosaïque.

$U = (u_i)_{1 \leq i \leq N}$ contient la position des germes associés à chaque cellule.

$T = (t_i)_{1 \leq i \leq N}$ est le vecteur des temps d'implantation des germes.

$V = (v_i)_{1 \leq i \leq N}$ regroupe les modules de vitesse des ondes sismiques.

2.1.1 Loi a priori :

Nous assumons a priori que les valeurs des paramètres et leur nombre sont indépendants.

$$P(m) = P(m|n)P(n) \quad (81)$$

où $P(n)$ est la loi a priori sur le nombre de cellules n . La mosaïque de Johnson-Mehl est construite à partir d'un processus ponctuel de Poisson. Par conséquent, n suit une loi de Poisson d'intensité θ :

$$P(n) = \frac{(\theta|D|)^n}{n!} \exp(-\theta|D|) \quad (82)$$

où $|D|$ est l'aire du domaine.

Le paramètre θ peut être interprété comme étant le nombre moyen de germes par unité d'aire (ou volume en 3D). Le choix de ce paramètre n'est pas évident à estimer, et découle d'un compromis entre le temps de calcul, la précision et la complexité du modèle. D'une part, il doit être suffisamment élevé pour permettre de reproduire finement les hétérogénéités du champ de vitesse dans les régions bien éclairées. D'autre part, il convient d'éviter de rajouter une paramétrisation superflue dans les zones peu ou pas éclairées. Nous examinons l'influence de ce paramètre sur le résultat final de l'inversion plus loin.

A chaque germe est associée une valeur réelle positive (marque) qui décrit le module de vitesse, ce qui forme un processus ponctuel marqué. Un processus ponctuel marqué est un processus pour lequel chaque point se voit attribuer en plus de sa position, des caractéristiques appelées marques. Pour plus d'information sur les processus ponctuel en général et les processus marqués voir (Van Lieshout, 2000; Chiu et al., 2013). La réalisation d'un processus ponctuel marqué est un ensemble dont les éléments possèdent deux composantes : une position et une marque.

Nous utilisons une loi uniforme pour la position des germes $(u_i)_{1 \leq i \leq n}$, le temps d'implantation $(t_i)_{1 \leq i \leq n}$ ainsi que pour le module de vitesses des ondes sismique $(v_i)_{1 \leq i \leq n}$.

$$\begin{aligned} P(u_i|n) &= \frac{1}{|D|} \\ P(t_i|n) &= \frac{1}{\Delta T} \mathbf{1}_{[t_{min}, t_{max}]}(t_i) \\ P(v_i|n) &= \frac{1}{\Delta V} \mathbf{1}_{[v_{min}, v_{max}]}(v_i) \end{aligned} \quad (83)$$

où: $\Delta T = t_{max} - t_{min}$; $t_{max}, t_{min} \in \mathbb{R}^+$ et $\Delta V = v_{max} - v_{min}$; $v_{max}, v_{min} \in \mathbb{R}^+$.

Finalement, en assumant que conditionnellement au nombre des cellules n , les paramètres sont a priori indépendants, $P(m|n)$ peut être exprimé comme le produit des lois a priori de chaque composante:

$$P(m|n) = P(U|n)P(T|n)P(V|n) = n! \prod_{i=1}^n P(u_i|n)P(t_i|n)P(v_i|n) \quad (84)$$

2.1.2 Inférence de la loi a posteriori :

Soit $m = (x, n)$ où $x = \{(u_1, v_1, t_1), \dots, (u_n, v_n, t_n)\}$ est l'état de la chaîne à une itération donnée. Dans notre travail, nous utilisons l'algorithme de Metropolis Hastings à saut réversible introduit dans le chapitre 2. Nous considérons ici trois type de perturbations proposées à chaque itération: l'ajout d'un germe, la suppression d'un germe, et le remplacement d'un germe par un autre. La naissance, la mort ou le remplacement d'un germe s'effectue selon les probabilités $j_b(m)$, $j_d(m)$ et $j_m(m)$ respectivement.

$$j_b(m) = \frac{\theta|D|}{\theta|D| + n + 1}$$

$$j_d(m) = \frac{n}{\theta|D| + n}$$

$$j_m(m) = 1 - j_b(m) - j_d(m)$$

Nous avons fait le choix d'utiliser des probabilités qui prennent en considération l'état actuel de la chaîne et qui permettent de mieux simuler le processus ponctuel.

Nous détaillons chaque mouvement :

1-La naissance :

Le nouveau germe $\xi^* = (u^*, v^*, t^*)$ est simulé indépendamment selon la loi $q(\xi^*) = q(u^*) \cdot q(v^*) \cdot q(t^*) = \frac{1}{|D|} \cdot \frac{1}{\Delta v} \cdot \frac{1}{\Delta t}$. Avec ce choix, le germe proposé a une position uniformément distribuée dans le domaine d'intérêt D , tandis que ses marques sont choisies indépendamment selon leur loi a priori.

La mort :

Le germe ξ^* à supprimer est sélectionné arbitrairement dans la configuration $q(x, \xi^*) = \frac{1}{n}$

La modification d'un germe:

Un germe à modifier η est choisi uniformément dans la configuration avec la probabilité $q(\eta) = \frac{1}{n}$, puis remplacé par un nouvel objet $\xi^* = (u^*, v^*, t^*)$ selon la densité

$q(\xi^*) = \frac{1}{|D|} \cdot \frac{1}{\Delta v} \cdot \frac{1}{\Delta t}$. Cette perturbation peut être vue comme une combinaison des deux précédentes à savoir la mort puis la naissance.

Remarque :

Pour effectuer toutes les perturbations, nous avons fait le choix d'utiliser les lois a priori. Ce choix n'étant évidemment pas le seul, il est néanmoins simple à implémenter et suffit pour assurer la convergence de l'algorithme.

La Figure 10 résume le déroulement d'une itération de l'algorithme.

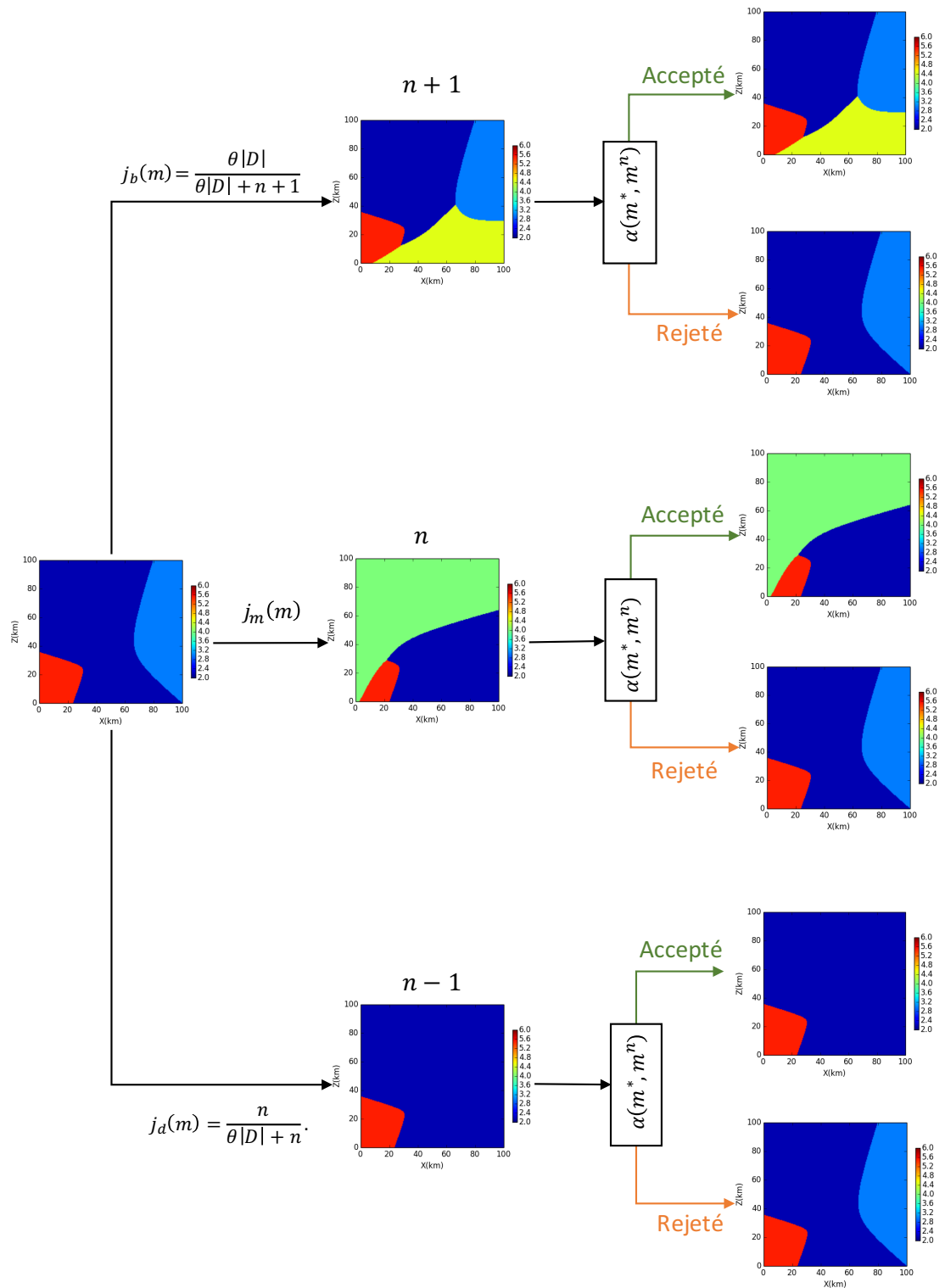
m^n m^* m^{n+1} 

Figure 10: Schéma résumant les principales étapes de l'algorithme, à gauche: l'état actuel de la chaîne, les figures au centre: modèle de vitesse construit à partir des paramètres proposés, à droite: modèle de vitesse retenu après acceptation/rejet

2.2 Les noyaux gaussiens

Le deuxième modèle paramétrique que nous proposons utilise des noyaux gaussiens comme fonctions de base. Il s'agit d'utiliser un modèle de vitesse constant et de considérer les variations de vitesse par rapport à cette valeur constante comme étant des « anomalies » modélisées par des noyaux gaussiens. Ce modèle de vitesse est donc adapté à la détection des anomalies de vitesse des ondes sismiques.

Nous considérons chaque anomalie comme une combinaison linéaire de ces fonctions.

$$v(u) = V_L + \sum_{1 \leq i \leq N} \alpha_i K \left(\left\| \frac{u - u_i}{\sigma_i} \right\|_2 \right) \quad (85)$$

Où V_L est la vitesse de fond, K est la fonction noyau, ici nous optons pour un noyau gaussien.

L'idée principale est de considérer l'anomalie (variations de vitesse par rapport à V_L) comme un ensemble de points (petites anomalies) $u_i, i = 1 \dots N$ qui vont diffuser une certaine amplitude $\alpha_i, i = 1, \dots, N$ (coefficient multiplicative), cette amplitude peut être négative ($\alpha_i < 0$) pour reproduire la chute de la vitesse dans le cas de l'anomalie négative (Figure 11a) ou positive ($\alpha_i > 0$) pour augmenter la vitesse dans la zone affectée par l'anomalie positive (Figure 11b). Cette diffusion prend place dans une zone définie par un noyau Gaussien et contrôlée par $\sigma_i, i = 1, \dots, N$.

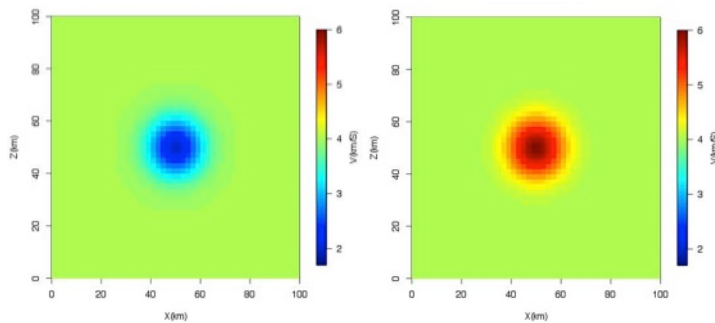


Figure 11: Exemple a) anomalie négative (lente) b) anomalie positive (rapide)

Le modèle de vitesse est complètement défini par le nombre des points (petites anomalies), leur position, la largeur et le coefficient de chaque fonction. Le vecteur des paramètres s'écrit sous la forme suivante :

$$m = (V_L, U, \alpha, \sigma, N) \quad (86)$$

où N est le nombre de fonctions de base, $U = (u_i)_{1 \leq i \leq N}$ le vecteur contenant la position des noyaux gaussiens, $\alpha = (\alpha_i)_{1 \leq i \leq N}$ est le vecteur des coefficients (amplitudes), finalement $\sigma = (\sigma_i)_{1 \leq i \leq N}$ comporte le paramètre d'échelle des fonctions de base.

2.2.1 La loi a priori:

Nous supposons tout d'abord que la vitesse de fond est indépendante du reste des paramètres.

$$P(m) = P(V_L)P(U, \alpha, \sigma, n) \quad (87)$$

La paramétrisation peut également être décrite comme un processus ponctuel marqué.

Dans notre cas, les noyaux Gaussiens sont implantés sur la réalisation du processus ponctuel de poisson et les marques associées correspondent à la variance et à l'amplitude de chaque gaussienne. La position de chaque noyau ainsi que ses marques constituent alors avec la vitesse de fond, les paramètres à déterminer conditionnellement aux données

Sous l'hypothèse de processus ponctuel marqué, les valeurs des paramètres (positions et marques) et le nombre des points sont indépendants:

$$P(m) = P(V_L)P(U, \alpha, \sigma|n)P(n) \quad (88)$$

où $P(n)$ est la loi a priori du nombre des noyaux gaussiens. Sous l'hypothèse de processus ponctuel, ce nombre dans un domaine D donné suit une loi de Poisson d'intensité θ

$$P(n) = \frac{(\theta|D|)^n}{n!} \exp(-\theta|D|) \quad (89)$$

où $|D|$ est l'aire du domaine.

Le paramètre θ correspond au nombre moyen de noyaux par unité de surface (ou de volume en 3D). Nous verrons plus tard l'influence de ce paramètre sur les résultats finaux.

Nous faisons l'hypothèse que conditionnellement au nombre des points n , les vecteurs U, σ, α sont indépendants:

$$P(m|n) = P(V_L)P(U|n)P(\alpha|n)P(\sigma|n) \quad (90)$$

$$n! \prod_{i=1}^n P(u_i|n)P(\alpha_i|n)P(\sigma_i|n)$$

Les points sont indépendamment distribués selon une loi uniforme. Les marques sont également distribuées de façon indépendante: les amplitudes suivent une loi uniforme et les écart-types suivent une loi exponentielle.

Pour chaque point $i, 1 \dots, N$

$$P(u_i|n) = \frac{1}{|D|} \quad (91)$$

$$P(\alpha_i|n) = \frac{1}{\Delta\alpha} 1_{[\alpha_{min}, \alpha_{max}]}(\alpha_i)$$

$$P(\sigma_i|n) = \frac{1}{\sigma_m} \exp\left(-\frac{\sigma_i}{\sigma_m}\right)$$

où: $\Delta\alpha = \alpha_{max} - \alpha_{min}$; $\alpha_{max}, \alpha_{min} \in R^+$ et $\sigma_m \in R^+$ est la moyenne a priori des écart-types.

Concernant la vitesse de fond, nous supposons qu'elle suit une loi a priori uniforme sur $[V_{min}, V_{max}]$ où $V_{min} > 0$ et $V_{max} > 0$. Ces bornes sont fixées en fonction des études sismiques et des connaissances géologiques préalables.

Remarque : Ce modèle paramétrique peut être généralisé en intégrant des effets d'anisotropie. Dans ce cas des paramètres supplémentaires sont à introduire pour contrôler l'anisotropie, à savoir une orientation ϕ_i et un second paramètre d'échelle σ_{z_i} .

$$v(u) = V_L + \sum_{1 \leq i \leq N} \alpha_i \exp\left(-\frac{1}{2} \left(\frac{(u_x - u_{x_i}) \cos(\phi_i) + (u_z - u_{z_i}) \sin(\phi_i)}{\sigma_{x_i}} \right)^2 + \left(\frac{-(u_x - u_{x_i}) \sin(\phi_i) + (u_z - u_{z_i}) \cos(\phi_i)}{\sigma_{z_i}} \right)^2 \right) \quad (92)$$

2.2.2 Inférence de la loi a posteriori :

Soit $m = (V_l, x, n)$ où $x = \{(u_1, \alpha_1, \sigma_1), \dots, (u_n, \alpha_n, \sigma_n)\}$, l'état de la chaîne à une itération donnée. Nous allons décrire, ici, les différents types de mouvements dans l'espace des paramètres permettant de proposer des candidats pour notre problème. En plus de la perturbation de la valeur de la vitesse de fond V_l , l'ajout, la suppression d'un noyau ou le remplacement de celui-ci par un autre sont également proposées.

- Avec une probabilité β , on perturbe la valeur de la vitesse de fond selon une marche aléatoire.
- Avec une probabilité $1 - \beta$, on ajoute, supprime ou remplace un point selon les probabilités $j_b(m)$, $j_d(m)$ et $j_m(m)$ respectivement. Nous avons fait le choix d'utiliser des probabilités qui prennent en considération l'état actuel de la chaîne et qui permettent de mieux simuler le processus ponctuel. $j_b(m) = \frac{\theta|D|}{\theta|D|+n+1}$, $j_d(m) = \frac{n}{\theta|D|+n}$, $j_m(m) = 1 - j_b(m) - j_d(m)$.

La naissance :

Un nouveau noyau gaussien $\xi^* = (u^*, \alpha^*, \sigma^*)$ est simulé indépendamment selon la loi $q(\xi^*) = q(u^*) \cdot q(\alpha^*) \cdot q(\sigma^*) = \frac{1}{|D|} \cdot \frac{1}{\Delta\alpha} \cdot \frac{1}{\sigma_m} e^{\frac{\sigma}{\sigma_m}}$. Par ce choix, le noyau proposé a une position uniformément distribuée dans le domaine d'intérêt D alors que ses marques sont choisies d'une façon indépendante selon leur loi a priori.

La mort :

Un noyau gaussien ξ^* à supprimer est sélectionné arbitrairement dans la configuration $q(x, \xi^*) = \frac{1}{n}$

La modification d'un germe:

Un noyau à modifier η est choisi uniformément dans la configuration avec la probabilité $q_1(\eta) = \frac{1}{n}$, puis remplacé par un nouveau noyau $\xi^* = (u^*, \alpha^*, \sigma^*)$ selon la densité $q_2(\xi^*) = \frac{1}{|D|} \cdot \frac{1}{\Delta\alpha} \cdot \frac{1}{\sigma_m} e^{\frac{\sigma}{\sigma_m}}$ de la même manière que pour la naissance. La loi instrumentale q est dans ce cas : $q(\eta, \xi) = \frac{1}{n} \cdot \frac{1}{|D|} \cdot \frac{1}{\Delta\alpha} \cdot \frac{1}{\sigma_m} e^{\frac{\sigma}{\sigma_m}}$.

Remarque :

Pour effectuer toutes les perturbations, nous avons fait le choix d'utiliser les lois a priori. Ces choix ne sont pas uniques, nous avons par conséquent privilégié des lois simples à implémenter.

La Figure 12 résume le déroulement d'une itération de l'algorithme.

3 Conclusion:

Dans ce chapitre, nous avons introduit deux nouvelles paramétrisations adaptées à l'approche bayésienne. Dans ce contexte, nous avons défini les lois a priori sur les différents paramètres et nous avons présenté la méthodologie adoptée pour inférer la distribution a posteriori pour chaque modèle paramétrique. Dans le chapitre suivant nous présentons des exemples d'applications qui permettent de valider notre méthodologie.

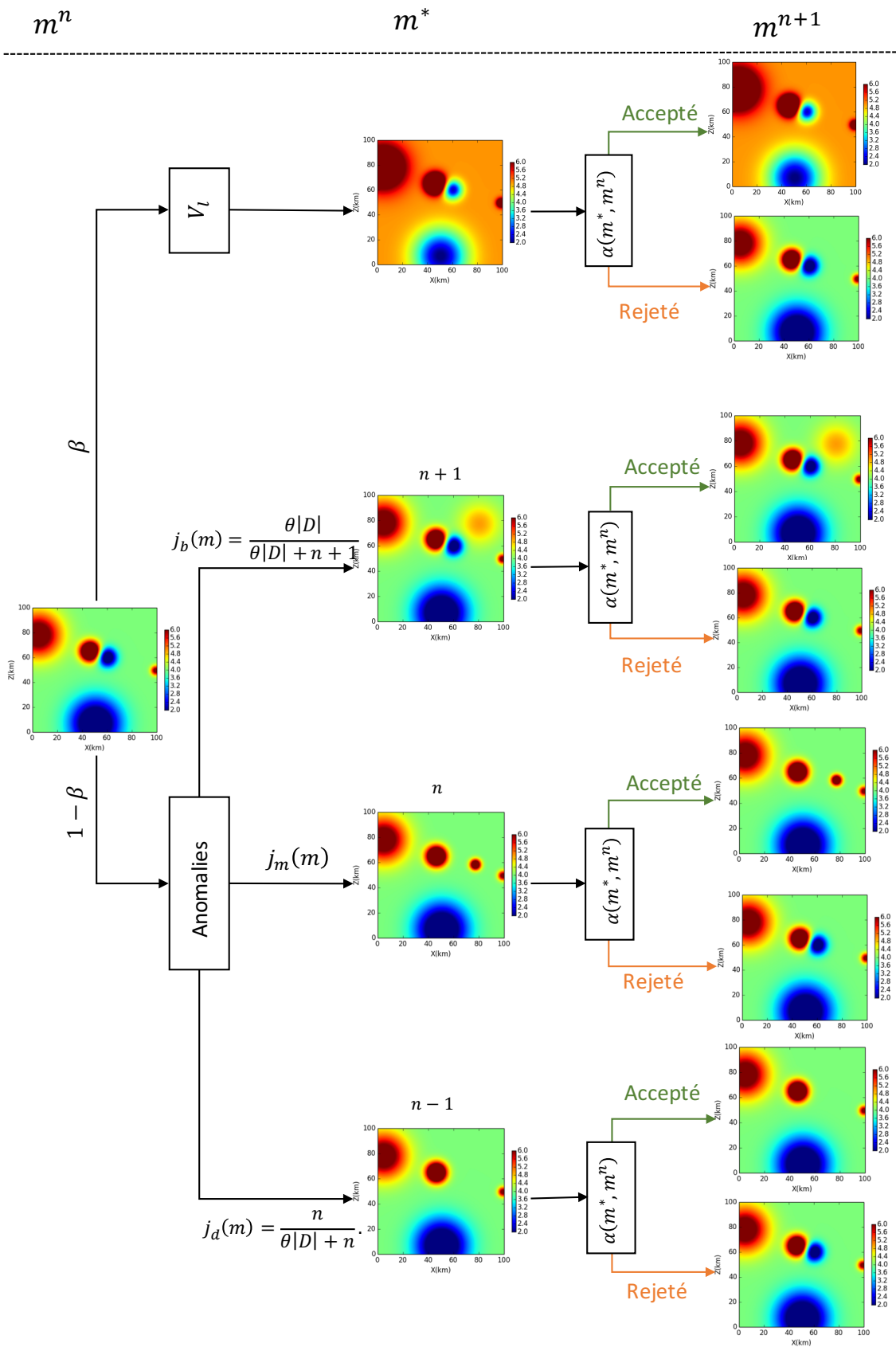


Figure 12: Schéma résumant les principales étapes de l'algorithme, à gauche: l'état actuel de la chaîne, les figures au centre: modèle de vitesse construit à partir des paramètres proposés, à droite: modèle de vitesse retenu après acceptation/rejet

Chapitre 4.

Applications

Afin d'évaluer la capacité des paramétrisations à reconstruire le modèle de vitesse, nous utilisons deux jeux de données simulés. Les travaux présentés dans ce chapitre ont fait l'objet d'un article à paraître prochainement dans la revue *Inverse Problems*. Une version préliminaire est reproduite en annexe.

Nous nous intéressons tout d'abord à une application simple qui nous permet d'évaluer l'influence de la loi a priori sur les résultats finaux. Nous illustrons les résultats de l'inversion avec les cartes de la moyenne a posteriori. Ce choix est notamment motivé par le fait qu'un tel estimateur prend en compte l'ensemble des réalisations et fournit une carte lisse qui atténue les discontinuités propres à la nature de la paramétrisation (discontinuités entre cellules par exemple), en ne gardant que les discontinuités propres au champ de vitesse. Ce choix d'estimateur peut s'avérer problématique si la distribution est multimodale. Nous évitons cependant cet écueil dans les applications que nous présentons. Chaque estimation sera accompagnée de la carte des écart-types qui permet d'évaluer visuellement le niveau d'incertitudes associé.

1 Analyse de sensibilité:

Nous considérons un modèle de vitesse synthétique contenant deux types d'anomalies : une anomalie lente en bleu et une anomalie rapide en rouge (Figure 13). Cet exemple permet, d'une part, de tester la capacité de la paramétrisation proposée à reconstruire le vrai modèle. D'autre part, il permet de voir l'impact que le nombre moyen de cellules a priori $\theta|D|$ a sur les performances de l'algorithme.

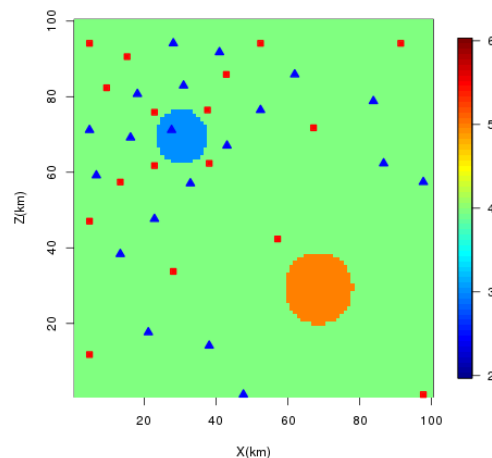


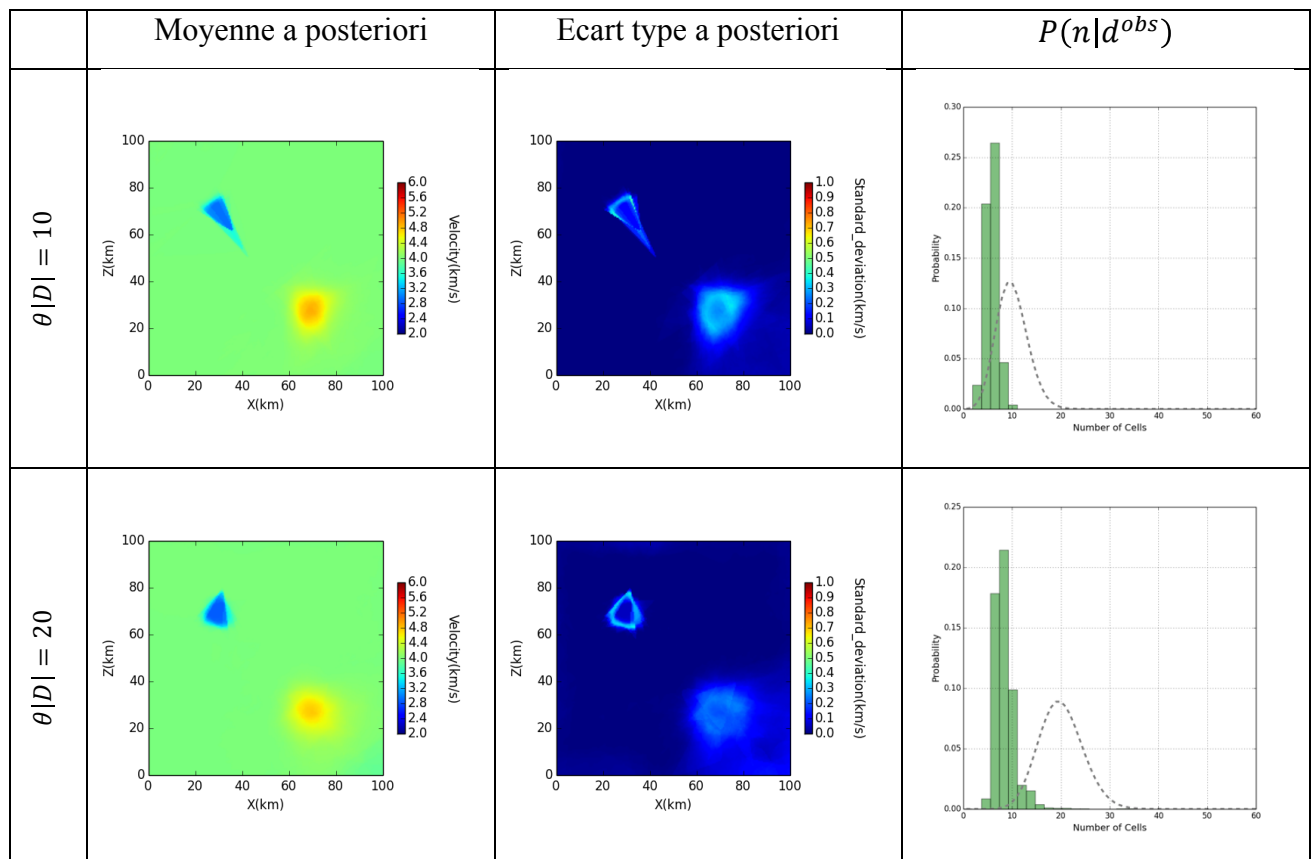
Figure 13 : Vrai modèle.

Les données synthétiques ont été générées sur une grille de taille 100km×100km discrétisée avec un pas constant de 2km dans chaque direction. Elles ont été calculées à partir d'un dispositif comprenant 17 sources et 20 récepteurs.

1.1 Mosaïque de Johnson-Mehl

Nous évaluons dans cette section l'impact de la moyenne a priori $\theta|D|$ sur les performances de l'algorithme. Le tableau 1 ci-après contient les résultats d'inversion pour différentes valeurs de $\theta|D|$. La première colonne contient les cartes du modèle moyen a posteriori, la deuxième colonne contient les cartes d'écart-types et la dernière colonne montre les histogrammes du nombre de cellules utilisées qui donnent la loi a posteriori $P(n|d^{obs})$ pour chaque inversion. Les courbes en pointillé, quant à elle, correspondent à la loi a priori sur ce nombre $P(n)$. Nous constatons d'ailleurs que la loi a posteriori $P(n|d^{obs})$ peut significativement différer de celle a priori $P(n)$. Concernant les autres résultats, à savoir le modèle de vitesse moyen et les cartes des écart-types, nous pouvons dire au premier abord que le paramètre θ n'a pas d'influence significative sur le résultat final et que l'algorithme s'adapte à la complexité du problème. Cependant, nous pouvons constater que le taux d'acceptation (Figure 14) dépend sensiblement de θ . En particulier, dans cet exemple, un choix de θ petit conduit à un taux d'acceptation trop faible révélateur d'une mauvaise adéquation de la connaissance a priori au problème.

Tableau 1 : Moyenne a posteriori, Ecart-type a posteriori et $P(n|d^{obs})$ pour différentes valeurs de $\theta|D|$



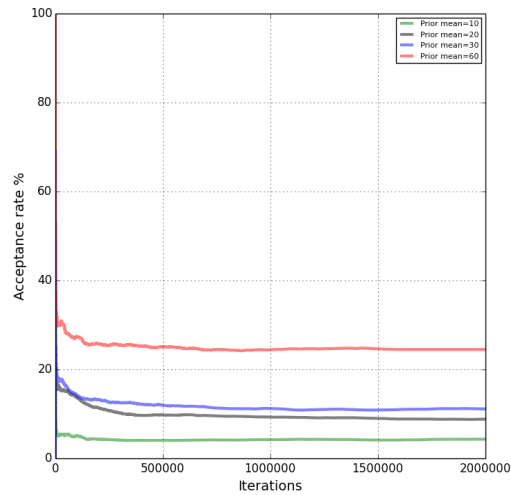
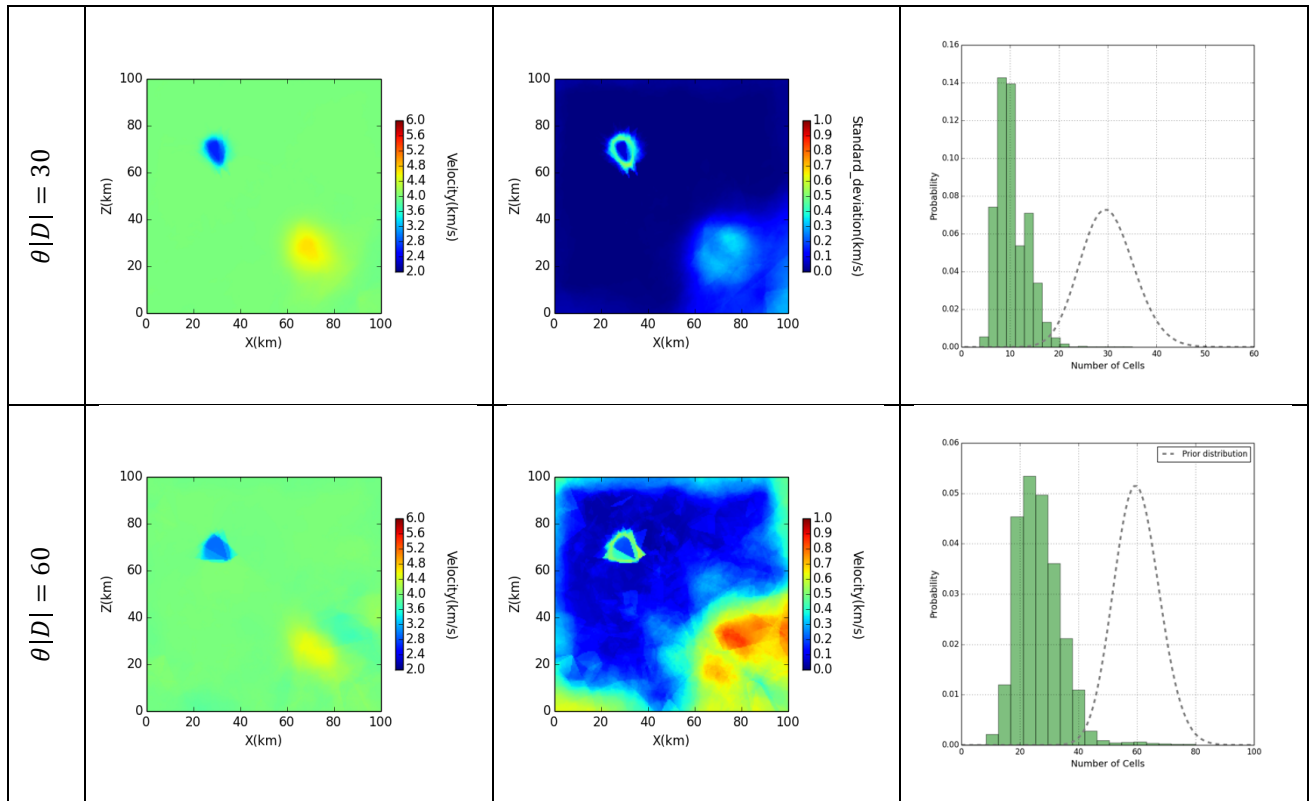


Figure 14: Taux d'acceptation moyen en fonction des différentes valeurs de θ

1.2 Noyaux gaussiens :

Nous étudions dans cette section l'impact de la moyenne a priori $\theta|D|$ et l'écart type moyen des noyaux gaussiens σ_m sur les performances de l'algorithme. Les tableaux ci-dessous (Tableau 2 et Tableau 3) contiennent les résultats d'inversion pour différentes valeurs de $\theta|D|$ et σ_m . De la même manière que pour la mosaïque de Johnson-Mehl, la première colonne

contient les cartes du modèle moyen a posteriori, la deuxième colonne contient les cartes d'écart-types et la dernière colonne montre les histogrammes du nombre de cellules utilisées qui donnent la loi a posteriori $P(n|d^{obs})$ pour chaque inversion. Les courbes en pointillé, quant à elle, correspondent à la loi a priori sur ce nombre $P(n)$.

