Tomographie des temps d'arrivée au-delà de la théorie des rayons

Raffaella Montelli

UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS - UFR SCIENCES
L'ÉCOLE DOCTORALE SCIENCES FONDAMENTALES ET APPLIQUEES

THÈSE

Présentée pour obtenir le titre de

Docteur en SCIENCES
de l'Université Nice-Sophia Antipolis

Discipline: Sciences
Mention: Sismologie

Raffaella MONTELLI

Tomographie des temps d'arrivée au-delà de la théorie des rayons.

A soutenir le 16 décembre devant le jury composé de

Jean Virieux (Professeur, UNSA)	Président
Guust Nolet (Professeur, Princeton University)	Directeur de thèse
Vincent Courtillot (Professeur, IPGP)	Rapporteur
Jean-Paul Montagner(Professeur, IPGP)	Rapporteur
Sébastien Chevrot(Chargé de Recherche, CNRS)	Examinateur

à 14h00 à Sophia Antipolis

UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS - UFR SCIENCES
L'ÉCOLE DOCTORALE SCIENCES FONDAMENTALES ET APPLIQUEES

THESIS

Submitted in accordance with the requirements for the degree of

Doctor of PHILOSOPHY
The Université Nice-Sophia Antipolis

Deparment of Earth Science
Speciality: Seismology

Raffaella MONTELLI

Seismic Tomography Beyond Ray Theory

The viva voce will take place on the 16th of December in front of the jury

Jean Virieux (Professor, UNSA)	President
Guust Nolet (Professor, Princeton University)	Thesis Advisor
Vincent Courtillot (Professor, IPGP)	Reader
Jean-Paul Montagner(Professor, IPGP)	Reader
Sébastien Chevrot(Researcher, CNRS)	Examiner

14h00, Sophia Antipolis

Résumé

La théorie de la diffraction prévoit que les temps de trajet des ondes avec une fréquence dominante finie sont affectés seulement par des hétérogénéités dans un volume autour du rayon obtenu par la théorie asymptotique fréquence infinie. Ce volume s'appelle la zone de Fresnel. À cause de la diffraction qui assure la continuité du front d'onde au travers des hétérogénéités, les hétérogénéités de taille inférieure à celle de la zone de Fresnel n'ont qu'un effet réduit sur le temps de trajet de l'onde arrivant à une station. Dahlen et al. (2000) montrent qu'il est possible de corriger des effets de cette diffraction par une méthode de tomographie volumique qui introduit de nouveau la notion de contenu fréquentiel dans l'estimation des temps de trajet et donc dans la reconstruction du milieu.

Dans cette thèse, j'étudie l'influence de cette technique d'imagerie à fréquence finie en tomographie globale sismologique des temps d'arrivée. Pour réduire les effets dû à la distribution des rayons dans la Terre, distribution qui n'est pas uniforme, j'introduis un maillage non-structuré pour représenter les vitesses dans la Terre.

Pour définir l'influence des phénomènes de diffraction dans les résultats tomographiques, j'utilise un nombre limité de données à longue période de très haute qualité obtenues par Guy Masters (Scripps). Je fais l'inversion en utilisant l'approche classique de tomographie des temps se fondant sur la théorie des rayons et l'approche nouvelle de tomographie des temps en se fondant sur l'approche à fréquence finie. La comparaison des deux tomographies montre que les anomalies de vitesse sont sous-estimées par la théorie des rayons d'un facteur 30–60% en fonction de la profondeur. C'est bien le signe que la théorie des rayons est une approximation trop grossière pour les données à basse fréquence que l'on utilise.

Pour augmenter la résolution de mon modèle tomographique, j'inverse conjointement les données à longue période déjà mentionnées avec des données haute-fréquence retraitées par Bob Engdahl (Boulder).

L'analyse de la compatibilité des deux bases de données est nécessaire avant de procéder à l'inversion conjointe. L'étude des temps eux-même (en fait, des écarts dans un modèle de référence) et les inversions des deux jeux de données independamment montrent un très bon niveau de compatibilité.

Tous les modèles obtenus montrent une très forte corrélation entre les anomalies de basse vitesse et la position d'un très grand nombre de points chauds déjà connus. Pour estimer la validité de ces anomalies, des tests de résolution ont été réalisés de façon à identifier la profondeur des panaches. À la suite de ces tests, on peut conclure positivement que des panaches profonds se trouvent au-dessous des structures suivantes: Ascension, Azores, Canaries, Easter, Hawaï, Samoa et Tahiti. Pour d'autres, l'origine profonde n'est pas confirmée et une origine plus superficielle est probable autour de 660 km comme la zone islandaise. D'autres n'ont jamais été vus avant comme au Seychelles et dans l'Ocean Indien. Enfin, dans la zone de la Mer de Coral et au Sud de Java, la tomographie a détectée des

panaches montant de la frontière entre le noyau et le manteau mais qui n'arrivent pas encore à la surface.

Abstract

Diffraction theory predicts that traveltimes of seismic waves with a finite dominant frequency are affected by heterogeneities within a narrow region around the ray-theoretical path, called Fresnel zone. Because of wavefront healing, heterogeneities whose scale-length is smaller than the size of the Fresnel zone have a reduced effect on the traveltime of a wave. Dahlen et al. (2000) show how is possible to correct for effects of wavefront healing using a new method, referred to as finite-frequency modeling.

In this thesis, we investigate the effects of using finite-frequency modeling in global seismic tomography. To reduce the consequence of the heterogeneous ray path coverage of the earth we also introduce an irregular model parameterization.

To establish the influence of diffraction phenomena on the final tomographic image, we use a high quality long-period data set obtained from Prof Masters (Scripps), that we invert both with standard ray theory and by using finite-frequency modeling. The comparison of the two tomographic images shows that the velocity anomalies are underestimated by $30 - 60\%$ when interpreted with classical ray theory.

In order to enhance the resolution of the tomographic image even further, we combine the long period arrival times, with the very best high-frequency measurements contained in Prof Engdahl's NEIC data set. A comparison of both data sets, as well as the results from separate inversions, show a high level of compatibility.

All the models show a strong correlation between low velocity anomalies and the location of a large number of known hotspots. The reliability of these anomalies was confirmed by extensive resolution analysis. As a result, we can confidently say that deep mantle plumes are located beneath Ascension, Azores, Canary, Easter, Hawaii, Samoa and Tahiti. Many others, among which is Iceland, are of more shallow origin. Newly discovered plume-like features are clearly visible beneath the Mid-Atlantic Ridge, the South-east Indian ridge, Seychelles and the Coral Sea.

Acknowledgments

Despite the interest that all the chapters of a thesis might arouse in a reader, we all know that the 'Acknowledgments' section is usually the most read; the section where everyone goes straight to look for who has been mentioned, and mainly to look for his/her name. Well, I am sure I have forgotten someone, but all of you have had a part during the years I spent here and there by working on my thesis. So if you don't find your name here don't be upset, because you are a part of it anyway!

First of all I would like to thank my two advisors: Guust Nolet in Princeton and Jean Virieux in Sophia Antipolis, France. Without them this thesis wouldn't have ever been possible. I am very grateful to both of you for having shared with me your deep knowledge; for having been always positive and supportive of me and for having always found the time in your busy schedules to discuss with me. Thank you also for having given me the chance to gain an international experience by allowing me to work both in France and in the USA.

Special thanks go to Tony Dahlen, this thesis would never have become so successful without his *banana-doughnut* kernels. But also I wish to thank you for your valuable discussions and explanations, and for your ever encouraging comments about my work.

I am also profoundly indebted to Guy Masters. Without his carefully processing and very fast reprocessing of the data this research would never have been possible. I would like also to thank Guy for the time I spent in San Diego collecting the data, for the patience he took in showing me how to process them, and for always answering promptly to my e-mails. Thank you and your wife also for the nice hospitality that you showed me during my stay there.

A particular thank you goes to Jason Morgan for the time he spent with me discussing about plumes and for your enthusiasm about my plumes!

I am grateful to Malcom Sambridge and Shu-Huei Hung for providing me with their softwares; parts of which are now pillars of my own software. Thanks also to Bob Engdahl for letting me use his latest reprocessed ISC data set and for his precious help.

I wish also to thank all other professors and researchers, present in both laboratories or with which I have been in contact, for their helpful and stimulating discussions, and for being an everyday example.

I want to thank the European Community, Princeton University, and the Department of Geosciences for the funding they paid for my studies...and not just for this.

As far as France goes, a person who deserves a special thank you is Bernhard. There are no words to describe how indebted I am to you. Without your help I cannot image how I could have survived solving all the French bureaucratic issues from Princeton. A special thank goes next to the secretaries in Sophia: Reine, Nathalie, Isabelle, Suzanne. Thanks also to all the Sophia/Villefranche crowds for

those great moments shared together...well, the World Cup 1998 and the European Championship 2000 were not great moment at all, but with all of you around they are still nice memories. In alphabetical order: Alessandra, Christophe and Marie Chantal, Diana, Etienne, Fred, Jean-Michel and Sandra, Jhi-Chuan, Loic, Marianne, Olivia, Olivier L., Olivier S. and Veronique.

As far as Princeton goes, special thanks go to Debbie (Smith): from the first day I met you, you made me always feel like I was at home in the department, thanks! And I am deeply indebted also to Laurie (Wanat), Mary Ann (Nicoletti), Nancy (Janos), Trish (Killian), Scheryl (Rickwell), Stephanie (Resko), and Robin (Pispecky). Thank you also to Bill Mott and Brian Mohr for the help dealing with my computers. Another special thank you goes to Laurel (Goodell) for having introduced me to the magical world of teaching. And a final thought goes to all the *princetonians* that have shared the last four years of my life, the good and the bad, in and out of the lab, I will always remember you ALL; but in particular Alex, Andrea, Clarisse, Francesco, Janette, Jonathan, Marianne and Christoph, Richard, Sara, Sergio, Sigal, Till and Tracey.

And finally in Italy I wish to thank for their continuous support during all these years through e-mails, and phone calls but mainly through our long friendship Chicca, Andrea U., Laura P. and Francesca Di Luccio.

I wish to thank also Suor Paola, a nun of strict enclosure, and Padre Savio, a missionary in Brazil, for our special friendship; but mainly for their amazing personality that inspire me in my everyday life.

Last, but not least I thank my family: mom, dad, Eddy and Laura, Marie Paule and Paul, Stanley and Agnes (my two sponsored children). Thank you for always being there.

I dedicate this thesis to Don Mario, a priest who loved physics and was profoundly fascinated by the earth's science.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Always expect the unexpected."
Anonymous

La tomographie sismique est actuellement la seule technique qui permet d'obtenir des images tridimensionnelles en grand détail de la structure interne de la Terre. De manière analogue à l'imagerie médicale qui permet de visualiser l'intérieur du corps humain, la tomographie sismique essaie d'imaginer l'interieur de la Terre en utilisant les informations que les ondes sismique portent à sa surface.

Diffèrentes techniques ont été développées depuis la première inversion des temps d'arrive (Aki & Lee, 1976; Dziewonski et al., 1977). Jusqu'à maintenant, la grande majorité des études de tomographie était fondée sur la théorie des rayons. Dans cette théorie, les ondes se propagent à fréquence infinie le long de courbes infiniment fines qui suivent la loi de Snell/Descartes pour être des rayons. Les temps de trajet sont affectés seulement par des anomalies de vitesse situées sur le rayon lui-même. Cette simplification rend le formalisme de la tomographie plus simple et explique la très large utilisation de cette approximation en sismologie.

Mais l'approximation des hautes fréquences ne prend pas en compte les effets dû aux phénomènes de diffraction sur les temps de parcours: les temps d'arrivée sont sensibles aux vitesses en dehors du rayon lui-même, dans un volume fini défini par la zone de Fresnel (Nolet, 1987, 1992; Woodward, 1992; Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000). L'intégrale le long du rayon pour calculer le temps de trajet se transforme en une intégrale de volume avec un noyau défini par le noyau de Fresnel.

Dahlen et al. (2000) ont dérivé une expression analytique du noyau de Fréchet pour les ondes à fréquence finie. Cette solution est obtenue en utilisant l'approximation paraxiale où les rayons voisins du rayon entre la source et la station ne dépend que des dérivées des vitesses sur le rayon en addition de la théorie de Born valable au premier ordre.

La modélisation numérique montre que la sensibilité des ondes de volume est

exactement zéro le long du rayon. Si la taille des structures anormales est plus petite que la longueur d'onde caractéristique des ondes propagées, elle n'affecte pas significativement les temps d'arrivée (Nolet & Dahlen, 2000). La sensitibilité maximale se trouve dans une région qui ressemble à une banane si on la regarde horizontalement entre la source et le récepteur et à un beignet si on la regarde verticalement. La forme du noyau de Fresnel est encore plus compliquée pour les ondes qui ont une réflexion à la surface de la Terre. Dans ce cas, l'onde rencontre une caustique qui produit un déphasage de 90 dégrée. Les changements de la région de sensibilité fait que cette région n'est plus nulle le long du rayon. Au contraire, pour l'approximation de la théorie des rayons, l'onde à fréquence infinie se rappelle toujours des anomalies qu'elle a rencontré lors de son trajet vers la station.

L'introduction de ces volumes d'intégration avec un noyau variant suivant les anomalies dans la tomographie des temps d'arrivée doit induire des changements dans la solution finale.

Tester les effets de cette imagerie où les fréquences finies sont prises en compte dans le cas de la tomographie en sismologie globale est le but principal de cette thèse. Lors des tomographies sismologiques réalisées, diffèrents ingrédients sont utilisés de façon à améliorer la résolution sur le résultat final. Ces ingrédients sont : (1) la paramétrisation du modèle de vitesse; (2) les données; (3) une relation qui permet de lier l'espace des modèles à l'espace des données; (4) la technique pour inverser le système d'équations résultant des point (1)–(3) et (5) le jugement du sismologue. Dans cette thèse, on présente chacun de ces ingrédients, avant de présenter les nouveaux modèles de la structure interne de la Terre.

La structure de la Terre n'est pas uniformément échantillonnée par les rayons sismiques. Les stations sont pour la plupart sur terre et, en majorité, dans l'hémisphère Nord. En même temps, les tremblements de terre sont concentrés sur les frontières de plaque. En plus, chaque phase sismique échantillonne une partie diffèrente de la Terre. Le choix de la paramétrisation du modèle et les données à utiliser est trés critique car les deux permettent de réduire les effets de l'illumination non uniforme des anomalies de vitesse et ainsi d'augmenter la résolution de l'image finale.

Le premier chapitre est consacré à la théorie de l'inversion du probléme tomographique (les points 3 et 4 déjà mentionnés). C'est une introduction aux notions de base qui permet de comprendre les chapitrés qui suivent. Il faut préciser que la théorie de tomographie prenant en compte les fréquences finies et la méthode d'inversion utilisés dans cette thèse ne sont pas originales, mais c'est la première fois qu'elles sont utilisées avec des données réelles sur la Terre entière.