Nous commençons tout d'abord par évaluer l'influence du nombre moyen a priori des noyaux gaussiens $\theta|D|$ en fixant la valeur de l'écart-type moyen a priori $\sigma_m = 6$ (Tableau 2).

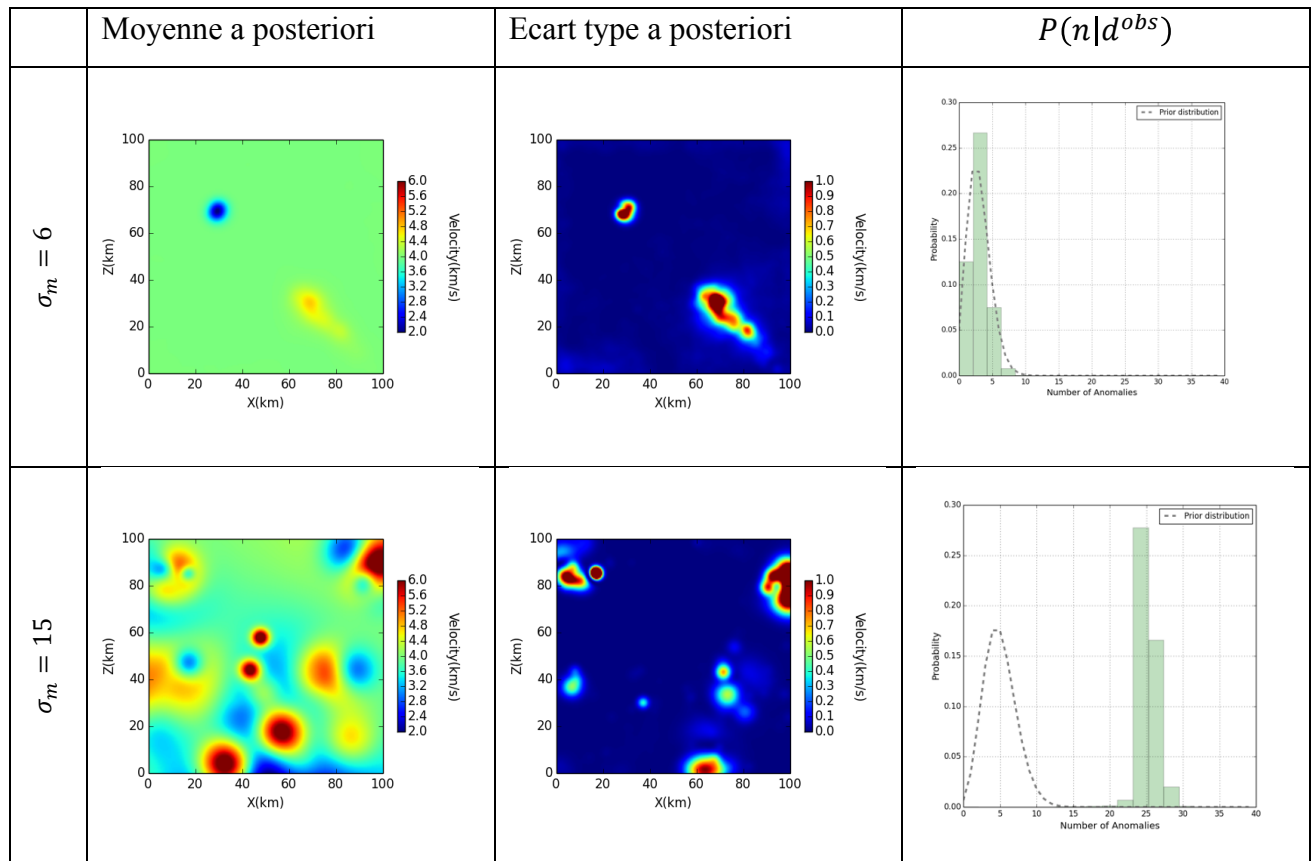
Tableau 2: Moyenne a posteriori, Ecart-type a posteriori et $P(n|d^{obs})$ pour différentes valeurs de $\theta|D|$ à $\sigma_m = 6$

	Moyenne a posteriori	Ecart type a posteriori	$P(n d^{obs})$
$\theta D = 5$			
$\theta D = 10$			

Nous constatons tout d'abord que la loi a posteriori $P(n|d^{obs})$ reste proche de la loi a priori $P(n)$. Concernant les cartes, les meilleurs résultats sont obtenus pour la valeur $\theta|D| = 5$. Nous constatons que les résultats sont très différents en fonction de $\theta|D|$, Le résultat final est sensible au choix du nombre et des caractéristiques des points poissonniens. Une légère modification du paramètre (ici le passage de $\theta|D| = 5$ à $\theta|D| = 10$) peut ainsi conduire à l'obtention de résultats différents. La Figure 15 montre le taux d'acceptation moyen de la chaîne pour les deux valeurs de θ . Le choix de θ doit se faire avec discernement, il a un impact non négligeable sur les performances de l'algorithme.

Nous évaluons maintenant l'influence de σ_m sur les résultats finaux en fixant la valeur de $\theta|D|$ à 5.

Tableau 3: Moyenne a posteriori, Ecart-type a posteriori et $P(n|d^{obs})$ pour différentes valeurs de σ_m à $\theta|D| = 5$.



L'algorithme est également sensible au choix du paramètre σ_m , un mauvais choix entraîne un mauvais comportement de l'algorithme. Là encore, le choix de σ_m doit se faire avec soin. En conclusion, le choix de θ et σ_m doit se faire avec discernement, il a un impact non négligeable sur les performances de l'algorithme. En pratique, la carte des vitesses issue d'une tomographie linéaire peut être utilisée comme information a priori. Elle permet d'estimer le nombre de noyaux gaussiens ainsi que leur étendu qui permettent de reproduire les variations de vitesse. En l'absence d'une telle information, le choix de ses deux paramètres n'étant pas évident, leur valeur devraient être déterminées par essai et erreur en examinant le taux d'acceptation moyen de la chaîne.

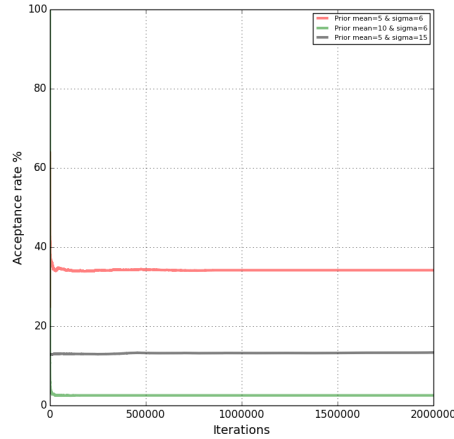


Figure 15: Taux d'acceptation moyen en fonction des différentes inputs

1.3 Conclusion

Les différents paramètres testés ici n'ayant pas d'interprétation physique, leur estimation au préalable n'est pas évidente. Il est nécessaire d'effectuer plusieurs tests au préalable afin de choisir les valeurs adéquates pour le problème qui assurent le bon fonctionnement de l'algorithme. Notons toutefois que la mosaïque de Johnson-Mehl est moins sensible aux paramètres et donc plus facile à régler. Après avoir pris connaissance des différentes difficultés liées à l'implémentation de chaque paramétrisation, nous testons notre méthodologie sur d'autres exemples plus complexes.

2 Données synthétiques 1:

Dans cette partie, nous utilisons un modèle de vitesse synthétique pour générer les temps de première arrivée. Le modèle est similaire à celui utilisé dans (Bodin et al., 2009). Il a des dimensions de 100 km de profondeur et 100 km de longueur et est composé de deux structures principales avec un fort contraste de vitesse: le triangle vert a une vitesse de 4 km/s et contient une anomalie positive où la vitesse augmente de 25%, tandis que la vitesse dans la zone orange est de 5 km/s avec la présence d'une petite anomalie lente se caractérisant par une chute de la vitesse de 20%.

Les données synthétiques ont été générées sur une grille de taille 100km×100km et discrétisée avec un pas constant de 2km dans chaque direction. Elles ont été calculées à partir d'un dispositif comprenant 17 sources (carrés rouges) et 20 récepteurs (triangles bleus). L'écart-type de l'ensemble des erreurs est : $\sigma_d = 0.4s$

Pour calculer des temps de première arrivée nous utilisons le solveur numérique par différences finies développé par (Noble et al., 2014)

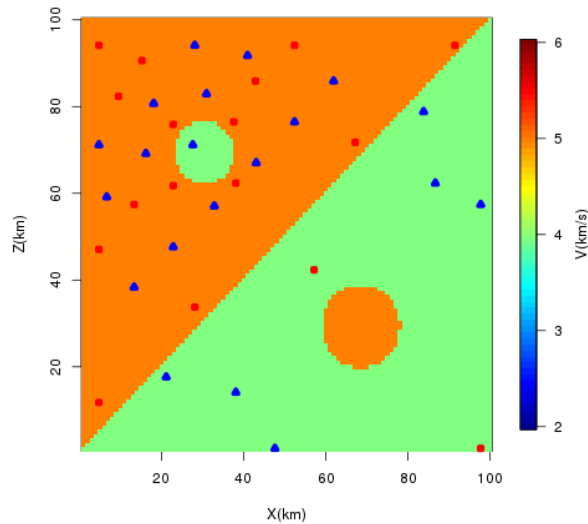


Figure 16: Vrai modèle

Les tableaux (Tableau 4 et Tableau 5) donnent l'ensemble des éléments nécessaires pour le réglage des lois a priori. Au vu des conclusions tirées précédemment dans la partie « Analyse de sensibilité » et après plusieurs tests, nous avons choisi de garder les valeurs suivantes qui assurent le bon fonctionnement de l'algorithme.

Concernant la mosaïque de Johnson-Mehl, le nombre de cellules moyen a priori retenu est de 30. La position de chaque germe est uniforme sur le domaine $[0,100] \times [0,100]$. La valeur de la vitesse en chaque cellule est uniforme sur $[2,6]$ enfin le temps d'implantation de chaque germe est uniforme sur $[0,20]$, cet intervalle est suffisant pour obtenir des cellules de tailles et de formes différentes.

Tableau 4: Lois a priori sur les paramètres du modèle m: Mosaïque de Johnson-Mehl.

Paramètres du modèle	La loi a priori	Paramètre de la loi a priori
Nombre de cellules	Poisson	Paramètre de Poisson 0.003
Coordonnées de chaque germe	Uniforme	Minimum 0
		Maximum 100
Vitesse v de chaque cellule	Uniforme	Minimum 2
		Maximum 6
Temps d'implantation des cellules	Uniforme	Minimum 0
		Maximum 20

Concernant les noyaux gaussiens, nous choisissons d'utiliser un nombre moyen de noyaux gaussiens a priori de 15. La position des noyaux gaussiens est uniforme dans le domaine $[0,100]$, leur amplitude est uniforme sur $[-4,4]$, l'écart type moyen est de 20. La valeur de la vitesse de fond, quant à elle, est uniforme sur $[2,6]$,

Tableau 5: Lois a priori sur les paramètres du modèle m: Noyaux gaussiens isotropes.

Paramètres du modèle	La loi a priori	Paramètre de la loi a priori	
La vitesse de fond	Uniforme	Minimum	2
		Maximum	6
Nombre de noyaux gaussiens	Poisson	Paramètre de Poisson	0.0015
Coordonnées de chaque anomalies	Uniforme	Minimum	0
		Maximum	100
Amplitude	Uniforme	Minimum	-4
		Maximum	4
Étendue	Exponentielle	Moyenne	20

Résultats :

Les résultats de l'inversion trans-dimensionnelle sont présentés ci-après. Les figures (Figure 17, Figure 18) représentent les modèles de vitesse moyens et les cartes des incertitudes associées (écarts-type) pour les cellules de Johnson-Mehl et les noyaux gaussiens isotropes respectivement. Ces résultats ont été obtenus après 1 million d'itérations pour la mosaïque de Johnson-Mehl et 2 millions d'itérations pour les noyaux gaussiens.

Nous remarquons tout d'abord que les principales structures ont été bien reconstruites, en particulier l'anomalie négative située dans la partie supérieure à gauche de l'image. En revanche, l'inversion semble moins précise dans la partie inférieure droite, autour de l'anomalie rapide. En effet, dans cette zone, la couverture des rais est faible et ne permet pas une estimation correcte du modèle de vitesse par l'algorithme de tomographie des temps de première arrivée. Ceci est clairement visible sur la carte des incertitudes (Figure 17b, Figure 18b), il existe une forte incertitude dans cette région, en particulier dans le cas de la mosaïque aléatoire de Johnson-Mehl. Enfin, il convient de noter que la paramétrisation de Johnson-Mehl arrive à reproduire la limite diagonale séparant les zones à fort contraste de vitesse avec plus de précision que les noyaux gaussiens. Cette remarque trouve son explication dans le fait

que les noyaux gaussiens utilisés comme fonction de base dans la deuxième paramétrisation sont des fonctions relativement lisses, peu adaptées à la reproduction de discontinuités.

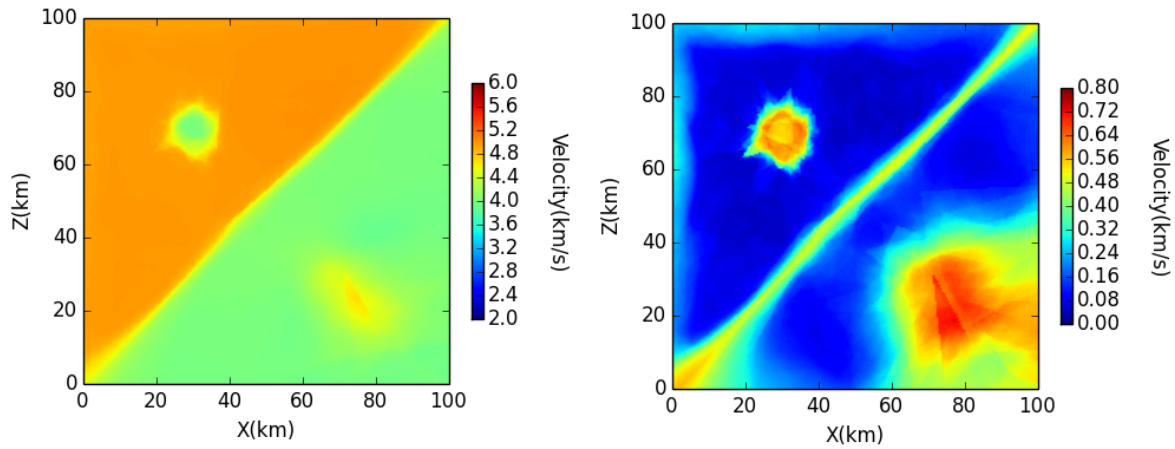


Figure 17 Johnson-Mehl : a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

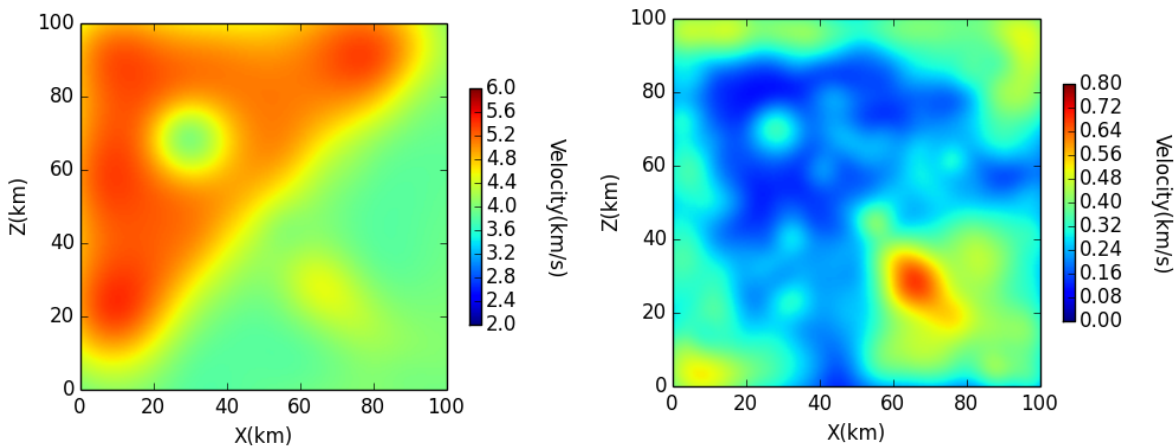


Figure 18 Noyaux gaussiens: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

Outre les cartes, des sections transversales à $Z=30\text{km}$ et $Z=70\text{km}$ ont été réalisées pour illustrer les incertitudes a posteriori. Les courbes vertes et rouges indiquent respectivement la moyenne a posteriori et le vrai modèle utilisé pour calculer des données synthétiques. Les lignes noires correspondent quant à elles aux intervalles de confiance à 95%. Les figures (Figure 19 et Figure 20) montrent que le modèle moyen a posteriori est presque identique au vrai modèle en particulier pour la mosaïque de Johnson-Mehl, sauf dans la zone voisine de la petite anomalie positive (Figure 19a, Figure 20a). La vitesse dans cette zone est nettement moins bien contrainte dû à la faible couverture des rais, toutefois la valeur réelle est comprise dans l'intervalle de confiance à 95%. A ce stade, on peut déjà constater que la mosaïque de Johnson-Mehl (JM), pour cet exemple, donne de meilleurs résultats en terme de modèle moyen et d'incertitudes.

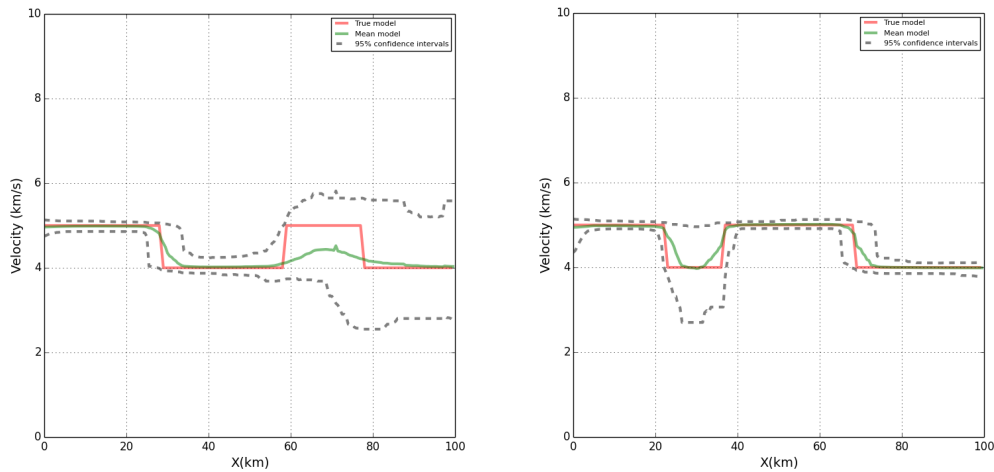


Figure 19 Johnson-Mehl: Sections transversales. (a) à Z= 30 km . (b) à Z=70 km. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%

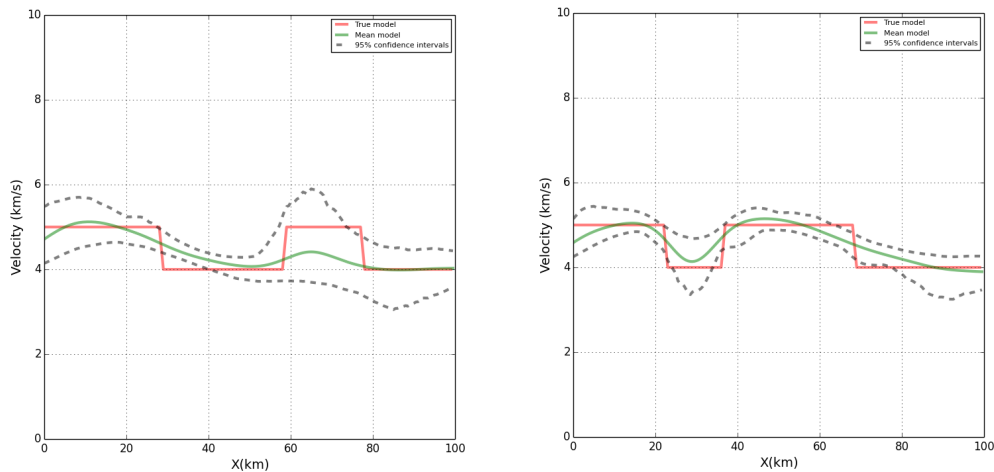


Figure 20 Noyaux gaussiens isotropes : Sections transversales. (a) à Z= 30 km . (b) à Z=70 km. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%

La Figure 21 présente les histogrammes du nombre de cellules Johnson-Mehl et noyaux gaussiens. Ces histogrammes fournissent une estimation de la distribution a posteriori du nombre de cellules Johnson-Mehl et des noyaux gaussiens $P(n|d^{obs})$. Le nombre de cellules de Johnson-Mehl est compris entre 8 et 30 avec un mode à 12; quant au nombre de noyaux gaussiens, celui-ci varie entre 12 et 38 avec un mode à 22. Il est évident que les noyaux gaussiens nécessitent plus de paramètres que la mosaïque de Johnson-Mehl. Il convient toutefois de mentionner que même pour les noyaux gaussiens, 88 paramètres au mode --étant donné que chaque noyau est caractérisé par quatre paramètres-- reste un nombre raisonnable. Enfin, il est important de noter que dans le cas de la mosaïque de Johnson-Mehl, la moyenne a posteriori est significativement différente de la valeur a priori. Ceci est en cohérence avec les conclusions tirées précédemment dans la section consacrée à l'analyse de sensibilité et confirme donc que le choix de la moyenne a priori n'a pas d'effet sur les résultats finaux dans

la mesure où ce nombre est suffisamment grand pour prendre en compte les complexités du modèle. Au contraire, le modèle paramétrique utilisant les noyaux gaussiens est relativement sensible à ce choix.

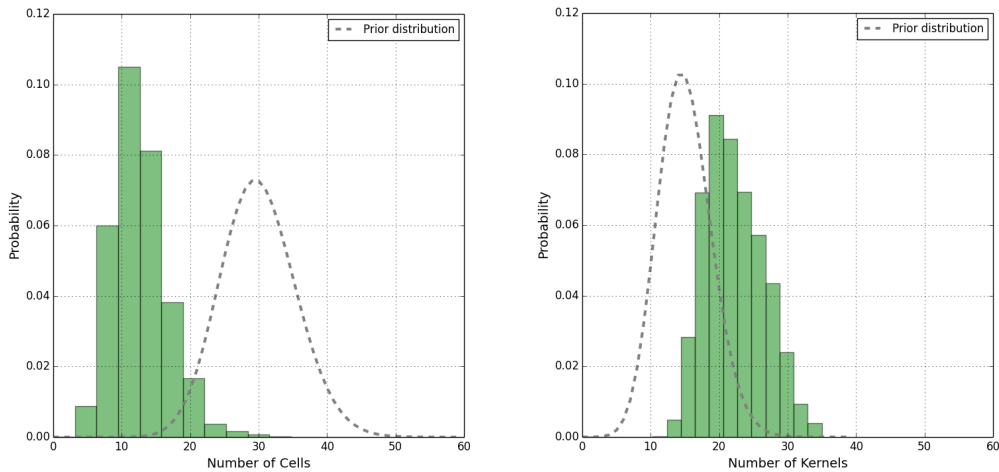


Figure 21 Histogramme du nombre de a) Cellules de Johnson Mehl b) Noyaux gaussiens

Enfin, une étape importante dans l'évaluation des résultats de tomographie est l'inspection des résidus. La Figure 22 montre l'histogramme des différences entre les temps observés et les temps d'arrivée calculés pour le modèle moyen a posteriori. La Figure 22 montre que la valeur absolue des résidus est inférieure à la valeur de l'écart-type des données $\sigma_d = 0.4s$, sauf pour quelques couples de source-récepteurs, particulièrement dans le cas de la mosaïque de JM. Ainsi, pour les deux paramétrisations le modèle moyen a posteriori permet d'expliquer suffisamment bien les données observées. Néanmoins, il convient de noter que les résidus pour la mosaïque de JM sont remarquablement faibles. Nous pouvons finalement affirmer au vu de ces résultats que la mosaïque de JM donne de meilleurs résultats.

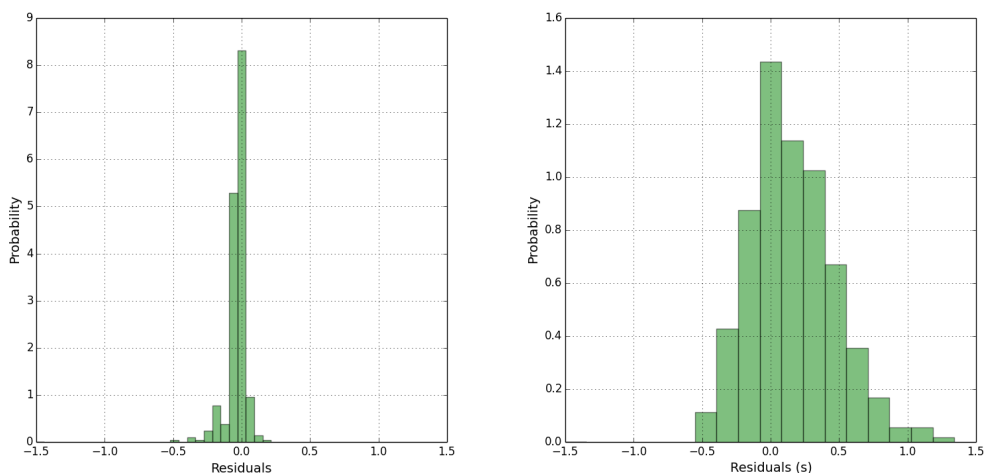


Figure 22 Histogramme des résidus a) Mosaïque de Johnson-Mehl b) Noyaux gaussiens

Il est intéressant de comparer les résultats de JM à ceux présentés dans (Bodin and Sambridge, 2009). On peut noter que le modèle de Johnson-Mehl introduit un paramètre supplémentaire par rapport au modèle de Voronoï, à savoir le temps d'implémentation du grain. Ce paramètre permet d'obtenir des grains avec des géométries plus complexes, là où le modèle de Voronoï produit uniquement des polytopes. En particulier, les cellules de JM ne sont pas nécessairement convexes. Ce paramètre supplémentaire est loin d'être superflu, puisqu'il permet de mieux reproduire certaines interfaces dans le champ de vitesse, contribuant ainsi à une amélioration des résultats obtenus. On notera à titre d'exemple que l'interface diagonale est mieux retrouvée. Nous présentons également le meilleur modèle (Figure 23) en vue de le comparer avec celui présenté dans (Bodin and Sambridge, 2009). Nous remarquons que les cellules de JM permettent de mieux reproduire les principales structures. En conclusion, on peut affirmer que pour cet exemple, bien que les cellules de JM utilisent un paramètre de plus que les cellules de Voronoï, leur forme générale permet de donner de meilleurs résultats.

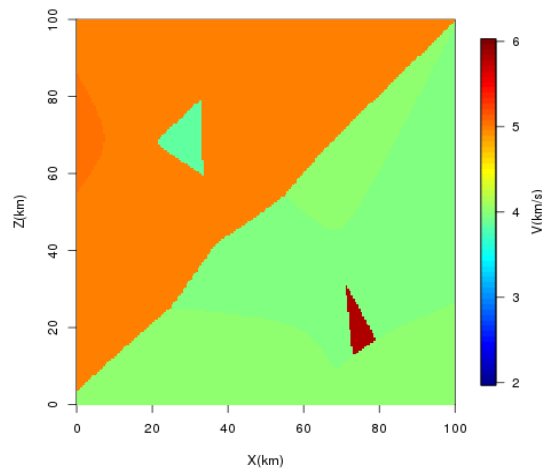


Figure 23: Meilleur modèle pour la mosaïque de Johnson-Mehl.

3 Données synthétiques 2:

Nous allons maintenant appliquer notre méthodologie à un modèle plus complexe et plus réaliste. Il s'agit d'une version modifiée du modèle développé à l'Amoco Tulsa Research Lab par O'brien en 1994. Le modèle présente un intérêt en raison de son utilisation comme benchmark dans l'industrie pétrolière. Le modèle de vitesse est purement acoustique et contient différents types d'hétérogénéités de proche-surface comme montré sur la Figure 24. Il s'agit d'un modèle complexe avec des variations de vitesse dans les deux directions horizontale et verticale.

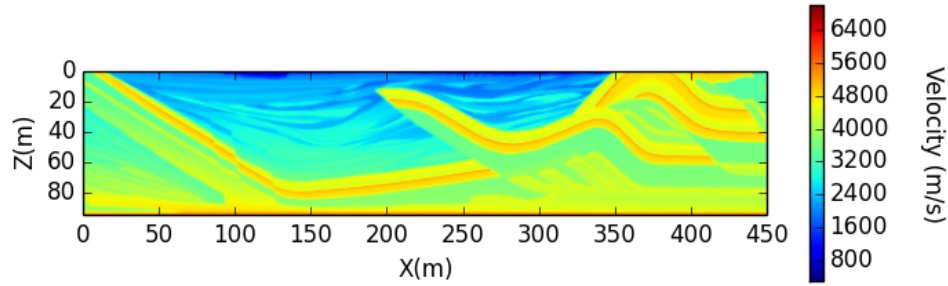


Figure 24 : Vrai modèle.

Temps de première arrivée :

Le modèle de vitesse mesure 450km de longueur et 95km de profondeur. Sur une grille discrétisée avec un pas constant de 5m, nous simulons une acquisition sismique en réfraction avec 23 sources et 44 récepteurs spatialement répartis à la surface du modèle. Les données observées (Figure 25) sont calculées en utilisant le solveur Eikonal développé par (Noble et al., 2014). Le nombre total de données observées est de 1012 et l'écart-type du bruit sur les données $\sigma_d = 0.005s$.

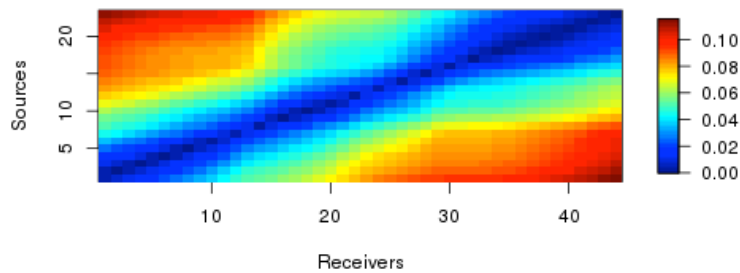


Figure 25: Carte des temps de premières arrivées.

Les tableaux (Tableau 6, Tableau 7) présentent l'ensemble des éléments nécessaires pour le réglage des lois a priori sur les paramètres du modèle. Au vu des conclusions tirées précédemment dans la partie « Analyse de sensibilité » et après plusieurs tests, nous avons choisi de garder les valeurs qui y sont indiquées, qui assurent le bon fonctionnement de l'algorithme.

Concernant la mosaïque de Johnson-Mehl, le nombre de cellules moyen a priori retenu est de 427. La position de chaque germe est uniforme sur le domaine $[0,450] \times [0,95]$. La valeur de la vitesse en chaque cellule est uniforme sur $[200,7000](m/s)$ enfin le temps d'implantation de chaque germe est uniforme sur $[0,20]$.

Tableau 6 : Les Lois a priori sur les paramètres du modèle m: Mosaïque de Johnson-Mehl.

Paramètres du modèle	La loi a priori	Paramètres de la loi a priori
Nombre de cellules	Poisson	Paramètre de poisson 0.01
Coordonnées de chaque cellule	Uniforme	Minimum 0
		Maximum 450 pour x et 95 pour z
Vitesse v de chaque cellule	Uniforme	Minimum 200
		Maximum 7000
Temps d'implantation des cellules	Uniforme	Minimum 0
		Maximum 20

Concernant les noyaux gaussiens, le nombre de cellules moyen a priori retenu est de 42. La position de chaque germe est tirée selon une loi uniforme sur le domaine $[0,450km] \times [0,95km]$, leur amplitude est uniforme sur $[-3000,3000]$. Nous utilisons ici des noyaux gaussiens anisotropes pour mieux gérer la complexité du modèle. Dans ce cas nous introduisons des paramètres supplémentaires pour contrôler l'anisotropie, à savoir une orientation (angle \emptyset) et un second paramètre d'échelle σ_{m_z} , leurs valeurs sont indiquées dans le tableau ci-dessous (Tableau 7).

Tableau 7: Les Lois a priori sur les paramètres du modèle m: Noyaux gaussiens anisotropes.

Paramètres du modèle	La loi a priori	Paramètres de la loi a priori
Vitesse de fond	Uniforme	Minimum 200
		Maximum 7000
Nombre d'anomalies	Poisson	Paramètre de Poisson 0.001
Coordonnées de chaque anomalies	Uniforme	Minimum 0
		Maximum 450 pour x et 95 pour z
Amplitude	Uniforme	Minimum -3000
		Maximum 3000
Étendue	Module	Exponentielle
		Mean 15

Angle	Uniforme	Minimum	0
		Maximum	π

Résultats :

Les résultats de l'inversion trans-dimensionnelle sont présentés dans les figures (Figure 26 et Figure 27) pour les deux types de paramétrisations la mosaïque de JM et les noyaux gaussiens anisotropes. Celle-ci montrent les cartes de la vitesse moyenne a posteriori et les incertitudes associées à chaque estimation (écart type). Ces résultats ont été obtenus après 1 million d'itérations.