Dans le deuxième chapitre, on décrit le problème du choix de la paramétrisation. Le champ de vitesse peux être représenté par une combinaison de cellules de la même taille ou par une distribution uniforme de points, techniques introduites par Aki & Lee (1976) and Thurber (1983), respectivement. Dans les deux cas, la vitesse peut changer sans affecter la convergence du système linéaire.

Dernièrement, des distributions irrégulières des points ont été proposées dans le contexte de la tomographie sismologique. Ces grilles permettent de prendre en considération la distribution irrégulière des rayons (comme dans les références suivantes Chiao & Kuo (2001) et Spakman & Bijwaard (2001)). Ces maillages permettent d'adapter librement les points en fonction de la résolution déduite des données utilisées. Les régions qui ne sont pas bien illuminées par les rayons pourront être échantillonnées par des cellules plus larges. Par contre, si la région est très bien résolue, on utilisera une grille plus fine localement. Pour construire ce maillage, on utilise une distribution de points connectés de manière à former un maillage de Delaunay (Watson, 1981, 1992; Sambridge et al., 1995). En géométrie 3D, une grille de Delaunay est constituée de tétraèdres. Dans ce chapitre, on expliquera comment on construit notre maillage Delaunay.

Dans le chapitre 4, on compare une tomographie obtenue avec cette nouvelle méthode de tomographie à fréquences finies et une tomographie obtenue avec la méthode plus standard à haute fréquence pour des observables temps déduites des donnés à longue période. Nos données sont constituées des temps d'arrivée des ondes P et PP–P à longue période, mesurés par cross-corrélation (Bolton & Masters, 2001). Dans cette méthode, les temps d'arrivée sont obtenus par comparaison de la forme d'onde avec la forme d'onde de référence. On améliore ainsi la précision sur le pointé des temps. On ne pointe pas le temps par détection du début du signal dans le sismogramme. Ces données, qui seront utilisées dans les chapitrés 5 et 6, ont été re-analysées et corrigées pour des erreurs d'instrumentation (seulement dernièrement découvertes) par Prof Guy Masters au Institute of Geophysics and Planetary Physics (IGPP) - University of California, San Diego.

Un grand nombre de temps d'arrivée est nécessaire pour obtenir un bon modéle global de la Terre. Le nombre de données à longue période disponible est relativement petit avec seulement 86400 temps d'arrivé. Pour obtenir une bonne image tomographique, on essaie d'utiliser le plus grand nombre des données possibles pour pouvoir avoir une distribution la plus uniforme possible. Les phases directes des ondes de volume ont une couverture limitée du champ de vitesse de la Terre, spécialement dans le manteau supérieur et dans le manteau moyen. En ajoutant des ondes qui se réfléchissent à la surface libre, comme les ondes PP ou les ondes pP, il est possible d'augmenter la résolution car ces ondes permettent l'échantillonnage des régions où il n'y a pas de stations et/ou de tremblements de terre.

Pour améliorer la résolution de notre image tomographique, dans le chapitre 5, nous combinons les données à longue période avec des données ISC à courte période re-analysées par Bob Engdahl. Nous sélectionnons seulement les données reconnues de haute précision en raison des la lecture sur le séismogramme. Nous sommes confiants que ces données représentent des arrivées de plus haute fréquence pour lesquelles la théorie des rayons est encore valable. Les données longue période sont plus sensibles aux structures à grande échelle. Au contraire, les ondes à courte période sont plus sensibles aux structures fines. La combinaison des deux diffèrent

jeux de données permet une meilleure illumination de la structure de la Terre avec les ondes courte période qui vont éclairer les régions qui ne sont pas échantillonnées par les données à longue période.

Les deux familles de données sont complètement diffèrentes en nature. Il est donc nécessaire d'étudier d'abord leur compatibilité. Dans le chapitre 5, nous analysons en détail l'uniformité des données et nous décrivons la technique utilisée pour les inverser conjointement.

Notre modelé tomographique montre, pour la première fois, des panaches profonds introduits par Morgan (1971, 1972) pour expliquer l'origine du volcanisme intra-plaque connu sous le nom de "points chauds". Les chaînons volcaniques de points chauds se trouvant dans les zones suivantes (Hawaï, Louisville, Réunion, Deccan), pourraient être expliqués par le mouvement de la plaque passant au dessus d'une source de magma trés profond dans le manteau. Récemment, l'hypothèse d'immobilité des points chauds était revue par plusieurs auteurs: Norton (1995) suggére que le coude observé dans la chaîne Hawaï-Empereur est plutôt l'enregistrement de la date à laquelle le panache est devenu fixe par rapport au manteau et Tarduno & Cottrell (1997), en utilisant des données paleomagnetiques, montre que la chaîn volcanique des Hawaï-Emperor montre que le point chaud s'est arrêté de se déplacer par rapport à la lithosphère Pacifique plutôt que le contraire.

On s'attend à ce que les panaches soient plutôt étroits, de l'ordre de quelques centaines de km de diamètre, plus petit que la région de Fresnel typique d'une onde P avec une période de 20 s qui est d'environ 1000 km pour une distance épicentrelle de 80 dégrée. Il est trés possible que l'image des panaches soit affectée si on ne considére pas les effecs dûs aux fréquences finies (Nataf & VanDecar, 1993). Corriger pour les effets de diffraction aide à améliorer la détection et à la résolution de ces structures étroites.

Le chapitre 6 est une description détaillée des panaches qu'on a découvert lors de notre reconstruction tomographique. Dans ce chapitre, on montre les résultats d'une analyse de la résolution de cette anomalie de basse vitesse qui nous permet de définir des limites de confiance pour la profondeur de la source et la largeur de chacun des 32 panaches visualisés lors de notre tomographie. Notons que quatre panaches n'ont pas été détectés auparavant car ils n'apparaissent dans aucune liste de points chauds publiés.

Les chapitrés 4–6 sont des versions modifiées des articles en cours de soumission (chapitre 5) ou déjà soumis pour être publier (chapitres 4 et 6). L'annexe C décrit une methode pour l'estimation de la matrice de résolution a utiliser pour l'adaptation des maillages. Cette technique n'a pas encore été introduit dans les outils algorithmiques de cette thèse mais qui devrait in fine permettre d'optimiser le maillage décritant le champ de vitesse. C'est un travail que j'ai fait pendant ma première année de thése en France et je crois qu'il doit apparaître dans cette thèse. Comme écrit dans les perspectives (chapitre 7), les outils de cette annexe seront utilisés dans mes études futures.

Our understanding of the Earth's dynamics relies strongly on the degree of knowledge of the internal structure of the Earth. Seismic tomography is the only tool available to map the three-dimensional structure of the Earth's interior in detail. In medicine the term computerized tomography (in Greek tomo = slice) appeared in the 1970's to denote that the 3D image of a human body was constructed from a combination of 2D slices (Hounsfield, 1973). In a similar manner, in *seismic* tomography one seeks to image the Earth's velocity structure by using the information brought to the surface by seismic waves traveling through the Earth.

Different approaches have been developed in seismic tomography since the first formal 3D traveltime inversions were performed (Aki & Lee, 1976; Dziewonski et al., 1977). Almost all global tomographic studies that use body waves are based on ray theory, in which waves propagate as rays that follow Snell's law, with traveltimes that are only influenced by anomalies located on this ray. This simplifies the mathematics, and is therefore widely used (to cite only few: Dziewonski, 1984; Inoue et al., 1990; Van der Hilst et al., 1997; Bijwaard et al., 1998). But ray theory neglects the fact that, due to diffraction effects, the traveltime is sensitive to velocity structures far away from the geometrical ray, within a volume known as the Fresnel zone (Nolet, 1987, 1992; Woodward, 1992; Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000).

Dahlen et al. (2000) derived an efficient expression for the Fréchet kernels of finite-frequency waves, based on the paraxial approximation to ray theory as well as the Born approximation. Hung et al. (2000) show a colorful collection of these kernels for different seismic phases. The sensitivity of a direct body wave turns out to be identically zero along the ray, and is confined to a banana-shaped volume surrounding the geometrical ray, the Fresnel zone. Because of wavefront healing, objects much smaller than the width of the Fresnel zone do not significantly affect the travel time of the wave (Nolet & Dahlen, 2000). By contrast, in ray theory, an infinite-frequency wave always "remembers" the shift accrued upon passage through an anomaly.

Figure 1.1 shows an example of these sensitivity regions computed for a P wave with 1 s dominant period (Figure 1.1a), with 20 s dominant period (Figure 1.1b), and for a PP wave with 20 s dominant period (Figure 1.1c). The sensitivity is significantly different between the short - and long - period waves. The higher the frequency of the wave, the thinner is the sensitivity region around the geometrical ray, with the sensitivity still being zero on the ray. The situation is even less intuitive for PP waves. Scattered waves may come in either earlier or later than the main phase and may or may not be Hilbert transformed. Therefore, surface reflected PP waves show a much more complicated shape of the sensitivity region, with the *banana-doughnut* shape replaced by a *saddle-shaped* sensitivity region upon passage of a caustic. It is clear that ray-theory provides a too crude approximation to the traveltimes, at least for the lower-frequency waves.

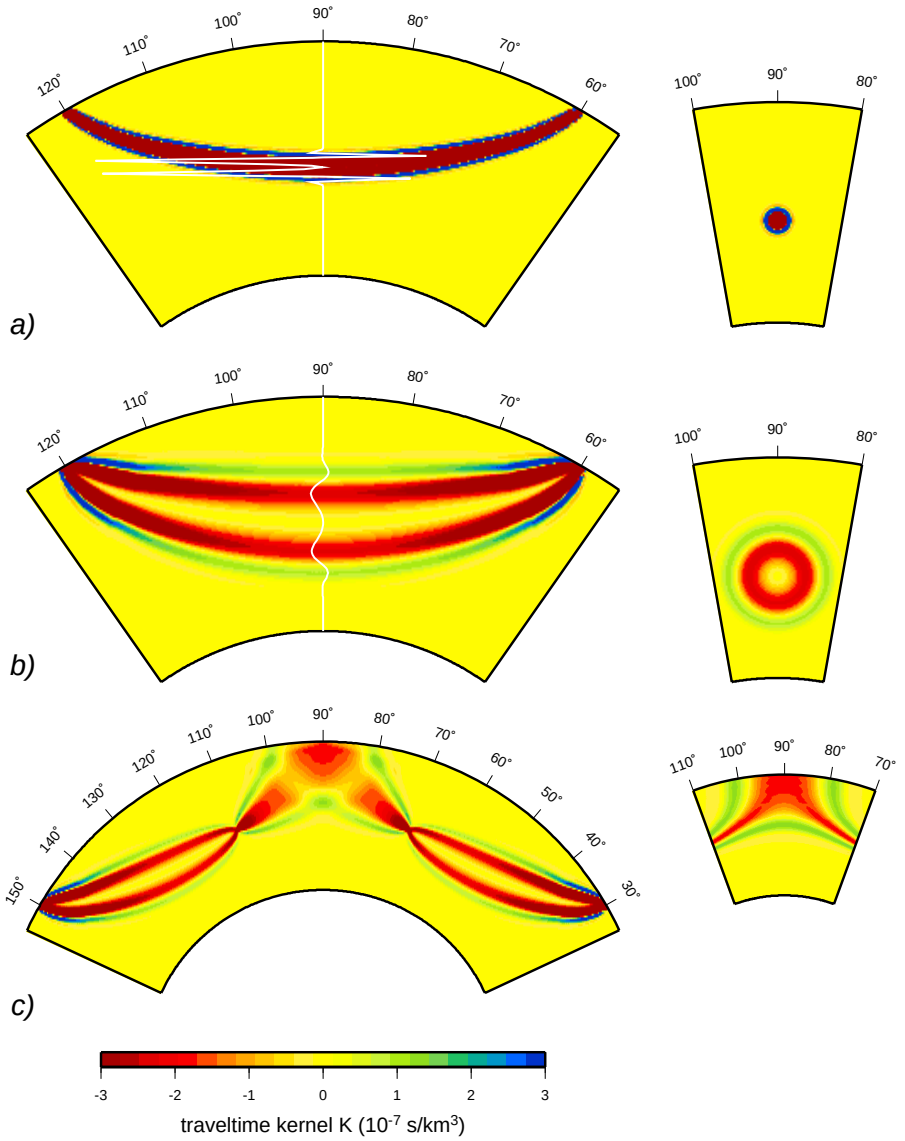


Figure 1.1 Sensitivity kernels for a direct P wave at an epicentral distance of 60° with dominant period: (a) 1 s, (b) 20 s and (c) for a PP wave at an epicentral distance of 120° with 20 s dominant period.

Therefore, the introduction of the finite-frequency modeling in seismic tomography must have a significant effect on the final images. This idea was the seed that brought to blossom this thesis.

Any tomographic study is based on a combination of ingredients all affecting the final result. Such ingredients are (1) the model parameterization: a set of basis functions used to map the velocity structure; (2) the set of data; (3) a functional to map the data space into the model space; (4) a technique to invert the resulting system of equations, and (5) last but not least, the subjective judgment of the seismologist. This thesis deals with each of these important steps before presenting a new set of images of the Earth's interior.

Seismic tomography suffers from the fact that earthquakes are not a controllable source of energy. They are concentrated mainly near plate boundaries and can occur at anytime. The uncertainties related to the hypocenter location introduces extra unknowns in the tomographic problem. Also, stations are not uniformly deployed around the globe, with the highest concentration found in the northern hemisphere. Finally, different seismic phases sample different parts of the Earth's structure. As result, the three-dimensional velocity structure of the Earth's interior in not uniformly illuminated with significant differences in the potential resolution of various regions. The choice of the model parameterization, point (1) of the above list; and the selection of the seismic data, point (2), are critical, since both help to improve the resolution of the final image.

In chapter 2, we give a brief overview of the theoretical framework of seismic tomography (points 3 and 4 of the above list are described in this chapter). We provide the reader with the essentials required to understand the chapters that follow. Though neither the "finite frequency" theory, nor the inversion technique is original, this is the first time both are combined and applied to real data.

In chapter 3 we tackle the problem of the choice of the model parameterization. The Earth can be represented by a combination of cells of the same size or grid points uniformly distributed, an approach introduced by Aki & Lee (1976) and Thurber (1983), respectively. In both cases, because of the regular character of the grid and the uneven distribution of seismic rays, there might be cells or points of the grid that are not illuminated at all. The velocity in these cells may freely vary without affecting the fit of the linear system. Recently, heterogeneous distributions of points (or cells) have been used in seismic tomography to better account for the uneven sampling of the Earth's velocity structure (most recent works are by Chiao & Kuo (2001) and Spakman & Bijwaard (2001)). The flexibility of such parameterizations allows us to adapt them to the resolving power of the data used, such that regions poorly sampled by seismic rays can be represented by larger cells, while densely sampled regions can be very finely discretized. Following Sambridge et al. (1995), we sample the Earth's velocity structure by means of a heterogeneous distribution of points connected to form a Delaunay mesh (Watson, 1981, 1992). In 3D a Delaunay mesh consists of a set of tetrahedrons. In this chapter we provide

a detailed description of these kind of grids, and describe how our global grid is generated.