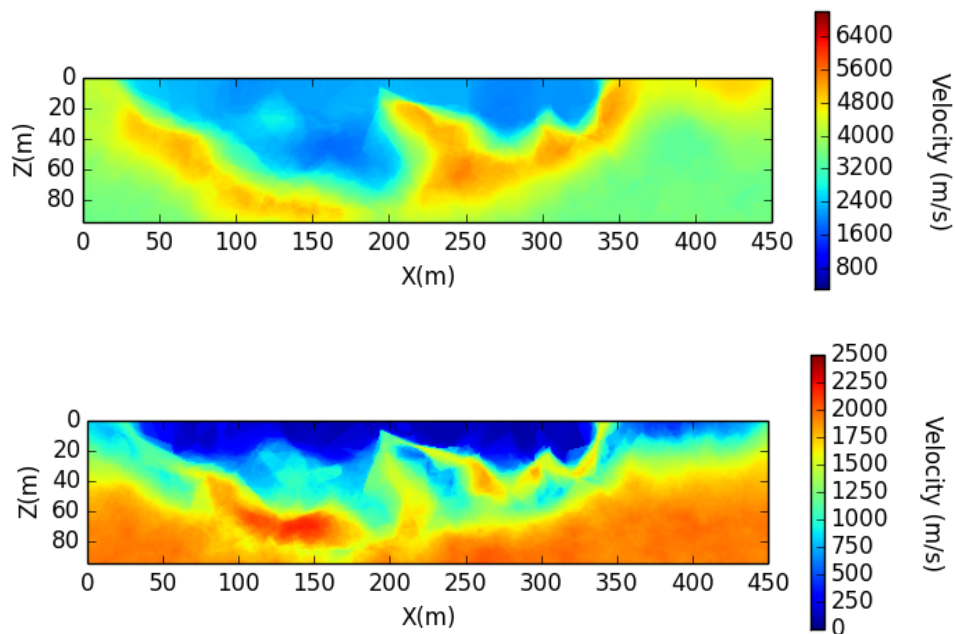


Figure 26 Johnson-Mehl : a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

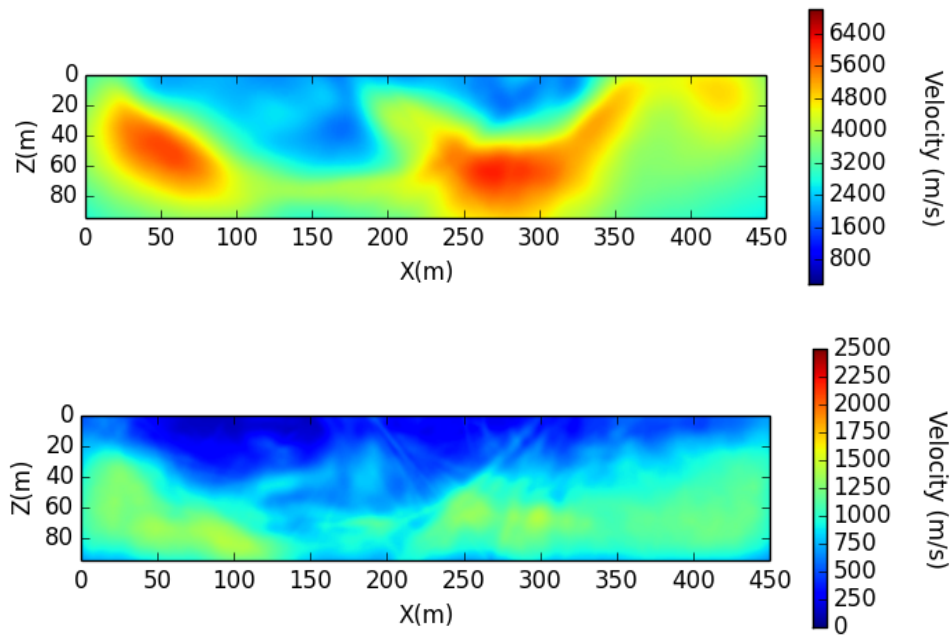


Figure 27 Noyaux gaussiens anisotropes: a) Carte de la moyenne a posteriori b) Carte des incertitudes (écart-type)

Outre les cartes, des sections transversales ont été réalisées à $Z=25\text{m}$ et $Z=65\text{m}$ pour illustrer les incertitudes a posteriori (Figure 28, Figure 29). Les courbes vertes et rouges indiquent respectivement la moyenne a posteriori et le vrai modèle utilisé pour calculer les données synthétiques. Les lignes noires correspondent quant à elles aux intervalles de confiance à 95%. Les figures (Figure 28 et Figure 29) montrent que le modèle moyen a posteriori est presque identique au vrai modèle en particulier près de la surface où les incertitudes sont relativement faibles. Ceci est clairement visible dans le cas de la mosaïque de JM, le modèle moyen est remarquablement proche du vrai modèle, les incertitudes estimées sont elles aussi cohérentes puisqu'elles sont plus grandes en profondeur dans la zone non couverte par les rais. Notons enfin que la valeur des incertitudes dans la partie non contrainte est plus forte dans le cas de la mosaïque de JM, ceci étant néanmoins principalement lié aux valeurs des lois a priori.

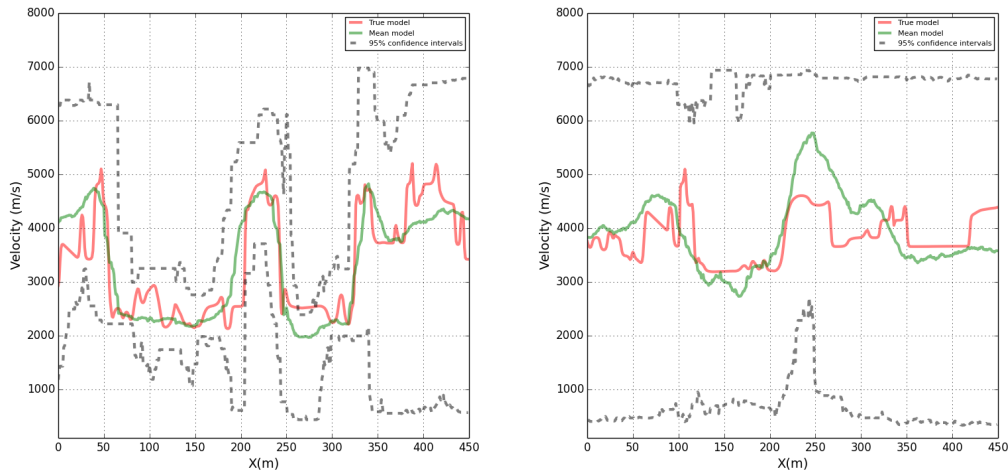


Figure 28 : Johnson-Mehl: Sections transversales. (a) à $Z=25m$. (b) à $Z=65m$. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%

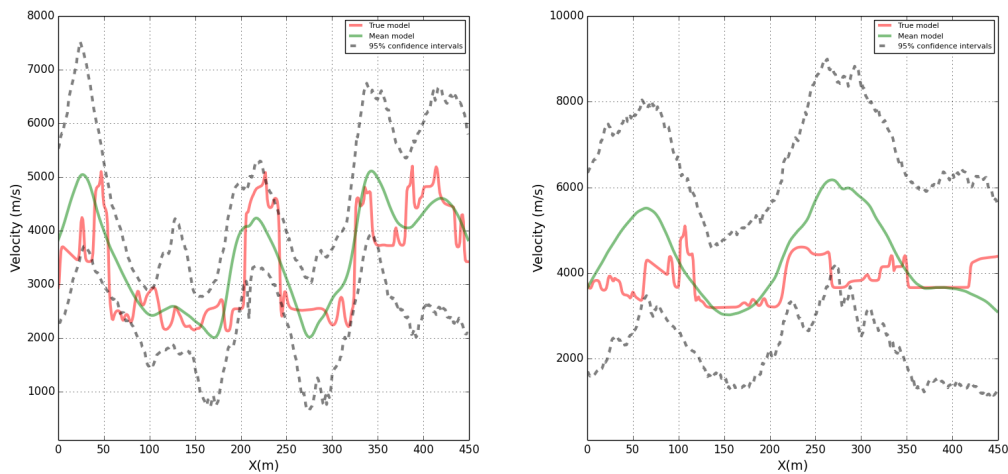


Figure 29 Noyaux gaussiens anisotropes: Sections transversales. (a) à $Z=25m$. (b) à $Z=65m$. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%

La Figure 30 représente les histogrammes du nombre de cellules Johnson-Mehl et de noyaux gaussiens. Le nombre de cellules de Johnson-Mehl est compris entre 200 et 340 avec un mode à 260, quand le nombre de noyaux gaussiens varie entre 30 et 70 avec un mode à 48.

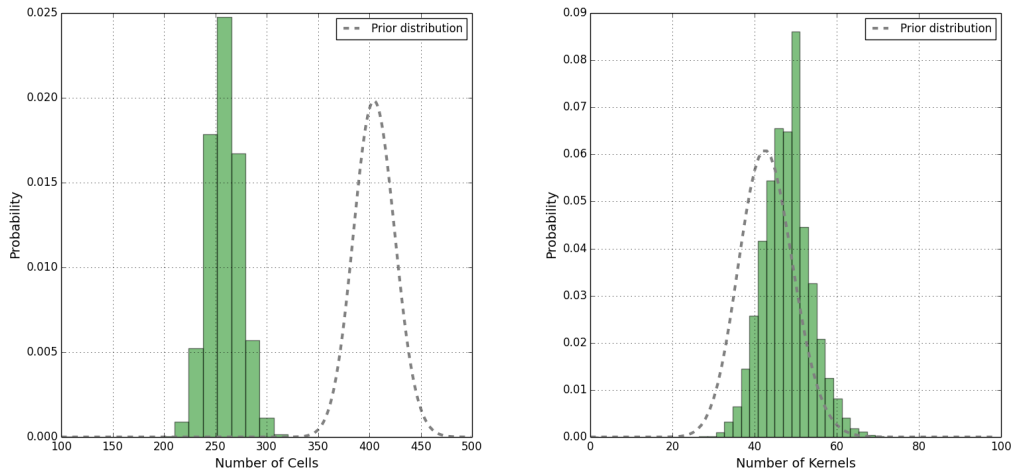


Figure 30 Histogramme du nombre de a) Cellules de Johnson Mehl b) Noyaux gaussiens anisotropes

Finalement, nous montrons dans la Figure 31 les histogrammes des résidus des temps entre les temps générés par le modèle moyen a posteriori et les données observées. La valeur absolue des résidus est inférieure à la valeur de l'écart-type des données $\sigma_d = 0.005s$, sauf pour quelques couples de source-récepteurs. Ainsi, pour les deux paramétrisations le modèle moyen a posteriori permet d'expliquer suffisamment bien les données observées.

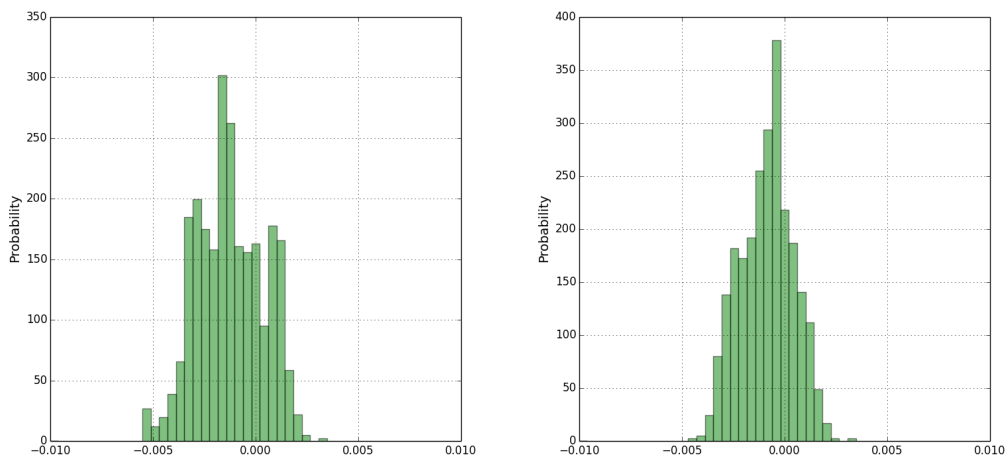


Figure 31 Histogramme des résidus en (ms) a) Mosaïque de Johnson-Mehl b) Noyaux gaussiens anisotropes

4 Conclusion :

Nous avons commencé ce chapitre par une analyse de sensibilité. Nous avons montré l'impact des différents paramètres de réglage sur les résultats finaux. Les résultats de la tomographie ont été ensuite illustrés pour un modèle synthétique, composé de deux structures principales avec un fort contraste de vitesse sur lesquelles se superposent deux petites anomalies. Nous avons finalement appliqué notre méthodologie à un modèle plus réaliste qui sert de benchmark dans l'industrie pétrolière. Les résultats montrent la capacité de notre algorithme à reconstruire le champ de vitesse des ondes sismiques en utilisant un nombre raisonnable de paramètres et ce même pour des modèles de sub-surfaces complexes. Les deux paramétrisations que nous avons proposées semblent donc bien adaptées à des applications en tomographie bayésienne des temps de premières arrivées. Dans les exemples présentés précédemment, la mosaïque de JM s'est remarquablement distinguée par sa capacité à reconstruire avec précision des discontinuités d'ordre 0 et par son réglage relativement facile et moins sensible aux choix des paramètres en comparaison des noyaux gaussiens. Les résultats montrés dans cette partie étant encourageants, il serait donc intéressant d'appliquer cette paramétrisation sur de grands jeux de données de sismique réfraction. Néanmoins le temps de calcul reste un des inconvénients majeurs de ces paramétrisations, il est de l'ordre de 2 jours pour 1 million d'itérations sur un ordinateur de bureau. Toutefois il serait possible d'éviter cet obstacle en parallélisant le code. En effet, étant donné que les calculs sont réalisés par point de tir, ces derniers peuvent être exécutés simultanément sur plusieurs cœurs permettant ainsi de réduire le temps de calcul.

Chapitre 5.

Modèle à couches

Nous nous intéressons dans ce chapitre à des modèles dédiés à la la paramétrisation dite à couches, plus particulièrement les couches horizontales. Nous illustrons tout d'abord les résultats de la tomographie pour un modèle synthétique réaliste. Les résultats que nous présentons sont obtenus pour le cas où le nombre de couches est fixé à l'avance. Cet exemple est l'occasion d'évaluer les performances de l'algorithme des chaînes en interaction PIR présenté dans le chapitre 2. Enfin, nous présentons une application sur un jeu de données réelles issues de la fracturation hydraulique. Nous utilisons dans ce contexte une approche trans-dimensionnelle et hiérarchique.

1 Données synthétiques:

Nous nous proposons dans cette section, d'illustrer la paramétrisation à couches horizontales à travers un exemple synthétique réaliste. Une étude similaire a fait l'objet d'une publication au congrès Petroleum Geostatistics 2015, Biarritz, France, dont une copie est présentée en annexe.

Le modèle est similaire à celui présenté dans (Gesret et al., 2015). Le modèle de vitesse s'étend sur 1200m de profondeur et 600 m de largeur. Il s'agit d'un modèle à 23 couches avec par endroit des gradients et des couches fines (Figure 32). Nous comparons les résultats d'une seule chaîne de Markov et ceux obtenus en utilisant les chaînes en interaction.

Le dispositif (Figure 33) comprend 15 sources (étoiles bleues) situées à une profondeur d'environ 1035m et 80 récepteurs placés dans deux puits distincts (triangles noirs) et la valeur de l'écart-type de la vraisemblance est $\sigma_d = 0.005s$.

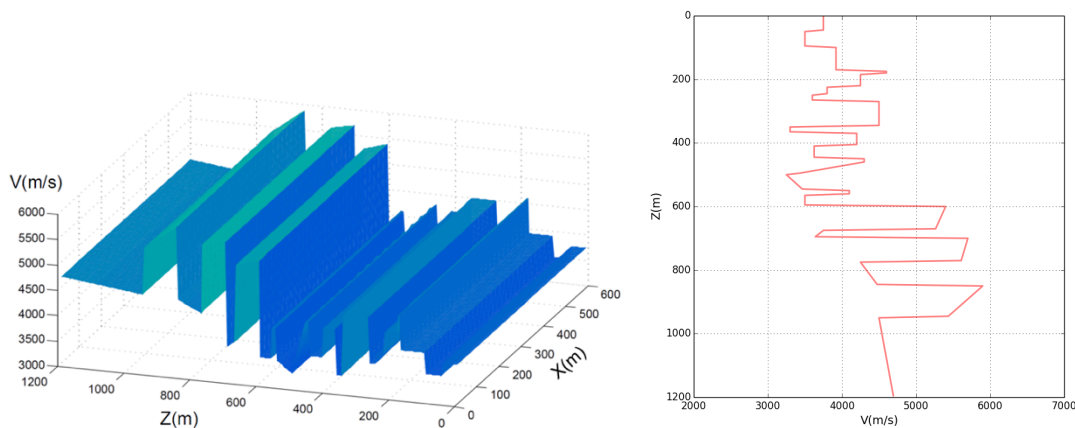


Figure 32: Vrai modèle a) une vue 3D b) profile vertical de vitesse en fonction de la profondeur

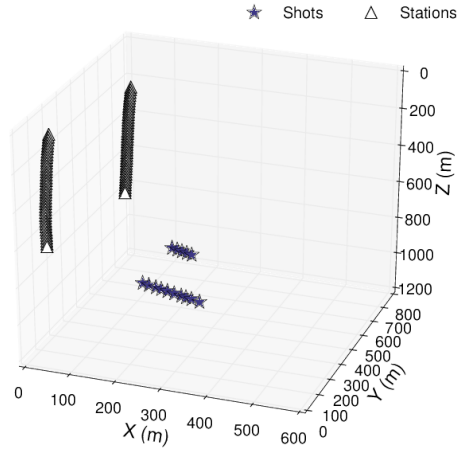


Figure 33: Géométrie d'acquisition

Les temps de première arrivée :

Les données synthétiques ont été générées sur une grille de taille 1200m×800m et discrétisée avec un pas constant de 5m dans la direction z et de 10m dans la direction x. Pour calculer ces temps de première arrivée nous utilisons le solveur numérique de (Noble et al., 2014). La Figure 34 montre la carte des temps de premières arrivées générés par le dispositif (Figure 33).

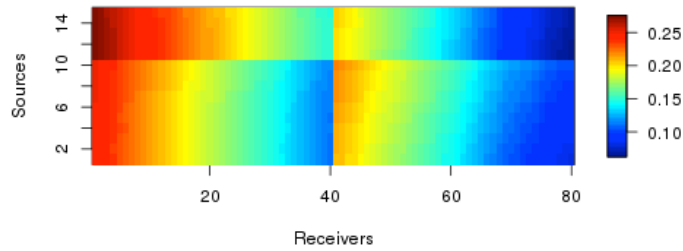


Figure 34: Carte des temps observés

Dans ce qui suit nous proposons d'inverser les données observées (temps de première arrivée synthétiques) issues du modèle à 23 couches avec un modèle à 10 couches horizontales. Le champ de vitesse en tout point du domaine peut s'écrire alors sous la forme suivante :

$$V_p(u) = v_{p_1} 1_{[0, z_1]}(z_u) + \sum_{i=2}^{10} v_{p_i} 1_{]z_{i-1}, z_i]}(z_u) \quad (93)$$

$1_{]z_{i-1}, z_i]}$ correspond à l'indicatrice de l'intervalle $]z_{i-1}, z_i]$.

Le modèle de vitesse est entièrement décrit par la profondeur z_i de chaque couche et par son module de vitesse des ondes P v_{p_i} .

Le vecteur des paramètres que l'on cherche à estimer conditionnellement aux données s'écrit comme suit:

$$m = (m_j)_{1 \leq j \leq 20} = (z_i, V_{p_i})_{1 \leq i \leq 10}$$

La loi a priori :

Nous faisons tout d'abord l'hypothèse que tous les paramètres sont indépendants. La loi a priori $P(m)$ s'exprime alors comme le produit des lois a priori sur chaque paramètre m_i :

$$P(m) = \prod_{1 \leq i \leq 20} P(m_i) = \prod_{1 \leq i \leq 10} P(z_i) \prod_{1 \leq i \leq 10} P(v_{p_i}) \quad (94)$$

Nous supposons a priori que tous les paramètres suivent une loi uniforme :

$$\forall i \in \llbracket 1, n \rrbracket \quad P(m_i) = \begin{cases} \frac{1}{(m_{i_{max}} - m_{i_{min}})} & \text{pour } m_{i_{min}} \leq m_i \leq m_{i_{max}} \\ 0 & \text{sinon} \end{cases} \quad (95)$$

Le tableau ci-dessous regroupe les valeurs choisies pour chaque paramètre. Les bornes ($z_{i_{min}}$ et $z_{i_{max}}$), des intervalles des lois a priori sur les profondeurs z_i , retenues permettent d'éviter le chevauchement des différentes couches. Nous utilisons des intervalles larges pour la valeur de la vitesse pour garantir le moins d'informations possibles.

Tableau 8: Les vitesses (m/s) et les profondeurs (m)

Vp ₁	2500	4500	Vp ₆	3500	6500
Z ₁	70	130	Z ₆	670	730
Vp ₂	3000	5000	Vp ₇	3500	6500
Z ₂	140	200	Z ₇	750	810
Vp ₃	3000	5000	Vp ₈	2500	6500
Z ₃	240	300	Z ₈	820	880
Vp ₄	3000	5000	Vp ₉	3500	7000
Z ₄	390	450	Z ₉	920	980
Vp ₅	2500	4500	Vp ₁₀	3000	6000
Z ₅	570	630	Z ₁₀	1190	1190

La Figure 35 ci-dessous présente quelques modèles de vitesse obtenus par simulation de la loi a priori des paramètres.

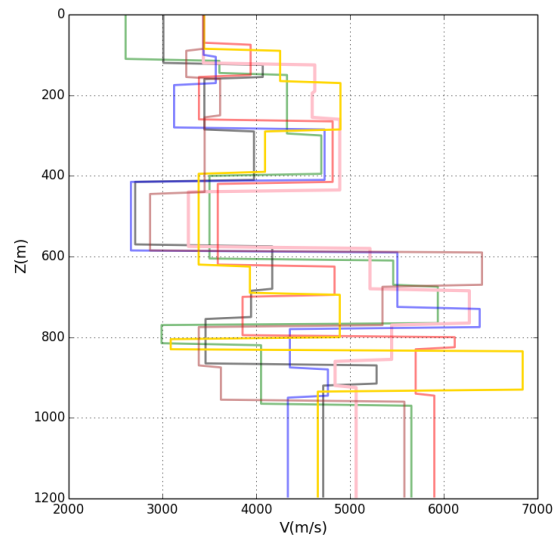


Figure 35: Réalisations à partir de la loi a priori

1.1 Une seule chaîne :

Nous présentons ici les résultats de l'inversion en utilisant l'algorithme Metropolis Hastings à marche aléatoire.

La Figure 36 présente les principaux résultats obtenus pour un nombre d'itérations de 50000². Le modèle moyen (en vert), suit bien le vrai modèle jusqu'à une certaine profondeur $z=800\text{m}$ au delà de laquelle, les deux modèles se séparent nettement l'un de l'autre. De plus les incertitudes semblent légèrement sous-estimées dans cette zone qui n'est pas contrainte par les données. En effet, la couche à faible vitesse à 800m ainsi que le gradient de vitesse pour les profondeurs supérieures à 1100m devraient être inclus dans l'intervalle de confiance. Le résultat obtenu dans cette partie du modèle n'est donc pas fiable.

² Ce nombre est faible, l'objectif étant principalement de comparer pour un temps de calcul égal, les performances de l'algorithme avec une seule chaîne et avec les chaînes en interaction.

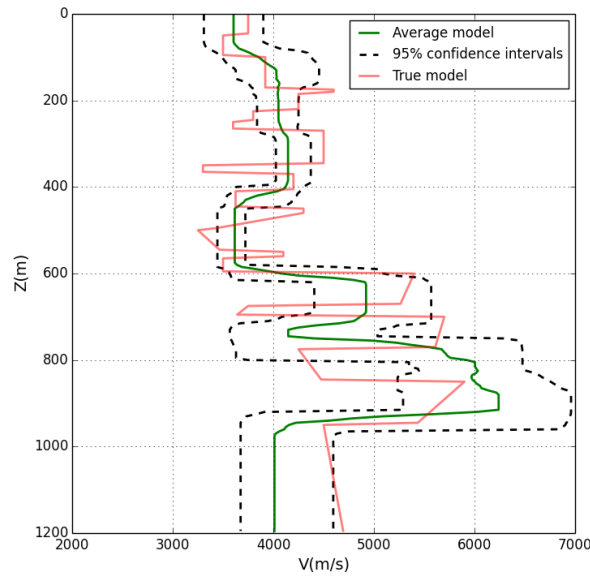


Figure 36: Résultats pour 50000 itérations. Vert: modèle moyen, rouge: vrai modèle, noir: intervalle de confiance à 95%

La Figure 37 montre l'énergie de la chaîne :

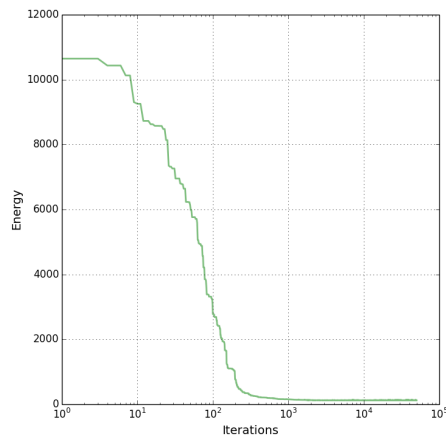


Figure 37: Énergie de la chaîne

Le nombre d'itérations (en abscisse) est affiché avec une échelle logarithmique afin de permettre une meilleure visualisation de la décroissance de l'énergie. Celle-ci correspond à la phase Burn-in (transitoire) que parcourt la chaîne avant d'atteindre son régime stationnaire

1.2 Chaînes en interaction :

Dans le but d'échantillonner la loi a posteriori, nous utilisons ici l'algorithme de Population Importance Resampling décrit dans le chapitre 3. Les résultats issus de cette inversion seront par la suite comparés à ceux obtenus dans le cas précédent où une seule chaîne est utilisée.

La mise en œuvre de cet algorithme repose fortement sur le choix des paramètres suivants

pour garantir une performance optimale.

- la construction de l'échelle de température et le nombre de chaînes.
- le choix du noyau en fonction de la température
- la probabilité de proposition d'un échange

Ces paramètres doivent être fixés par l'utilisateur : En absence de règles théoriques, le lecteur trouvera dans (Romary, 2010) un certain nombre de conseils pratiques et règles empiriques. Dans ce qui suit, nous allons présenter la méthodologie adoptée dans ce travail de thèse pour ajuster ces différents paramètres.

Réglages :

- La construction de l'échelle de température :

La température maximale : Le premier paramètre à régler est la température maximale, ce choix repose essentiellement sur des règles heuristiques similaires aux méthodes d'initialisation de température pour le recuit simulé. Nous adoptons ici, une méthode inspirée de celle présentée dans (Brus and Heuvelink, 2007).

Nous lançons dans un premier temps avec un échantillonneur indépendant, un calcul où toutes les transitions sont acceptées. Nous obtenons ainsi un échantillon contenant un certain nombre d'états de la chaîne. Soit χ le taux d'acceptation de la chaîne à T_{max} . Le calcul de la température maximale se fait en utilisant la formule suivante :

$$T = \frac{\Delta E^+}{\ln(\chi)} \quad (96)$$

où ΔE^+ correspond à la moyenne des écarts négatifs d'énergies :

$$\Delta E^+ = \frac{\sum_{i \in I^+} (E_{i+1} - E_i)}{\text{card}(I^+)} \quad (97)$$

avec $I^+ = \{i, E_{i+1} \leq E_i\}$, E_i correspond à l'énergie de l'état i dans l'échantillon obtenu précédemment.

En prenant un taux d'acceptation de 0.6 et en effectuant 1000 réalisations, la température maximale retenue est finalement de $T_{max} = 7400$.

Le nombre de chaînes : Le nombre de chaînes résulte d'un compromis : celui-ci doit être suffisamment élevé pour assurer les échanges entre les chaînes, mais il ne doit pas être trop grand pour ne pas ralentir la descente de l'information des chaînes à haute température vers

basses.

Nous avons utilisé 5 chaînes avec une répartition géométrique des températures.

$$T_k = T_0 \cdot \left(\frac{T_4}{T_0}\right)^{\frac{k}{4}}$$

$$T_0 = 1 \leq T_1 = 9.27487 \leq T_2 = 86.0233 \leq T_3 = 797.855 \leq T_4 = 7400$$

- Noyaux de transition :

L'efficacité de l'algorithme dépend également du type d'échantillonneur utilisé pour chaque chaîne lors des transitions MH classiques :

Aux plus hautes températures, la distribution cible est aplatie sous l'effet de la température et par conséquent facile à parcourir. Pour assurer une exploration large de l'espace d'état, nous utilisons un échantillonneur global du type échantillonneur indépendant IMH. Concernant les chaînes à basse température, nous utilisons un échantillonneur local de type marche aléatoire. Ce comportement permet d'échantillonner plus précisément les régions pertinentes qui ont été découvertes par les chaînes à haute température lors des itérations précédentes. Afin d'améliorer le mélange de chaque chaîne, nous choisissons un échantillonneur de type marche aléatoire à accroissement symétrique dont le pas dépend de la valeur de la température³.

- Probabilité d'échange :

Le dernier point à relever est la probabilité d'échange. Ce choix est un compromis, cette valeur ne devant pas être trop petite pour que les échanges puissent avoir lieu. Elle ne doit pas être trop élevée pour éviter de renvoyer souvent la chaîne à des états déjà proposés dans le passé. Nous choisissons ici une valeur de 25%.

Résultats :

La Figure 38 montre les principaux résultats de l'inversion, ces résultats ont été obtenus après 10000 itérations soit le même temps de calcul que pour le cas d'une seule chaîne utilisée plus haut.

Le modèle moyen, en vert sur la Figure 38, a un comportement identique au vrai modèle, il apparaît comme une version lisse de ce dernier. Les incertitudes sont en cohérence avec la géométrie d'acquisition : les intervalles sont étroits dans les régions bien contraintes tandis qu'ils deviennent plus larges en profondeur, dans les zones mal contraintes où il n'y a pas de capteurs.

³ Le pas de la marche aléatoire est une fonction croissante de la température.

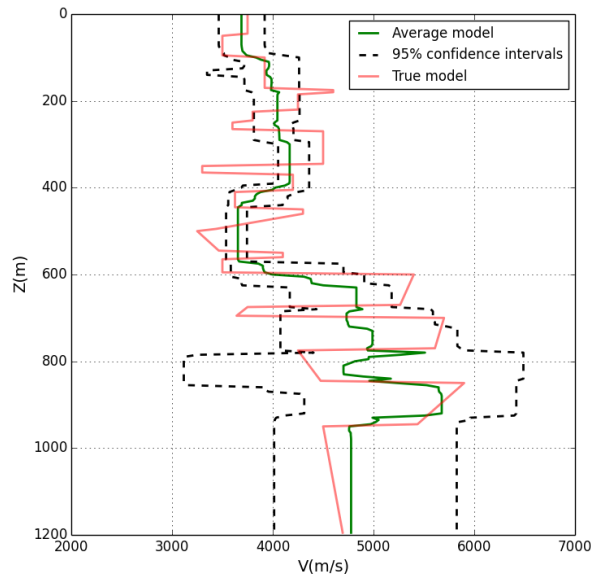


Figure 38: Résultats pour 10000 itérations. Vert: modèle moyen, rouge: vrai modèle, noir: intervalle de confiance à 95%

La Figure 39a montre les énergies des chaînes en parallèle en fonction des itérations. La figure montre également des zones de chevauchement assez importantes entre chaque couple de températures adjacentes. La présence de ces zones est indispensable pour assurer l'échange d'informations. Par ailleurs, nous remarquons que ces échanges ont permis à la chaîne d'intérêt T_0 d'atteindre son régime stationnaire très rapidement (Figure 39b).

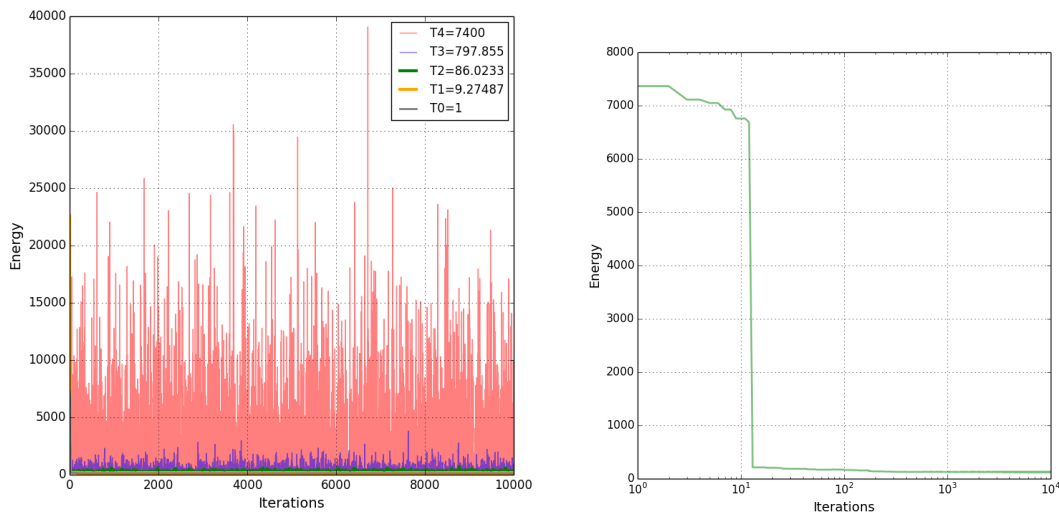


Figure 39: a) Énergies des 5 chaînes b) Énergie pour la chaîne T_0

Comparaison :

Les échanges permettent d'améliorer le mélange des chaînes et par conséquent, les résultats. En effet, à un temps de calcul égal, les performances de la chaîne d'intérêt sont nettement

meilleures dans le cas des chaînes en interaction. Les échanges d'états ont permis à la chaîne d'accéder à des états qu'elle n'a pas pu atteindre quand il n'y avait pas d'échange, comme en témoignent les histogrammes de la vitesse dans la dernière couche dans la Figure 40a. Cette figure illustre bien l'intérêt majeur de l'utilisation des chaînes en interaction qui est de pouvoir échapper à l'attraction d'un seul mode. Il convient également de noter que les échanges ont permis de réduire la durée du Burn-in. La chaîne d'intérêt T_0 atteint son régime stationnaire beaucoup plus rapidement que dans le cas d'une seule chaîne (Figure 40b). En conclusion, le modèle moyen est beaucoup plus proche du vrai modèle que pour une seule chaîne, en particulier pour $z > 700m$. Dans cette zone, les incertitudes paraissent plus fiables également. Ainsi, la couche à faible vitesse à 800m et le gradient de vitesse pour les profondeurs supérieures à 1100m sont inclus dans les intervalles de confiance tandis qu'ils ne l'étaient pas pour une seule chaîne.

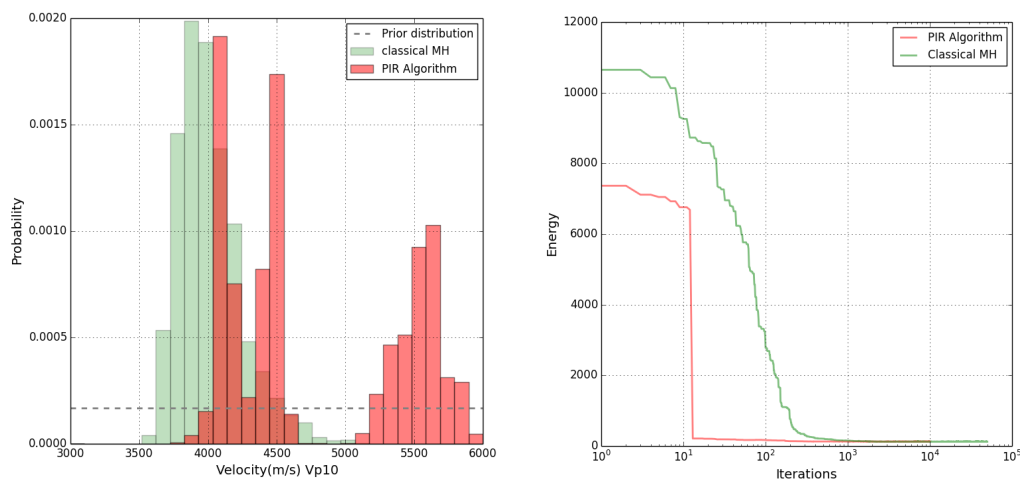


Figure 40: a) Histogrammes de la vitesse dans la dernière couche. b) Les énergies pour le cas classique et pour les chaînes en interaction

Les résidus des temps pour les chaînes en interaction sont affichés dans la Figure 41. Pour la majorité des couples source-récepteur la valeur du résidu est inférieure à $\sigma_d = 0.005s$. Le modèle moyen obtenu s'ajuste suffisamment bien aux données.

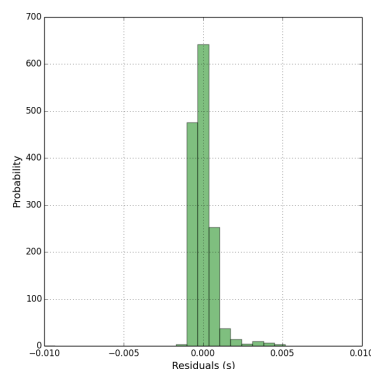


Figure 41: Résidus des temps de première arrivée pour tous les couples source-récepteur

2 Données réelles:

Nous nous intéressons maintenant à un jeu de données réelles acquises dans le contexte de la fracturation hydraulique. Il s'agit des mêmes données que celles présentées dans (Gesret et al., 2015; Bottero et al., 2016). La micro-sismicité induite représente un moyen de suivi des champs en exploitation dans les réservoirs pétroliers. Elle permet de caractériser la géométrie des fractures et de localiser les ruptures. Il importe donc d'avoir des localisations qui sont précises, cela suppose une bonne connaissance du modèle de vitesse ainsi que des incertitudes associées à ce modèle de vitesse. En effet, l'essentiel de la contribution aux erreurs de localisation est attribué à une mauvaise prise en compte des incertitudes associées au modèle de vitesse. Dans le contexte de la micro-sismicité induite, le modèle à couches est souvent utilisé. Classiquement, ce modèle est tout d'abord construit à partir des logs sismiques (Figure 42), il est ensuite calibré grâce aux tirs de calibration (Eisner et al, 2009 ; Maxwell, 2010). (Gesret et al., 2015) ont montré qu'il était possible de prendre en compte l'incertitude du modèle de vitesse en la propageant aux incertitudes de localisation des événements sismiques. A cette fin, les auteurs ont appliqué une tomographie bayésienne au jeu de données (Figure 33, Figure 43) en inversant pour un modèle à nombre de couches et bruit sur les données fixés. Dans notre travail, afin de mieux estimer la vitesse des ondes tout en quantifiant de manière fiable les incertitudes associées à cette estimation, nous développons une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique. Cette approche vise à traiter efficacement la complexité du modèle à couches en limitant l'intervention de l'utilisateur ainsi le bruit sur les données sera considéré comme une inconnue du problème.

La Figure 42 montre le log sismique des ondes P acquis dans la zone d'étude:

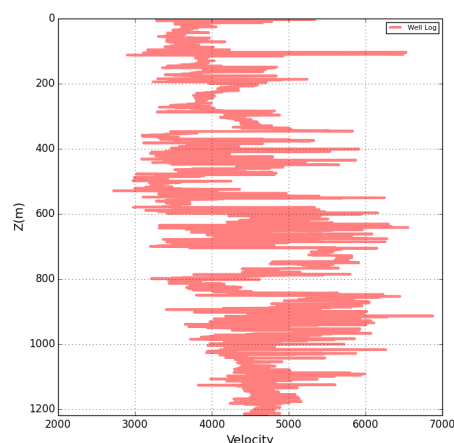


Figure 42: Log sismique des ondes P

Les temps de premières arrivées :

La géométrie d'acquisition est similaire à celle utilisée dans l'exemple synthétique (Figure 33). La Figure 43 montre la carte des temps de premières arrivées des ondes P obtenus pour toutes les sources et tous les récepteurs disponibles. Les zones blanches correspondent aux temps de première arrivée qui n'ont pu être pointés.

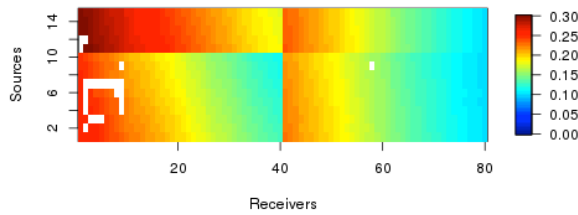


Figure 43: carte des temps de premières arrivées des ondes P

2.1.1 Résultats avec bruit fixé :

Le choix du nombre de couches est un élément important dans l'inversion. En effet, l'utilisation d'un nombre très élevé de couches peut introduire une complexité superflue dans la description du problème. Il peut également induire une surestimation des incertitudes. Dans le cas contraire, un nombre de couches faible est incapable de prendre en compte la complexité du problème et peut induire une sous estimation des incertitudes. Dans cette partie, à la différence du cas traité précédemment, nous ne fixons pas le nombre de couches. Nous augmentons ainsi le degré de liberté de l'algorithme en intégrant le nombre de couches aux inconnues du problème.

Le modèle de vitesse s'écrit alors sous la forme générale :

$$V_p(u) = v_{p_1} 1_{[0, z_1]}(z_u) + \sum_{i=2}^K v_{p_i} 1_{[z_{i-1}, z_i]}(z_u) \quad (98)$$

$1_{[a,b]}$ correspond à l'indicatrice de l'intervalle $[a, b]$.

Posons $Z = (z_i)_{1 \leq i \leq K}$ et $V_p = (v_{p_i})_{1 \leq i \leq K}$, K étant le nombre de couches.

Le modèle de vitesse est entièrement décrit par le nombre des couches K , la profondeur z_i de chaque couche et par son module de vitesse des ondes P v_{p_i} .

Le vecteur des paramètres que l'on cherche à estimer conditionnellement aux données s'écrit comme suit :

$$m = (Z, V_p, K)$$

La loi a priori :

Nous faisons l'hypothèse a priori que les valeurs des paramètres et leur nombre sont conditionnellement indépendants.

$$P(m) = P(m|k)P(k) \quad (99)$$

où $P(k)$ est la loi a priori sur le nombre de couches k .

- Vitesse des ondes sismiques P :

Nous utilisons également une loi uniforme pour les valeurs de vitesse des ondes sismique $P(v_{p_i})_{1 \leq i \leq n}$.

$$P(v_{p_i}|k) = \frac{1}{\Delta V} 1_{[v_{min}, v_{max}]}(v_{p_i}), \quad (100)$$

où $\Delta V = v_{max} - v_{min}$; $v_{max}, v_{min} \in \mathbb{R}^+$

Dans notre application, $v_{min} = 2000 \text{ m/s}$ et $v_{max} = 7000 \text{ m/s}$.

- Position des interfaces :

Nous considérons une loi uniforme entre $z_{min} = 0$ et $z_{max} = 1100\text{m}$ en imposant une épaisseur de couche minimale : $h_{min} = 20\text{m}$. Respecter cette condition nous mène à définir $\Delta Z_i = z_{max} - z_{min} - 2ih_{min}$ qui représente l'espace encore valable pour placer l'interface i quand $(i - 1)$ interfaces sont déjà placées. La loi a priori de z_i est $P(z_i|n) = \frac{1}{\Delta Z_i}$.

Sachant que il existe $(k - 1)!$ manières d'ordonner les interfaces, la loi a priori sur le vecteur des positions des interfaces peut finalement s'écrire sous la forme suivante :

$$P(Z|k) = \prod_{i=1}^{k-1} \frac{(k-1)!}{\Delta Z_i} \quad (101)$$

- Nombre de couches :

Nous choisissons d'utiliser une loi uniforme entre $k_{min} = 2$ et $k_{max} = 30$.

$$P(k) = \begin{cases} \frac{1}{(k_{max} - k_{min})} & \text{pour } k_{min} \leq k \leq k_{max} \\ 0 & \text{sinon} \end{cases} \quad (102)$$

La valeur k_{max} est choisie pour garantir qu'à chaque fois qu'un mouvement d'ajout est accepté, que l'on puisse placer une interface en respectant $h_{min} = 20$, ceci est toujours possible si $\Delta Z_{k_{max}} \geq 20m$.

Finalement, $P(m|k)$ peut être exprimé comme le produit des lois a priori de chaque composante:

$$P(m|k) = P(Z|k)P(V|k) \quad (103)$$

Inférence de la loi a posteriori :

La paramétrisation utilisée dans cette partie considère un nombre aléatoire de paramètres. Pour inférer la distribution a posteriori, nous utilisons l'algorithme de Metropolis Hastings à saut réversible (Reversible Jump) (Green, 1995) présenté dans le chapitre 2. Nous utilisons ici un schéma où trois perturbations sont possibles : l'ajout d'une couche, la suppression d'une couche, et le remplacement d'une couche par une autre. La naissance, la mort ou le remplacement d'une couche s'effectue selon les probabilités $j_b(m) = \frac{1}{4}$, $j_d(m) = \frac{1}{4}$ et $j_m(m) = \frac{1}{2}$ respectivement. Ces valeurs assurent le bon mélange de la chaîne et permettent d'avoir un taux d'acceptation moyen de l'ordre de 24%.

La naissance : La nouvelle couche (z^*, v^*) est simulée indépendamment selon la loi $q(z^*, v^*) = q(z^*) \cdot q(v^*) = \frac{1}{\Delta Z_n} \cdot \frac{1}{\Delta V} \mathbf{1}_{[v_{min}, v_{max}]}(v^*)$.

La mort : La couche d'indice i à supprimer est sélectionnée arbitrairement dans la configuration $q(i) = \frac{1}{n-1}$

La modification d'une couche: La modification se fait en perturbant soit la profondeur soit la vitesse. La perturbation dans chacun des deux cas s'effectue selon une marche aléatoire.

Résultats :

La Figure 44 présente les principaux résultats d'inversion. Ces résultats ont été obtenus pour un nombre d'itérations $N=500000$ itérations. L'inversion a duré 3 jours sur un ordinateur de bureau. Le modèle moyen en vert est en cohérence avec les logs sismiques. Les incertitudes représentées par les courbes en pointillé sont calculées pour les différents états générés par la chaîne de Markov. Elles sont en adéquation avec la géométrie d'acquisition présentée dans la Figure 33: les intervalles sont relativement étroits dans les régions bien contraintes tandis qu'ils deviennent plus larges en profondeur, dans les zones mal contraintes où l'on n'a pas de capteurs.

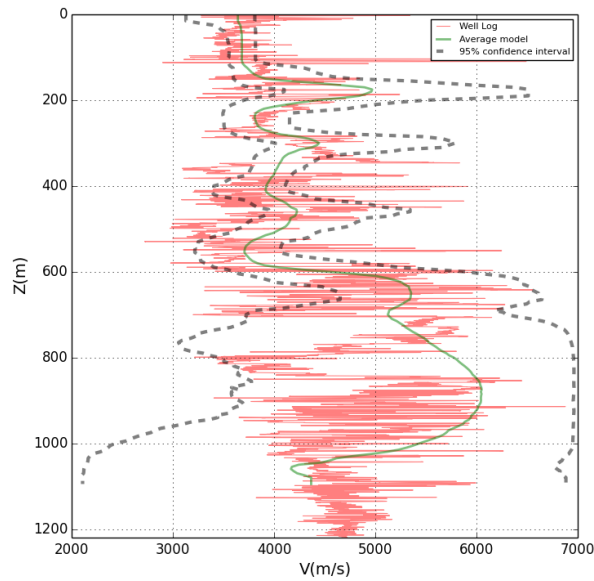


Figure 44: Résultats pour 500000 itérations. Vert: modèle moyen, rouge: vrai modèle, gris: intervalle de confiance à 95%

La Figure 45a suivante montre le nombre de couches a posteriori. Cet histogramme montre qu'il y a une majorité de modèles de vitesse qui ont un nombre faible de couches ne dépassant pas les 15 couches. Bien que la loi a priori soit uniforme sur un intervalle large ici [2,30], rares sont les modèles qui ont un nombre de couches supérieur à 25. Ce résultat met en lumière le principe selon lequel l'approche bayésienne favorise naturellement les modèles parcimonieux, sous la condition qu'ils permettent d'expliquer suffisamment bien les données.

Nous montrons également sous forme d'histogramme, les résidus des temps entre les données observées et les temps calculés en utilisant le modèle moyen (Figure 45b). La valeur absolue des résidus est inférieure à la valeur de l'écart-type du bruit sur les données $\sigma_d = 0.0025s$, sauf pour quelques couples de source-récepteurs. Au vu de tous les résultats montrés dans cette partie, nous pouvons affirmer que le modèle moyen explique suffisamment bien les données observées.

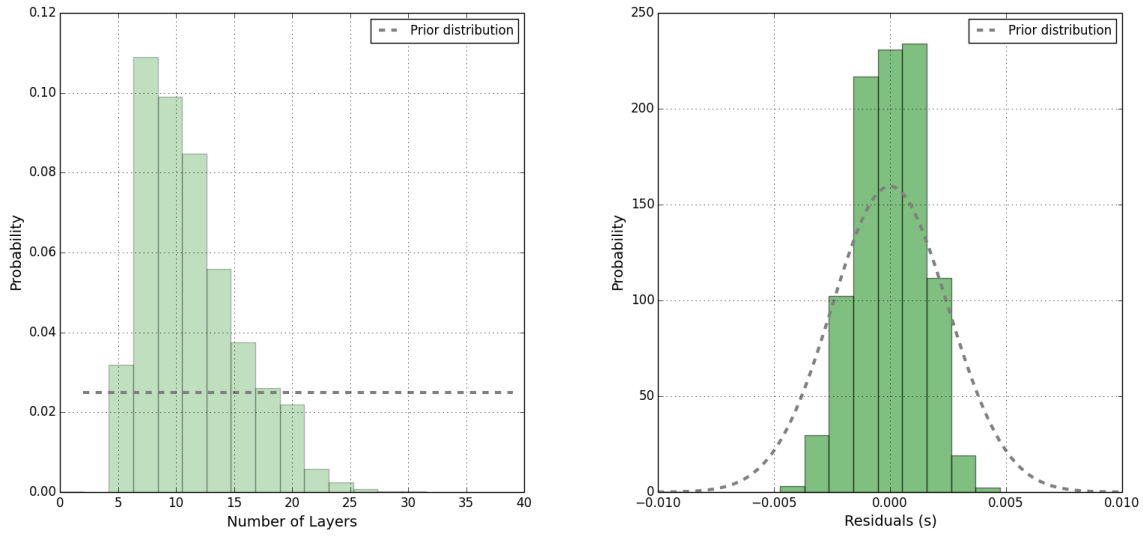


Figure 45 : a) Histogramme du nombre de couches a posteriori b) Histogramme des résidus.

2.1.2 Résultats avec bruit inconnu :

La complexité de la solution dépend également de l'écart-type des données σ_d ((Agostenni et al 2010) et (Bodin et al, 2012)). En pratique, dans la plupart des cas, sa valeur est difficile à déterminer. Il est alors judicieux de le considérer comme un paramètre inconnu (hyper-paramètre) du problème inverse. Le modèle bayésien hiérarchique (Raudenbush and Bryk, 2002; Goldstein, 2011) introduit en géophysique par (Malinverno and Briggs, 2004; Malinverno and Parker, 2006) permet de traiter ce problème.

Loi a priori :

Nous utilisons ici une méthodologie similaire à celle présentée précédemment pour le cas où l'écart-types σ_d est fixé et nous assumons que la loi a priori sur l'écart-types σ_d est uniforme.

$$P(\sigma_d) = 1_{[\sigma_{d_{min}}, \sigma_{d_{max}}]}(\sigma_d) \quad (104)$$

avec $\sigma_{d_{min}} = 0.0005s$ et $\sigma_{d_{max}} = 0.01s$. Les intervalles sont larges pour garantir le caractère non-informatif de la loi a priori.

Inférence loi a posteriori :

La loi a posteriori des paramètres est donnée par :

$$P(m, \sigma_d | d^{obs}) \propto \frac{1}{(\sigma_d)^{N_d}} \cdot \exp\left(-\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2\right) \cdot P(m) \cdot P(\sigma_d) \quad (105)$$

Le principe général de l'algorithme est le suivant :

- avec une probabilité $\beta = 0.25$, on perturbe la valeur de σ_d selon une marche aléatoire.
- avec une probabilité $1 - \beta$, on effectue une itération RJ-MCMC de la même manière que dans le cas précédant pour un bruit fixe. La valeur de β est choisie pour assurer le bon fonctionnement de l'algorithme.

Résultats :

La Figure 46 présente les principaux résultats d'inversion. Ces résultats ont été obtenus pour un nombre d'itérations $N=500000$ itérations. L'inversion a duré 3 jours sur un ordinateur de bureau. Les courbes vertes et rouges indiquent respectivement la moyenne a posteriori et le log sismique. Les lignes grises en pointillé correspondent quant à elles aux intervalles de confiance à 95%. La Figure 46 montre que le modèle moyen est en cohérence avec le log sismique. Les incertitudes représentées par les courbes en pointillé sont calculées pour les différents états générés par la chaîne de Markov. Elles sont en adéquation avec la géométrie d'acquisition présentée dans la Figure 33: les intervalles sont étroits dans les régions bien contraintes tandis qu'ils deviennent plus larges en profondeur, dans les zones mal contraintes où il n'y a pas de capteurs. Nous remarquons que les incertitudes dans la région entre 20m et 700m sont beaucoup plus faibles que dans le cas où sigma est fixé à 0.0025s, les régions bien contraintes sont mieux distinguées. L'algorithme donne de bons résultats sans avoir à fixer la valeur de l'écart-type sur les données à l'avance et en utilisant un intervalle large pour la loi uniforme sur ce paramètre.

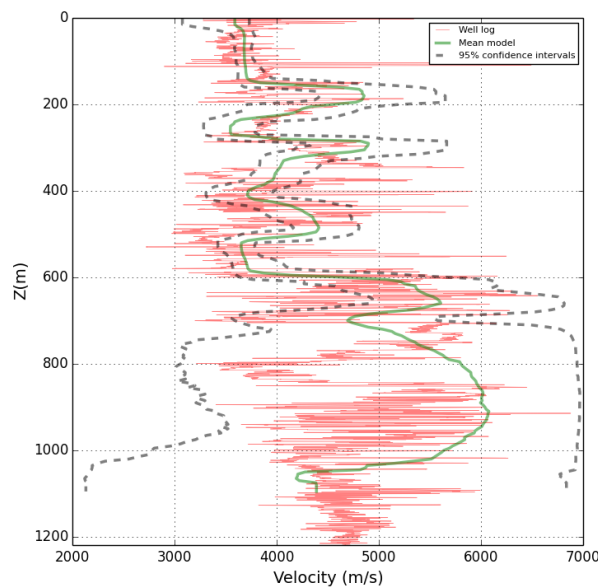


Figure 46: Résultats pour 500000 itérations. Vert: modèle moyen, rouge: vrai modèle, gris: intervalle de confiance à 95%

Nous affichons ci-dessous l’histogramme du bruit σ_d , $P(\sigma_d|d^{obs})$. Nous pouvons constater que la valeur de σ_d se stabilise autour de 0.0013s, la valeur que nous avons utilisé précédemment, à savoir 0.0025s, était donc surestimée.

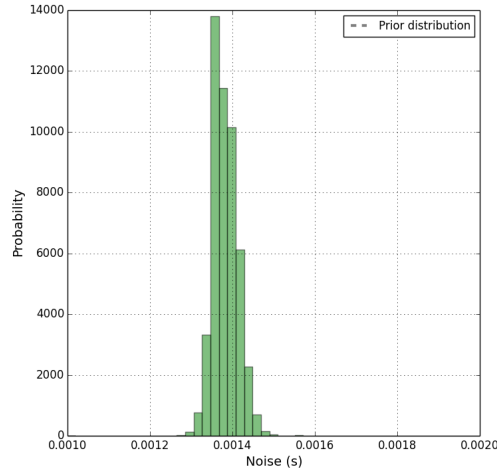


Figure 47: Histogramme de σ_d .

La Figure 48 présente le nombre de couches a posteriori. La majorité des modèles de vitesse ont un nombre de couches inférieur à 24. Nous constatons également que la proportion des modèles ayant un nombre de couches supérieur à 15 est plus élevée par rapport au cas précédent où le bruit est fixé en témoigne la Figure 48b. Dans le cas précédent, la valeur de sigma était surestimée, l’algorithme tendait à s’ajuster à des modèles qui étaient trop simples par rapport à la vraie complexité du modèle.

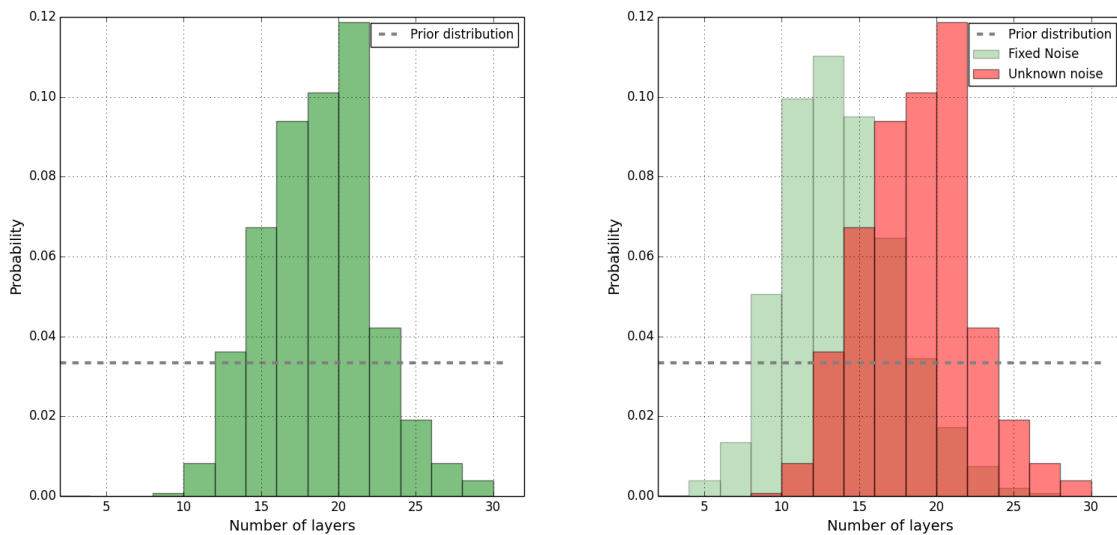


Figure 48: a) Nombre de couches pour le modèle hiérarchique, b) Histogramme des couches pour le modèle hiérarchique (rouge) et pour σ_d connu (vert).

La Figure 49 présente l'histogramme des différences entre les temps observés et les temps d'arrivée calculés pour le modèle moyen a posteriori. Nous constatons que les résidus sont distribués selon une gaussienne d'écart-type à $\sigma_d = 0.0013s$. Ceci est donc cohérent avec le résultat obtenu pour l'écart-type $\sigma_d = 0.0013s$.

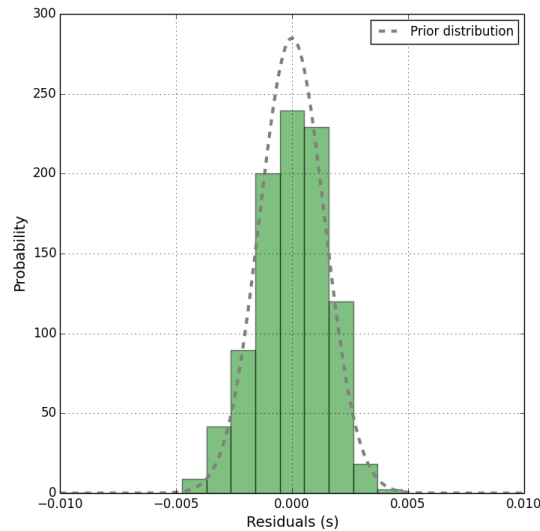


Figure 49: Histogramme des résidus

3 Conclusion :

Ce chapitre a été consacré à la paramétrisation à couches horizontales. Nous avons tout d'abord illustré les résultats de la tomographie sismique avec un nombre de couches fixe pour un modèle synthétique réaliste. Nous avons pu également évaluer les performances de l'algorithme Population Importance Re-sampling (PIR) et nous avons montré que les échanges permettaient d'améliorer le mélange des chaînes. À temps de calcul égal, les performances de la chaîne sont nettement meilleures dans le cas des chaînes en interaction. Nous nous sommes également intéressés à un jeu de données réelles dans le cadre de la micro-sismicité induite. Nous avons développé dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique en intégrant le nombre de couches et la valeur de l'écart-type du bruit sur les données aux inconnues du problème inverse. Les résultats obtenus montrent que cette approche permet de traiter efficacement la complexité du modèle à couches. Les résultats de l'algorithme PIR dans le cas de la dimension fixe étant encourageants, ils ouvrent une nouvelle perspective pour l'utilisation de cet algorithme dans le cas trans-dimensionnel.

Chapitre 6.

Compressed Sensing appliqué à la tomographie

Dans ce chapitre, nous présentons un modèle de tomographie bayésienne qui repose sur la théorie du compressed sensing (Candès et al., 2006; Candes et al., 2006; Donoho, 2006). En pratique, la plupart des signaux couramment traités (son, image, vidéo) sont compressibles dans certaines familles de fonctions mathématiques, comme les bases de Fourier ou les bases d'ondelettes. En mesurant un signal de taille N dans l'espace physique puis en le représentant dans une base adaptée, on peut ainsi fortement le compresser, c'est à dire l'exprimer en fonction d'un nombre K de fonctions de base vérifiant $K \ll N$. Cette propriété, largement empirique, est à la base des techniques de compression utilisées en traitement d'image. Dans le cadre de la tomographie bayésienne, il semble légitime d'attendre des champs de vitesse qu'ils se prêtent bien à ces méthodes de compression.

Le compressed sensing, ou acquisition comprimée en français, est une théorie qui permet de définir des acquisitions linéaires de signaux mathématiques sous une forme directement comprimée. Cette théorie permet donc, dans le cadre d'un problème de tomographie sismique, de proposer une représentation parcimonieuse du champ de vitesse entièrement non-adaptative: la seule connaissance a priori introduite dans le choix de la représentation est qu'il existe une base de fonctions qui permet de représenter le champ de vitesse avec un nombre restreint de coefficients. Dans ce chapitre, nous supposons ainsi que le champ de vitesse considéré est compressible dans une base de cosinus discrets.

Dans ce chapitre, nous commençons par exposer les bases de la théorie du compressed sensing. Un exposé complet de cette théorie dépassant très largement le cadre de cette thèse, nous nous contenterons d'énoncer les principaux résultats et de donner quelques clefs de compréhension de la théorie. Nous référons le lecteur intéressé par de plus amples détails à la référence (Davenport et al., 2011) dont s'inspire largement la première partie du présent chapitre. Nous présentons dans un second temps un algorithme d'inversion Bayésienne basé sur le compressed sensing, dérivé des algorithmes présentés dans les chapitres qui précèdent. Enfin, nous concluons en présentant les résultats de l'algorithme sur un cas test, et en évoquant plusieurs pistes de recherches futures qu'il nous paraît intéressant d'explorer.

1 Compressed sensing :

Le compressed sensing vise à développer des méthodes de mesures linéaires et non-adaptatives permettant d'acquérir un signal de manière parcimonieuse. Commençons par

préciser que les signaux auxquels nous nous intéressons sont des vecteurs de taille n . Nous nous placerons donc dans l'espace vectoriel \mathbb{R}^n , que nous supposerons muni de sa structure d'espace hilbertien induit par le produit scalaire usuel.

Définition 1.1 Un vecteur x de \mathbb{R}^n est dit k -parcimonieux ou k -sparse si

$$\|x\|_0 \leq k \quad (106)$$

où $\|x\|_0$ correspond au nombre de coefficients non nuls de x .

Dans la suite, nous noterons Σ_k l'ensemble des vecteurs k -sparses de \mathbb{R}^n . Il est intéressant de noter que Σ_k n'est pas un espace vectoriel. Il suffit pour cela de remarquer que la somme de deux vecteurs de Σ_k ayant des coefficients non nuls à des indices différents deux à deux résulte en un vecteur ayant nécessairement $2k$ coefficients non nuls. La propriété de stabilité par addition n'est donc pas vérifiée par Σ_k . On notera néanmoins que la somme de deux vecteurs de Σ_k appartient nécessairement à l'espace Σ_{2k} .

Supposons à ce stade que l'on souhaite définir p mesures linéaires et non-adaptatives qui permettent de reconstituer de façon univoque n'importe quel vecteur de $\Sigma_k \subset \mathbb{R}^n$. D'un point de vue mathématique, compte-tenu de l'aspect linéaire des mesures, cette opération peut s'exprimer sous la forme

$$y = Ax \quad (107)$$

où $x \in \Sigma_k \subset \mathbb{R}^n$, $y \in \mathbb{R}^p$, et A est une matrice de taille p par n , que nous appellerons matrice de mesure. Deux questions se posent naturellement à ce stade:

- Quelles conditions la matrice de mesure doit-elle vérifier afin qu'il soit possible de retrouver n'importe quel vecteur de x de Σ_k de façon univoque?
- En pratique, comment peut-on définir un algorithme de complexité polynomiale qui permette de retrouver x ?

1.1 Matrice de mesure

Intéressons-nous pour commencer à la première de ces questions. Dans la suite de cette partie, nous commencerons par présenter des résultats théoriques garantissant que la matrice de mesure permet de retrouver le vecteur x initial. Nous considérerons également une généralisation au cas, plus réaliste en pratique, de signaux compressibles à défaut d'être exactement sparse, avant d'examiner la question des problèmes de stabilité de la mesure par rapport au bruit. Nous nous contenterons d'énoncer et de discuter les principaux résultats sans nécessairement donner de démonstrations, l'idée étant de d'esquisser pour le lecteur les contours de la théorie, tout en le renvoyant à (Davenport et al., 2011) pour une description plus en profondeur.

Une condition essentielle pour qu'il soit possible de retrouver $x \in \Sigma_k$ à l'issue de la mesure est qu'il y ait unicité de la mesure associée. Ainsi, pour tout couple de vecteurs x_1, x_2 de vecteurs de Σ_k , on doit vérifier

$$A(x_1 - x_2) \neq 0 \quad (108)$$

On a vu plus haut que la somme de deux vecteurs de $x \in \Sigma_k$ était nécessairement un vecteur de l'espace Σ_{2k} . Par conséquent, une condition équivalente à (108) est que l'intersection du noyau $\ker(A)$ de la matrice de mesure et de l'espace Σ_{2k} , des vecteurs $2k$ -sparses soit réduit au vecteur nul. On peut démontrer assez aisément (voir (Davenport et al., 2011)) que cette condition a une implication immédiate quant à la structure de la matrice A .

Théorème 1.1 Soit y un vecteur de \mathbb{R}^p . Il existe au plus un vecteur x de Σ_k tel que $y = Ax$ si, et seulement si $\text{spark}(A) > 2k$, où $\text{spark}(A)$ correspond au plus petit nombre de colonnes de A qui forment une famille liée dans \mathbb{R}^p .

En présence de vecteurs exactement k -parcimonieux, le théorème 1.1 donne une condition nécessaire et suffisante sur la matrice de mesure pour qu'il soit, au moins en théorie, possible de retrouver n'importe quel vecteur x de Σ_k à l'issue de la mesure. On notera qu'il est nécessaire que la condition $p > 2k$ portant sur la dimension de l'espace de mesure (i.e sur le nombre de mesures à réaliser) soit vérifiée.

En pratique, les hypothèses sous-jacentes au théorème 1.1 sont peu rencontrées en pratique. Les vecteurs que nous chercherons à mesurer sont rarement k -parcimonieux. Ils restent cependant suffisamment «proches» de vecteurs k -parcimonieux pour qu'ils soient correctement approchés par ces derniers. Afin de formaliser cette notion de «proximité», nous introduisons la définition qui suit:

Définition 1.2 Soit $\epsilon > 0$. Un vecteur x de \mathbb{R}^n est dit k -compressible si

$$\min_{\hat{x} \in \Sigma_k} \|x - \hat{x}\|_2 < \epsilon \quad (109)$$

Il est clair que les conditions édictées par le théorème 1.1 ne sont pas suffisantes pour permettre de reconstruire un vecteur x k -compressible à partir du résultat $y = Ax$ de la mesure. Le théorème 1.1 se généralise cependant au cas des vecteurs compressibles si on impose à la matrice A de vérifier la *null space property* (NSP).

Définition 1.3 La matrice A vérifie la NSP d'ordre k si il existe une constante $C > 0$ telle que, pour tout vecteur $h \in \text{Ker}(A)$ et pour tout ensemble discret d'indices Λ de cardinal fini $|\Lambda| \leq k$. On note Λ^c le complémentaire de Λ i.e $\Lambda^c = \{1, 2, \dots, n\} \setminus \Lambda$

$$\|h_\Lambda\|_2 \leq C \frac{\|h_{\Lambda^c}\|_1}{\sqrt{k}} \quad (110)$$

Dans cette expression, la notation h_Λ est utilisée pour désigner le vecteur h dont on n'a gardé non nuls que les coefficients dont les indices correspondent aux éléments de Λ .

La NSP doit s'interpréter comme une généralisation de la condition de non-intersection du noyau de A avec l'espace Σ_k au cas de vecteurs non plus sparses mais uniquement compressibles. Remarquons à cet égard que si h est un vecteur exactement k -sparse de $\text{Ker}(A)$, alors $\|h_{\Lambda^c}\|_1 = 0$. Par conséquent, d'après la NSP, le vecteur h est nécessairement nul. La NSP permet bien de retrouver le fait que le seul vecteur k -sparse de $\text{Ker}(A)$ est le vecteur nul. Intuitivement, la NSP impose une condition sur les vecteurs de $\text{Ker}(A)$ qui impose à ces vecteurs de ne pas être concentrés sur un trop petit nombre d'indices.

Intéressons-nous à ce stade aux garanties de reconstruction fournies par la NSP. Soit $\Delta: \mathbb{R}^p \rightarrow \mathbb{R}^n$ un algorithme qui permet de reconstruire un vecteur x k -compressible à partir d'une mesure $y = Ax$. On considère la garantie de reconstruction suivante:

$$\|\Delta(Ax) - x\|_2 \leq C \frac{\min_{\hat{x} \in \Sigma} \|x - \hat{x}\|_2}{\sqrt{k}} \quad (111)$$

Intuitivement, cette relation garantit que la reconstruction permet de retrouver exactement les vecteurs k -sparse à l'issue de la mesure, et que l'erreur de reconstruction pour les vecteurs k -compressible ne dépend que de la compressibilité du vecteur considéré. Le théorème qui suit fait le lien entre la garantie de reconstruction (111) et la NSP.