Chapter 4 represents the core of this thesis. It provides a direct comparison of a finite-frequency tomography with the results of an inversion obtained by using standard ray theory. Our data set consists of long-period P and PP–P traveltimes measured by cross-correlation (Bolton & Masters, 2001). Measurements obtained by cross-correlation provide an integrated arrival-time difference between two waveforms, not simply the difference between the onset times. Therefore the ray theoretical description for a traveltime along the ray may be no longer valid, and finite-frequency modeling may be necessary. These data, also used in chapter 5 and 6, have been re-analyzed and corrected for several instrumental timing errors (only recently discovered) by Prof. Guy Masters at Institute of Geophysics and Planetary Physics (IGPP) - University of California, San Diego.

To obtain a good image one aims to use a large set of data that sample the mantle as uniformly as possible. Direct phases have a limited ray path coverage, especially in the upper and mid-mantle. Incorporating reflected arrivals, such as PP or pP, can help because they sample regions of the Earth not sampled by direct waves. However, the long period data set used in chapter 4 is still relatively small (only a total of 86,400 arrival times). Recently, we see an increasing number of tomographic images obtained by inversion of a combination of different seismic phases (for instance, Van der Hilst et al. (1991, 1997), Masters et al. (1996), Su & Dziewonski (1997), Mégnin & Romanowicz (2000) and Vasco et al. (2003)). To enhance the resolution of our tomographic model, in chapter 5 we combine the long-period data set with very-high quality short-period arrival times extracted from the ISC data set and re-processed by Prof. Bob Engdahl at the University of Colorado. By selecting only the data with two digits decimal precision from his data set we are confident that these measurements correspond to the highest frequency wave, for which ray theory may likely be still valid. Long-period data are only sensitive to large-scale structure, while short-period data are also sensitive to small scale structure. The combination of the two provides us with a good illumination of the Earth's velocity structure, with the short-period data filling the gap left by the long-period data. However, the two classes of measurements are completely different in their nature, and a study of their compatibility is required before we perform a joint inversion. In chapter 5 we provide a detailed analysis of the consistency of the two data sets used and a description of the technique we use to jointly invert them.

To our own surprise, our tomographic images show for the first time clear, distinct low velocity anomalies rising from the deep mantle that resemble the plumes hypothesized by Morgan (1971, 1972) as possible explanation for the origin of the intraplate volcanism known as "hotspots". In what is called the "plume model", the long linear chain of volcanoes characteristic of several hotspots such as Hawaii, Louisville, Reunion - Deccan, could be explained by the moving plate passing over

a relatively fixed upwelling of hot material rising from great depth in the mantle. Figure 1.2 shows the Hawaiian-Emperor volcanic chain as seen in the topography of the sea-floor. Hawaii-Emperor is the best example of a volcanic chain on Earth. It is characterized by a 43 my old bend, that separates the youngest westward trending Hawaiian islands from the northward trending Emperor seamounts, and which is thought to be due to an abrupt change of the Pacific plate motion (Richards & Lithgow-Bertelloni, 1996). This hypothesis is based on the assumption that plumes are relatively fixed with respect to the plate motion. Steinberger & O'Connell (1998) also adopted this assumption for their time-dependent modeling of mantle circulation; though in their model single plumes are able to swing in the large-scale mantle flow beneath the moving plate as a consequence of convective motion or "mantle wind". Recently, the hypothesis of plumes's fixity has been challenged by several authors: Norton (1995) suggest that the bend is the record of the time when the Hawaiian plume became fixed in the mantle and Tarduno & Cottrell (1997) present paleomagnetic data to suggest that the track records differences in motion of the Hawaiian hotspot relative to the Pacific lithosphere, rather than the other way around (for a review of hotspot fixity and mantle wind we refer to Christensen (1998)).

Plumes are expected to be rather narrow, of the order of a hundred km of diameter or so, smaller than the typical width of the Fresnel zone of a 20 s - period wave which extends to 1000 km for an epicentral distance of 80°. Therefore their image may be affected by a failure to model finite-frequency effects (Nataf & VanDecar, 1993). Finite-frequency tomography enhances the capability to detect such narrow structures.

Chapter 6 focuses on the plume-like structures revealed by our tomographic study. It presents the results of a detailed resolution analysis which allow us to put constraints on the depth and width of the source region for 32 plumes, at least four of which are newly discovered.

Note to the readers

This thesis contains papers that are about to be submitted or have already been submitted for publication. Chapters that contain them are self-explanatory and therefore some concepts are repeated. Appendix C contains material that has not yet been used in the tomographic inversions described here. However, it represents the work I have done in my first first year spent in France, and therefore I believe it deserves a place in this thesis. As described in the "Afterwords" (chapter 7), the content of this Appendix is going to be used in future research projects.

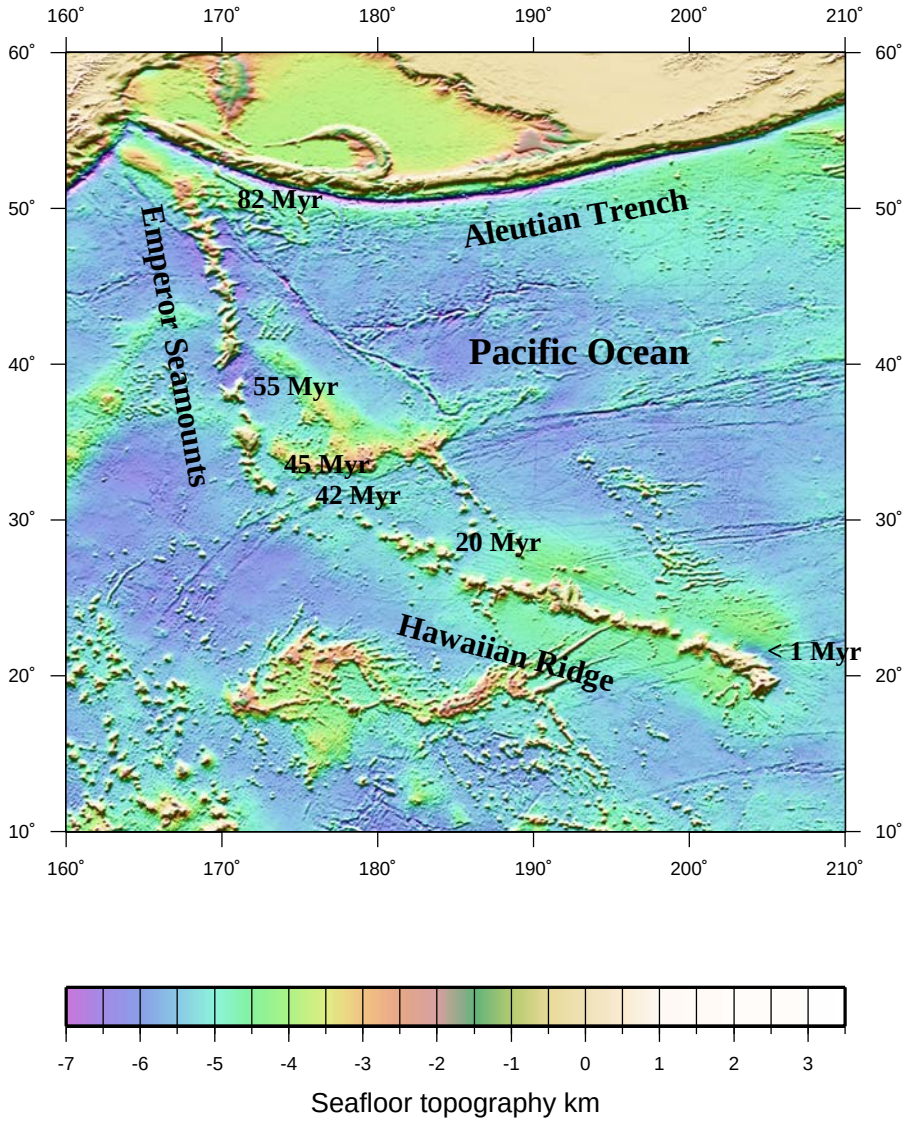


Figure 1.2 Seafloor topography in the Hawaiian-Emperor region on the Pacific plate, after Smith & Sandwell (1997)

Chapter 2

Theoretical framework

2.1 Résumé

Dans ce chapitre, je présente des notions de base de la tomographie sismique des écarts des temps qu'on utilise dans cette thèse. La méthode tomographique des écarts se base sur des temps de propagation théoriques des ondes sismiques calculés pour un milieu terrestre à symétrie sphérique. Ce temps théorique est comparé au temps observé. La différence est un résidu de temps de propagation représentatif des perturbations de vitesse subies par les ondes traversant la vraie structure de la Terre.

L'inversion de ces résidus permet d'obtenir une représentation tri-dimensionnelle de la structure interne de la Terre. Le problème est fortement non-lineaire, vu que le parcours des rayons utilisé pour le calcul des temps théoriques dépend lui aussi du champ de vitesse que l'on essaie de reconstruire. L'hypothèse que le rayon dans la vraie Terre ne différe pas de manière significative de la trajectoire dans le modèle initial permet de linéariser le problème et de le réduire à une forme matricielle $Ax = b$ où le vecteur b contient les résidus de temps. Le vecteur x représente les perturbations en vitesse et les écarts de position et de temps origine de la source sismique qui est aussi inconnue. L'inversion est faite en manière à obtenir à la foirs les changements des vitesses mais aussi les perturbations de la source du seisme et de son temps d'origine. Le vecteur x contient donc la solution pour les changements de vitesse x_{c} et les sources x_{h}. La matrice A est constituée par la partie relative au modèle qu'on construit soit en utilisant la théorie des rayons ou en utilisant la modélisation à fréquence finie et la partie relative aux paramètres de la source qu'on construit avec la théorie des rayons.

La solution x est trouvée en minimisant une mesure des anomalies au sens des moindres carrés qui conduit au système $Ax = b$ (Nolet, 1987). En général, même si on a un nombre plus important de données que de paramètres, le système tomographique qu'on doit résoudre est sous-déterminé car la distribution des rayons n'est pas uniforme. Cela va dire qu'il y a des solutions x non nulles du système

linéaire $Ax = b$ qui satisfont $Ax = 0$. Ces solutions constituent l'espace nul de la matrice A. Quelles combinaisons linéaires des vecteurs dans l'espace nul peuvent être ajoutées à une solution particulière pour donner une nouvelle solution qui satisfasse les données d'une manière identique. Pour limiter ce problème, on applique des régularisations au système, méthode que l'on désigne sous l'expression des moindres carrés pondérés. On parle alors d'amortissement qui limite la distance entre la solution finale et la solution initiale, ce qui a pour conséquence de réduire les amplitudes des anomalies pour les vitesses et qui réduit aussi les déplacements des sources ainsi que les changements des temps origine. Un lissage est aussi appliqué pour rendre les variations de vitesse plus lisses dans chaque direction (Nolet, 1987).

La statistique du χ^2 fournit une moyenne quantitative pour l'estimation de la solution plausible. J'utilise cette méthode pour déterminer la meilleure solution et pour comparer les solutions obtenues avec les differentes méthodes utilisées au cours de cette thèse.

2.2 Introduction

This chapter provides a brief description of the equations that are fundamental to every tomographic study. It is not meant to be an exhaustive description, but a rather brief overview of the essentials of the theory behind delay time tomography. For a more detailed analysis I refer the reader to the work by Nolet (1987), Spakman & Nolet (1988), Menke (1989) and Parker (1994).

2.3 The system of equations

In the ray approximation, the traveltime of a ray in a three-dimensional Earth is expressed by:

$$T = \int_L c(\mathbf{x})^{-1} dl \tag{2.1}$$

where L is a ray path and $c(\mathbf{x})$ the velocity.

We have a fairly good knowledge of the radially symmetric structure of the Earth $c(r)$. Given an Earth reference model, one can compute synthetic traveltimes to compare to the measured one. The difference between the two values is the delay time δT and can be expressed as:

$$\delta T = T^{\mathsf{obs}} - T^{\mathsf{syn}} = \int_L c(\mathbf{x})^{-1} dl - \int_{L_0} c_0(\mathbf{r})^{-1} dl \tag{2.2}$$

The inverse problem is non-linear, since the ray path L depends on the unknown velocity. Assuming that the ray in the real earth does not differ significantly from the trajectory in the starting model (Fermat's principle), this expression can be linearized.

The corresponding perturbation with respect to the predicted time for a set of sources and receivers is given by:

$$\delta T_i = -\int_{L_0} \frac{\delta c(\mathbf{x})}{c_0(\mathbf{r})^2} dl, \;\; i = 1, ...N. \tag{2.3}$$

However, Marquering et al. (1998), Dahlen et al. (2000), Hung et al. (2000) and Nolet & Dahlen (2000) show that the approximation of infinitely narrow rays does not model diffraction effects. In their theory, rays have a finite thickness depending on the frequency content of the wavefield. The 1D integral along the geometrical ray is replaced by a 3D volume integral (Dahlen et al., 2000):

$$\delta T_i = \int_{\oplus} K_i(\mathbf{x}) \frac{\delta c}{c_0} d^3\mathbf{x} \tag{2.4}$$

over the entire earth $\oplus$. The quantity $K(\mathbf{x})$ is the 3D Fréchet kernel of a finite-frequency traveltime shift δT that has been measured by cross-correlation of a broadband waveform with a spherical earth synthetic.