Théorème 1.2 Si la matrice de mesure A et l'algorithme de reconstruction Δ vérifient la garantie de reconstruction (111), alors A vérifie nécessairement la NSP d'ordre $2k$. La NSP apparaît ainsi comme une condition nécessaire à l'obtention de la garantie de reconstruction. On verra plus loin qu'elle est également une condition suffisante.

Nous avons à ce stade présenté deux conditions, en l'occurrence le *spark* et la NSP, qui garantissent la possibilité de reconstruire tout vecteur compressible à partir d'un nombre restreint de mesures. Ces conditions excluent cependant l'importante question de la robustesse au bruit. En vue de formaliser cette question, nous nous donnons une matrice de mesure $A: \mathbb{R}^n \rightarrow \mathbb{R}^p$ et un algorithme de reconstruction $\Delta: \mathbb{R}^p \rightarrow \mathbb{R}^n$.

Définition 1.4 La paire (A, Δ) est dite C -stable si, $\forall x \in \Sigma_k, \forall e \in \mathbb{R}^n$, on vérifie

$$\|\Delta(Ax + e) - x\|_2 \leq C \|e\|_2 \quad (112)$$

En langage courant, si la mesure est contaminée par un bruit quelconque, l'impact de ce bruit sur la qualité de la reconstruction ne peut pas être arbitrairement large.

La NSP ne fournit pas une condition assez forte pour garantir la stabilité de la reconstruction. Cette condition est cependant fournie par la restricted isometry property (RIP), définie comme suit:

Définition 1.5 Une matrice A vérifie la RIP d'ordre k s'il existe δ_k , $0 \leq \delta_k \leq 1$, tel que $\forall x \in \Sigma_k$:

$$(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2 \quad (113)$$

Intuitivement, si A vérifie la RIP d'ordre $2k$, alors A préserve en quelque sorte la distance entre deux vecteurs de Σ_k . Le théorème qui suit permet de relier la RIP à la NSP.

Théorème 1.3 Si A vérifie la RIP d'ordre $2k$, avec $\delta_{2k} < \sqrt{2} - 1$ alors A vérifie la NSP d'ordre $2k$ avec la constante

$$C = \frac{\sqrt{2}\delta_{2k}}{1 - (1 + \sqrt{2})\delta_{2k}} \quad (114)$$

1.2 RIP et matrices aléatoires

Dans la section précédente, nous avons présenté un certain nombre de critères que la matrice de mesure doit absolument vérifier dans le contexte de l'acquisition comprimée. Cependant, en pratique, il est extrêmement difficile de contrôler que ces critères sont effectivement vérifiés lorsqu'on construit la matrice de mesure A . En effet, pour nous en assurer, il convient d'examiner $\binom{n}{k}$ sous-matrices de A , ce qui fait de cette question un problème combinatoire.

Par chance, il est assez aisé d'obtenir des matrices de mesure en considérant des matrices gaussiennes aléatoires. Nous ne rentrerons pas dans le détail de la théorie et nous contenterons d'énoncer l'important résultat qui suit:

Théorème 1.4 Soit A une matrice de taille p par n . On suppose que les entrées de A sont choisies aléatoirement en suivant une distribution gaussienne. Alors, sous la condition

$$p = O\left(k \log\left(\frac{n}{k}\right)\right) \quad (115)$$

il existe une constante $C > 0$ telle que la matrice A vérifie la RIP d'ordre $2k$ avec une probabilité au moins supérieure à $1 - \exp(-Cp)$.

En pratique, nous nous appuyerons donc sur des matrices aléatoires pour le design des matrices de mesure utilisées.

1.3 Reconstruction :

Dans cette sous-section, nous nous intéressons plus précisément à l'algorithme qui permet de reconstruire $x \in \mathbb{R}^n$ à partir du vecteur de mesures $y \in \mathbb{R}^p$. Rappelons que y s'exprime en fonction de x via le système linéaire :

$$y = Ax \tag{116}$$

où A vérifie la RIP d'ordre $2k$. Comme $p \ll n$, ce système est très largement sous-déterminé, et possède par conséquent une infinité de solutions. Nous avons cependant fait l'hypothèse que le vecteur x est k -parcimonieux. Par conséquent, nous pouvons obtenir une solution unique en sélectionnant la solution du système qui comporte le moins de coefficients non nuls. Du point de vue mathématique, cela revient à résoudre le problème d'optimisation.

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_0 \text{ sous la contrainte } \|Ax - y\|_2 = 0 \tag{117}$$

Cette approche a cependant deux inconvénients. En premier lieu, elle ne tient pas compte du potentiel bruit de mesure. On peut remédier à ce problème en remplaçant la contrainte dans (116) par une contrainte inégalité

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_0 \text{ sous la contrainte } \|Ax - y\|_2 < \epsilon \tag{118}$$

où ϵ est du même ordre de grandeur que le bruit. Le second inconvénient de l'approche est que le problème d'optimisation considéré n'est pas convexe. Il est cependant possible d'obtenir un problème d'optimisation convexe équivalent en remplaçant la semi-norme $\|\cdot\|_0$ par la norme $\|\cdot\|_1$ dans (118). Rappelons que la norme L_1 d'un vecteur de \mathbb{R}^n est définie par

$$\|x\|_1 = \sum_{i=0}^{n-1} |x_i| \tag{119}$$

Le problème d'optimisation considéré devient:

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_1 \text{ sous la contrainte } \|Ax - y\|_2 < \epsilon \tag{120}$$

Il convient de noter que pour des problèmes d'optimisation convexes, la norme L_1 conduit naturellement à des solutions parcimonieuses. Cette propriété découle directement des caractéristiques de la boule unité L_1 . On peut intuitivement le comprendre à partir de l'exemple de la figure qui suit (Figure 50). Le problème auquel nous nous intéressons consiste à déterminer le point x vérifiant $\|Ax - y\|_2 = 0$, de norme L_1 minimale. Supposons que l'équation $\|Ax - y\|_2 = 0$ soit décrite (schématiquement) par la courbe rouge. Dans ce cas, la solution qui nous intéresse s'obtient en faisant "grandir" la boule unité jusqu'à ce qu'elle intersecte la courbe rouge. On voit que du fait de la géométrie de la boule unité de la norme

L_1 , cette intersection va en général se faire sur l'un des angles. Or, les angles correspondent précisément à des vecteurs où seule une des coordonnées est non nulle. À l'inverse, on note que la boule unité associée à la norme L_2 est parfaitement isotrope. Par conséquent, il n'y a pas a priori de point privilégié pour l'intersection.

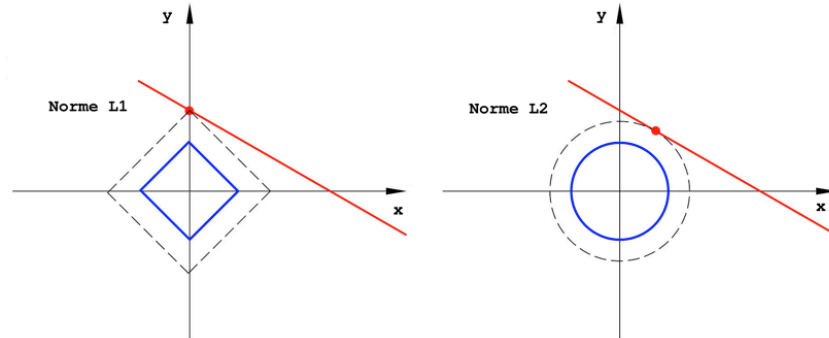


Figure 50: Contrainte L_1 et solution parcimonieuse.

Sous la forme (120), on peut démontrer qu'il est possible de reconstruire le champ de vitesse à condition que la matrice de mesure vérifie la RIP, la reconstruction satisfaisant à la propriété de stabilité exprimée par la définition 1.4.

2 Application à la paramétrisation du champ de vitesse :

Dans la première section de ce chapitre, nous avons brièvement présenté les résultats principaux de la théorie du compressed sensing. Nous présentons dans cette section une application à la paramétrisation des champs de vitesse en tomographie. Dans la suite, nous notons v le champ de vitesse discrétisé sur une grille de taille $N_z \times N_x$. Nous représentons le champ de vitesse sous la forme d'un vecteur de taille N . Nous faisons l'hypothèse que le champ de vitesse est compressible dans une base de cosinus discrets.

La matrice de mesure est obtenue en échantillonnant le champ de vitesse en M points sélectionnés aléatoirement. La mesure prend donc la forme d'une matrice A de taille $M \times N$, où les coefficients sur chaque ligne sont nuls à l'exception du coefficient correspondant au point échantillonné, fixé à 1. Le vecteur de mesures $m \in \mathbb{R}_+^M$ s'exprime donc en fonction du champ de vitesse v sous la forme:

$$m = Av \tag{121}$$

À ce stade, on note que le vecteur v n'est pas sparse, ni même compressible: par hypothèse, la vitesse n'est pas compressible dans l'espace physique mais dans une base de cosinus discrets. La transformée en cosinus discrets est cependant une opération linéaire, qu'on peut donc représenter par une matrice Ψ . Le champ de vitesse s'exprime ainsi aisément en fonction des coefficients de projection c dans une base DCT via la relation

$$v = \Psi^{-1}c \quad (122)$$

La mesure s s'exprime donc en fonction du vecteur compressible c de la manière qui suit :

$$m = A\Psi^{-1}c \quad (123)$$

Afin de retrouver la décomposition c du champ de vitesse dans une base DCT, on est par conséquent amené à résoudre le problème d'optimisation qui suit:

$$\hat{c} = \arg \min_{c \in \mathbb{R}^N} \|c\|_1 \text{ sous la contrainte } \|A\Psi^{-1}c - m\|_2 < \epsilon \quad (124)$$

où la valeur ϵ est fixée en fonction de l'amplitude du bruit. En pratique, on considérera le problème de minimisation équivalent suivant:

$$\hat{c} = \arg \min_{c \in \mathbb{R}^N} \|A\Psi^{-1}c - m\|_2 + \lambda \|c\|_1 \quad (125)$$

connu dans la littérature sous le nom de LASSO. La paramètre λ se déduit (non trivialement) du paramètre ϵ . En pratique, ce paramètre sera fixé de manière empirique. On voit qu'il contrôle le compromis entre la qualité de l'approximation et la parcimonie de la solution. Une valeur élevée de λ tend ainsi à favoriser des solutions plus parcimonieuses.

Une illustration de l'algorithme de reconstruction est présentée ci-dessous. Un champ de vitesse "test" est présenté sur la Figure 51. Ce champ de vitesse est obtenu en décomposant le champs de vitesse benchmark présenté dans le chapitre 4 (Figure 24), sous-échantillonné à une taille 90×19 , sur une base de cosinus discrets, puis en reconstruisant en ne gardant que les $k = 30$ coefficients de plus grande amplitude en valeur absolue. Les résultats de la reconstruction par l'algorithme de compressive sensing décrit ci-dessus sont présentés sur la Figure 51. Pour former le vecteur de mesure, on utilise $M = 180$ valeurs du champ de vitesse mesurées en des points sélectionnés aléatoirement sur le champ de vitesse test selon une loi uniforme. La valeur du paramètre λ qui intervient dans l'expression du problème d'optimisation est $\lambda = 0.1$. Pour les tailles considérées, la reconstruction prend quelques centièmes de seconde sur un ordinateur de bureau, ce qui autorise a priori son intégration dans un algorithme MCMC. Le problème d'optimisation est résolu avec un solveur de la librairie scikit-learn (Pedregosa et al., 2011).

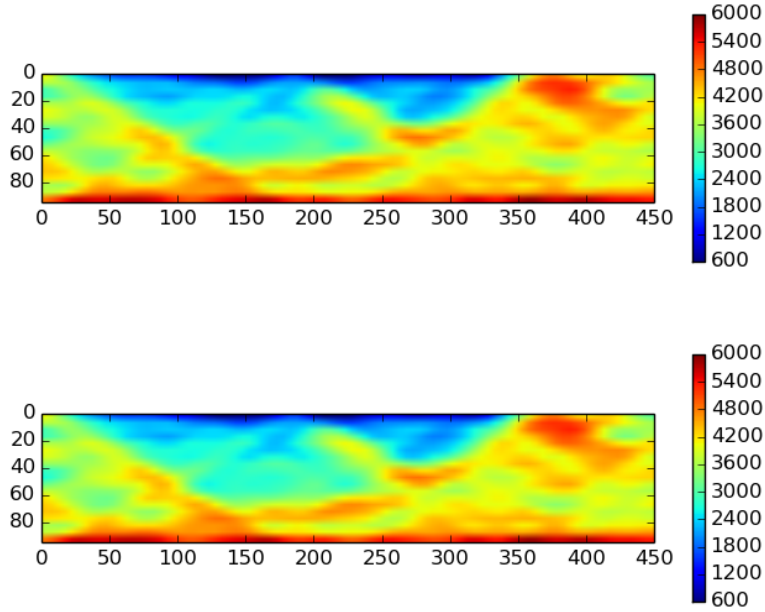


Figure 51: Champ de vitesse test, de taille 95×19 (haut). Reconstruction à partir de $m = 180$ valeurs du champ de vitesse (bas). On note que l’algorithme permet d’obtenir une reconstruction quasi-parfaite du champ de vitesses initial.

2.1 Intégration dans une approche bayésienne

L’exemple de la section précédente montre qu’il est possible de reconstruire un champ de vitesse de taille N à partir d’un nombre $M \ll N$ de mesures de vitesses, sous l’hypothèse que le champ de vitesse s’exprime de façon parcimonieuse dans une base de cosinus discrets. Ce constat peut être utilisé pour proposer une paramétrisation parcimonieuse du champ de vitesse dans le cadre d’un algorithme d’inversion bayésienne.

Dans ce qui suit, nous supposons que le champ de vitesse que nous cherchons à inverser est compressible dans une base de cosinus discrets. En d’autres termes, il existe un champ de vitesse exactement k -sparse qui constitue une bonne approximation du champ de vitesse que l’on cherche à identifier, au sens de la définition 1.2. Notons N la taille du champ de vitesses, et v le vecteur correspondant aux valeurs des vitesses en chaque point. La mesure que l’on effectue va consister à échantillonner $M \ll N$ valeurs de vitesse, en des points sélectionnés aléatoirement selon une loi uniforme. On obtient ainsi une matrice de mesure A de taille M par N , qui possède uniquement un 1 sur chaque ligne, le reste des valeurs étant égales à 0. L’indice correspondant à la valeur 1 correspond à l’un des points du champ de vitesse qui est mesuré. D’un point de vue mathématique, on aboutit donc à un processus de mesure décrit par l’équation :

$$m = Av \tag{126}$$

Ici, le vecteur v représente la valeur de la vitesse en chaque point, qui n'est pas sparse. Si on note c la transformée de v dans une base de cosinus discrets et Ψ^{-1} la matrice décrivant la transformée en DCT inverse, cette équation devient

$$m = A\Psi^{-1}c$$

On rappelle que, connaissant m , il est possible de reconstruire c dans le cadre dessiné par la théorie du compressed sensing avec la technique décrite dans la section précédente.

Le principe de l'inversion bayésienne va être d'échantillonner le vecteur v en essayant de déterminer les coefficients de vitesse aux points de mesure m . On ne cherchera donc pas à déterminer la valeur exacte du champ de vitesse en chaque point, mais uniquement sa valeur en un nombre plus restreint de points. Supposons que le champ de vitesse soit k -parcimonieux dans une base de cosinus discret. Dans ce cas, un résultat théorique sur le compressive sensing indique qu'il faut de l'ordre de $O(k\sqrt{N})$ coefficients pour la reconstruction. On constate donc qu'il est possible de considérablement réduire la dimension du problème considéré. Une fois fixée la matrice de mesure, le vecteur des paramètres que nous cherchons à échantillonner correspond au vecteur de mesures m . Chaque mesure correspond à la vitesse en un point du champ. Nous choisirons donc comme loi a priori des paramètres une loi uniforme entre deux valeurs v_{min} et v_{max} .

Pour échantillonner le vecteur de paramètres, nous utilisons un algorithme MCMC classique de type marche aléatoire. A chaque étape de l'algorithme, le champ de vitesse v est reconstruit à partir du vecteur de mesures m . A chaque étape, la loi de transition que nous proposons consiste à sélectionner aléatoirement un des points de mesure et à perturber la valeur de la vitesse en ce point selon une loi gaussienne de moyenne la valeur actuelle de la vitesse et de variance fixée.

Remarque :

On notera que l'approche proposée est robuste. En effet, le fait que la matrice de mesure vérifie la RIP implique que la reconstruction est stable au sens de la définition 1.4. En d'autres termes, si le vecteur de mesure se rapproche du vecteur de mesure qu'on aurait obtenu avec le champ de vitesses réel, alors le champ reconstruit par l'algorithme se rapproche également du champ de vitesses réel.

2.2 Résultats et discussion :

Nous présentons dans cette section une illustration de l'algorithme décrit précédemment sur deux problèmes de tomographie. Nous commençons par appliquer l'algorithme à un champ de vitesse synthétique naturellement sparse, obtenu à partir du modèle de benchmark présenté dans la partie 4. Le second exemple s'appuie sur le modèle de

benchmark présenté dans le chapitre 4, l'objectif étant ici est de tester la méthodologie sur un exemple complexe.

2.2.1 Inversion d'un champ de vitesse parcimonieux dans une base de cosinus discrets :

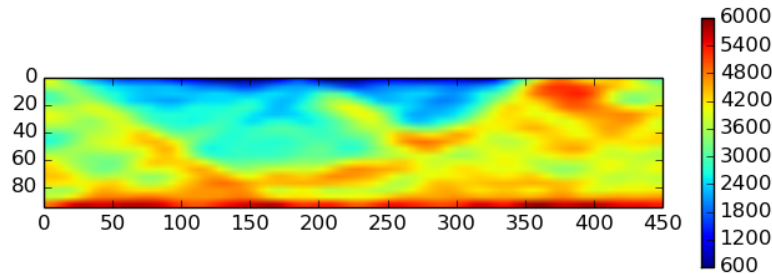


Figure 52: Vrai modèle

Nous nous intéressons dans un premier temps à l'inversion d'un champ de vitesse possédant une représentation parcimonieuse dans une base de cosinus discrets. Ce champ de vitesse est obtenu en décomposant le champ de vitesse du modèle de benchmark présenté dans la partie 4, sous-échantillonné à une taille 90×19 , sur une base de cosinus discrets, puis en reconstruisant en ne gardant que les $k = 30$ coefficients de plus grande amplitude en valeur absolue. Nous avons vu précédemment qu'en utilisant des techniques d'acquisition comprimée, il était possible de reconstruire ce champ avec $M = 180$ mesures.

Sur une grille de taille (90×19) , nous simulons une acquisition sismique en réfraction avec 23 sources et 44 récepteurs spatialement répartis à la surface du modèle. Les données observées sont calculées en utilisant le solveur Eikonal développé par (Noble et al., 2014). Le nombre total de données observées est de 1012 et l'écart-type du bruit sur les données $\sigma_d = 0.005s$.

Loi a priori :

Nous utilisons un nombre de points $M = 200$ points tirés aléatoirement dans le domaine (Figure 53).

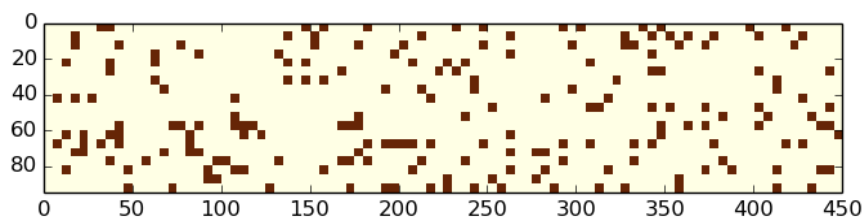


Figure 53: Position des points sélectionnés (en rouge) pour reconstruire le champ de vitesse

Nous prenons $v_{min} = 600m/s$ et $v_{max} = 6000m/s$. L'intervalle de la loi a priori est large assurant le caractère non-informatif de la loi a priori.

Résultats

Les résultats de l'inversion sont présentés dans la Figure 54. Celle-ci montre la carte de la vitesse moyenne a posteriori et les incertitudes associées à cette estimation (écart type). Nous utilisons le logarithme pour afficher les valeurs des incertitudes. Ces résultats ont été obtenus après 1 millions d'itérations.

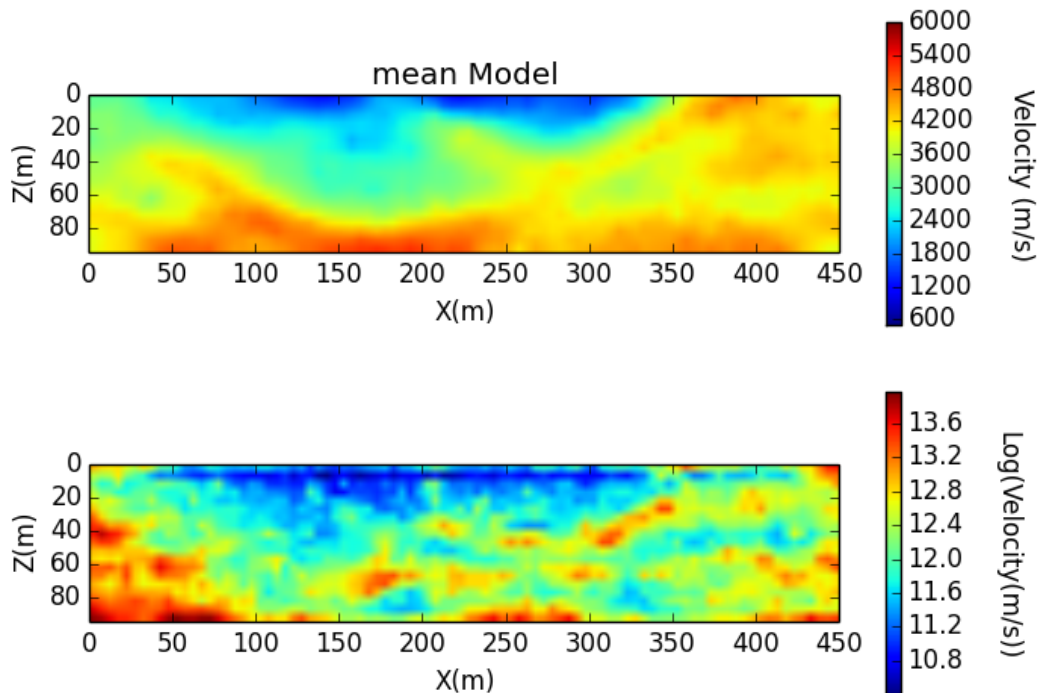


Figure 54: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

Nous remarquons tout d'abord que le modèle moyen est presque identique au vrai modèle. L'algorithme a réussi à reconstruire avec précision le vrai modèle à partir d'un nombre limité de points tirés aléatoirement dans le domaine et avec des lois a priori très larges. Néanmoins, la carte des incertitudes est relativement peu fiable. On note en effet que l'incertitude est faible dans les zones qui ne sont pas contraintes par les données. Ceci s'explique essentiellement par le choix de la base: les cosinus sont des fonctions à support global. Lorsqu'on modifie un point dans une zone non contrainte, on modifie donc nécessairement les coefficients de la décomposition dans la base DCT. Lorsqu'on reconstruit le champ de vitesse à partir des coefficients DCT, on impacte donc de manière globale le champ de vitesse, y compris dans les zones contraintes. Ainsi, même si la valeur de vitesse modifiée se situe dans une zone non contrainte, la modifier peut conduire à des modifications significatives du champ de vitesse dans les zones contraintes.

2.2.2 Modèle réaliste:

Nous appliquons notre méthodologie au modèle de benchmark présenté dans le chapitre 4. Nous effectuons l'inversion avec un nombre de points $M = 300$.

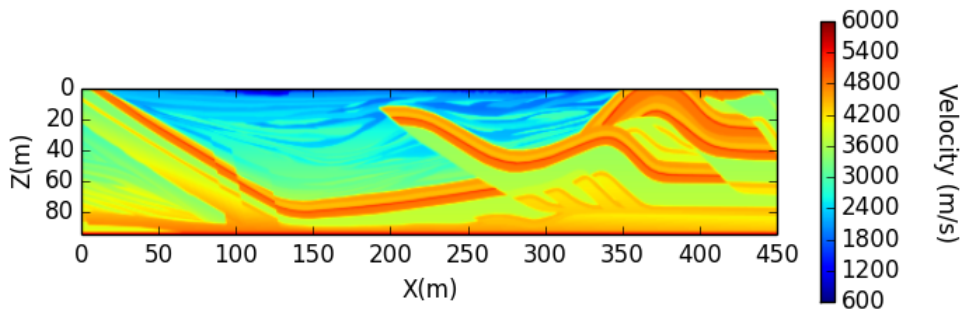


Figure 55: Vrai modèle

Nous utilisons la même géométrie d'acquisition que le cas précédent. La carte des temps de premières arrivées est donc identique à celle présenté dans le chapitre 4 (Figure 56).

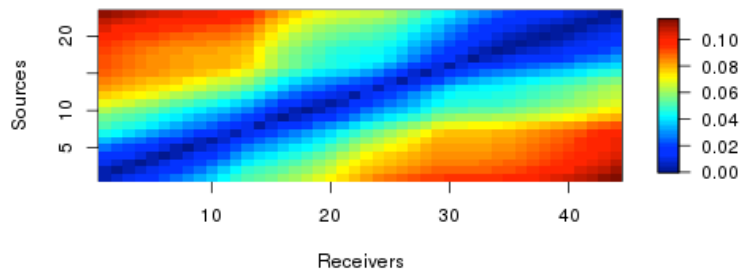


Figure 56: Carte des temps de première arrivée

Résultats :

Les résultats de l'inversion sont présentés dans la Figure 57 Celle-ci montre la carte de la vitesse moyenne a posteriori et les incertitudes associées à cette estimation (écart-type). Ces résultats ont été obtenus après 1 million d'itérations.

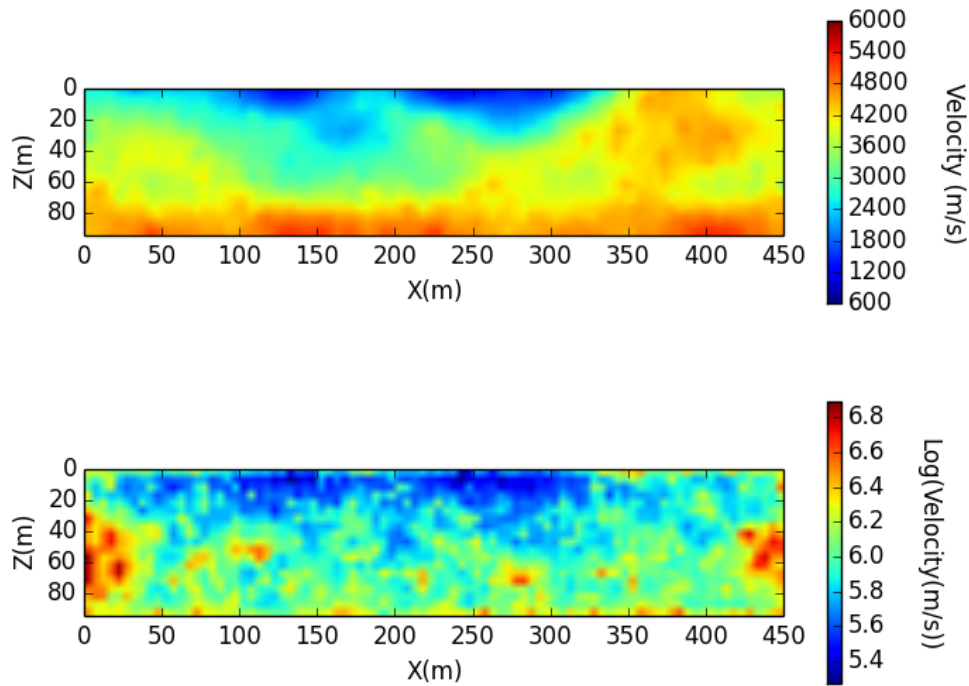


Figure 57: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

Nous remarquons tout d’abord que le modèle moyen est différent du vrai modèle. Notons cependant que le champ de vitesse est sous-échantillonné à une taille 90×19 . Avec cette résolution, il est clair que l’hypothèse de base qui nous a conduit à considérer l’acquisition comprimée, à savoir que le champ de vitesse est compressible dans une base de cosinus discrets, n’est pas valide. Compte-tenu du nombre de points utilisés pour l’échantillonnage, il était complètement illusoire d’espérer reconstruire les structures fines du modèle de vitesse. Comme on l’a vu dans l’exemple précédent, avec le nombre de points considérés, on s’attend au mieux à pouvoir reconstruire un champ de vitesse k -sparse, avec k de l’ordre de 30 (Figure 52). On pouvait donc a priori s’attendre à ce que l’algorithme retourne un champ de vitesse proche du meilleur signal k -parcimonieux approximant le champ de vitesse. Il était donc légitime d’espérer obtenir un résultat similaire au champ de vitesse présenté sur la Figure 52. En pratique, on constate bien que le champ de vitesse renvoyé par l’algorithme est proche de ce champ. Pour les mêmes raisons évoquées dans le cas précédent, la base DCT ne permet pas d’obtenir une carte des incertitudes cohérente avec la géométrie d’acquisition.

Afin de partiellement remédier à ce problème, la solution que nous proposons consiste à lancer plusieurs calculs MCMC en parallèle, avec à chaque calcul une matrice d’échantillonnage différente, et donc a fortiori des points d’échantillonnage différents. De cette manière, on force les chaînes de Markov à converger vers des états stationnaires donnant des résultats similaires entre eux dans les zones contraintes, mais différents dans les zones non contraintes. Les résultats de l’inversion sont présentés dans la Figure 58.

Celle-ci présente la carte de la vitesse moyenne a posteriori et les incertitudes associées à cette estimation (écart type). La carte de vitesse est obtenue en intégrant les résultats obtenus avec 12 chaînes en parallèle. Chaque chaîne a été lancée avec un million d'itérations. Le burn-in retenu est de 2.0×10^5 .

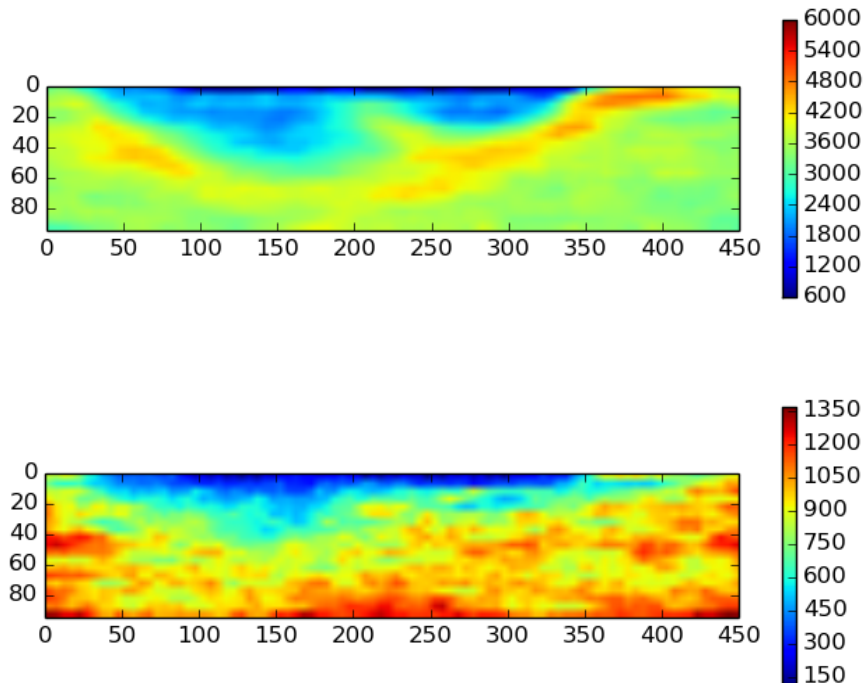


Figure 58: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)

3 Conclusion et perspectives :

Nous avons présenté dans ce chapitre un modèle paramétrique basé sur la théorie du compressed sensing. A notre connaissance, l'utilisation du compressed sensing dans ce contexte est une nouveauté. Les premiers résultats obtenus avec cette technique sont encourageants, dans le sens où le modèle est capable de reconstruire avec précision les champs de vitesses à partir d'un nombre limité de points et en utilisant des lois a priori qui sont larges (peu d'information), dès lors que le champ de vitesse que l'on cherche à déterminer est suffisamment parcimonieux. Un certain nombre de problèmes restent cependant ouverts à l'issue de ce travail de thèse, qui nous paraissent ouvrir d'intéressantes perspectives pour de futures recherches.

Dans le cadre de la présente étude, nous avons uniquement considéré des méthodes d'acquisition comprimée basées sur la transformée en cosinus discrets. Ces méthodes présentaient un intérêt double: d'un point de vue pratique en premier lieu puisqu'il est facile

de trouver des routines numériques permettant un calcul efficace de la transformée en cosinus discret; d'un point de vue physique en second lieu, puisque travailler avec une base de cosinus discrets nous permet de travailler avec une matrice de mesure très simple, qui se contente d'échantillonner le champ de vitesses en un nombre restreint de points. Il pourrait néanmoins être très intéressant de généraliser l'algorithme proposé dans cette thèse à de nouvelles bases de fonctions, comme par exemple les ondelettes, en particulier les shearlets. L'intérêt de telles fonctions est double dans le problème qui nous intéresse. D'une part, on peut s'attendre à ce que le champ de vitesse soit beaucoup mieux compressible dans une base d'ondelettes locales que dans une base de cosinus discrets. D'autre part, et il convient d'insister sur ce point, les cosinus discrets sont des fonctions dont le support s'étend sur le champ de vitesse tout entier. Cela n'est pas sans conséquences sur l'estimation de l'incertitude, comme discuté dans le cas du premier champ de vitesse. En décomposant le champ de vitesses sur une base de fonctions localisées spatialement comme les ondelettes, on peut potentiellement espérer améliorer significativement les résultats des cartes d'incertitudes.

Une autre voie d'amélioration des premiers résultats présentés dans ce chapitre est leur application à des champs de vitesse de taille beaucoup plus importante. Considérons ainsi un champ de taille N^2 . Supposons que ce champ soit k -sparse dans une base de fonctions orthogonales. Pour N suffisamment grand, on observe souvent (empiriquement) que $k \ll N^2$. Dans ce cadre, le nombre de coefficient nécessaire à la reconstruction du champ de vitesse évolue en $O\left(k \log\left(\frac{N^2}{k}\right)\right)$. A l'inverse, pour d'autres modèles comme les mosaïques de Voronoï ou de Johnson-Mehl, on peut s'attendre à ce que le nombre de paramètres à considérer dans le modèle est de l'ordre de $O(N^2)$. Ces résultats asymptotiques laissent à penser que des paramétrisations basées sur le compressed sensing pourraient s'avérer d'autant plus adaptées que les autres méthodes pour de grands champs de vitesse. Parvenir à passer à l'échelle l'algorithme proposé reste cependant une tâche difficile: la résolution du problème d'optimisation (120) basé sur le LASSO peut en effet demander un temps conséquent. Ce travail de mise à l'échelle ne saurait donc être effectué sans un travail d'implémentation efficace de la méthode.