Since the earthquake parameters influence the travel time as well, we shall simultaneously derive the velocity perturbations $\delta c(\mathbf{x})$ and the hypocenter parameters (origin time, longitude, latitude and depth) from a discrete number of delay times. Each delay time can be seen as the sum of a delay δT_c caused by the velocity perturbation and a delay δT_h due to source mislocation and error in the origin time, plus a measurement error ϵ:

$$\delta T = \delta T_c + \delta T_h + \epsilon \tag{2.5}$$

For computational simplicity we describe the velocity $c(\mathbf{x})$ by a finite number m of parameters, i.e. cells, grid points, spherical harmonics.

$$c(\mathbf{r}) = \sum_{k}^{m} c_k h_k(\mathbf{r}), \tag{2.6}$$

where h_k denotes a set of basis functions spanning a subspace of the (Hilbert) space of all possible earth models $c(\mathbf{x})$. Similarly is done for the velocity perturbation $\delta c(\mathbf{x})$. After substitution of the expression 2.6 for the velocity perturbation $\delta c(\mathbf{x})$ in either equation 2.4 or 2.3, for the delay time δT_c we obtain:

$$\delta T_i = \sum_{j=1}^{m} A'_{ij} \delta c_j, \;\; i = 1...., n \tag{2.7}$$

where A'_{ij} has the form:

$$A'_{ij} = -\int_{L^0} \frac{h_j(\mathbf{x})}{c_0^2} dl \tag{2.8}$$

in the ray approximation, and:

$$A'_{ij} = \int_{\oplus} K_i(\mathbf{x}) \frac{h_j(\mathbf{x})}{c_0(\mathbf{x})^2} dl \tag{2.9}$$

in the finite-frequency approximation. In both formulations, the tomographic problem reduces to a similar discrete system that in matrix notation can be written as:

$$\mathsf{A}\,\mathsf{x} = [\mathsf{A}'\,\mathsf{H}]\mathsf{x} = \mathsf{b} \tag{2.10}$$

with b the vector of the delay times δT; $\mathsf{x} = [\mathsf{x_c}, \mathsf{x_h}]^\mathsf{T}$ a vector of the unknown velocity perturbations $\mathsf{x_c}$ and source corrections $\mathsf{x_h}$, A the matrix mapping the model space into the data space. H is the matrix of partial derivatives with respect to the source parameters that we compute by using standard ray theory:

$$\begin{aligned}
\frac{\partial T}{dT_0} &= -1; \\
\frac{\partial T}{dx} &= -\sin i \, \frac{\sin \xi}{c}; \\
\frac{\partial T}{dy} &= -\sin i \, \frac{\cos \xi}{c}; \\
\frac{\partial T}{dz} &= -\frac{\cos \xi}{c}
\end{aligned} \tag{2.11}$$

with x longitude, y latitude and z depth of the hypocenter, T_0 origin time and i and ξ incident angle and source azimuth, respectively(Lee & Stewart, 1981).

2.3.1 Solving a tomographic inverse problem

In general we have many more data then unknowns, yet due to the sparseness of the ray distribution, the inverse problem has a partly underdetermined nature. Also, because of data errors, the system of equations is inconsistent and no exact solution exists. The standard procedure is to minimize a measure of the discrepancy in the system $\mathsf{Ax} = \mathsf{b}$ in a least square sense (Nolet, 1987):

$$min \sum_{i=1}^{n} \mid \sum_{j=1}^{m} A_{ij} x_j - b_i \mid^2 . \tag{2.12}$$

Differentiating with respect to x_j yields the system:

$$\mathsf{A}^\mathsf{T}\mathsf{A}\mathsf{x} = \mathsf{A}^\mathsf{T}\delta\mathsf{T}, \tag{2.13}$$

known as *normal equations*. If $\mathsf{A}^\mathsf{T}\mathsf{A}$ is not singular, then the normal equations have exactly one solution:

$$\mathsf{x} = (\mathsf{A}^\mathsf{T}\mathsf{A})^{-1}\mathsf{A}^\mathsf{T}\delta\mathsf{T}, \tag{2.14}$$

This solution minimizes the sum of squared errors. Because of the underdetermined nature of the tomographic problem, there are solutions x of the linear system $\mathsf{Ax} = \mathsf{b}$ that satisfy $\mathsf{Ax} = \mathsf{0}$. These solutions define the nullspace of the matrix A. In terms of the least squares system, the existence of a non-empty nullspace of A means that $\mathsf{det}(\mathsf{A}^\mathsf{T}\mathsf{A})$ is zero. Any linear combination of vectors belonging to the nullspace can be added to a particular solution to provide a new solution that fits the data equally well. If the size of the matrix A is small enough to be stored in a computer memory, the nullspace can be determined exactly by diagonalizing the matrix $\mathsf{A}^\mathsf{T}\mathsf{A}$, and a solution can be found by truncating the eigenvalues spanning the nullspace, a method called "Truncated Singular Value Decomposition". Alternatively one may use a damped least square approach (e.g. Tarantola & Valette, 1982). Extra equations $\varepsilon x_i = 0$ are added to the system based on our a priori expectations:

$$\begin{bmatrix} A \\ \varepsilon I \end{bmatrix} \mathsf{x} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{2.15}$$

By again solving the system in a least square sense, we obtain:

$$\mathsf{x} = [\mathsf{A}^\mathsf{T}\mathsf{A} + \epsilon^2 \mathsf{I}]^{-1} \mathsf{A}^\mathsf{T}\mathsf{b} \tag{2.16}$$

The tunable parameter ε has the function to shift the eigenvalues λ_i^2 to $\lambda_i^2 + \varepsilon^2$. Clearly, the matrix $[\mathsf{A}^\mathsf{T}\mathsf{A} + \varepsilon^2\mathsf{I}]^{-1}$ is not singular. However, the larger the value of ε, the more it degrades the data fit, since it biases $\mathbf{x}$ to a lower amplitude solution. In simultaneous inversion for velocity perturbations and hypocentral parameters, we apply a damping to the source corrections $\mathsf{x_h}$ to limit changes in origin time and hypocentral location coordinates. A Laplacian damping is usually applied to render the velocity variations smooth in every direction as in (Nolet, 1987).

The level of the regularization is somewhat subjective. The final tomographic image is usually the result of many inversions in which one experiments with different value of the tunable parameters (norm damping and smoothing). Some criterion for selecting the final result is required. The theory of maximum likelihood estimation gives us a good guideline.

2.3.2 The maximum-likelihood estimate of the model parameter

The least squares problems in seismic tomography can be rephrased in a statistical sense. Given a set of data, with known standard deviation, we can ask which is the most likely model from which these data could have arisen. A basic assumption is that data errors are independent and Gaussian distributed, so that the joint likelihood P of the data set occurring (the quantity we want to maximize) is the product of the individual probabilities:

$$P = \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(b_i - (Ax)_i)^2 / 2\sigma_i^2}. \tag{2.17}$$

Maximizing the probability is equivalent to minimizing the exponent, or:

$$\min \chi^2 = \sum_{i=1}^{n} \left(\frac{(Ax)_i - b_i}{\sigma_i} \right)^2 . \tag{2.18}$$

For very large tomographic problems, involving a large number n of degree of freedom, the expected value of χ^2 for the true Earth $\mathsf{x}_{\mathsf{true}}$ approaches n. This follows from the central limit theorem which predicts that the probability density function for χ^2 will itself become normally distributed with mean n and standard deviation $\sqrt{2n}$. The χ^2 statistic provides a reliable quantitative assessment of the plausibility of models which we use instead of the "variance reduction" criterion. Variance reduction is as much a measure of the adequacy of the starting model as it is a measure of fit of the final solution and does not take into account for "a priori" estimates of data errors.

Chapter 3

Seismic tomography with heterogeneous model parameterization.

3.1 Résumé

Le choix de la paramétrisation du modèle est critique. Elle permet de réduire la taille de l'espace nul de la matrice A. On peut augmenter la taille des cellules jusqu'au moment où elles seront toutes illuminées par des rayons mais la grille finale ne fournit pas une bonne représentation du champ de vitesse de la Terre. Des grilles très fines sont donc utilisées et des méthodes de régularisation utilisées pour contrôler la non-unicité de la solution.

Mais, au lieu d'augmenter la taille de toutes les cellules, on pourrait augmenter seulement la taille des cellules qui sont très mal déterminées et réduire la taille de celles qui sont sous-déterminées pour pouvoir égaliser l'illumination des points de la grille. À cette fin, des maillages irréguliers ont été récemment introduits dans le problème tomographique. Ces grilles ont l'avantage de pouvoir s'adapter à la limite de résolution des données utilisées. Elles permettent donc de contrôler la densité des rayons et donc l'illumination du modèle comme l'illustrent récemment certains articles (Chiao & Kuo (2001); Spakman & Bijwaard (2001)). Les régions qui ne sont pas bien échantillonées pourraient maintenant être représentées par de larges cellules et, au contraire, des régions très bien illuminées pourraient être échantillonées avec des distributions de points plus denses. Dans ce cas, le nombre des inconnues est reduite et le système à inverser est plus stable. Dans cette thèse, je suis l'exemple de Sambridge et al. (1995) et j'utilise un maillage de Delaunay, introduit par Watson (1981). Dans ce chapitre je décris ce type de grille et je décris et je construis la maille que j'ai utilisée dans mes inversions. En géométrie 3D, une maille de Delaunay est constituée par des tétrahèdres. Les points sont distribués de manière irrégulière avec une densité qui varie avec la profondeur. La Terre est

divisée en quatre coquilles. La première correspond au manteau superieur où la distance entre les points de la maille est d'environ 200 km. Les points entre 660 km et 1400 km sont distants d'environ 400 km. À la base du manteau, la distance entre points atteint 600 km. Même si nous n'utilisons pas des données dans le noyau terrestre, pour la définition de la grille, nous ajoutons des points aussi dans le noyau à une distance d'environ 1000 km entre eux. Enfin, ce type de maillage sont toujour convexes. Les points d'un maillage Delaunay sont toujours enveloppés par une surface convexe appelée la coque convexe. Pour générer cette surface convexe, j'utilise la même stratégie utilisée dans le code TERRA pour générer les mailles des éléments finis. Le maillage final utilisé dans les inversions décrites dans cette thèse contient 19270 points. La vitesse ailleurs que sur les points de la grille est définie par interpolation linéaire.

3.2 Introduction

Many tomography studies model the Earth with cells of the same size (Aki et al., 1977; Evans & Achauer, 1993; Spakman, 1991), others make use of a regular distribution of grid points (Thurber, 1983; Eberhart-Phillips, 1986). In both cases, because the regular character of the grid cannot take into account the often very heterogeneous raypath distribution, there might be no rays crossing a particular cell. The velocity in this cell may freely vary without affecting the fit of the linear system but may generate artifacts in the final tomographic images. Some tomographic models expand the Earth's velocity field into a finite number of spherical harmonics (Dziewonski, 1984; Woodhouse & Dziewonski, 1984; Dziewonski & Woodhouse, 1987; Morelli & Dziewonski, 1987; Tanimoto, 1990; Boschi & Dziewonski, 1999). An advantage of spherical harmonics is that it provides easy control over the smoothness of the solution by reducing the highest angular order.

If some nodes of the grid are illuminated by few raypaths or if they are not illuminated at all, the matrix A of the linear tomographic system $\mathsf{Ax} = \mathsf{b}$ is singular. One could solve the problem by increasing the size of the cells until they all get illuminated, i.e. until the rank of the matrix A (i.e. the number of linearly independent rows) equals the number of cells. However, such a coarse grid would not provide a convenient representation of the earth's velocity field. Thus, fine grids are used and regularization (i.e. smoothing or damping) of the system of equations introduced to deal with the non-uniqueness of the solution.

Instead of increasing the size of each cell, one could increase the size only of the cell that are poorly determined, while reducing the size of the overdetermined ones to equalize the illumination. With the aim to better suppress the influence of the heterogeneous ray path coverage, and to be able to adapt the parameterization to some measure of the resolving power of the applied data, heterogeneous grids have been introduced recently to represent the velocity structure (Michelena & Harris, 1991; Vesnaver, 1994; Michelini, 1995; Sambridge et al., 1995; Sambridge

& Gudmundsson, 1998; Böhm et al., 1997; Böhm & Vesnaver, 1999; Böhm et al., 2000; Curtis, 1999; Kissling et al., 2001; Wéber, 2001; Chiao & Kuo, 2001; Spakman & Bijwaard, 2001). The flexibility of these grids to adapt to the raypath distribution, together with powerful new interpolation schemes, make them the best choice to reduce or zeroed the size of the null-space of the tomographic system. Regions poorly sampled by seismic rays can be now represented by larger cells, while nicely sampled regions can be very finely discretized. This reduces the number of unknowns and leads to a better conditioning of the inverse problem.

3.3 Definition of Delaunay meshes

In our tomographic study, the earth's velocity structure is modeled by using a 3D Delaunay mesh (Watson, 1981; Sambridge et al., 1995). Given a set of points irregularly distributed, a Delaunay mesh represents a set of triangles in two dimensions or tetrahedrons in three dimensions. In two dimensions, a Delaunay triangulation is built by following the *empty circumcircle criterion*. This criterion states that the three vertices of the triangle must lie on one circle. Other points can lie on the same circle but no other nodes of the grid can lie within that circle (Figure 3.1a). If more than three points of the grid lie on one circle, the mesh is still a Delaunay mesh but it is a non-unique one. Circles passing through one of these vertices establish all the natural neighbors of that vertex. The natural neighbors about any point is the set of closest surrounding nodes of that particular point. A pair of neighbors share a mutual interface of a triangle in a Delaunay mesh, or share a region that is closer to those two points than to any other point of the grid (Watson, 1981, 1992; Sambridge et al., 1995). These set of regions, one per node and consisting of the part of the plane nearest to that node, are called Voronoi cells and are unique in any number of dimensions. Delaunay triangles are formed by simply connecting the nodes whose Voronoi cells have common boundaries. Point L in Figure 3.1b has neighbors A, C, D, E, F, G, M and K. They are all in neighboring Voronoi cells, or equivalently are all connected by the sides of Delaunay triangles. The triangles defined in this way are the most equiangular possible. The extension to three dimensions is straightforward and the Delaunay mesh becomes an aggregation of irregular tetrahedron. We build the Delaunay connections by using *qhull*, a package distributed by the Geometry Center of Minneapolis (Barber et al., 1996).

3.4 Generating a global Delaunay mesh

We generate a distribution of points whose distance varies with depth.

The Earth is divided in four shells. The top shell corresponds to the upper mantle and distance among points is about 200km. Points of the grid between

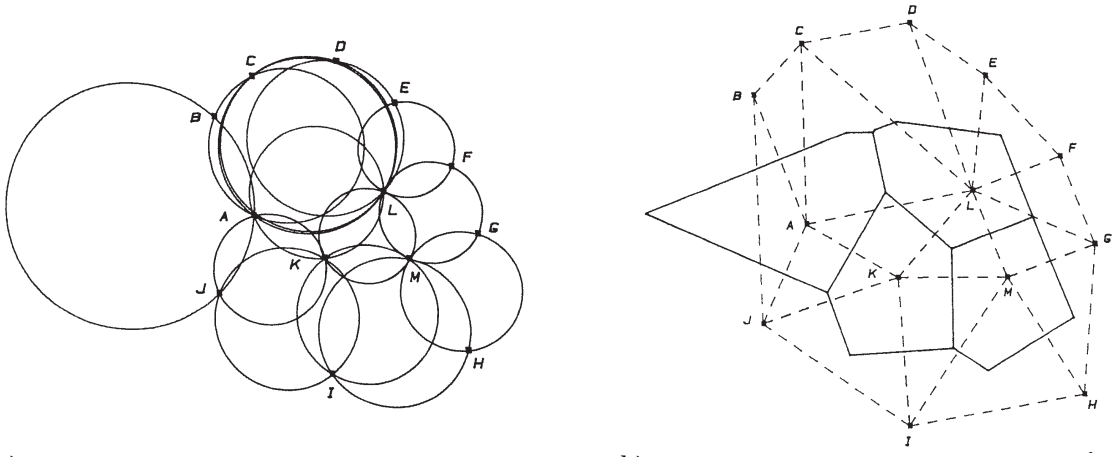


Figure 3.1 (a) Any 2D Delaunay mesh is constructed by satisfying the *empty circumcircle property* which states that any three points of the grid forming a Delaunay triangle lie on one circle and no other point of the grid can lies withing that circle. (b) Voronoi tessellation (solid line) and Delaunay triangulation (dashed line) are dual. Voronoi cells divide the plane into a set of regions, one for each node, such that any point in that region is closer to that node than to any other node of the grid. Voronoi cells are unique in any number of dimensions and their boundaries consist of perpendicular bisectors between points. Delaunay triangles are formed by simply connecting the nodes whose Voronoi cells have common boundaries. A Delaunay triangulation is unique unless four nodes lie on the same circle, or five on the same sphere (Figure from Watson, 1992).

660km and 1400km depth are about 400km apart. From the mid-mantle to the core-mantle boundary, average distance increases to about 600km. Finally, although we are not inverting for the structure of the core, points within this region are required for consistency of the Delaunay mesh. They are distributed about 1000 km apart.

Points are initially generated within each shell. Then, they are merged together. In this process, a control on the distance is introduced to avoid that nodes on the shell's boundary are too close. Whenever nodes are closer than the distance characteristic of that depth, the node is rejected. Figure 3.2 shows the distribution of points in the grid at different depths.

Polygonal tessellation of natural neighbors are always convex. Thus in 3D, a Delaunay mesh is wrapped by a convex surface called *convex hull*. We construct the convex hull by following the approach used in the TERRA code for generating the finite element mesh (Baumgardner, 1983). We include the earth within a regular icosahedron defined by 20 points. We connect the points to form the triangulation. New points are added by recursively subdividing the edges of the triangles. Each edge is divided by finding the middle point on the great circle connecting the two vertices. We stop when the incenter of the new triangle, i.e. the center of the triangle's incircle, is still above the earth's surface. The convex hull used in this study contains 642 points. The distance between the incenter of each triangle forming the convex hull and the center of the earth is 6377 km. Points on the convex hull are 885 km apart and 6340 km away from the center of the earth (Figure 3.3).

3.5 Basis functions

The velocity at any point in the model is defined by linear interpolation (Figure 3.4) using the data values at the four vertices of the tetrahedron containing that point. The location of a generic point $\mathbf{x}$ within a tetrahedron is described by the barycentric coordinates with respect to the four vertices $\mathbf{x}_k, k = 1, 4$ of the tetrahedron:

$$\mathbf{x} = h_1\mathbf{x}_1 + h_2\mathbf{x}_2 + h_3\mathbf{x}_3 + h_4\mathbf{x}_4 \quad with \quad h_1 + h_2 + h_3 + h_4 = 1. \tag{3.1}$$

Equation 3.1 is a linear system of four equations that we solve for the barycentric coordinates h_i as a function of $\mathbf{x}$. These coordinates give the weight by which the data values at the vertices of the tetrahedron contribute to the linear interpolation at point $\mathbf{x}$. The velocity at each vertex of a tetrahedron is known and the velocity c at any point $\mathbf{x}$ within a tetrahedron is given by:

$$c(\mathbf{x}) \;=\; h_1(\mathbf{x})c_1 + h_2(\mathbf{x})c_2 + h_3(\mathbf{x})c_3 + h_4(\mathbf{x})c_4 \;=\; \sum_{k=1}^{4} h_k c_k \tag{3.2}$$

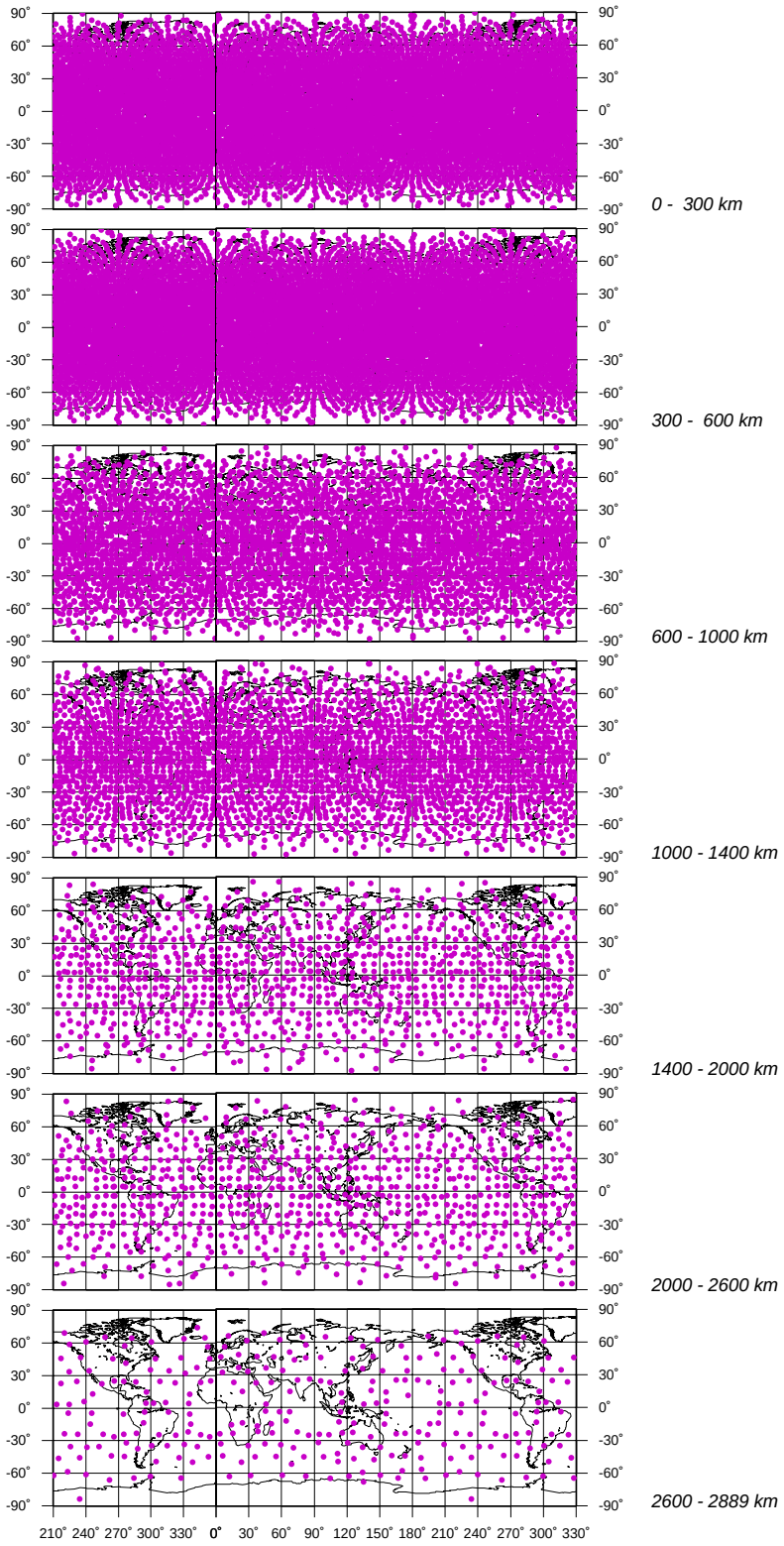


Figure 3.2 Distribution of points in the global grid as a function of depth.

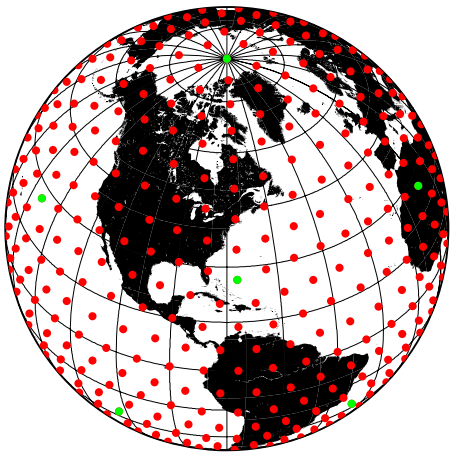
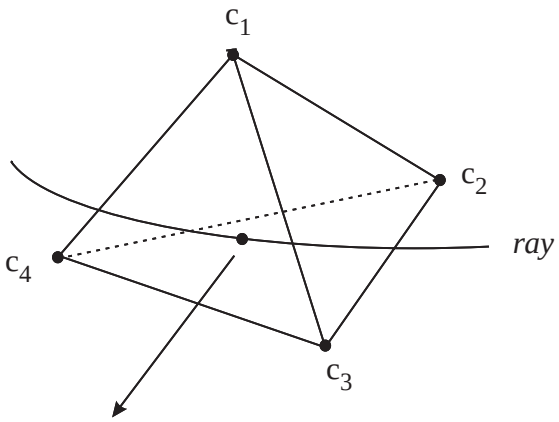


Figure 3.3 Convex hull of our global Delaunay mesh. Points seems located on the earth's surface, but they are actually 6400 km away from the center of the Earth. Green points are the initial points of the icosahedron. Red points are the result of the subdivision of the icosahedron.



$$c(\mathbf{x}) = h_1(\mathbf{x})\, c_1 \; + \; h_2(\mathbf{x})\, c_2 \; + \; h_3(\mathbf{x})\, c_3, + h_4(\mathbf{x})\, c_4$$

Figure 3.4 Interpolation scheme on a tetrahedron element of a Delaunay mesh. The velocity of a point within the tetrahedron is computed by linear interpolation of the velocity values at the vertices of the tetrahedron.

where h_k are the interpolant. These coefficients are important for the computation of the elements A_{ik} of the matrix A, as shown in equations 2.6–2.9.

Chapter 4

Global P and PP traveltime tomography: rays versus waves

A modified version of this chapter has been submitted for publication in the Geophysical Journal International as: Montelli R., Nolet. G., Masters G., Dahlen F.A., Hung S.-H.: Global P and PP traveltime tomography: rays versus waves

4.1 Résumé

Dans ce chapitre, je discute en detail les différentes formulations mathématiques du problème tomographique avec les deux méthodes envisagées: la théorie des rayons et la modélisation aux frequences finies. La différence entre les deux méthodes est principalement dans la matrice A du système à inverser. Dans la théorie des rayons, chaque ligne de la matrice correspond à un rayon reliant une source à une station. Les éléments A_{ij} sont des poids d'interpolation intègrant sur la section du rayon i contenue dans tous les tétraèdres qui ont le point j de la grille comme sommet commun. Dans la modélisation aux frequences-finies, chaque ligne de la matrice représente un noyau de Fréchet qui connecte une source à une station. Chaque élément A_{ij} peut être considéré comme une somme des effets du noyau i contenu dans tous les tétraèdres qui ont le point j de la grille comme sommet commun.

Pour tester la difference entre les deux approches tomographiques, j'utilise la base de données de Guy Masters à longue période. Je sélectionne 66238 P et 20167 PP-P temps d'arrivée. Les mesures ont été refaites par Guy Masters pour tenir en compte des erreurs instrumentales des stations seulement récemment decouvertes. Ces mesures de temps sont obtenues par des techniques de cross-corrélation. Les residus montrent un décalage en temps de 4 s pour les ondes P et d'environ 1 s pour les PP-P. Vu qu'on peut bien assumer que les mesures des PP-P sont indépendantes de la source, je considère l'excentrage observé dans les PP-P dû à des imprécisions dans le modèle de référence. Je corrige donc légèrement le modèle de référence *iasp91* dans la région des discontinuités (entre 400 km et 600 km), qui est connue

pour être la moins bien définie afin d'éliminer le décalage d'une seconde observée dans le temps PP-P. J'utilise le modèle de référence résultant pour recalculer les résidus des arrivées P. Cette correction est simplement technique. L'origine de ce décalage n'est pas encore très claire et sera recherchée lors d'études futures.

Le chapitre rend compte des résultats des inversions des données obtenues par les deux méthodes. Les resultats montrent que la théorie des rayons sous-estime les anomalies de vitesse par 30–60% en fonction de la profondeur par rapport à la modélisation aux fréquences finies. L'approximation à haute fréquence est une approximation trop brutale pour les données à longue période. Pour cela, les phénomènes de diffraction doivent être pris en compte. Pour analyser les effets dûs au maillage, j'ai fait une inversion en utilisant une grille très fine. Les images tomographiques qu'on obtient montrent qu'il n'y pas un effet très significatif de la densité du maillage.

Les modèles obtenus montrent une très claire correlation entre les anomalies de basse vitesse et les positions connues de plusieurs points chauds. Par exemple, on retrouve Easter Island, Tahiti, Hawaii et Kerguelen comme exemple de points chauds alimentés par des panaches profonds. Des points chauds comme Afar, Iceland et Reunion semble plutôt avoir leur origine dans la région de transition à 660 km. La tomographie jusqu'à maintenant n'était pas vraiment capable de montrer des images capables de convaincre que les panaches profonds existaient comme suggéré initialement par Morgan (1972). À cause de cette lacune de l'imagerie sismologique, un autre courant de pensée s'est développé pour lequel les points chauds sont générés simplement par des fractures de la lithosphere et n'ont donc pas une origine profonde. Pour vérifier la fiabilité de cette détection d'anomalies profondes, je montre les résultats d'un test de résolution. Le but est de vérifier que les panaches profonds ne sont pas dûs à un manque de résolution dans le manteau supérieur et sont bien un signal propre près de l'interface noyau-manteau. De même, ce test montre que les panaches superficiels ne sont pas définis en raison d'un manque de résolution à la base du manteau.

4.2 Abstract

This chapter presents the first implementation of finite-frequency global traveltime tomography for compressional waves. Our data set consists of long-period P and PP-P traveltimes measured by cross-correlation. We compare results of a finite-frequency analysis with an inversion obtained using standard ray theory. The traveltime of a finite-frequency wave is sensitive to velocity structure off the geometrical ray. The broadband P traveltime is sensitive to anomalies in a hollow banana-shaped region surrounding the unperturbed path, with the sensitivity being zero on the ray. Because of the minimax nature of the surface reflected PP wave, its sensitivity is much more complicated. We compute the 3D traveltime sensitivity efficiently by using the paraxial approximation in conjunction with ray theory and

the Born approximation (Dahlen et al., 2000). We compare tomographic models with the same χ^2 fit, for both ray theory and finite-frequency analysis. Depending on the depth and size of the anomaly, the amplitudes of the velocity perturbations in the finite-frequency tomographic images are 30%–60% larger than in the corresponding ray-theoretical images, demonstrating a major shortcoming of ray theory, and indicating that wavefront healing cannot be neglected when interpreting long-period seismic waves. The images obtained provide, for the first time, unambiguous evidence that a limited number of hotspots are fed by plumes originating in the lower mantle.