Enfin, il est intéressant de conclure en resituant l'algorithme proposé par rapport à l'état de l'art en manière de tomographie sismique. Des travaux récents (Hawkins and Sambridge, 2015) ont consacré l'utilisation de nouvelles bases, et en particulier de bases d'ondelettes, pour paramétrer le champ de vitesse. Ces méthodes visent à exploiter le caractère parcimonieux du champ de vitesse dans une base donnée, en cherchant à déterminer les coefficients de décomposition du champ de vitesse dans cette base. En théorie, en déterminant la valeur des $k \ll n$ plus grands coefficients de la décomposition et en reconstruisant, il est possible d'obtenir une excellente approximation du champ de vitesse. Le problème inhérent à ces méthodes est qu'en général, aucune information a priori n'est disponible sur les k coefficients à considérer, de sorte qu'il faut explorer l'espace des coefficients tout entier

pour les déterminer, ce qui s'avère extrêmement coûteux. La solution proposée dans (Hawkins and Sambridge, 2015) consiste à explorer les coefficients niveau par niveau dans une décomposition multi-résolution. Le compressed sensing fournit un cadre beaucoup plus général et surtout non-adaptif, qui permet a priori de considérer n'importe quelle base de fonctions orthogonales pour compresser le champ de vitesse.

Chapitre 7.

Conclusion générale

Nous nous sommes intéressés au cours de ce projet de thèse à étudier la tomographie des temps de première arrivée dans un cadre bayésien. Compte tenu des objectifs de la thèse, il importait de disposer de modèles paramétriques qui permettent de décrire avec un nombre limité de paramètres le champ de vitesse des ondes sismiques. Cette problématique nous a conduit à nous intéresser à de nombreux domaines scientifiques, allant de la tomographie sismique à la géométrie stochastique et le traitement d'images, en passant par la simulation géostatistique et l'inférence bayésienne.

L'approche bayésienne nécessite d'échantillonner suivant la loi a posteriori des paramètres. Cela nous a amené à considérer des méthodes de simulation stochastique. Plus particulièrement, les méthodes de type Monte Carlo par chaîne de Markov (MCMC). Nous avons passé en revue dans le Chapitre 2, les principales variantes de l'algorithme Metropolis-Hastings.

Le Chapitre 3 s'intéressait à la paramétrisation, après avoir dressé un bref état de l'art de la littérature sur les paramétrisations existantes, nous avons proposé des modèles parcimonieux du modèle de vitesse des ondes sismiques. Dans un premier temps, nous avons proposé un modèle de mosaïque aléatoire, le modèle de Jonhson-Mehl, dérivé des mosaïques de Voronoï déjà proposées en tomographie bayésienne. Le second modèle que nous avons proposé est, quant à lui, décrit par une combinaison linéaire de fonctions gaussiennes, centrées sur la réalisation d'un processus ponctuel de Poisson. Pour chaque modèle, nous avons présenté un exemple de validation sur des champs de vitesse simulés en plus d'un modèle synthétique plus complexe qui sert de benchmark dans l'industrie pétrolière. Ces résultats ont été présentés dans le Chapitre 4. La paramétrisation est à même de reconstruire dans les deux cas de manière satisfaisante le champ de vitesse des ondes sismiques en utilisant un nombre raisonnable de paramètres. Les résultats prouvent donc que les deux paramétrisations que nous avons proposées sont bien adaptées à des applications en tomographie bayésienne des temps de première arrivée.

Nous nous sommes intéressés, dans le Chapitre 5, à un jeu de données réelles acquises dans le contexte de la fracturation hydraulique. Nous avons développé dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique qui nous a permis de traiter efficacement la complexité du modèle à couches en intégrant le nombre de couches et le bruit sur les données aux paramètres à estimer. Les résultats obtenus sont cohérents avec le log sismique acquis indépendamment sur le site. Ces résultats ouvrent une nouvelle perspective

visant à propager les incertitudes associées au modèle de vitesse aux incertitudes de localisation.

Nous nous sommes enfin intéressés dans le Chapitre 6 à un modèle paramétrique, encore imparfait, qui est basé sur la théorie du compressed sensing. Les quelques travaux réalisés sont encourageants et montrent que l'application du compressed sensing à la tomographie sismique peut donner lieu à des développements ambitieux. Au vu des tests effectués, il semble bien que la priorité est d'adopter une base de fonctions localisées spatialement. En effet, dans notre travail, nous avons uniquement considéré des méthodes d'acquisition comprimée basées sur la transformée en cosinus discrets. Nous avons vu que l'utilisation de cette base ne permet pas d'obtenir une estimation fiable des incertitudes. Pour surmonter cet écueil, il conviendrait de recourir à des bases de fonctions localisées spatialement comme les ondelettes. D'autres perspectives sont évoquées et discutées dans le Chapitre 6.

Les algorithmes utilisés dans cette thèse sont coûteux en temps de calcul car ils nécessitent la résolution du problème direct à chaque itération. Il est possible de remédier à ce problème en parallélisant le code. En effet, étant donné que les calculs sont réalisés par point de tir, ces derniers peuvent être exécutés simultanément sur plusieurs cœurs permettant ainsi de réduire le temps de calcul. Cet objectif pourra être atteint également en améliorant le mélange de la chaîne de Markov. Parmi les moyens permettant d'améliorer les performances de la chaîne figurent notamment l'utilisation des chaînes en interaction. Nous avons montré, dans le Chapitre 5, à travers un exemple synthétique que les résultats de l'algorithme PIR dans le cas de dimension fixe sont encourageants. En effet, les échanges permettent d'améliorer le mélange des chaînes et par conséquent, les résultats. À un temps de calcul égal, les performances de la chaîne sont nettement meilleures dans le cas des chaînes en interaction. Il serait donc intéressant d'étendre l'utilisation de l'algorithme PIR au cas trans-dimensionnel. Enfin, il importe de noter que sur le plan théorique, le passage au 3D s'effectue sans difficultés notables pour les différentes paramétrisations proposées dans ce travail de thèse. Il s'accompagne cependant d'une augmentation significative du nombre de paramètres à considérer de sorte que le temps de calcul devient difficilement acceptable. La parallélisation du code pourrait donc rendre possible la généralisation de nos paramétrisations à des applications tridimensionnelles dans le cadre d'une tomographie passive comme dans (Agostinetti et al., 2015).

Annexes

New Parametric Models For Bayesian Seismic Tomography

Jihane Belhadj¹, Thomas Romary¹, Alexandrine Gesret¹, Mark Noble¹

¹ Mines ParisTech, PSL Research University, centre de géosciences, 35 rue St Honoré, F-77300 Fontainebleau, France.

Abstract

First arrival time tomography aims at inferring the seismic wave propagation velocity using experimental first arrival times. This constitutes an inverse problem that is usually ill-posed. In particular, the unicity of the solution is not guaranteed. In our study, we rely on a Bayesian approach to estimate the wave velocity and the associated uncertainties. This approach incorporates the information provided by the data and the prior knowledge of the velocity model. Bayesian tomography allows for a better estimation of wave velocity as well as associated uncertainties. However, this approach remains fairly expensive, and MCMC algorithms that are used to sample the posterior distribution are efficient only as long as the number of parameters remains within reason. Therefore, a first step toward an efficient implementation of the Bayesian approach is to properly parameterize the model to reduce its dimension and to select adequate prior distribution for the parameters. In this article, we introduce new parsimonious parameterizations enabling to accurately reproduce the wave velocity field with the associated uncertainties.

1 Introduction

Seismic tomography is one of the main methods for imaging the subsurface of the Earth using seismic waves. This technique has been used for many years to infer the propagation velocity of seismic waves from their experimental arrival times (the measurement). A recent review is given by (Rawlinson et al., 2010). This is an inverse problem that is usually nonlinear and ill-posed. The solution of an ill posed problem may not exist or may not be unique. Furthermore, it may be unstable. In other words small errors in the data may produce large errors in the solution. A classical approach toward its resolution is to rely on iterative linearized approaches (Taillandier et al., 2009). However, these methods present several issues. In particular, they predict wrong seismic velocities when the problem is highly nonlinear and non-convex. In fact, the iterative algorithms may converge to a local minimum instead of the global minimum. Hence, they are not able to properly quantify the uncertainties associated with the solution they provide.

To account for the uncertainties in the experimental data and for the non-linearities between the data and the model parameters, Bayesian approaches coupled with Markov chain-Monte Carlo (MCMC) algorithms have been considered in the literature (Tarantola and Valette, 1982). This approach enables to aggregate information provided by the data and prior knowledge of the velocity model, thus yielding a better estimation of the wave velocities and of the associated uncertainties, that can be e.g. propagated to uncertainties regarding the location of an earthquake's hypocentre as shown by (Agostinetti et al., 2015; Gesret et al., 2015).

Although high performance computing resources are easy to come by these days, the Bayesian approach remains fairly expensive, and MCMC algorithm are efficient only as long as the number of parameters remains within reason. This requires a careful reflection both on the parameterization of the velocity model, in order to reduce the problem's dimension, and on the definition of the prior distribution of the parameters.

The velocity model is usually represented as a linear combination of basis functions: local cells and spherical harmonics are among the most commonly used (Nolet et al., 2008). In classical approaches both basis functions and their number are user defined. The size of the cells is thus chosen in advance. Several improvements were made to classical parameterizations to better deal with spatial heterogeneities of seismic tomography. The objective is to use the data to refine the mesh (see (Sambridge and Rawlinson, 2005) for a recent review). It is referred to as adaptive grids. These are for example built using cubic B-splines as in the work of (Micheline, 1995) or genetic algorithm see (Curtis and Snieder, 1997). It is also possible to use Delaunay tetrahedra and Voronoi polyhedral, introduced by (Sambridge et al., 1995) for tomographic problems using a Neighbourhood Algorithm, and popularized in Bayesian approach by (Bodin et al., 2009; Bodin and Sambridge, 2009; Bodin et al., 2012).

The first parametric model we propose uses a random Johnson-Mehl tessellation. The model was proposed by Johnson and Mehl (Johnson and Mehl, 1939) to describe crystallization processes in metallurgy. It has various fields of applications such as metallurgy (Frost et al., 1992), biology (Wolk, 1975) and crystallography [Kolomogrov, 1937]. To the best of our knowledge, this model has not been investigated in geophysics. Unlike Voronoï tessellation, Johnson-Mehl tessellation contains cells that are not necessarily convex polytopes, therefore providing smooth boundary between cells.

The second one uses Gaussian kernels as basis functions. It is especially adapted to the detection of seismic wave velocity anomalies. Each anomaly is considered to be a linear combination of these basis functions.

The two proposed parametric models can accommodate a random number of parameters. To infer the posterior probability distribution, a Reversible Jump Metropolis-Hastings algorithm

(Green, 1995) is used. It is a generalization of the classical Metropolis-Hastings algorithm that allows the simulation of the unknown space dimension. This algorithm was introduced in the field of geophysics by (Malinverno, 2002) as part of the electric inversion of a layered model, and this trans-dimensional approach has since gained popularity amongst the geophysics community. (see e.g (Agostinetti and Malinverno, 2010; Dettmer et al., 2010)). A few years later, (Bodin and Sambridge, 2009) widened its scope to include two-dimensional problems. The first three-dimensional application was very recently put forth by (Agostinetti et al., 2015).

The manuscript is organized as follows. In section 2 we briefly introduce the forward problem. A presentation of the Bayesian tomography is given in section 3. In section 4, we introduce new parameterizations that can be used to represent the wave velocity model. Following a presentation of the models in their most general form, we briefly outline the prior distribution of the parameters used in each case. In section 5 we describe MCMC methods and the Reversible Jump Metropolis-Hastings algorithm used to sample from the posterior probability distribution. In section 6 we first illustrate the tomography results with a synthetic velocity model which contains two small anomalies. We then apply our methodology to a more advanced and more realistic synthetic model that serves as a benchmark in the oil industry. Conclusions are drawn in the last section.

2 Forward modeling:

In seismic traveltime tomography the data are traveltimes. Hence the forward modeling relies on the Eikonal equation. This equation relates the wave velocity model to the gradient of traveltimes. It describes the travel time behavior of the waves in the high frequency approximation.

$$\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 + \left(\frac{\partial t}{\partial z}\right)^2 = s^2(x, y, z) \quad (1)$$

Where x, y, z are spatial coordinates, t is the travel time and s is the slowness (inverse of velocity). Several methods have been proposed to solve the Eikonal equation (1) numerically using finite-difference methods.

In our work, we use an Eikonal solver developed by (Noble et al., 2014). The special feature of this algorithm is the use of both a spherical approximation near the source and a plane wave approximation further from the source. In order to better handle complex wave propagation, the algorithm makes use of a fast sweeping method and relies on a variable grid spacing in all three directions of space.

3 Bayesian approach:

Seismic tomography consists in estimating a velocity model m from a noisy measurement vector d^{obs} .

$$d^{obs} = F(m) + \varepsilon \quad (2)$$

Where $d^{obs} = (d_1^{obs}, \dots, d_N^{obs}) \in R^N$ the measurement vector, $m = (m_1, \dots, m_M) \in R^M$ is the unknown parameter vector, $F(m)$ is the forward model used to create synthetic data and ε is an additive noise including theoretical errors.

In this work we formulate the inverse problem within a Bayesian framework. In a Bayesian approach all information is represented in probabilistic terms. The approach aims at estimating the posterior distribution of the parameters through a combination of prior knowledge of the velocity model and information provided by the data. This posterior distribution is obtained using Bayes theorem:

$$P(m|d^{obs}) = \frac{P(d^{obs}|m) \times P(m)}{P(d^{obs})} \quad (3)$$

Where

$P(d^{obs} | m)$ is the likelihood function, which is the conditional probability of observing the data given a model m .

$P(m)$ is the a priori probability density of the parameters m , that is, what is known about the model m before collecting the data d^{obs} .

$P(d^{obs})$ is the evidence which is the probability of the data independent of the parameters m . It does not depend on any particular state of the system. It can therefore be considered as a normalization constant. The posterior distribution thus expresses:

$$P(m|d^{obs}) \propto P(d^{obs}|m) \cdot P(m) \quad (4)$$

3.1 The likelihood

The likelihood represents how well a model (synthetic data) fits the data and requires thus solving the forward problem. Assuming the additive noise in (2) is Gaussian with zero mean and variance σ_d^2 , the likelihood is:

$$P(d^{obs}|m) \propto \exp\left(\frac{-1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|_2^2\right) \quad (5)$$

Where $\|a\|_2 = \sqrt{a^t a}$, $a \in R^n$ is the standard l_2 norm.

3.2 The prior

The prior contains all information about the parameters m , which is available independently from the data. In practice, this information could come for example from geological knowledge. The formulation of $P(m)$ depends on the circumstances of the case, and on the parameterization chosen.

4 The parameterization

In seismic tomography, the parameterization of the velocity model is a key issue. The challenge lies in providing relevant and parsimonious parametric models.

In the next section, we introduce new parameterizations that can be used to represent the wave velocity model. Following a presentation of the models in their most general form, we briefly outline the prior distribution of the parameters used in both cases.

4.1 Johnson Mehl tessellation:

The Johnson-Mehl tessellation was introduced by (Johnson and Mehl, 1939) as a spatial model to describe crystallization processes. A spatial tessellation is a subdivision of a domain into non-overlapping regions (cells). The most commonly random tessellations used in literature are the Voronoï tessellation and the Delaunay triangulation (Okabe et al., 2009). To get an extensive presentation of random tessellations, the reader is referred to the books of (Matheron, 1975; Schneider and Weil, 2008; Chiu et al., 2013). The Johnson Mehl tessellation can be seen as a generalization of the Voronoï tessellation where the nuclei are augmented additively with time. All cells grow then isotropically with a constant speed around their nuclei. The growth of cells boundaries is stopped when they meet, and globally when the domain is filled. Hence, all points falling in an existing cell are systematically removed. The steps of the construction of the Johnson-Mehl tessellation are displayed in . More details can be found in the article of (Møller, 1995), which provides a unified exposition of Random Johnson-Mehl tessellations.

The Figure 1 displays the steps of the construction of the Johnson-Mehl tessellation.

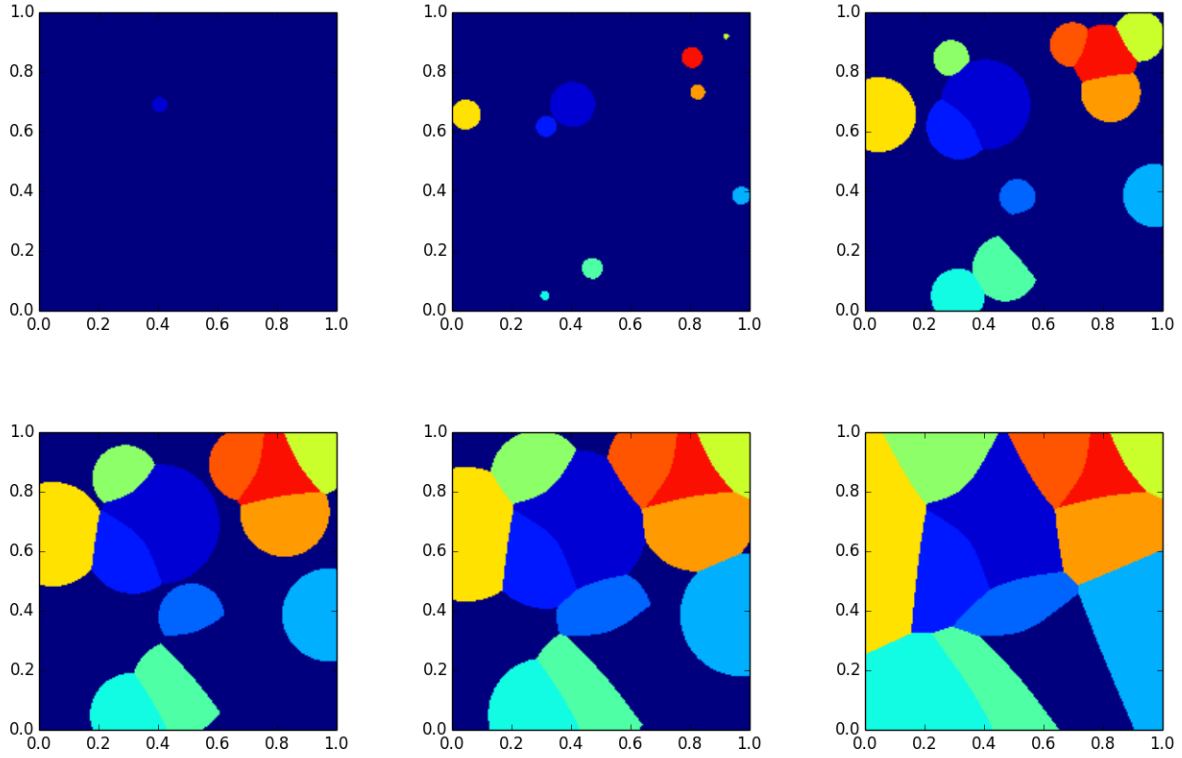


Figure 1 The steps of the construction of the Johnson-Mehl tessellation

From a mathematical standpoint, a Johnson-Mehl tessellation is built from a sequential Poisson point process: The points (nuclei) $u_i, i = 1 \dots N$ are implanted additively at a time $t_i, i = 1, \dots, N$. The cells $C_i, i = 1, \dots, N$ corresponding to the points $u_i, i = 1 \dots N$ are defined by

$$C_i = \{u / \forall j, j \neq i, t_i + \frac{\|u - u_i\|_2}{v_c} \leq t_j + \frac{\|u - u_j\|_2}{v_c}\} \quad (6)$$

Note that when all nuclei are implanted at the same time, we obtain the classical Poisson-Voronoi tessellation model.

In this article, we use Johnson-Mehl tessellation cells as basis functions to represent the velocity model:

$$v(u) = \sum_{1 \leq i \leq n} v_i \phi_i(u) \begin{cases} \phi_i(u) = 1 & u \in C_i \\ \phi_i(u) = 0 & \text{otherwise} \end{cases} \quad (7)$$

The velocity model is completely described by the number of cells, the positions of their nuclei, their time of implementation and the wave velocity of each cell. The parameters vector can be written as:

$$m = (U, V, T,)$$

Where n is number of cells.

$U = (u_i)_{1 \leq i \leq n}$ contains the position of cell nuclei.

$T = (t_i)_{1 \leq i \leq n}$ is a the vector of implantation time.

And $V = (v_i)_{1 \leq i \leq n}$ contains the wave velocity value.

4.1.1 The prior:

We assume a priori that the parameters values and their number are independent.

$$P(m) = P(m|n)P(n) \quad (8)$$

where $P(n)$ is the prior on the number of cells. The Johnson-Mehl tessellation is built from a Poisson point process. Therefore, n follows a Poisson distribution with parameter θ :

$$P(n) = \frac{(\theta|D|)^n}{n!} \exp(-\theta|D|) \quad (9)$$

where $|D|$ is the domain area.

The parameter θ can be interpreted as the average number of nuclei per unit area (or volume in 3D). The choice of this parameter is a compromise between the accuracy of the model and computational cost. On the one hand, the number of cells must be sufficiently high to reproduce velocity heterogeneities with accuracy, especially in the well-observed regions. On the other hand, it is computationally costly to rely on a fine discretization in the regions that are poorly observed.

To each implanted point, we attach a positive real value, which will represent the velocity. That forms a marked point process.

We use a uniform distribution for the positions of the cell nuclei $(u_i)_{1 \leq i \leq n}$ and for their marks (velocity $(v_i)_{1 \leq i \leq n}$ and implantation time $(t_i)_{1 \leq i \leq n}$).

$$\begin{aligned}
P(u_i|n) &= \frac{1}{|D|} \\
P(t_i|n) &= \frac{1}{\Delta T} 1_{[t_{min}, t_{max}]}(t_i) \\
P(v_i|n) &= \frac{1}{\Delta V} 1_{[v_{min}, v_{max}]}(v_i)
\end{aligned} \tag{10}$$

where: $\Delta T = t_{max} - t_{min}$; $t_{max}, t_{min} \in \mathbb{R}^+$ and $\Delta V = v_{max} - v_{min}$; $v_{max}, v_{min} \in \mathbb{R}^+$.

Finally assuming that conditionally to the number of cells n , the parameters are a priori independent, hence $P(m|n)$ can be expressed as the product of the prior distributions of each of the components:

$$P(m|n) = P(U|n)P(T|n)P(V|n) = n! \prod_{i=1}^n P(u_i|n)P(t_i|n)P(v_i|n) \tag{11}$$

4.2 Gaussian kernels:

Here we propose to use Gaussian kernels as basis functions. The purpose is to provide a way to detect velocity anomalies of seismic wave. Each velocity anomaly is considered to be a linear combination of the basis functions.

$$v(u) = V_L + \sum_{1 \leq i \leq N} \alpha_i K \left(\left\| \frac{u - u_i}{\sigma_i} \right\|_2 \right), \tag{12}$$

where V_L is the velocity of the background medium, K is the kernel function, here we use a Gaussian kernel.

The key is to imagine the anomaly as a set of points (small anomalies) $u_i, i = 1 \dots N$ that diffuse a ‘‘characteristic amplitude’’ $\alpha_i, i = 1, \dots, N$ (multiplicative coefficient). This amplitude can be negative ($\alpha_i < 0$) (Figure 2a), to reproduce the velocity drop of a negative anomaly, or positive ($\alpha_i > 0$) (Figure 2b), to increase the velocity in the area around a positive anomaly. The diffusion takes place in a location defined by a Gaussian kernel and controlled by $\sigma_i, i = 1, \dots, N$.

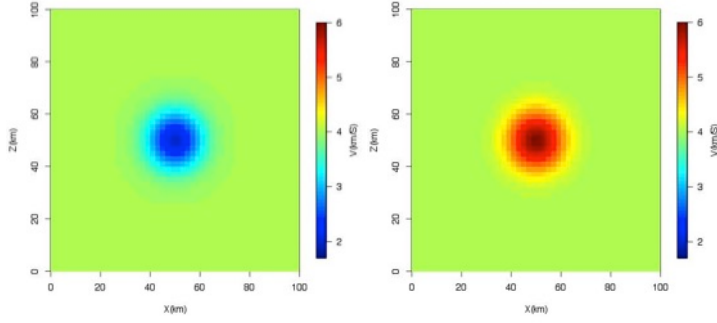


Figure 2: a) negative anomaly. 2b) positive anomaly

The velocity model is completely described by the number of points (small anomalies), their position, and the width and the multiplicative coefficient of the kernel function implanted at each point. The parameters vector can be written as

$$m = (V_L, U, \alpha, \sigma,),$$

where N is number of basis functions, $U = (u_i)_{1 \leq i \leq N}$ contains the localisation of Gaussian Kernels, $\alpha = (\alpha_i)_{1 \leq i \leq N}$ is the vector of multiplicative coefficients (amplitudes), and $\sigma = (\sigma_i)_{1 \leq i \leq N}$ contains the width of the basis functions.

4.2.1 The prior:

Here, we can also represent the parameterization as a marked point process. Gaussian kernels are localized at the realization of the point process. The marks of the point process are the width and the multiplicative coefficient of each basis function. They are the parameters to be conditioned to the data together with the number of points.

First of all, we suppose that the background velocity is independent from all the other parameters. In addition we assume that the marks and the location of points (kernels) are independent from their number:

$$P(m) = P(V_L)P(U, \alpha, \sigma|n)P(n), \quad (13)$$

where $P(n)$ is the prior on the number of point. Here, n follows a Poisson distribution with a parameter θ :

$$P(n) = \frac{(\theta|D|)^n}{n!} \exp(-\theta|D|), \quad (14)$$

where $|D|$ is the domain area.

The parameter θ is the average number of points (kernels) per unit area (or volume in 3D). For example, the velocity map from linearized tomography can be used as prior information.

It provides an estimation of the size of the anomaly and therefore the number of kernels to be used. In the event that no such information is available, the value can be determined by trial and error.

We assume that conditionally to the number of points n , all parameter vectors U , σ , α are independent:

$$P(m|n) = P(V_L)P(U|n)P(\alpha|n)P(\sigma|n) \quad (15)$$

$$n! \prod_{i=1}^n P(u_i|n)P(\alpha_i|n)P(\sigma_i|n)$$

For each point $i, 1, \dots, N$.: the position u_i and the amplitude α_i are considered to follow a uniform distribution, while the width is assumed to follow an exponential distribution.

$$P(u_i|n) = \frac{1}{|D|}$$

$$P(\alpha_i|n) = \frac{1}{\Delta\alpha} 1_{[\alpha_{min}, \alpha_{max}]}(\alpha_i) \quad (16)$$

$$P(\sigma_i|n) = \frac{1}{\sigma_m} \exp\left(-\frac{\sigma_i}{\sigma_m}\right)$$

where: $\Delta\alpha = \alpha_{max} - \alpha_{min}$; $\alpha_{max}, \alpha_{min} \in R^+$ and $\sigma_m \in R^+$.

As regards the background velocity, we use a uniform distribution on $[V_{min}, V_{max}]$ where $V_{min} > 0$ and $V_{max} > 0$. The endpoints are set according to seismic studies and prior geological knowledge.

5 MCMC and RJMCMC:

The solution of a Bayesian seismic tomography is represented by the posterior distribution. This posterior distribution describes all the a posteriori information available about the parameters m i.e about the velocity model. This information is usually summarized by quantities such as the maximum a posteriori model, a posteriori mean model and variance or other posterior quantiles.

These quantities can be described by an integral of the form:

$$E[h(m)|d^{obs}] = \int h(m)P(m|d^{obs})dm, \quad (17)$$

where h is some function of interest, e.g $h(m) = m$ for the posterior mean.

The implementation of the Bayesian approach can be difficult because the evaluation of the integral (17), is most of the time intractable. A practical way to tackle this issue is to use Monte Carlo methods. The idea is to generate an ensemble of candidate solutions which is distributed according to the posterior probability density function. Hence the integral (17) can be easily approximated by the sample mean:

$$E[h(m)] \approx \frac{1}{N} \cdot \sum_{n=1}^N h(m^n), \quad (18)$$

where samples m^n , $n = 1..N$ are drawn from the posterior (5).

While direct simulation techniques like rejection are not efficient when the dimension of the problem is large and the target distribution is complex, a technique to obtain these samples is to build a Markov chain that behaves asymptotically as the posterior distribution after an initial transient regime called burn-in period.

5.1 Metropolis-Hastings for fixed dimension:

Metropolis-Hastings is one of widely used Markov Chain Monte Carlo (MCMC) methods see (Metropolis et al., 1953). It requires the choice of a proposal distribution q . This proposal distribution is used to propose potential transitions for the Markov chain. Given a current model m , a candidate m^* is accepted with probability $\alpha(m, m^*)$ defined as:

$$\alpha(m, m^*) = \min \left\{ \frac{P(m^*|d^{obs}) \cdot q(m|m^*)}{P(m|d^{obs}) \cdot q(m^*|m)}, 1 \right\}. \quad (19)$$

5.2 Metropolis-Hastings for variable dimension:

Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995), also known as trans-dimensional MCMC, is an extension of classical MCMC methods to models with varying dimension of the parameter space. This method is thus particularly suited to the conditioning of the parametric models described above.

Suppose the chain is in state m with dimension d and we want to move to a new state m^* with with dimension d^* . To that end, we first choose a move type i with probability $j_i(m)$. Then we define two auxiliary variables u and u^* with dimension r and r^* respectively, so that the transformation $(m, u) \rightarrow (m^*, u^*)$ is a diffeomorphism. These variables are drawn from two proposal distributions q and q^* and allow to cover the difference of dimension between m and m^* . $d + r = d^* + r^*$

Finally, the new state m^* is obtained from a deterministic and reversible function h .

$$h(m, u) = m^* \quad (20)$$

The Green ratio is defined so that the chain is reversible:

$$r = \frac{P(m^*|d^{obs}).j_i(m^*).q^*(u^*)}{P(m|d^{obs}).j_i(m).q(u)} \left| \frac{\partial(m^*, u^*)}{\partial(m, u)} \right|, \quad (21)$$

where: $\left| \frac{\partial(m^*, u^*)}{\partial(m, u)} \right|$ is the determinant of the Jacobian matrix of the transformation from m to m^* .

Johnson-Mehl tessellation:

In our work, we adopt a scheme where three move types are present: At each iteration, a candidate is proposed by either – adding a new cell nuclei (birth step), - deleting a cell nuclei (death step) or changing an existing cell nuclei (move step). We assume that the probabilities corresponding to each move type depend on the current state dimension. With a probability $j_b(m) = \frac{n}{\theta|D|+n+1}$ we add a new nuclei u^* to the current state m . This nuclei is drawn independently from the prior $q(u^*) = \frac{1}{|D|} \cdot \frac{1}{\Delta T} \cdot \frac{1}{\Delta V}$. With a probability $j_d(m) = \frac{\theta|D|}{\theta|D|+n}$ we delete a nuclei which is picked randomly $q(u^*) = \frac{1}{n}$. Finally with a probability $j_m(m) = 1 - j_b(m) - j_d(m)$, we pick randomly an existing nuclei and we replace it with a new one. As for the birth step, the new nuclei is drawn independently from the prior. The proposal distribution for this step is $q(u^*) = \frac{1}{n} \cdot \frac{1}{|D|} \cdot \frac{1}{\Delta T} \cdot \frac{1}{\Delta V}$. It is worth noticing that this kernel is invariant for the Poisson distribution, thus respecting the prior distribution.

Gaussian kernels:

With a probability β , we perturb the background velocity value using a Gaussian distribution.

With a probability $1 - \beta$, in much the same way the Johnson-Mehl tessellations are handled, we propose either to:

- Add a new point with a probability $j_b(m) = \frac{n}{\theta|D|+n+1}$
- Delete a cell nuclei with a probability $j_d(m) = \frac{\theta|D|}{\theta|D|+n}$
- Change an existing point with a probability $j_m(m) = 1 - j_b(m) - j_d(m)$

The other parameters, namely the marks of the point process are handled similarly as above.

6 Inversion of Synthetic Data:

To test the algorithm proposed in this article, we employ two synthetic datasets.

6.1 Model 1:

We first consider a test case where the true velocity model used for calculating the synthetic data is similar to the one used in (Bodin and Sambridge, 2009; Bodin et al., 2009). It has dimensions of 100 km in length and depth. As shown in Figure 3 two main areas with strong velocity contrast can be distinguished: the green triangle has a velocity of 4km/s and contains a positive anomaly where the velocity increases by 25% whereas the orange one has a velocity of 5km/s and contains a small low anomaly where the velocity drops by 20%.

The observed data consists of P first travel times recorded at 23 positions from a set of 17 calibration shots as indicated by the red squares (shots) and blue triangles (receivers).

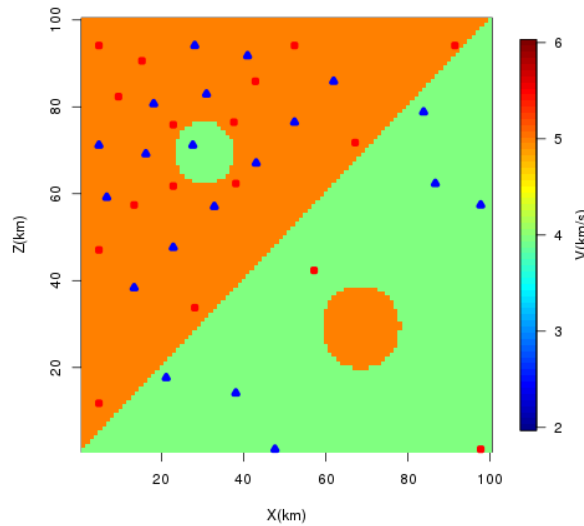


Figure 3: True model

Table 1 and 2 list the numerical values chosen for the parameters of the prior distributions. For Johnson-Mehl tessellation, the average number of cells used is 30. Germs are uniformly distributed in the domain $[0,100] \times [0,100]$. The wave velocity in each cell is uniformly distributed $[2,6]$, finally the time of implantation is uniformly distributed in $[0,20]$.

Table 1 Prior distributions of model parameters m: Johnson-Mehl tessellation

Model parameter	Prior distribution	Prior parameter
Number of cells	Poisson	Poisson parameter 0.003
Spatial coordinates of nuclei	Uniform	Minimum 0

		Maximum	100
Velocity v	Uniform	Minimum	2
		Maximum	6
Time of implantation	Uniform	Minimum	0
		Maximum	20

Concerning the Gaussian kernels, the average number of kernels chosen is 20, their position is uniformly distributed in the domain $[0,100] \times [0,100]$, their amplitude follows an uniform distribution on $[-6,6]$, the mean width is 15. Finally, the background velocity is uniformly distributed in $[2,6]$.