4.3 Introduction

Global P-wave tomographic models have so far been obtained by applying the approximation of ray theory. Waves propagate as rays only in the high-frequency limit of the elastodynamics equations of motion. All scattering interactions of the waveform with the heterogeneities of the propagation medium are neglected under the assumption that the velocity field varies slowly on the scale of the wavelength. Rays might bend and be deviated by the velocity structure, but energy is conserved along the ray and is only influenced by the earth's properties along an infinitesimally narrow path that follows Snell's law. The information contained in the P or S wave is reduced to a single number, i.e. the first break, which is assumed to correspond to the arrival of the highest-frequency observable wave. This simplifies the mathematics, but it is quite far from the physical reality where rays have a given thickness depending on the frequency content of the propagated wavefield (Kravtsov, 1988).

If the scale length of heterogeneities is comparable to the width of the Fresnel zone, finite-frequency effects are important. In regions where shadow zones or strong diffractors are present, scattering and diffraction phenomena may occur, and the validity of ray theory also breaks down. In ray theory, waves preserve the time shifts accrued upon passage through an anomaly somewhere along its path. Because of an intrinsic diffraction phenomenon called wavefront healing, finite-frequency wavefronts do not. Diffraction acts to fill in or heal irregularities in the wavefront. Also, diffracted waves of significant amplitude might interfere with the direct wave and introduce a bias in the traveltime measurements. Consequently, the traveltime of a finite-frequency seismic wave is sensitive to velocity anomalies off the geometrical ray. Simple diffraction theory shows that a narrow region around the ray path affects the traveltimes, whereas structure far from the ray paths play minor roles. Such a volume surrounding the geometrical ray path is called the first Fresnel zone and is loosely defined as the region where significant constructive interference of seismic energy takes place (Wielandt, 1987; Nolet, 1987, 1990; Müller et al., 1992; Nolet & Dahlen, 2000).

The widespread availability of broadband digital data has led to the recent de-

velopment of accurate techniques for traveltime measurements using cross-correlation of an observed body-wave phase with the corresponding spherical-earth synthetic phase (Bolton & Masters, 2001). Cross-correlation methods have also been used to measure differential traveltime of two phases at the same station (Kuo et al., 1987; VanDecar & Crosson, 1990; Woodward & Masters, 1991; Su et al., 1994). The measurement obtained in this way provides an integrated arrival-time difference between two waveforms, not simply the difference between onset times. Therefore the ray theoretical description for a traveltime along the ray may no longer be valid, and finite-frequency waveform modeling may be required. A major aim of this study is to investigate the necessity of finite-frequency theory for broadband seismic data.

The idea of using a ray of non-zero width to bridge the gap between rays and waves dates back at least to Hagendoorn (1954). He introduced the concept of beam width defined as the region falling within the first Fresnel zone. Various later attempts have been made to compute Fresnel zones for bandlimited seismic traveltimes in two and three dimensions (Gelchinsky, 1985; Woodward, 1992; Yomogida, 1992; Cardimora & Garmany, 1993; Stark & Nikolayev, 1993; Vasco & Majer, 1993; Li & Tanimoto, 1993; Marquering et al., 1998, 1999; Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000). They use the single scattering or first-order Born approximation, the Rytov approximation, or the Kirchhoff approximation to compute Fréchet sensitivity kernels that relate traveltime perturbations to velocity anomalies.

Because of the computational difficulties that accompany a three-dimensional formulation of sensitivity kernels for P or S waves, little effort has been put so far into using them for seismic tomography, at least outside the geophysical exploration community. Yomogida (1992) uses single scattering together with the paraxial ray approximation to model the first Fresnel zone, and implements the 2D Fréchet kernels in a synthetic inversion for teleseismic waves. Castle et al. (2000) and Husen & Kissling (2001) use what they call *fat* rays for the shear wave speed anomalies at the base of the mantle, and a tomographic study of the Antofagasta area (Northern Chile), respectively. However, fat rays account only approximately for wavefront healing effects, which may lead to incorrect results. Finally, Yoshizawa (2002) uses Fresnel-area ray tracing to compute the zone of influence about surface wave paths, and uses such kernels to extract a new Australian upper mantle model.

Marquering et al. (1998, 1999) and Zhao et al. (2000) present exact expressions for the Fréchet kernel for delay times, obtained by summing surface wave and normal modes, respectively. Zhao et al. (2001) uses such expressions in a 3D tomographic study of the Western Pacific region. The normal-mode theoretical kernel is by far the most general description of the sensitivity kernel around the unperturbed ray. However, since mode summations involve extensive computation, the implementation of the exact kernel in a global inversion of large data sets is not presently feasible. Dahlen et al. (2000) provide an alternative procedure to eco-

nomically compute the Fréchet kernels of a finite-frequency traveltime measured by cross-correlation of a broad-band waveform with a spherical-earth synthetic seismogram. The Green's function, the response to the point scatterer in Born theory, is represented as a sum of rays rather than modes. The paraxial approximation renders the computation of such kernels much more efficient.

In this chapter we present the first global traveltime tomography of finite-frequency waves based upon the 3D Fréchet kernel formalism of Dahlen et al. (2000), and compare the result with the one obtained with classical ray theory. We refer to their paper for a complete derivation of the Fréchet kernels.

4.4 Ingredients of the inverse problem

4.4.1 Data and reference model

We use arrival times of P and PP waves with 20 s dominant period (Bolton & Masters, 2001). We invert 66238 P traveltimes and 20167 PP–P differential traveltimes. Measurements have been corrected for data logger time errors.

Absolute times are mainly affected by noise and errors due to source mislocation. They are measured by cross-correlation of an observed pulse with a synthetic which is constructed by convolving the instrument response with a t^* attenuation operator (Bolton & Masters, 2001); the attenuation time t^* is kept constant at 1 s for P waves.

Differential PP–P times are obtained by cross-correlation of the Hilbert transformed direct P phase with the reflected PP phase (Woodward & Masters, 1991). Here the t^* operator accounts for the different attenuation histories of the direct and reflected phase, respectively. Differential traveltimes have the advantage of eliminating source and receiver bias, and are sensitive to shallow structure in the vicinity of the bounce point, thus allowing us to constrain regions of the world where there are no sources or receivers.

Predicted absolute and differential times are computed using the *iasp91* velocity model (Kennett & Engdahl, 1991). We correct for the signal due to ellipticity and for the effect of the crust (including topography). Crustal corrections are obtained using the $2^\circ \times 2^\circ$ global crustal model CRUST2.0 (model available through the Rem web site: *http://mahi.ucsd.edu/Gabi/rem.html*). The remaining residuals show a baseline shift of about –1 s for PP–P and about +4 s for P (Figure 4.1a). Although the origin of these offsets is not very clear, major candidates to explain them are the use of NEIC locations and imperfections in the 1D reference model. Since differential times such as PP–P are quite insensitive to source mislocations, the –1 s offset is best explained by the inadequacy of the 1D reference velocity model used or by the inadequacy of t^*. We eliminated this offset by making a slight change to the *iasp91* velocity model in the transition region (Fig. 4.2). This correction is purely technical and allows us to eliminate the PP–P offset for a range of distances.

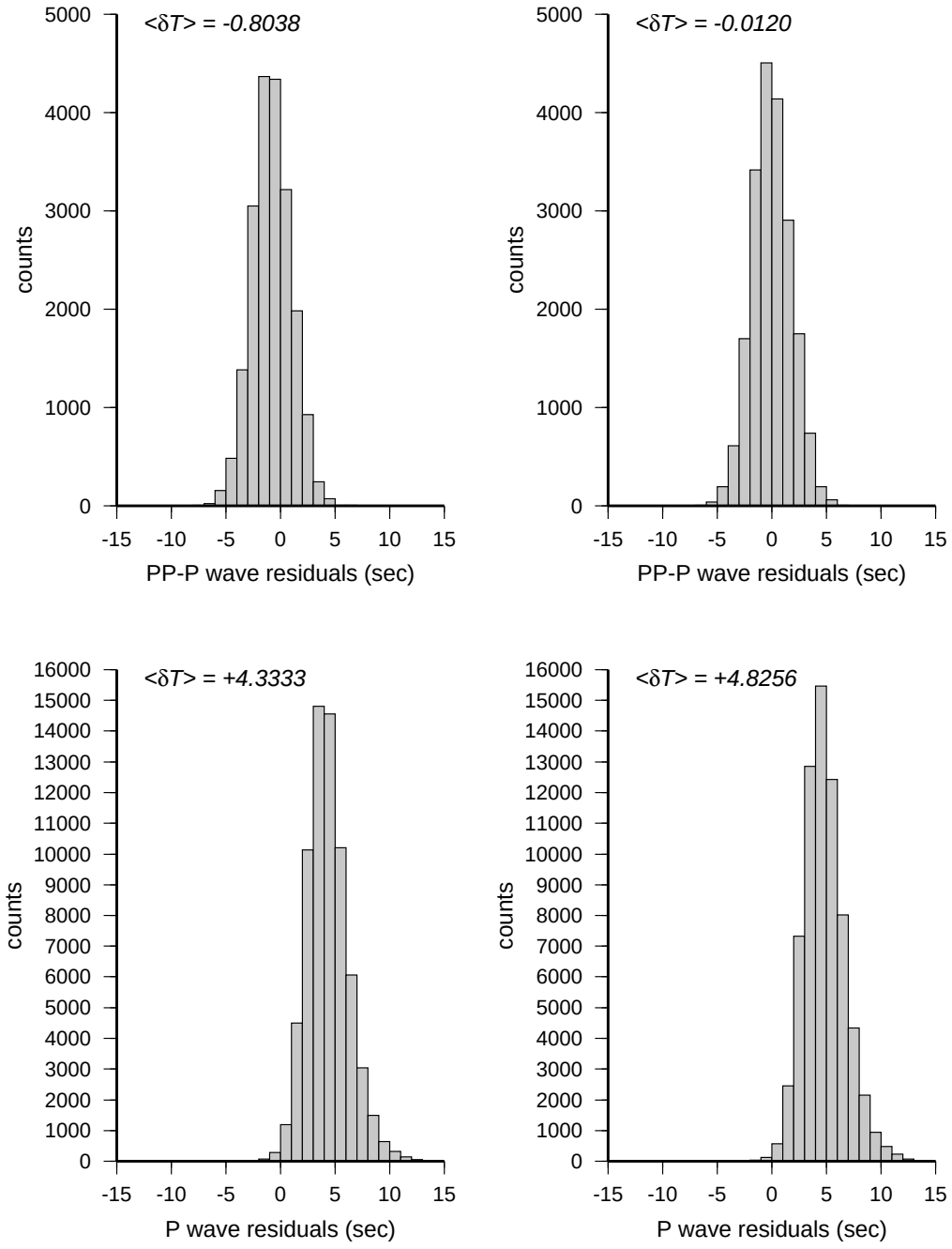


Figure 4.1 Top: Histograms of PP–P residuals of data used in this study. Bottom: P residuals computed with (left) *iasp91* velocity model and (right) a modified version of the *iasp91* velocity model shown in Figure 4.2 $< \delta T >$ indicates the average value of delay times. The offset of $\sim +5$ sec is discussed in the text.

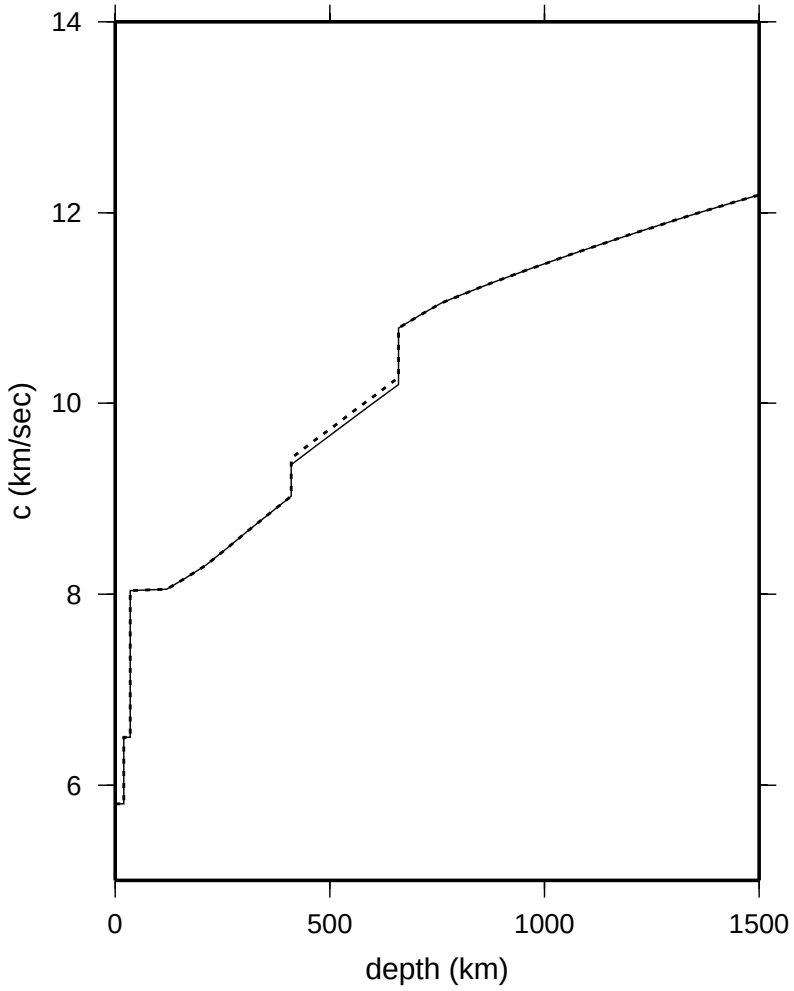


Figure 4.2 Comparison of the *iasp91* velocity model (solid line) with a model obtained by applying a slight perturbation (0.7%) to the velocity of the *iasp91* velocity model (dotted line) between 400 and 660km. The modified *iasp91* model removes the bias in the PP-P data.

After centering the PP–P times to zero, the offset in the P data increases to about +5 s (Fig. 4.1b). The origin of this bias is not fully understood, but we suspect it is related to the fact that we use origin times established for high frequency waves, whereas the measurements are done at 20 s with a simple source pulse in the cross-correlation. This shortcoming is inherent to our data set. In order to minimize the effect, we have applied a constant correction to all P traveltimes such that the average $<\delta T_{\mathrm{P}}>=0$. This allows us to damp perturbations in origin times during the actual inversion.

4.4.2 Model parameterization

We sample the velocity structure by using an irregular distribution of points to form a Delaunay mesh (Watson, 1981, 1992; Sambridge et al., 1995). In 3D, a Delaunay mesh is an aggregation of space filling, disjoint, irregular tetrahedrons uniquely defined. We build the Delaunay connections by using *qhull*, a package distributed by the Geometry Center of Minneapolis (Barber et al., 1996).

Node spacing is adapted to the expected resolving length of our data and ranges from about 200 km in the upper mantle to about 600 km in the lower mantle. The total number of nodes we use to model the global Earth is $m = 19,279$.

The velocity c at any point $\mathbf{x}$ in the model is defined by linear interpolation within each tetrahedron spanned by this mesh, formally expressed as:

$$\delta c(\mathbf{x}) = \sum_k \delta c_k h_k(\mathbf{x}), \tag{4.1}$$

where h_k denote the interpolation functions, k being an index over the four nodes of the tetrahedron that contains $\mathbf{x}$.

4.4.3 Delay times tomography: rays and waves

To investigate the effects of wavefront healing, we shall compare ray-theoretical tomography with finite-frequency wave tomography. In the following two sections we briefly review the analytical description of both formulations, which in the end reduce to a similar discrete system $\mathsf{Ax=b}$ of n traveltime shifts b_i measured by cross-correlation and m velocity perturbations x_j, which we solve in both cases in a least-squares sense (Paige & Saunders, 1982; Nolet, 1985). The inversion technique that we use is well established and is described in detail by Nolet (1987) and Spakman & Nolet (1988).

Rays

In the ray approximation, a measured traveltime residual is given by a 1D line integral along the unperturbed spherical-earth ray:

$$\delta T = -\int_{\mathrm{ray}} c(r)^{-2} \delta c(\mathbf{x})\, dl, \tag{4.2}$$

where dl is the differential arclength along the ray, $c(r)$ is the background wavespeed at radius r (i.e modified *iasp91*), and $\delta c(\mathbf{x})$ is the 3D heterogeneity that one is seeking to image. Fermat's Principle allows us to use the raypath computed for the background velocity $c(r)$ (e.g. Nolet, 1987). By virtue of the linear interpolation on the tetrahedron structure (equation 4.1), the expression for the traveltime shift δT becomes:

$$\delta T_i = \sum_j A_{ij} \frac{\delta c_j}{c_j}, \tag{4.3}$$

where $c_j = c(r_j)$ at node j, and the elements of the matrix A are given by:

$$A_{ij} = -\int_{i\text{thpath}} dl\, c^{-1}(r) h_j(\mathbf{x}), \tag{4.4}$$

with i the datum number, and j the node index.

In the case of a differential traveltime PP–P, the delay time $\delta T^{\mathrm{PP-P}}$ is given by $\delta(T^{\mathrm{PP}} - T^{\mathrm{P}})$ and the elements of the matrix A are simply the difference between the matrix elements of the two phases individually.

Finite-frequency waves

In finite-frequency tomography the 1D integral along the geometrical ray is replaced by a 3D volume integral:

$$\delta T = \int_{\oplus} K(\mathbf{x}) \frac{\delta c}{c} d^3\mathbf{x} \tag{4.5}$$

over the entire Earth $\oplus$ in which the wave-speed perturbation is non-zero, $\delta c/c \neq 0$. The quantity $K(\mathbf{x})$ is the 3D Fréchet kernel of a finite-frequency traveltime shift δT that has been measured by cross-correlation of a broadband waveform with a spherical earth synthetic.

Following Dahlen et al. (2000) the 3D Fréchet kernel $K(\mathbf{x})$ is expressed by a double ray sum over all scattered body waves. This formula reduces to an easily computable expression by invoking the paraxial approximation which eliminates the need to conduct repeated two-point ray tracing. By ignoring all forward scattering rays that are not of the same type as the unperturbed path, the Fréchet kernel reduces to a compact expression given by (Dahlen et al., 2000):

$$K(\mathbf{x}) = -\frac{1}{2\pi c} \sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|} \frac{\displaystyle\int_0^\infty \omega^3 \left|\dot{m}(\omega)\right|^2 \sin\Phi \, d\omega}{\displaystyle\int_0^\infty \omega^2 \left|\dot{m}(\omega)\right|^2 d\omega}, \tag{4.6}$$

where

$$\Phi = \frac{1}{2}\omega \mathbf{q}^{\mathrm{T}} \cdot (\mathbf{M}' + \mathbf{M}'') \cdot \mathbf{q} - [\mathrm{sig}(\mathbf{M}' + \mathbf{M}'') - 2]\frac{\pi}{4} \ . \tag{4.7}$$

The matrices $\mathbf{M}'$ and $\mathbf{M}''$ are the forward and backward 2×2 traveltime Hessians along the central ray, and $\mathbf{q}$ is the location vector of a scatterer at $\mathbf{x}$, in ray

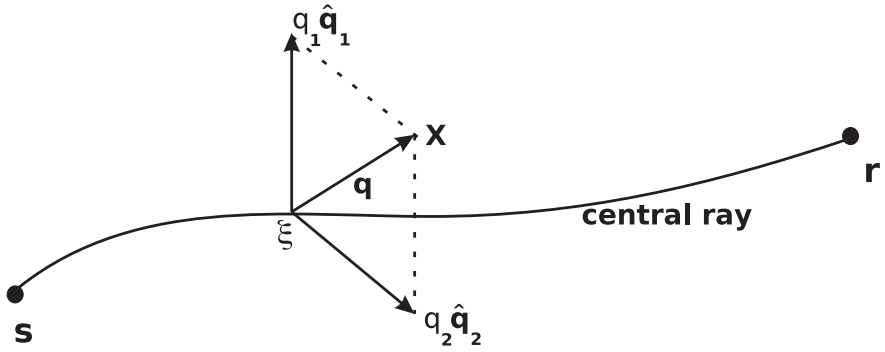


Figure 4.3 Perpendicular projection of a scatterer $\mathbf{x}$ onto the paraxial point ξ, situated on the central geometrical ray from the source $\mathbf{s}$ to the receiver $\mathbf{r}$. The off-path difference vector is expressed in terms of two orthogonal unit vectors: $q = q_1\hat{\mathbf{q}}_1 + q_2\hat{\mathbf{q}}_2$. The ray centered coordinates of the scatterer are $\mathbf{x} = (q_1, q_2, l)$ where l is the arclength along the central ray Dahlen et al. (2000).

coordinates (see Figure 4.3). The symbols det and sig denote the determinant and the signature, or the number of positive minus the number of negative eigenvalues of $\mathbf{M}' + \mathbf{M}''$, respectively; ω is the angular frequency, and $c = c(r)$ is the background spherical-earth velocity. The kernel for a single, well-isolated seismic phase depends only upon the sum $\mathbf{M}' + \mathbf{M}''$ of forward and backward traveltime Hessians along the central geometrical ray. The quantity Φ (eq. 4.7) represents the phase delay of the wave scattered from $\mathbf{x}$. The quantity $|\dot{m}(\omega)|^2$ is the power spectrum of the attenuated synthetic (see section 4.4.1) and specifies the frequency content of the cross-correlated arrivals. This is a reminder that $K(\mathbf{x})$ is the Fréchet kernel of a finite-frequency traveltime measurement δT. We have ignored a possible bias in dominant frequency caused by the correlation operator emphasizing the early part of the waveform rather than a full period.

Written out explicitly, the 3D integral for the traveltime shift δT is given by:

$$\begin{aligned} \delta T &= -\frac{1}{2\pi}\int_0^L dl \iint_{-\infty}^{\infty} dq_1 dq_2 (1 + q_k \partial_k c) c^{-2} \delta c \\ &\times \sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|} \frac{\int_0^\infty \omega^3 \, |\dot{m}(\omega)|^2 \sin\Phi \, d\omega}{\int_0^\infty \omega^2 \, |\dot{m}(\omega)|^2 \, d\omega}. \end{aligned} \tag{4.8}$$

The limit $\pm\infty$ on the transverse integrals over q_1, q_2 are purely formal; in practice, the kernel $K(\mathbf{x})$ is negligible except within the first one or two Fresnel zones about the central ray.

Again, by virtue of the linear interpolation on the tetrahedron structure (equa-

tion 4.1) the expression for the traveltime shift δT becomes:

$$\delta T_i = \sum_j A_{ij} \frac{\delta c_j}{c_j}, \tag{4.9}$$

where the elements of the matrix A are given by:

$$\begin{aligned} A_{ij} &= -\frac{1}{2\pi} \int_{i\text{thpath}} dl \iint_{-\infty}^{\infty} dq_1 dq_2 (1 + q_k \partial_k c) c^{-1} h_j(\mathbf{x}) \\ &\times \sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|} \frac{\int_0^{\infty} \omega^3 \left|\dot{m}(\omega)\right|^2 \sin\Phi d\omega}{\int_0^{\infty} \omega^2 \left|\dot{m}(\omega)\right|^2 d\omega}, \end{aligned} \tag{4.10}$$

with $i = 1, 2, ..., n$ the datum number, $j = 1, 2, ..., m$ the node index and $k = 1, 2$ the ray coordinate index.

In the case of differential traveltimes $\delta T^{\mathrm{PP-P}} = \delta(T^{\mathrm{PP}} - T^{\mathrm{P}})$, which is related to the difference of the individual Fréchet kernels: $K^{\mathrm{PP-P}}(\mathbf{x}) = K^{\mathrm{PP}}(\mathbf{x}) - K^{\mathrm{P}}(\mathbf{x})$.

From a mathematical point of view the difference between the ray- theoretical and finite-frequency approach resides in the elements of the matrix A. In the ray-theoretical formulation, each row of the matrix A represents the geometrical ray connecting the source $\mathbf{s}$ to the receiver $\mathbf{r}$. The elements A_{ij} are interpolation weights integrated along the arclength of the ray i contained in all tetrahedrons having node j as common vertex. In the finite-frequency modeling, each row of A represents one Fréchet kernel connecting the source $\mathbf{s}$ to the receiver $\mathbf{r}$. Therefore each elements of the matrix A_{ij} can be seen as the integrated effect of the kernel i contained in all tetrahedrons having node j as common vertex.

Typical widths of the sensitivity region (i.e. the diameter of the ring around the unperturbed geometrical ray) at the turning point of a direct P wave range from about 1000 km to about 1300 km for a 60° and 80° epicentral distance respectively. Delay times are relatively insensitive to velocity perturbations close to the geometrical ray. This region of insensitivity is smaller near the source and receiver but can extend to about 400 km near the turning point of a P wave at 80° epicentral distance, giving the characteristic shape of a doughnut to the banana kernel (Figure 4.4a). PP waves show a much more complicated shape of the sensitivity region than direct P (Figure 4.4b). The PP wave from a source to a receiver passes through a source-to-receiver caustic where it experiences a non-geometrical $\pi/2$ shift; the backward wave from receiver to source passes through the corresponding receiver-to-source caustic. Upon passage through these caustics the shape of the kernel changes drastically and does not resemble a hollow banana any more. The on-ray PP sensitivity is still identically zero between the source and the source-to-receiver caustic, and between the receiver and the receiver-to-source

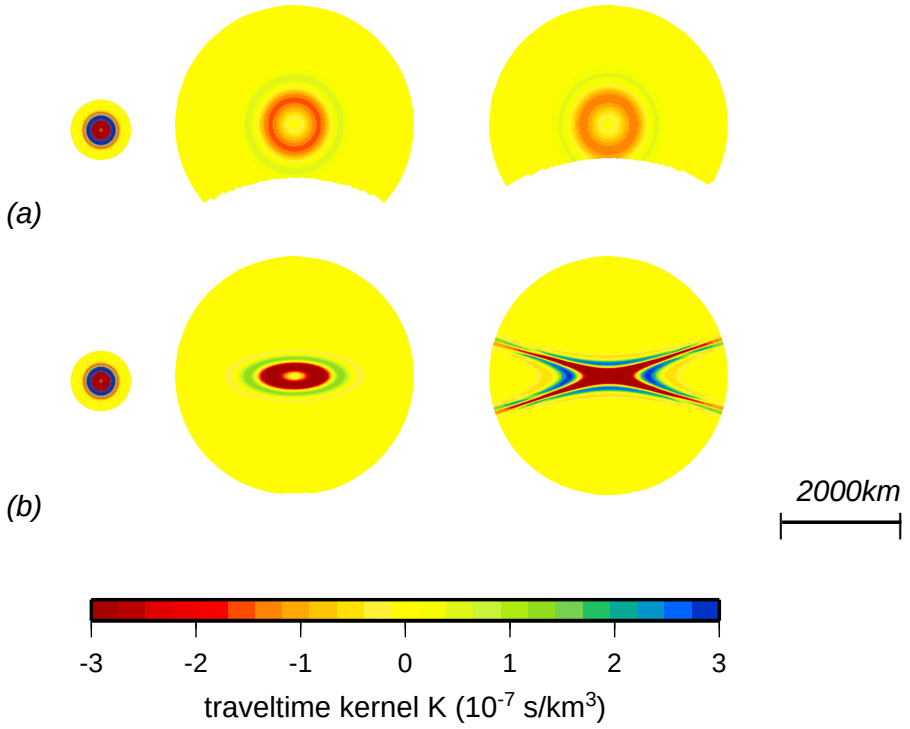


Figure 4.4 Ray-perpendicular cross sections of the Fréchet kernel (a) for a P wave between the source and the turning point and (b) for a PP wave between the source and the bounce point. The shape of the PP kernel changes drastically upon passage of a caustic.

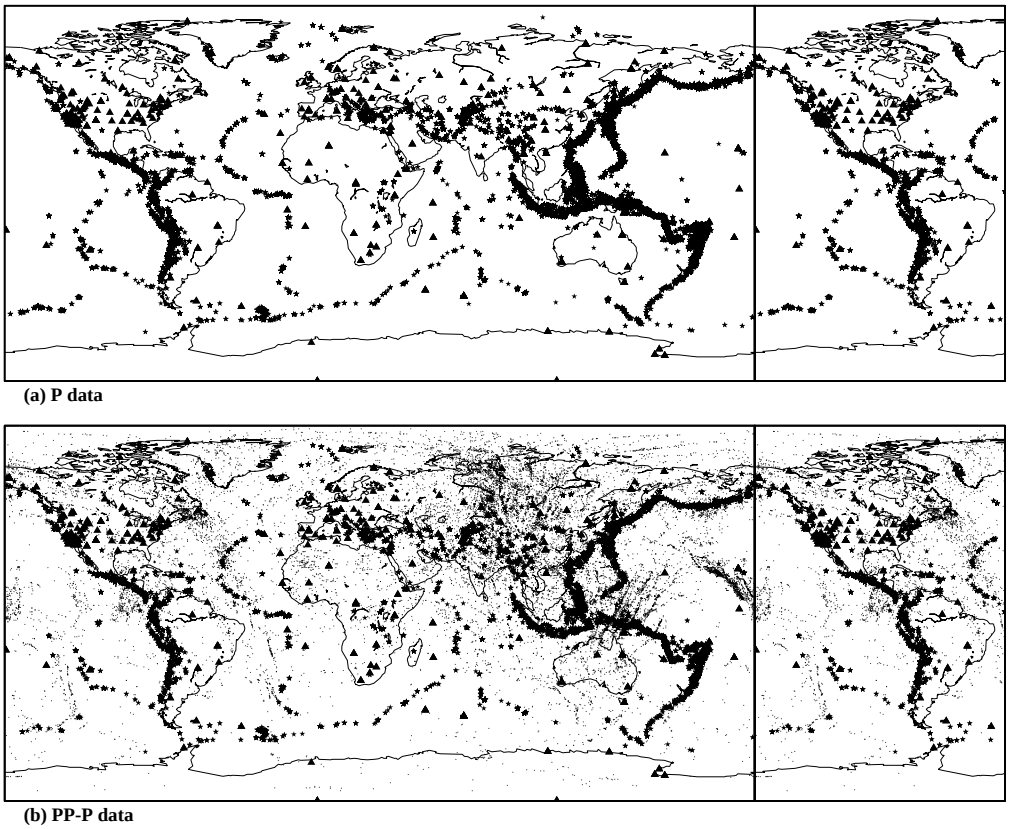


Figure 4.5 (a) distribution of sources (stars) and receivers (triangle) for P data; (b) distribution of sources (stars), stations (triangles) and bounce points (dots) for PP data.

caustic; however, it is nonzero between the two caustics (Fig. 4.4b). The characteristic zero-to-maximal-to-zero sensitivity variation of the PP waves along the geometrical ray is due to the jumps in the term $\mathrm{sig}(\mathbf{M}' + \mathbf{M}'')\pi/2$ of equation (4.7) (Dahlen et al., 2000; Hung et al., 2000). Steps of integration in the computation of the matrix A have been dynamically adapted along the kernel to take into account changes in the size and shape of the sensitivity region.