Table 2 Prior distributions of model parameters m : Isotropic Gaussian kernels

Model parameter	Prior distribution	Prior parameter	
The background velocity	Uniform	Minimum	2
		Maximum	6
Number of kernels	Poisson	Poisson parameter	0.002
Spatial coordinates of kernels	Uniform	Minimum	0
		Maximum	100
Amplitude (multiplicative coefficient)	Uniform	Minimum	-6
		Maximum	6
Width	Exponential	Mean	15

The main results of the inversion are displayed in Figure 4 and 5. Figure 4,5 represent the average velocity and the estimated errors (standard deviation) maps for Johnson Mehl tessellation and isotropic Gaussian kernels respectively. These results are obtained on the base of 1000 models by taking a 1/ 100 systematic sample from 1 million after a burn-in period of 10000 iterations.

The main areas are well reconstructed, in particular the negative anomaly in the upper left part. On the other hand, the inversion is less accurate around the positive anomaly in the lower right part. The poor recovery of this anomaly is related to the low coverage of seismic data in this portion of the domain. This can be clearly seen in the errors map (Figure 4b, 5b)

as the uncertainty is high in that region, especially for the Johnson Mehl parameterization. Finally, the diagonal boundary is reconstituted more precisely in the case of the Johnson Mehl parameterization than in the case of Gaussian kernels. This can be explained by the regularity of the basis functions used in the second parameterization.

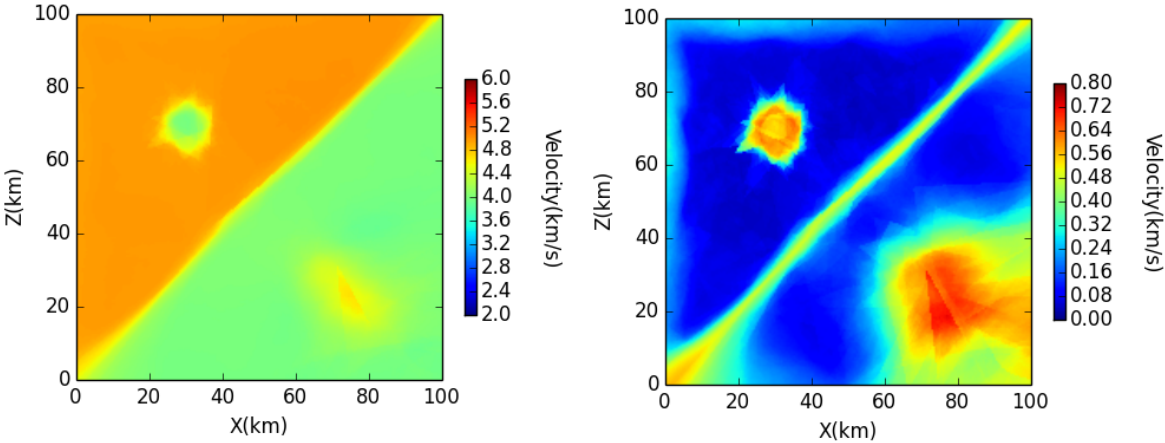


Figure 4: Johnson-Mehl results: a) the posterior mean map b) the estimated error map (standard deviation)

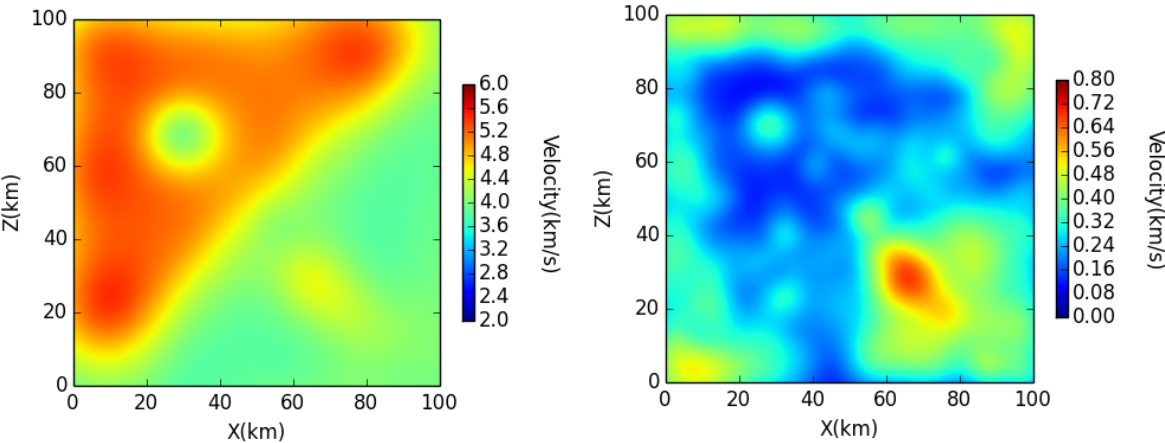


Figure 5: Gaussians kernels: a) the posterior mean map b) the estimated errors map (standard deviation)

In addition to maps, the posterior uncertainties are also illustrated along cross-sections in Figure 6 and Figure 7. Green and red lines show the posterior mean model and the true model used to compute synthetic data. The black lines correspond to the 95% confidence interval. As shown in Figure 6,7 the posterior mean model is almost identical to the true model except around the positive anomaly (Figure 6a, Figure 7a). The value of the positive anomaly is less well constrained, nonetheless the true value is included in the 95% confidence interval. At this stage, it may be observed that better results are achieved using JM tessellation.

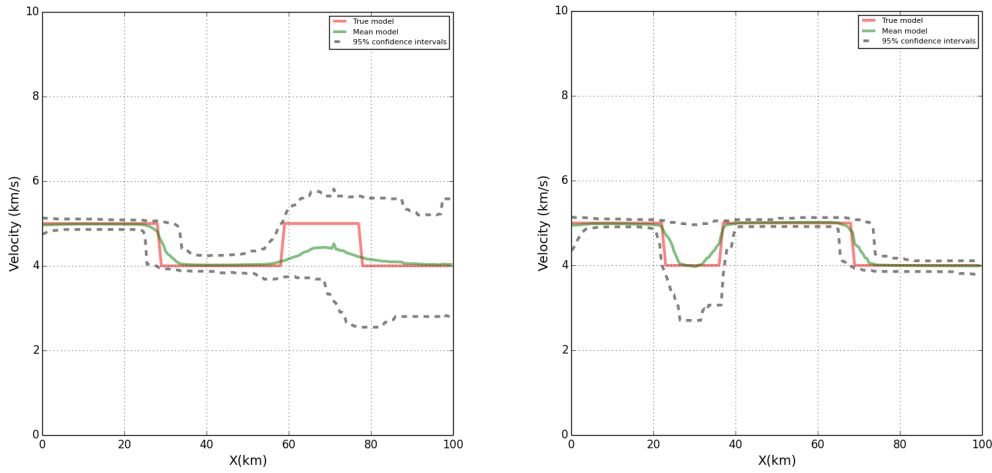


Figure 6: Johnson-Mehl: Seismic velocity profiles at different depth. (a) 30 km depth. (b) 70 km depth. The green curve is the posterior mean profile, the red curve is the true one, The black lines correspond to the 95% confidence interval

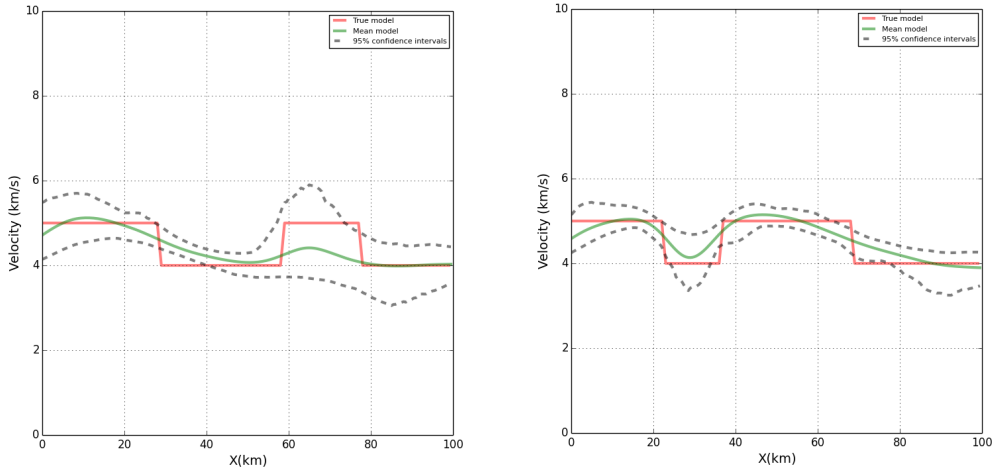


Figure 7: Gaussian kernels: Seismic velocity profiles at different depth. (a) 30 km depth. (b) 70 km depth. The green curve is the posterior mean profile, the red curve is the true one, The black lines correspond to the 95% confidence interval

Figure 8 displays the histograms of the number of Johnson-Mehl cells and Gaussian kernels. These histograms provide an estimation of the posterior distribution of the number of Johnson-Mehl cells/ kernels $P(n|d^{obs})$. For Johnson-Mehl tessellation, the number of cell nuclei is between 8 and 30 with a mode at 12, whereas, the number of Gaussian kernels varies between 12 and 38 with a mode at 22. It is clear that Gaussian kernels require more parameters than Johnson-Mehl tessellation. Nevertheless, it must be mentioned that even for Gaussian kernels 88 parameters (each kernel is characterized by four parameters) is still reasonable. Finally, it is important to note that in the case of the Johnson-Mehl tessellation, the posterior mean is significantly different from the prior value. Hence the choice of the prior mean has no effect on the final results. On the contrary, the parameterization using Gaussian kernels is sensitive to this choice. An optimal value is tuned by trial and error procedure.

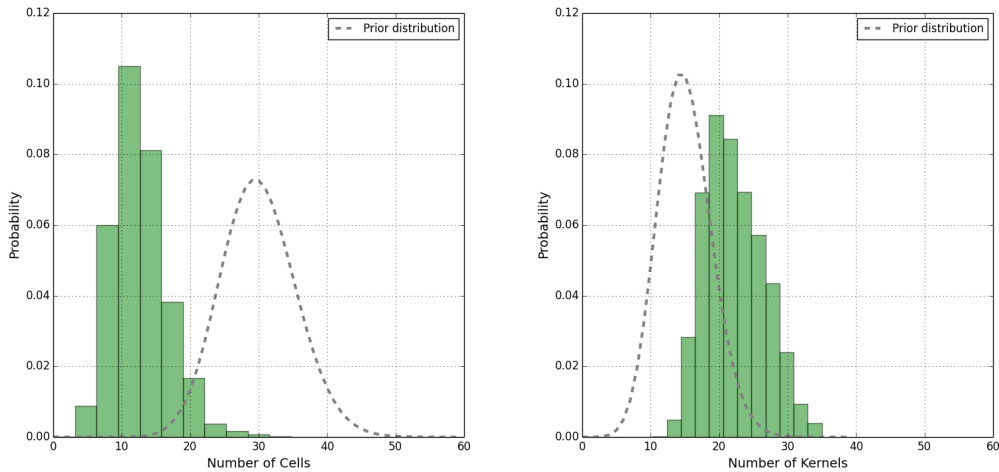


Figure 8: Histogram of the number of a) Johnson Mehl cells b) gaussian kernels

Finally, to inspect residuals we display in Figure 9 the histogram of the differences between observed and posterior arrival times in the inversion. The posterior data are computed for the posterior mean model. Figure 9 shows that the residuals are bounded within the assumed value of data errors of 0.4s except only for a few couple source receivers. Thus, for both parameterizations the posterior mean model explains the observed data well. It should be noted that, in the case of JM tessellation, the residuals are strikingly small. In the light of these results, we can say that JM tessellation leads to better results.

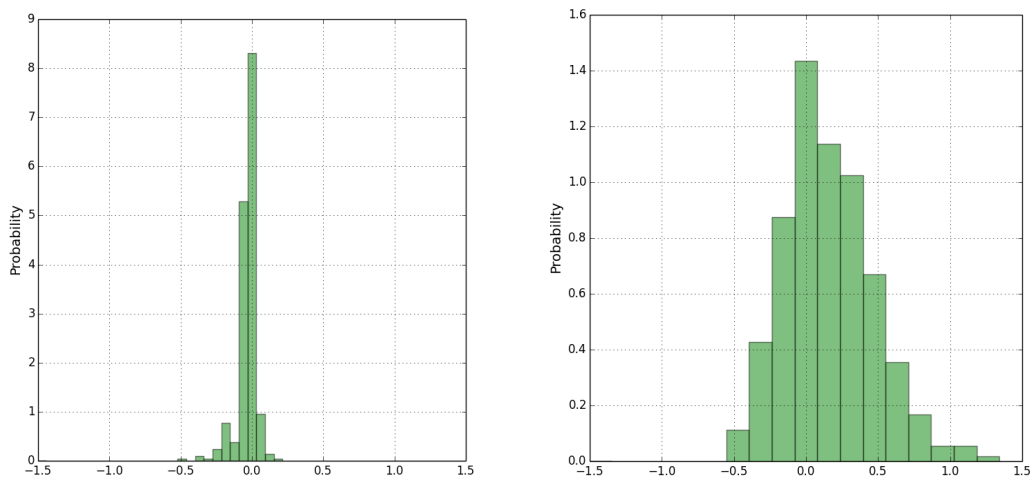


Figure 9: Histogram of the residuals a) Johnson-Mehl tessellation b) Gaussian kernels

It is worthwhile comparing JM tessellation results with those obtained by (Bodin and Sambridge, 2009). In this context, it may be noted that the Johnson-Mehl Model introduces an additional parameter, namely the time of implantation. This parameter provides grains with

general shapes whereas voronoï model considers only polytopes. In particular, JM cells are not necessarily convex. This additional parameter is useful since it allows to better reproduce certain interfaces that improve the results. One will for instance note that the diagonal boundary is reconstituted more precisely in the case of the Johnson Mehl model. We also display the best model (Figure 10) to compare it with the one presented in (Bodin and Sambridge, 2009). We note that the main areas are better reconstructed. In conclusion, we can state that, for this example, although the Johnson-Mehl Model introduces an additional parameter, their general shape leads to better results.

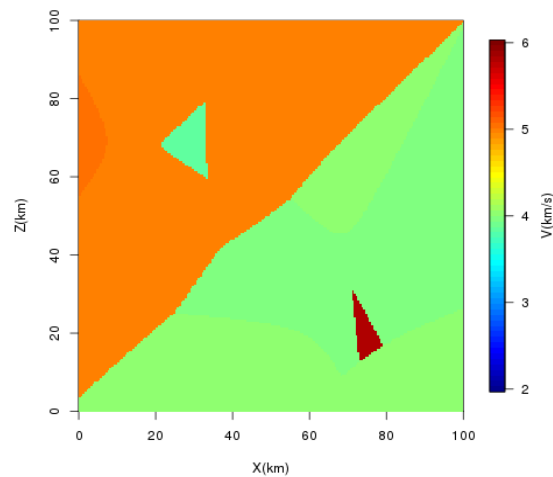


Figure 10: The best model

6.2 Model 2:

We now apply our methodology to a more realistic synthetic model. Amoco Tulsa Research Lab designed the model in question in 1994. This model is of interest due to its use as a benchmark in the oil industry. The velocity model is 2D, purely acoustic and contains many different types of near-surface geology as shown in Figure 11. It is a complex model with velocity variations in both horizontal and vertical directions.

The observed data consists of P-waves first travel times recorded at 44 positions from a set of 23 calibration shots, both located on the topography with 20 m shot interval and 10 m receiver interval. The domain is 95m in depth and 450m wide.

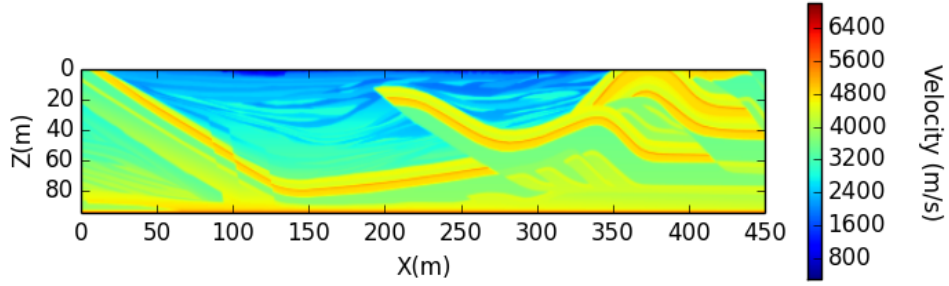


Figure 11: The true velocity model

Table 3 and 4 list the necessary numerical values of the parameters of the prior distributions. For Johnson-Mehl tessellation, the mean number of cells used is 427. Germs are uniformly distributed in the domain $[0,450] \times [0,95]$. The wave velocity in each cell is uniformly distributed $[200,7000]$, finally the time of implantation is uniformly distributed in $[0,20]$.

Table 3: Prior distributions of model parameters m : Johnson-Mehl tessellation

Model parameter	Prior distribution	Prior parameter
Number of cells	Poisson	Poisson parameter 0.01
Spatial coordinates of nuclei	Uniform	Minimum min in model domain
		Maximum max in model domain
Velocity v	Uniform	Minimum 200
		Maximum 7000
Time of implantation	Uniform	Minimum 0
		Maximum 20

To better deal with the complexity of the model, we opted for using anisotropic Gaussian kernels which suppose additional parameters to control anisotropy, namely an orientation and a second scale parameter. The mean number of kernels chosen is 40, their position is uniformly distributed in the domain $[0,450] \times [0,95]$, their amplitude follows a uniform distribution on $[-3000, 3000]$, the mean width is 20. Finally, the background velocity is uniformly distributed in $[200,7000]$.

Table 4: Prior distributions of model parameters m : Anisotropic Gaussian kernels

Model parameter	Prior distribution	Prior parameter
The background velocity	Uniform	Minimum 200

			Maximum	7000
Number of kernels		Poisson	Poisson parameter	0.001
Spatial coordinates of kernels		Uniform	Minimum	min in model domain
			Maximum	max in model domain
Amplitude (multiplicative coefficient)		Uniform	Minimum	-3000
			Maximum	3000
Width	Modulus	Exponential	Mean	15
	Angle	Uniform	Minimum	0
			Maximum	π

The inversion was performed using Johnson-Mehl tessellations and anisotropic Gaussian kernels. The results are presented in Figures 12 and 13. These figures show the average velocity profiles and the estimated errors (standard deviation) map for each parameterization. These results are obtained by taking a 1/ 1000 systematic sample from 1 million after a burn-in period of 10000 iterations.

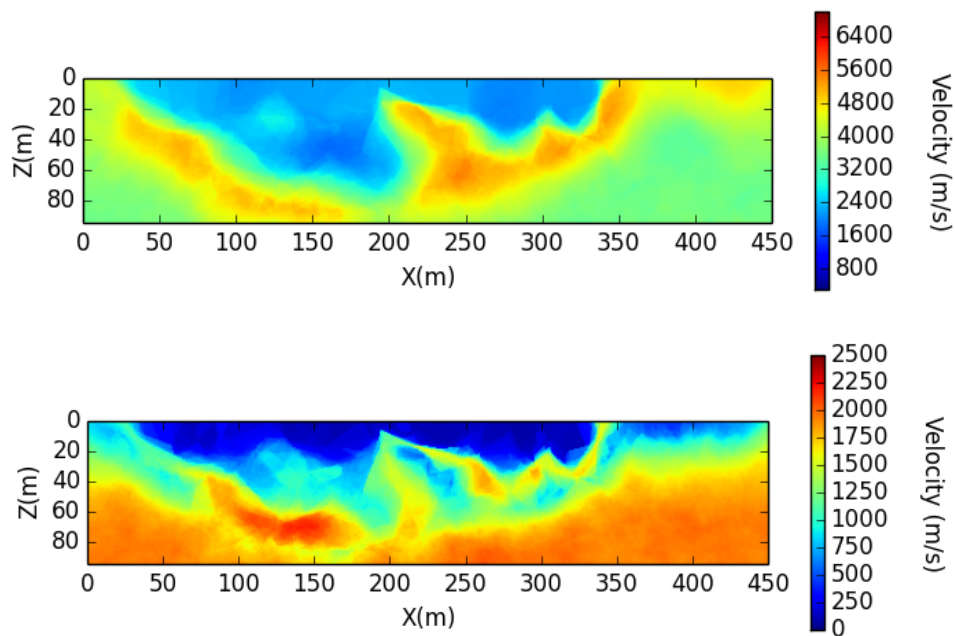


Figure 12: Johnson-Mehl results: a) posterior mean map b) estimated error map (standard deviation)

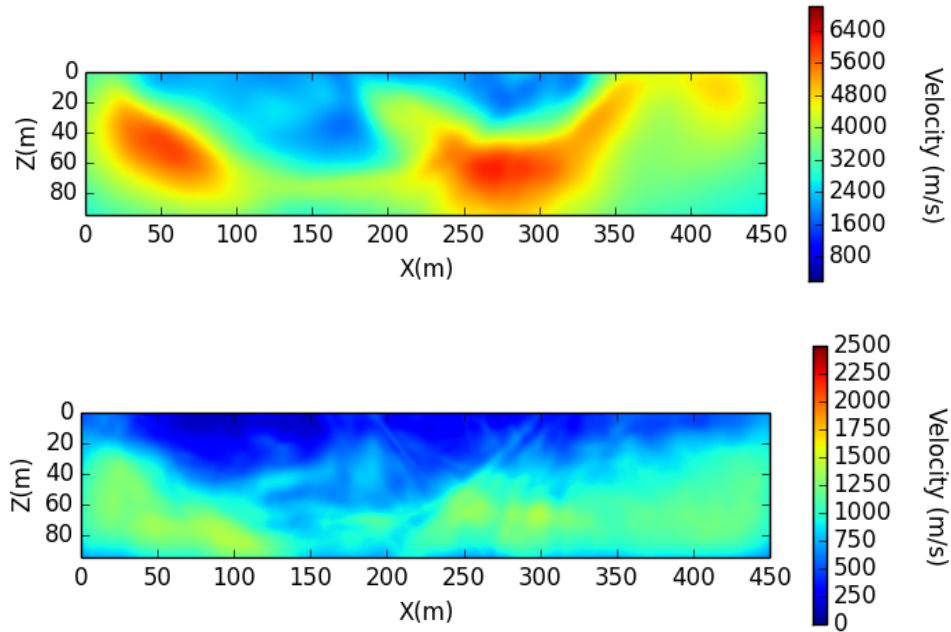


Figure 13: Anisotropic Gaussians kernels: a) posterior mean map b) estimated errors map (standard deviation)

In both models, the main structures are well reproduced particularly in the near surface where the ray coverage is denser. The bottom of the velocity model is poorly constrained by data; consequently the uncertainty is high in this part of the model as shown in Figure 12b and 13b. Figure 12b shows a high uncertainty near the upper boundary even if this area is well constrained. It is mainly caused by the irregularity of Johnson-Mehl tessellation that yields rough transitions between cells.

Several velocity profiles for both parameterizations at different locations are shown in Figures 14 and 15. Green and red lines show the posterior mean model and the true model used to compute synthetic data. The black lines correspond to the 95% confidence interval. The posterior mean model fits closely, especially in the near surface where the standard deviation errors are lower. This is clearly seen in the case of JM, the mean model is remarkably similar to the true one. The uncertainties are also consistent with the geometry acquisition; they are high in the bottom of the model. Finally, we note also that the uncertainty value, in the non constrained area, is higher in the case of Johnson mehl model. This is mainly due to the prior values.

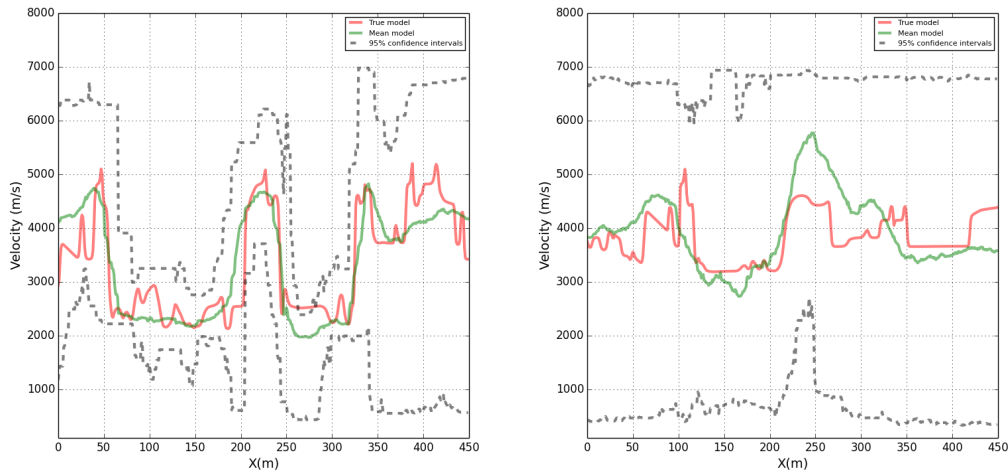


Figure 14: Johnson-Mehl: Seismic velocity profiles at different depth. (a) 25 m depth. (b) 65 m depth. The green curve is the posterior mean profile, the red curve is the true one, The black lines correspond to the 95% confidence interval

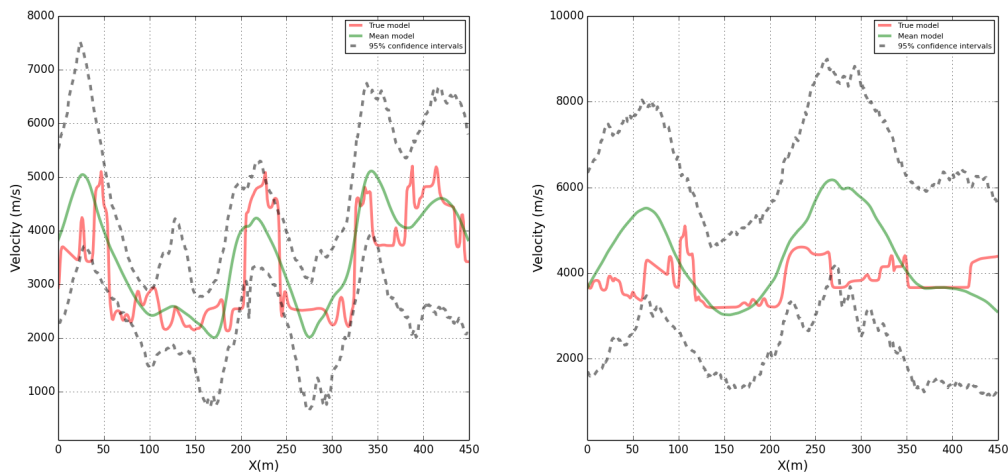


Figure 15: Anisotropic Gaussian kernels: Seismic velocity profiles at different depth. (a) 25 m depth. (b) 65 m depth. The green curve is the posterior mean profile, the red curve is the true one, The black lines correspond to the 95% confidence interval

Figure 16 displays the histograms number of cells/ kernels used in the inversion for both parameterizations. For Johnson-Mehl tessellation, the number of cell nuclei ranges from 200 and 340 with a mode at 260, whereas for Gaussian kernels, the number of small anomalies varies between 30 and 70 with a mode at 48. Unlike Gaussian kernels, the choice of the prior mean does not influence the final results in the case of Johnson-Mehl.

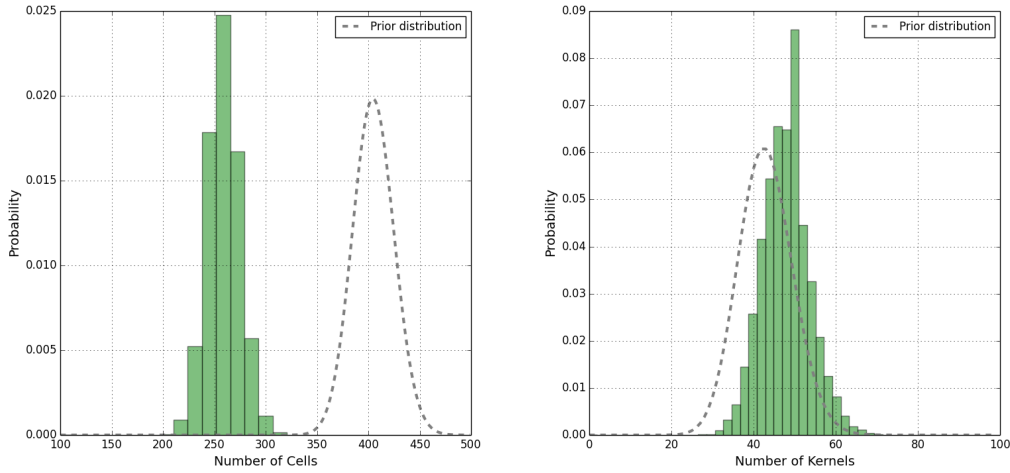


Figure 16: Histogram of the number of a) Johnson-Mehl cells b) Anisotropic Gaussian kernels

Finally, we display in Figure 17 the histogram of the differences between observed and posterior arrival times in the inversion. The posterior data are computed for the posterior mean model. Figure 17 shows that the posterior residuals are mostly less than the assumed value of data errors of 0.005s. Hence, the posterior mean model explains the observed data well for both parameterizations.

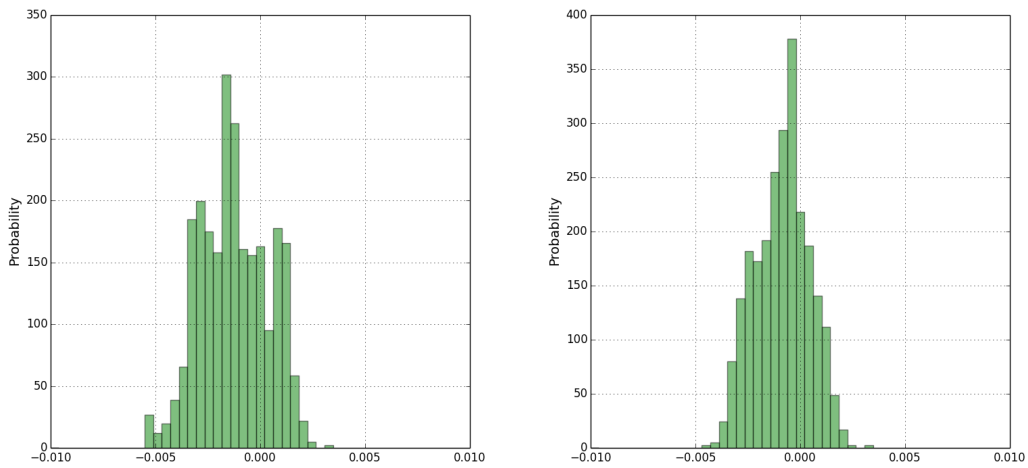


Figure 17: Histogram of the residuals a) Johnson-Mehl tessellation b) Anisotropic Gaussian kernels

7 Conclusion:

In this article we presented two new parsimonious parameterizations that can be used to represent the wave velocity model. The first parametric model uses Gaussian kernels as basis functions. The second one uses a random Johnson-Mehl tessellation. Johnson-Mehl tessellation contains cells that are not necessarily convex polytopes, therefore providing

smooth boundary between cells. The forward problem is solved by an Eikonal solver (Noble et al., 2014), which allows to obtain accurate first arrival times in strongly heterogeneous velocity models. The inversion is performed using a Bayesian approach coupled with a Reversible Jump Markov chain-Monte Carlo (RjMCMC). This algorithm allows simultaneous inference on both model and parameter space. Designing relevant reversible jump samplers is important to ensure an efficient trans-dimensional inversion. It is clear that the acceptance rate of the algorithm is linked to the choice of proposal schemes. Details of acceptance criterion are discussed in more details in (Green, 2003). (Sisson, 2005) provides efficient methods of constructing samplers, and convergence diagnostics. First, the tomography results were illustrated with a synthetic, horizontally layered velocity model, which contains strong velocity contrast and two small anomalies. Then, the methodology was applied to a realistic synthetic model that serves as a benchmark in the oil industry. The tomography results reveal the ability of our algorithm to map the velocity heterogeneities with precision using few parameters. However, the inversion lasted for approximately 3 days on a desktop computer for 100000 iterations. The algorithm is time-consuming because it requires to solve the forward direct problem at each iteration. This issue can be addressed by parallelization. Another way to reduce run-time is by reducing the number of iterations. This can be achieved by improving the mixing properties of the generated markov chain, for instance using interacting Markov chains (Romary, 2009; Dosso et al., 2014). These MCMC algorithms simulate several chains at the same time which exchange information.

Finally it should be noted that in this work, it was assumed that the standard deviation of the data noise σ_d was known. In practice this is not always the case. Data noise has a significant effect on the tomography results. For this reason, it has to be treated appropriately. When its value is poorly known, it must be considered as a parameter of the model. A way to tackle this lack of knowledge is to use Hierarchical Bayes formulation (Raudenbush and Bryk, 2002; Goldstein, 2011). Examples of its application in geophysics are (Malinverno and Briggs, 2004; Malinverno and Parker, 2006; Bodin et al., 2012; Agostinetti et al., 2015).

Bibliography

Agostinetti, N.P., and Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophys. J. Int.* *181*, 858–872.

Agostinetti, N.P., Giacomuzzi, G., and Malinverno, A. (2015). Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling. *Geophys. J. Int.* *201*, 1598–1617.

Bodin, T., and Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophys. J. Int.* *178*, 1411–1436.

Bodin, T., Sambridge, M., and Gallagher, K. (2009). A self-parametrizing partition model approach to tomographic inverse problems. *Inverse Probl.* *25*, 55009.

- Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. (2012). Transdimensional tomography with unknown data noise. *Geophys. J. Int.* *189*, 1536–1556.
- Chiu, S.N., Stoyan, D., Kendall, W.S., and Mecke, J. (2013). *Stochastic geometry and its applications* (John Wiley & Sons).
- Curtis, A., and Snieder, R. (1997). Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics* *62*, 1524–1532.
- Dettmer, J., Dosso, S.E., and Holland, C.W. (2010). Trans-dimensional geoacoustic inversion. *J. Acoust. Soc. Am.* *128*, 3393–3405.
- Dosso, S.E., Dettmer, J., Steininger, G., and Holland, C.W. (2014). Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation. *Inverse Probl.* *30*, 114018.
- Frost, H., Thompson, C., and Walton, D. (1992). Simulation of thin film grain structures—II. Abnormal grain growth. *Acta Metall. Mater.* *40*, 779–793.
- Gesret, A., Desassis, N., Noble, M., Romary, T., and Maisons, C. (2015). Propagation of the velocity model uncertainties to the seismic event location. *Geophys. J. Int.* *200*, 52–66.
- Goldstein, H. (2011). *Multilevel statistical models* (John Wiley & Sons).
- Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* *82*, 711–732.
- Green, P.J. (2003). Trans-dimensional markov chain monte carlo. *Oxf. Stat. Sci. Ser.* 179–198.
- Johnson, W.A., and Mehl, R.F. (1939). Reaction kinetics in processes of nucleation and growth. *Trans Aime* *135*, 396–415.
- Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophys. J. Int.* *151*, 675–688.
- Malinverno, A., and Briggs, V.A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics* *69*, 1005–1016.
- Malinverno, A., and Parker, R.L. (2006). Two ways to quantify uncertainty in geophysical inverse problems. *Geophysics* *71*, W15–W27.
- Matheron, G. (1975). *Random sets and integral geometry* (John Wiley & Sons).
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* *21*, 1087–1092.
- Micheline, A. (1995). An adaptive-grid formalism for travelttime tomography. *Geophys. J. Int.* *121*, 489–510.
- Møller, J. (1995). Generation of Johnson-Mehl crystals and comparative analysis of models for random nucleation. *Adv. Appl. Probab.* 367–383.

- Noble, M., Gesret, A., and Belayouni, N. (2014). Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophys. J. Int.* *199*, 1572–1585.
- Nolet, G., and others (2008). A breviary of seismic tomography. *Imaging Inter.*
- Okabe, A., Boots, B., Sugihara, K., and Chiu, S.N. (2009). *Spatial tessellations: concepts and applications of Voronoi diagrams* (John Wiley & Sons).
- Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Sage).
- Rawlinson, N., Pozgay, S., and Fishwick, S. (2010). Seismic tomography: a window into deep Earth. *Phys. Earth Planet. Inter.* *178*, 101–135.
- Romary, T. (2009). Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space. *Comput. Geosci.* *13*, 103–122.
- Sambridge, M., and Rawlinson, N. (2005). Seismic tomography with irregular meshes. *Seism. Earth Array Anal. Broadband Seism.* 49–65.
- Sambridge, M., Braun, J., and McQueen, H. (1995). Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophys. J. Int.* *122*, 837–857.
- Schneider, R., and Weil, W. (2008). *Stochastic and integral geometry* (Springer Science & Business Media).
- Sisson, S.A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. *J. Am. Stat. Assoc.* *100*, 1077–1089.
- Taillandier, C., Noble, M., Chauris, H., and Calandra, H. (2009). First-arrival traveltime tomography based on the adjoint-state method. *Geophysics* *74*, WCB1–WCB10.
- Tarantola, A., and Valette, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys.* *20*, 219–232.
- Wolk, C.P. (1975). Formation of one-dimensional patterns by stochastic processes and by filamentous blue-green algae. *Dev. Biol.* *46*, 370–382.

First Arrival Travel Time Tomography – Bayesian Approach

J. Belhadj* (Mines ParisTech), T. Romary (Mines ParisTech), A. Gesret (Mines ParisTech) & M. Noble (Mines ParisTech)

Summary

First arrival time tomography aims at determining the propagation velocity of seismic waves from experimental measurements of their first arrival time. This problem is usually ill-posed and is classically tackled by considering various iterative linearized approaches. However, these methods can yield wrong seismic velocity for highly nonlinear cases and they fail to estimate the uncertainties associated to the model. In our study, we rely on a Bayesian approach coupled with an interacting Markov chain-Monte Carlo (MCMC) algorithm to estimate the wave velocity and the associated uncertainties. The main difficulty associated to this approach is that traditional MCMC algorithms can be inefficient when multimodal probability distributions or complex velocity models involving a great number of parameters come into play. Therefore, a first step toward an efficient implementation of the Bayesian approach is to properly parameterize the model to reduce its dimension and to select adequate prior distribution for the parameters. In this paper, we present a ten layers probabilistic model for the velocity, that we illustrate on tomography results.

Introduction

First arrival time tomography aims at inferring the seismic wave's propagation velocity (the velocity model) from experimental first arrival time (the measurement). This is an inverse problem, and usually an ill-posed one, that would traditionally be solved using various iterative linearised approaches as for example in Taillandier et al., (2009). But if the problem is highly non-linear, these methods can yield wrong seismic velocities. They also fail to estimate reliably the uncertainties associated to the model.

To take into account uncertainty regarding experimental data, as well as the nonlinear relations between data and model parameters, geophysicist have taken up probabilistic methods see Tarantola et al., (1982), and especially Bayesian approach coupled with Markov chain-Monte Carlo (MCMC) algorithms. This approach involves quantifying the information provided by the data on the prior knowledge of the velocity model. Bayesian tomography allows for a better estimation of wave velocity as well as associated uncertainties, that can be propagated to uncertainties regarding the location of an earthquake's hypocentre as shown by Gesret et al., (2012).

However, with multimodal probability distributions or complex velocity models involving a great number of parameters, traditional MCMC algorithms can be inefficient: the Markov chain progresses slowly (it is said to mix badly) and has difficulty in moving. As a

consequence, it can be stuck in the vicinity of a local mode. To solve this problem, and to improve the sampling of the state space, numerous variations can be made to the traditional samplers. More recently, the scientific community has taken an interest in interacting Markov chain. These are MCMC algorithms that simulate several chains at the same time, allowing for information exchange between them.

Although high performance computing resources are easy to come by these days, the Bayesian approach remains fairly expensive, and MCMC algorithm are efficient only as long as the number of parameters remains within reason. This requires a careful reflection both on the parameterization of the velocity model, in order to reduce the problem's dimension, and on the definition of the prior distribution of the parameters. Here, we document a probabilistic approach to the velocity model and illustrate the tomography results with a ten layers model.

Theory

The velocity model to be inverted is defined as a set of parameters $\mathbf{m}=(\mathbf{m}_i)$ whose values characterize the system. These parameters are related to the measurements \mathbf{d}^{obs} by:

$$\mathbf{d}^{obs} = F(\mathbf{m}) + \varepsilon \quad (1)$$

Where ε is an additive noise (theoretical and experimental).

In a Bayesian approach all information is represented in a probabilistic term. This approach involves quantifying the information provided by the data on the prior knowledge of the velocity model.

Bayes' rule Bayes,(1763) is used to give the posterior probability density function.

$$P(\mathbf{m}|\mathbf{d}^{obs}) = \frac{P(\mathbf{d}^{obs}|\mathbf{m}) \times P(\mathbf{m})}{P(\mathbf{d}^{obs})} \quad (2)$$

$\mathbf{P}(\mathbf{d}^{obs} | \mathbf{m})$ is the likelihood function, which is the probability of observing the data given a model \mathbf{m} .

$$P(\mathbf{d}^{obs} | \mathbf{m}) \propto \exp \left(-\frac{1}{2} \left\| \frac{F(\mathbf{m}) - \mathbf{d}^{obs}}{\sigma_d} \right\|^2 \right) \text{ Gaussian likelihood}$$

σ_d is the estimated standard deviation of the data.

$\mathbf{P}(\mathbf{m})$ is *a priori* probability density of \mathbf{m} , that is, what is known about the model \mathbf{m} before collecting the data \mathbf{d}^{obs} .

$\mathbf{P}(\mathbf{d}^{obs})$ is the evidence. It does not depend on any particular state of the system. It can therefore be considered as a normalization constant. The problem to be solved thus expresses:

$$P(m|d^{obs}) \propto \exp\left(-\frac{1}{2} \left\| \frac{F(m) - d^{obs}}{\sigma_d} \right\|^2\right) \cdot P(m)$$

This is the Bayesian formulation of the inverse problem. To solve (4) we try to generate an ensemble of candidate solutions whose density is distributed according to the posterior probability density function. We focus here on Markov Chains Monte-Carlo methods for their universality and the ease of their implementation, for more details see Robert et al., (2004)

Metropolis-Hastings is one of widely used Markov Chain Monte Carlo (MCMC) methods see Metropolis et al., (1954). It requires the choice of a proposal distribution q . This proposal distribution allows to proposing potential transitions for the Markov chain. Given a current model m , a candidate m^* is accepted with probability $\alpha(m, m^*)$ defined as :

$$\alpha(m, m^*) = \min\left\{1, \frac{P(m^*|d^{obs}) \cdot q(m|m^*)}{P(m|d^{obs}) \cdot q(m^*|m)}\right\}$$

In practice, the use of classical MCMC methods is problematic in high dimensional model spaces especially when the posterior is high dimensional and/or multimodal. Many techniques have been proposed to improve the sampling of the state space including among others parallel interacting Markov Chains introduced by Geyer (1991).

The principle of parallel interacting Markov Chains is to simulate several chains, each at different temperature and to allow exchange information, swapping their current states.

At each iteration n , a proposal is made to swap the parameters of a pair of adjacent chains i and $i + 1$. The proposal is accepted according to some probability θ .

- If the swap is rejected, a classical MH transition is performed.
- Otherwise, a state $m^{(i+1,j)}$ $j < n$ is proposed among all states generated by the chain $i+1$ in the past. The exchange is then accepted with probability :

$$\beta(m^{(i,n)}, m^{(i+1,j)}) = \min\left(1, \frac{r^{(l)}(m^{(i+1,j)})}{r^{(l)}(m^{(i,n)})}\right)$$

Where: $r^{(l)}(m) = \frac{P_l(m|d^{obs})}{P_{l+1}(m|d^{obs})}$ called the importance function.

As the temperature increases, the tempered distributions becomes flatter, the high temperatures chains can sample coarsely large volumes of the model space. Low temperature chains can then, by means of swaps access to states that they could not achieve if these exchanges did not take place. Therefore the first chain associated with the target distribution will be able to escape from local modes.

Inversion of Synthetic Data

In this section, we show inversion results using Interacting MCMC for a synthetic 2D velocity model. The exact model is composed of 23 layers where the velocity is either constant or increasing with depth as shown in figure 1. The observed data consists of P first travel times recorded at 48 positions from a set of 10 calibration shots as indicated by the red squares (shots) and black triangles (receivers) in Figure 2. The domain is 1200m in depth and 600m laterally. It is discretized on a rectangular grid (nz=240, nx=60).

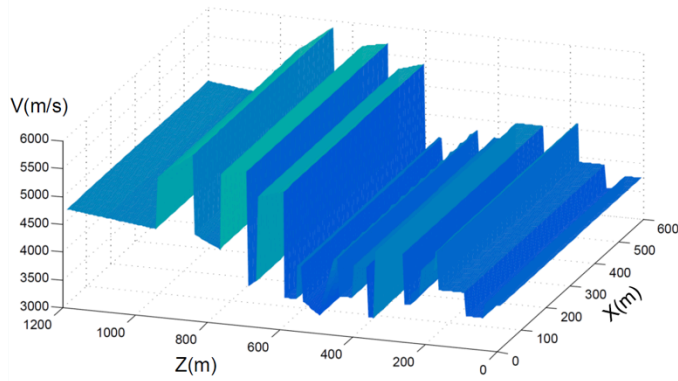


Figure 1: Exact velocity model for 2D inversion the receivers, the red squares are the sources

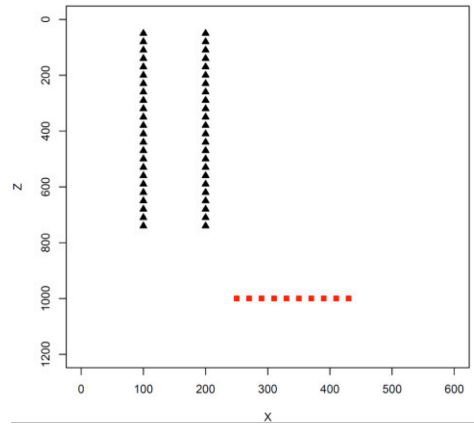


Figure 2: Geometrical setting. The black triangles represent the receivers, the red squares are the sources

Parameterization chosen: The parameterization consists of a 1-D P wave velocity profile as a function of depth made from 10 layers. Each layer has a velocity parameter and an unknown interface depth.

$$\mathbf{m} = (z_1, \dots, z_{10}, v_p^1, \dots, v_p^{10})$$

The prior: We assume a priori that the velocities and the depth of the interfaces are independent. The a priori PDF is uniform in all variables: For each parameter m_i we define an interval $[m_{i\min}, m_{i\max}]$ on which m_i can take any value with the same probability.

The Interacting MCMC algorithm has been run for 50 000 iterations, with five chains. We construct the temperature ladder distributing geometrically between $T_0 = 1$ and $T_4 = 3000$. The chosen features of the run are shown in table 1. We note $\sigma_{\mathbf{m}} = (\sigma_{m_1}, \sigma_{m_2}, \dots, \sigma_{m_{20}})$ and $\sigma_{\mathbf{m}}^2 = (\sigma_{m_1}^2, \sigma_{m_2}^2, \dots, \sigma_{m_{20}}^2)$.

Number $K + 1$ of chains	5				
Probability θ of proposing a swap	25%				
Sequence $(T)_i$ of temperatures	:				
	$T_0=1$	$T_1 = 3000^{1/4}$	$T_2 = 3000^{1/2}$	$T_3 = 3000^{3/4}$	$T_{max} = 3000$
Samplers at each temperature	SIMH	SIMH	SIMH	SIMH	IMH
$q(m^*lm)$	$N(m, I_{20} \sigma_m^2)$	$N(m, I_{20} \sigma_m^2)$	$N(m, I_{20} \sigma_m^2)$	$N(m, I_{20} \sigma_m^2)$	$P(m)$
σ_m	Linear decreasing function of T : $aT+b$				

Numerical results:

Figure 3 shows the energy of the states generated by the five chains in function of the number of the iteration. It also shows that each couple of chains at adjacent temperatures has an overlapping energy profiles, allowing for information exchange between them.

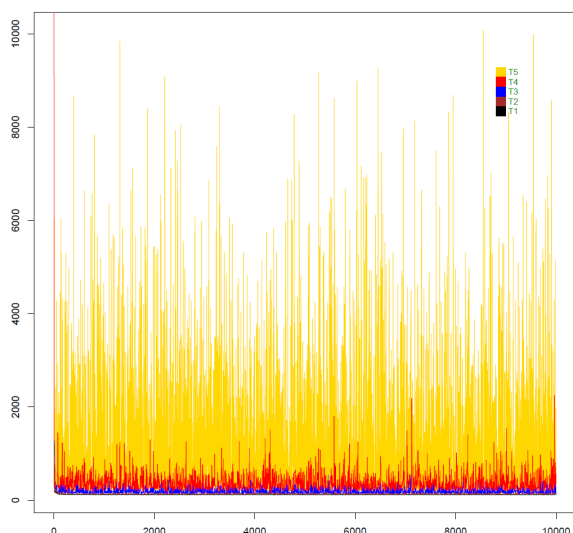


Figure 3: Energy of the states of the five chains

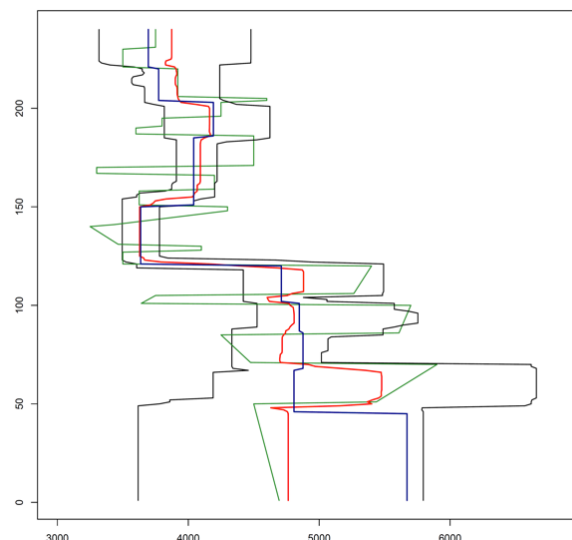


Figure 4: Posterior mean profile (red), 95 % confidence intervals (black), MAP model (blue) and the exact model (green)

Fig. 4 shows the vertical profile for $x=100m$. The posterior mean profile, the MAP model, the exact model and 95 % confidence intervals are shown. These intervals enable the identification of the regions where velocities are well or poorly constrained.

Conclusion :

We have described an Interacting MCMC algorithm for Bayesian inversion. The key element is to sample the a posteriori distribution using several chains at the same time, allowing for information exchange between them. These provide a good exploration of the state space by the target chain. This approach allows for a better estimation of wave velocity as well as associated uncertainties.

References :

Bayes, M., and M. Price, 1763. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS, Philosophical Transactions (1683-1775), pp. 370–418,.

Gesret, A., N. Dessassis and T.Romary, 2012. Microseismic Monitoring - Consequences of Velocity Model Uncertainties on Event Location Uncertainties, presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013, United Kingdom.

Geyer, C. J., 1991. Markov chain Monte Carlo maximum likelihood.

Robert, C. P., and G. Casella, 2004. Monte Carlo statistical methods, vol. 319. Citeseer.

Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009. First-arrival traveltime tomography based on the adjoint-state method, Geophysics, vol. 74, no. 6, pp.

Tarantola, A., and B. Valette, 1982. Generalized nonlinear inverse problems solved using the least squares criterion, Reviews of Geophysics, vol. 20, no. 2, pp. 219–232.

Bibliographie

Agostinetti, N.P., and Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophys. J. Int.* *181*, 858–872.

Agostinetti, N.P., Giacomuzzi, G., and Malinverno, A. (2015). Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling. *Geophys. J. Int.* *201*, 1598–1617.

Aki, K., and Lee, W. (1976). ARRAY USING FIRST P ARRIVAL TIMES FROM LOCAL EARTHQUAKES. *J. Geophys. Res.*

Aki, K., and Richards, P.G. (1980). *Quantitative Seismology: Theory and Methods* (WH Freeman and Company, San Francisco).

Andrieu, C., and Thoms, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* *18*, 343–373.

Aster, R., Borchers, B., and Thurber, C. (2005). Preface. *Int. Geophys.* *90*, xi–xii.

Atchadé, Y.F., and Perron, F. (2007). On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics* *41*, 77–84.

Bayes, M., and Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, first communicated by mr. price, in a letter to john canton, amfrs. *Philos. Trans. 1683-1775* 370–418.

Bertero, M., and Boccacci, P. (1998). *Introduction to inverse problems in imaging* (CRC press).

Bodin, T., and Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophys. J. Int.* *178*, 1411–1436.

Bodin, T., Sambridge, M., and Gallagher, K. (2009). A self-parametrizing partition model approach to tomographic inverse problems. *Inverse Probl.* *25*, 55009.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P. (2012a). Transdimensional tomography with unknown data noise. *Geophys. J. Int.* *189*, 1536–1556.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., and Rawlinson, N. (2012b). Transdimensional inversion of receiver functions and surface wave dispersion. *J. Geophys. Res. Solid Earth* *117*.

- Bottero, A., Gesret, A., Romary, T., Noble, M., and Maisons, C. (2016). Stochastic seismic tomography by interacting Markov chains. *Geophys. J. Int.* 207, 374–392.
- Brus, D.J., and Heuvelink, G.B. (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138, 86–95.
- Candes, E.J., Romberg, J.K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* 59, 1207–1223.
- Candès, E.J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* 52, 489–509.
- Chiu, S.N., Stoyan, D., Kendall, W.S., and Mecke, J. (2013). *Stochastic geometry and its applications* (John Wiley & Sons).
- Curtis, A., and Snieder, R. (1997). Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics* 62, 1524–1532.
- Daubechies, I., and others (1992). *Ten lectures on wavelets* (SIAM).
- Davenport, M.A., Duarte, M.F., Eldar, Y.C., and Kutyniok, G. (2011). *Introduction to compressed sensing*. Preprint 93, 2.
- Dettmer, J., Dosso, S.E., and Holland, C.W. (2010). Trans-dimensional geoacoustic inversion. *J. Acoust. Soc. Am.* 128, 3393–3405.
- Dettmer, J., Molnar, S., Steininger, G., Dosso, S.E., and Cassidy, J.F. (2012). Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models. *Geophys. J. Int.* 188, 719–734.
- Donoho, D.L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* 52, 1289–1306.
- Fisher, R.A. (1956). *Statistical methods and scientific inference*.
- Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 721–741.
- Gesret, A., Desassis, N., Noble, M., Romary, T., and Maisons, C. (2015). Propagation of the velocity model uncertainties to the seismic event location. *Geophys. J. Int.* 200, 52–66.
- Geyer, C.J. (1991). Markov chain Monte Carlo maximum likelihood.
- Geyer, C.J., and Thompson, E.A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Assoc.* 90, 909–920.
- Gilbert, E. (1962). Random subdivisions of space into crystals. *Ann. Math. Stat.* 33, 958–972.

- Goldstein, H. (2011). *Multilevel statistical models* (John Wiley & Sons).
- Graps, A. (1995). An introduction to wavelets. *IEEE Comput. Sci. Eng.* 2, 50–61.
- Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
- Green, P.J., and Mira, A. (2001). Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* 88, 1035–1053.
- Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.* 14, 375–396.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* 223–242.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hawkins, R., and Sambridge, M. (2015). Geophysical imaging using trans-dimensional trees. *Geophys. J. Int.* 203, 972–1000.
- Iba, Y. (2001). Population Monte Carlo algorithms. *人工知能学会論文誌* 16, 279–286.
- Improta, L., Zollo, A., Herrero, A., Frattini, R., Virieux, J., and Dell’Aversana, P. (2002). Seismic imaging of complex structures by non-linear traveltime inversion of dense wide-angle data: application to a thrust belt. *Geophys. J. Int.* 151, 264–278.
- Jasra, A., Stephens, D.A., and Holmes, C.C. (2007). On population-based simulation for static inference. *Stat. Comput.* 17, 263–279.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, (The Royal Society), pp. 453–461.
- Jeffries, H. (1961). *Theory of probability* (Clarendon Press, Oxford).
- Johnson, W.A., and Mehl, R.F. (1939). Reaction kinetics in processes of nucleation and growth. *Trans Aime* 135, 396–415.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., and others (1983). Optimization by simulated annealing. *Science* 220, 671–680.
- Kou, S., Zhou, Q., and Wong, W.H. (2006). Discussion paper equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Stat.* 1581–1619.
- Laplace, P.S. (1921). *Essai philosophique sur les probabilités* (Gauthier-Villars).

- Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophys. J. Int.* *151*, 675–688.
- Malinverno, A., and Briggs, V.A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics* *69*, 1005–1016.
- Malinverno, A., and Parker, R.L. (2006). Two ways to quantify uncertainty in geophysical inverse problems. *Geophysics* *71*, W15–W27.
- Mallat, S. (1999). *A wavelet tour of signal processing* (Academic press).
- Marinari, E., and Parisi, G. (1992). Simulated tempering: a new Monte Carlo scheme. *EPL Europhys. Lett.* *19*, 451.
- Matheron, G. (1975). *Random sets and integral geometry* (John Wiley & Sons).
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* *21*, 1087–1092.
- Micheline, A. (1995). An adaptive-grid formalism for traveltimes tomography. *Geophys. J. Int.* *121*, 489–510.
- Milson, J., and Eriksen, A. (2011). *Field geophysics* (John Wiley & Sons).
- Møller, J. (1992). Random Johnson-Mehl tessellations. *Adv. Appl. Probab.* 814–844.
- Mosegaard, K., and Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *J. Geophys. Res. Solid Earth* *100*, 12431–12447.
- Noble, M., Gesret, A., and Belayouni, N. (2014). Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophys. J. Int.* *199*, 1572–1585.
- Nolet, G. (1985). Solving or resolving inadequate and noisy tomographic systems. *J. Comput. Phys.* *61*, 463–482.
- Nolet, G., and others (2008). *A breviary of seismic tomography*. Imaging Inter.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* *12*, 2825–2830.
- Podvin, P., and Lecomte, I. (1991). Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophys. J. Int.* *105*, 271–284.
- Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Sage).

- Rawlinson, N., Pozgay, S., and Fishwick, S. (2010). Seismic tomography: a window into deep Earth. *Phys. Earth Planet. Inter.* *178*, 101–135.
- Robbins, H., and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* 400–407.
- Robert, C. (2006). *Le choix bayésien: Principes et pratique* (Springer Science & Business Media).
- Robert, C.P.C.P., and Casella, G. (2004). *Monte Carlo statistical methods*.
- Roberts, G.O., Gelman, A., Gilks, W.R., and others (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* *7*, 110–120.
- Romary, T. (2009). Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space. *Comput. Geosci.* *13*, 103–122.
- Romary, T. (2010). Bayesian inversion by parallel interacting markov chains. *Inverse Probl. Sci. Eng.* *18*, 111–130.
- Sambridge, M. (2013). A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophys. J. Int.* ggt342.
- Sambridge, M., and Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Rev. Geophys.* *40*.
- Sambridge, M., and Rawlinson, N. (2005). Seismic tomography with irregular meshes. *Seism. Earth Array Anal. Broadband Seism.* 49–65.
- Sambridge, M., Braun, J., and McQueen, H. (1995). Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophys. J. Int.* *122*, 837–857.
- Schneider, R., and Weil, W. (2008). *Stochastic and integral geometry* (Springer Science & Business Media).
- Sheriff, R.E., and Geldart, L.P. (1995). *Exploration seismology* (Cambridge university press).
- Taillandier, C., Noble, M., Chauris, H., and Calandra, H. (2009). First-arrival traveltimes tomography based on the adjoint-state method. *Geophysics* *74*, WCB1–WCB10.
- Tarantola, A. (1987). *Inverse problem theory: Methods for data fitting and parameter estimation* (Elsevier, Amsterdam).
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (siam).
- Tarantola, A., and Valette, B. (1982). Generalized nonlinear inverse problems solved using

the least squares criterion. *Rev. Geophys.* 20, 219–232.

Tikhonov, A.N., and Arsenin, V.Y. (1977). *Solutions of ill-posed problems.*

Van Lieshout, M. (2000). *Markov point processes and their applications* (World Scientific).

Vidale, J. (1988). Finite-difference calculation of travel times. *Bull. Seismol. Soc. Am.* 78, 2062–2076.

Zelt, C.A., Azaria, A., and Levander, A. (2006). 3D seismic refraction travelttime tomography at a groundwater contamination site. *Geophysics* 71, H67–H78.

Liste des figures

Figure 1: Réfraction pour un terrain à deux couches horizontales.....	18
Figure 2: Mouvements des particules d'un milieu dans lequel se propagent les ondes de volume : a) compression b) cisaillement.....	20
Figure 3: Distributions tempérées.	38
Figure 4: Exemple papier (Gesret et al, 2015).	46
Figure 5: Exemple papier (Agostinetti and Malinverno, 2010)	47
Figure 6 Cellules de Voronoï	47
Figure 7: Exemple ondelette de Haar	50
Figure 8: Exemple ondelette de Daubechies	50
Figure 9: étapes de construction de la mosaïque de Johson-Mehl	53
Figure 10: Schéma résumant les principales étapes de l'algorithme, à gauche: l'état actuel de la chaîne, les figures au centre: modèle de vitesse construit à partir des paramètres proposés, à droite: modèle de vitesse retenu après acceptation/rejet	58
Figure 11: Exemple a) anomalie négative (lente) b) anomalie positive(rapide)	59
Figure 12: Schéma résumant les principales étapes de l'algorithme, à gauche: l'état actuel de la chaîne, les figures au centre: modèle de vitesse construit à partir des paramètres proposés, à droite: modèle de vitesse retenu après acceptation/rejet	63
Figure 13 : Vrai modèle.	65
Figure 14: Taux d'acceptation moyen en fonction des différentes valeurs de θ	67
Figure 15: Taux d'acceptation moyen en fonction des différentes inputs	70
Figure 16: Vrai modèle	71
Figure 17 Johnson-Mehl : a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)	73
Figure 18 Noyaux gaussiens: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type).....	73

Figure 19 Johnson-Mehl: Sections transversales. (a)à Z= 30 km . (b) à Z=70 km. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%	74
Figure 20 Noyaux gaussiens isotropes : Sections transversales. (a)à Z= 30 km . (b) à Z=70 km. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%. 74	74
Figure 21 Histogramme du nombre de of a) Cellules de Johnson Mehl b) Noyaux gaussiens 75	75
Figure 22 Histogramme des résidus a) Mosaïque de Johnson-Mehl b) Noyaux gaussiens	75
Figure 23: Meilleur modèle pour la mosaïque de Johnson-Mehl.....	76
Figure 24 : Vrai modèle.	77
Figure 25: Carte des temps de premières arrivées.....	77
Figure 26 Johnson-Mehl : a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type)	79
Figure 27 Noyaux gaussiens anisotropes: a) Carte de la moyenne a posteriori b) Carte des incertitudes (écart-type).....	80
Figure 28 : Johnson-Mehl: Sections transversales. (a)à Z= 25m . (b) à Z=65m. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%.....	81
Figure 29 Noyaux gaussiens anisotropes: Sections transversales. (a)à Z= 25m . (b) à Z=65m. Vert : modèle moyen, rouge : le vrai modèle, noir : intervalle de confiance à 95%.....	81
Figure 30 Histogramme du nombre de of a) Cellules de Johnson Mehl b) Noyaux gaussiens anisotropes.....	82
Figure 31 Histogramme des résidus en (ms) a) Mosaïque de Johnson-Mehl b) Noyaux gaussiens anisotropes	82
Figure 32: Vrai modèle a) une vue 3D b) profile vertical de vitesse en fonction de la profondeur	85
Figure 33: Géométrie d'acquisition	86
Figure 34: Carte des temps observés.....	86
Figure 35: Réalisations à partir de la loi a priori.....	88
Figure 36: Résultats pour 50000 itérations. Vert: modèle moyen, rouge: vrai modèle, noir: intervalle de confiance à 95%	89
Figure 37: Énergie de la chaîne.....	89

Figure 38: Résultats pour 10000 itérations. Vert: modèle moyen, rouge: vrai modèle, noir: intervalle de confiance à 95%	92
Figure 39: a) Énergies des 5 chaînes b) Énergie pour la chaîne T0.....	92
Figure 40: a) Histogrammes de la vitesse dans la dernière couche. b) Les énergies pour le cas classique et pour les chaînes en interaction.....	93
Figure 41: Résidus des temps de première arrivée pour tous les couples source-récepteur	93
Figure 42: Log sismique des ondes P.....	94
Figure 43: carte des temps de premières arrivées des ondes P.....	95
Figure 44: Résultats pour 500000 itérations. Vert: modèle moyen, rouge: vrai modèle, gris: intervalle de confiance à 95%	98
Figure 45 : a) Histogramme du nombre de couches a posteriori b) Histogramme des résidus.	99
Figure 46: Résultats pour 500000 itérations. Vert: modèle moyen, rouge: vrai modèle, gris: intervalle de confiance à 95%	100
Figure 47: Histogramme de σ_d	101
Figure 48: a) Nombre de couches pour le modèle hiérarchique, b) Histogramme des couches pour le modèle hiérarchique (vert) et pour σ_d connu (rouge).....	101
Figure 49: Histogramme des résidus.....	102
Figure 50: Contrainte L1 et solution parcimonieuse.....	109
Figure 51: Champ de vitesse test, de taille 95×19 (haut). Reconstruction à partir de $m = 180$ valeurs du champ de vitesse (bas). On note que l'algorithme permet d'obtenir une reconstruction quasi-parfaite du champ de vitesses initial.....	111
Figure 52: Vrai modèle	113
Figure 53: Position des points sélectionnés (en rouge) pour reconstruire le champ de vitesse	113
Figure 54: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type). 114	
Figure 55: Vrai modèle	115
Figure 56: Carte des temps de première arrivée.....	115
Figure 57: a) La carte de la moyenne a posteriori b) La carte des incertitudes (écart-type). 116	

Liste des tableaux

Tableau 1 : Moyenne a posteriori, Ecart-type a posteriori et Pndobs pour différentes valeurs de θ_D	66
Tableau 2: Moyenne a posteriori, Ecart-type a posteriori et Pndobs pour différentes valeurs de θ_D à $\sigma_m = 6$	68
Tableau 3: Moyenne a posteriori, Ecart-type a posteriori et Pndobs pour différentes valeurs de σ_m à $\theta_D = 5$	69
Tableau 4: Lois a priori sur les paramètres du modèle m: Mosaïque de Johnson-Mehl.....	71
Tableau 5: Lois a priori sur les paramètres du modèle m: Noyaux gaussiens isotropes.....	72
Tableau 6 : Les Lois a priori sur les paramètres du modèle m: Mosaïque de Johnson-Mehl. .	78
Tableau 7: Les Lois a priori sur les paramètres du modèle m: Noyaux gaussiens anisotropes.	78
Tableau 8: Les vitesses (m/s) et les profondeurs (m).....	87

La tomographie des temps de première arrivée vise à retrouver un modèle de vitesse de propagation des ondes sismiques à partir des temps de première arrivée mesurés. Cette technique nécessite la résolution d'un problème inverse afin d'obtenir un modèle sismique cohérent avec les données observées. L'approche bayésienne permet d'estimer la distribution spatiale de la vitesse de propagation des ondes sismiques. Il en résulte une meilleure quantification des incertitudes associées. Cependant l'approche reste relativement coûteuse en temps de calcul, les algorithmes de Monte Carlo par chaînes de Markov (MCMC) classiquement utilisés pour échantillonner la loi a posteriori des paramètres n'étant efficaces que pour un nombre raisonnable de paramètres. Elle demande, de ce fait, une réflexion à la fois sur la paramétrisation du modèle de vitesse afin de réduire la dimension du problème et sur la définition de la loi a priori des paramètres. Le sujet de cette thèse porte essentiellement sur cette problématique.

Le premier modèle que nous considérons est basé sur un modèle de mosaïque aléatoire, le modèle de Johnson-Mehl, dérivé des mosaïques de Voronoï déjà proposées en tomographie bayésienne. Le deuxième modèle est, quant à lui, décrit par une combinaison linéaire de fonctions gaussiennes, centrées sur la réalisation d'un processus ponctuel de Poisson. Pour chaque modèle, nous présentons un exemple de validation sur des champs de vitesse simulés. Nous appliquons ensuite notre méthodologie à un modèle synthétique plus complexe qui sert de benchmark dans l'industrie pétrolière. Nous proposons enfin, un modèle de vitesse basé sur la théorie du compressive sensing pour reconstruire le champ de vitesse. Ce modèle, encore imparfait, ouvre plusieurs pistes de recherches futures.

Dans ce travail, nous nous intéressons également à un jeu de données réelles acquises dans le contexte de la fracturation hydraulique. Nous développons dans ce contexte une méthode d'inférence bayésienne trans-dimensionnelle et hiérarchique afin de traiter efficacement la complexité du modèle à couches.

Mots Clés

Tomographie sismique, problème inverse, approche bayésienne, paramétrisation, MCMC, trans-dimensionnel, hiérarchique

First arrival time tomography aims at inferring the seismic wave propagation velocity using experimental first arrival times. In our study, we rely on a Bayesian approach to estimate the wave velocity and the associated uncertainties. This approach incorporates the information provided by the data and the prior knowledge of the velocity model. Bayesian tomography allows for a better estimation of wave velocity as well as associated uncertainties. However, this approach remains fairly expensive, and MCMC algorithms that are used to sample the posterior distribution are efficient only as long as the number of parameters remains within reason. Hence, their use requires a careful reflection both on the parameterization of the velocity model, in order to reduce the problem's dimension, and on the definition of the prior distribution of the parameters. In this thesis, we introduce new parsimonious parameterizations enabling to accurately reproduce the wave velocity field with the associated uncertainties.

The first parametric model that we propose uses a random Johnson-Mehl tessellation, a variation of the Voronoï tessellation. The second one uses Gaussian kernels as basis functions. Each anomaly is considered to be a linear combination of these basis functions localized at the realization of a Poisson point process. We first illustrate the tomography results with a synthetic velocity model, which contains two small anomalies. We then apply our methodology to a more advanced and more realistic synthetic model that serves as a benchmark in the oil industry. Finally, we propose a new parametric model based on the compressed sensing techniques. The first results are encouraging. However, the model still has some weakness related to the uncertainties estimation.

In addition, we analyse real data in the context of induced microseismicity. In this context, we develop a trans-dimensional and hierarchical approach in order to deal with the full complexity of the layered model.

Keywords

Seismic tomography, inverse problem, Bayesian approach, parameterization, MCMC, transdimensional, hierarchical.