The introduction of volume kernels in the inversion significantly improves the "ray path" coverage. Also, it is clear from Figure 4.4 that a 1D line integral along a ray is an extremely crude approximation of the complex sensitivity region of a PP wave. PP waves are particularly useful in global tomography because they provide constraints on regions where there are no sources or receivers. Figure 4.5a shows the sources and stations distribution of the P waves contained in our data set. Significant parts of the globe such as the southern hemisphere are badly covered by paths. Because of the bounce points, the introduction of PP waves in the inversion significantly enhances the path coverage in the upper mantle.

Because of the wide span of the sensitivity regions, finite-frequency waves sam-

ple a larger volume of the model than do the rays. As result, the matrix A for finite-frequency waves is a factor of 16 less sparse than the one constructed with rays for our parameterization. However, the sensitivity is significant only in a limited region around the geometrical ray and many of the matrix elements are small.

In Figure 4.6 we show a comparison between the column density for the two matrices. We define the density for a given node of the grid as the sum of all the elements of the matrix in the column correspondent to that node. As expected, the density is larger and broader for the finite-frequency matrix.

4.5 Technical aspects of the inversion

We simultaneously invert for perturbations in velocity $\delta c/c$ and in hypocentral parameters (origin time, longitude, latitude, depth). Our system of inversion becomes $[\mathsf{A}\,\mathsf{H}]\mathsf{x} = \mathsf{b}$ where H is the matrix of partial derivatives with respect to the source parameters and where $\mathsf{x} = [\mathsf{x_c}, \mathsf{x_h}]^{\mathrm{T}}$ now contains both unknown velocity perturbations $\mathsf{x_c}$ and source corrections $\mathsf{x_h}$. The quantity b on the right side is the vector of the delay times δT.

We have $5,738$ sources and the grid consists of $19,279$ points giving a total of $42,231$ unknowns and $n = 86,405$ observations (see section 4.4.1). Changes in origin time and source location are computed with respect to NEIC values.

We have many more data than unknowns, yet due to the sparseness of the ray distribution, the problem has a partly underdetermined nature. Also, because of errors, the system of equations is inconsistent. To regularize the inversion, we apply norm damping to the velocity perturbation $\mathsf{x_c}$, which biases to a lower amplitude solution, and to the source corrections $\mathsf{x_h}$, which limits changes in origin times and hypocentral location coordinates: the strength of this $\|\mathsf{x_c}\| \to \mathsf{0}$ and $\|\mathsf{x_h}\| \to \mathsf{0}$ norm damping is controlled by two tunable parameter ϵ_c and ϵ_h. To supplement the norm damping, we also apply second derivative damping, $\|\mathsf{S x_c}\| \to \mathsf{0}$, governed by a parameter ϵ_S; this biases the solution toward smooth velocity variations in every direction (latitudinal, longitudinal and radial). Due to the irregular nature of the grid, our smoothing operator is not truly a canonical second derivative ∇^2 since it averages on the total number of node neighbors. In finite-frequency modeling, the effect of the uneven sampling is reduced by the implementations of kernel volumes. Because of their intrinsic averaging nature, a natural smoothing is implicitly introduced into the inversion reducing the degree to which we have to implement the "ad hoc" smoothing scheme in this case.

How are we going to compare two models obtained with two different techniques? Assuming a Gaussian distribution of the data errors, the maximum-

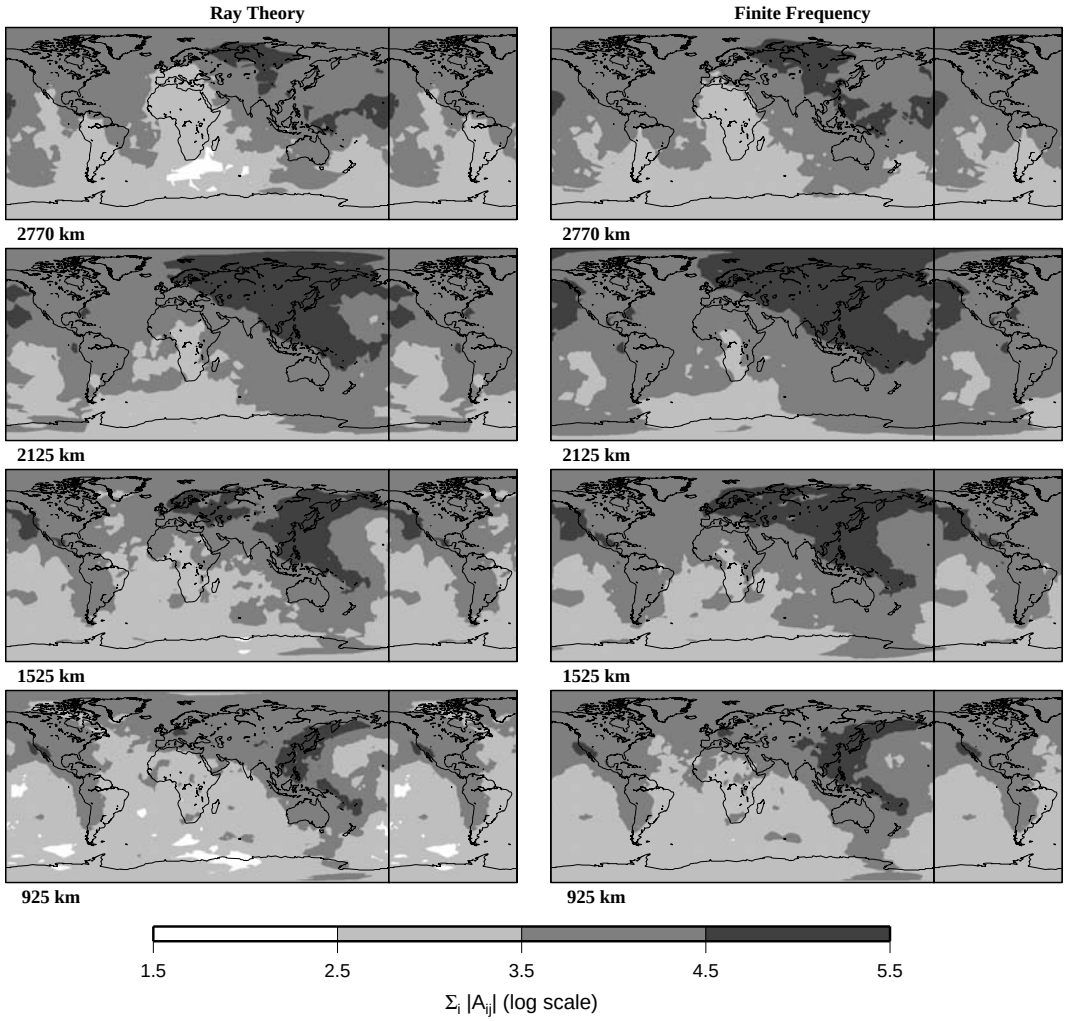


Figure 4.6 Sections at different depth of the density of the matrix A for ray theory (left) and finite-frequency waves (right) expressed as sum of the absolute values of the elements of each column of the matrix A. Note that the maps have been "wrapped around" to aid in the visualization of patterns in the vicinity of the Greenwich meridian.

likelihood estimate of the model parameters is obtained by minimizing χ^2:

$$\sum_i \left(\frac{\sum_j A_{ij} x_j - b_i}{\sigma_i} \right)^2 = \chi^2 \tag{4.11}$$

where σ_i are the estimated errors in the data. Each measurement b_i is characterized by a grade A, B, and C based on the confidence of the pick and how well the waveforms match. Each grade corresponds to a different error σ_i. The choice of σ_i in the inversion is in a sense a subjective one. A priori estimates of the σ_i values given to the grading were slightly raised on the basis of the misfits obtained in preliminary inversions for P and PP-P separately. While this introduces a certain amount of circular reasoning, the σ_i adopted are close enough to the a priori estimates to provide a reasonably objective yardstick for inversion. We assigned an error σ of 0.44, 0.53, 0.79s to A, B and C, respectively, for P and 0.75, 0.95, 1.15, respectively, for PP–P residuals.

To the extent that these errors are normally distributed, the quantity χ^2 is a sum of n squares of normally distributed quantities, each normalized to unit variance (Press et al., 1992). Thus a typical value of χ^2 for a good fit is $\chi^2 = N$. More precisely, the χ^2 statistic has a mean n and a standard deviation $\sqrt{2n}$. In the model space this motivates us to look for solutions which lie on the boundary of the allowable misfit region, i.e. where $\chi^2 = n$. Ray theoretical and finite-frequency tomography will have different regions of allowable misfit. We compare models which have exactly the same χ^2, with $\chi^2 \simeq n$. We shall not be dogmatic in requiring χ^2 to equal n precisely, since our a priori estimates for σ_i are somewhat uncertain. In our inversion, the initial value of χ^2 is about $7.5n$ and the final χ^2 is kept constant at $102,500$, i.e., about $1.2n$.

The least-square system we are minimizing can finally be expressed as:

$$\chi^2 + \epsilon_c \|\mathsf{x}_c\|^2 + \epsilon_h \|\mathsf{x}_h\|^2 + \epsilon_S \|\mathsf{S}\mathsf{x}_c\|^2 = \text{minimum}, \tag{4.12}$$

where ϵ_S and S are the smoothing factor and smoothing operator, respectively.

The three damping factors (ϵ_c for the model norm, ϵ_h for the hypocentral parameters, and ϵ_S for the smoothing) define a three-dimensional space. Each point in this space corresponds to a particular solution, and models with the same χ^2 span 2D surfaces. We experimented with changing both the damping ϵ_c and the smoothing ϵ_S while keeping the damping factor ϵ_h constant. Solutions confirm that, because of their differential nature, PP-P data are insensitive to source parameters. Also, because of the implicit smoothing, finite-frequency inversion requires much smaller damping and smoothing parameters to achieve the same χ^2.

Figure 4.7 shows the tradeoff between the squared model velocity norm $\|\mathsf{x_c}\|^2$ and the model roughness $\|\mathsf{S}\mathsf{x_c}\|^2/\|\mathsf{x_c}\|^2$ for $\chi^2 = 1.2n$. Finite-frequency solutions have a larger model norm. The labels RT (Ray Theory) and FF (Finite Frequency) in the plot indicate the two solutions we compare in the following section.

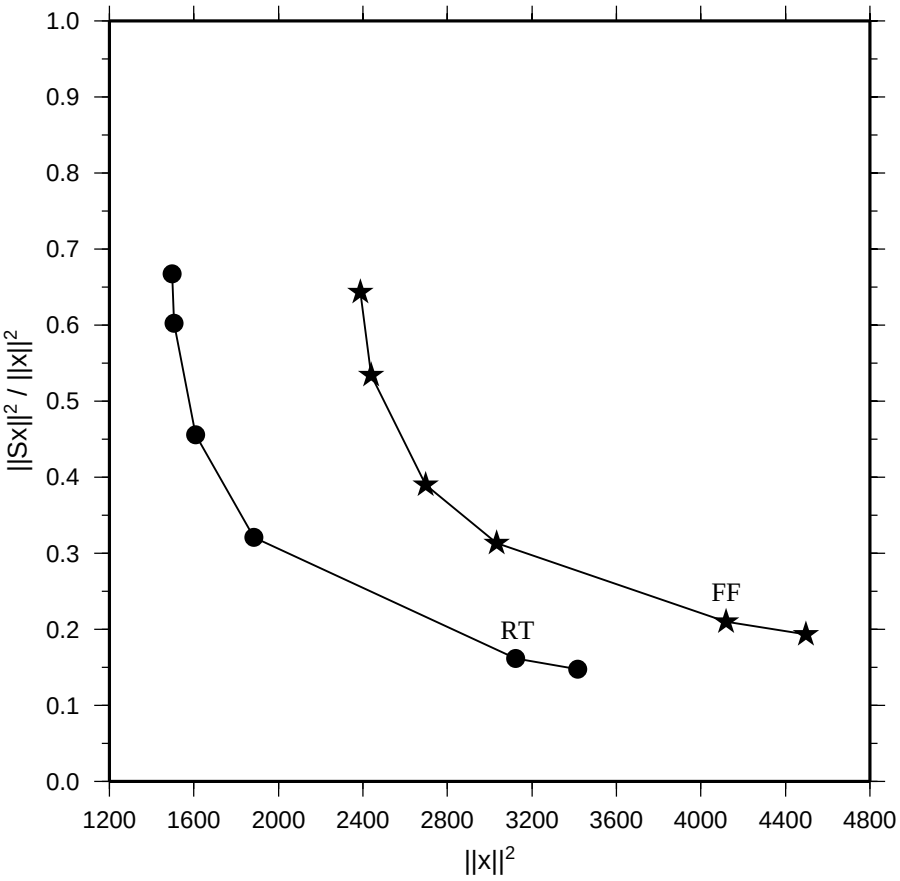


Figure 4.7 Model norm versus roughness of the solution for finite-frequency (stars) and ray-theoretical (dots) tomography with $\chi^2 = 1.2N$. Labels RT and FF indicate locations of finite-frequency and ray theory models compared in Figures 4.8– 4.10.

4.6 Inversion results

In our inversion, the norm damping (ϵ_h) applied to the hypocentral part of the solution is such that 95% of the changes in the origin time lie between ± 1.2 sec and 95% of the hypocentral coordinates between ± 14 km.

Figure 4.8 shows P wave velocity perturbations with respect to our background model for ray theory (left) and finite-frequency (right) at different depths. It is evident that amplitudes of the velocity anomalies are larger in the finite-frequency tomography than they are in the ray tomography. As can be seen in Figure 4.7 the two solutions show heterogeneities of the same scale length, have similar roughness and are characterized by very minimal norm damping.

Figure 4.9 is a quantitative measure of the ratio between velocity changes in the wave versus ray tomography at different depths. To compute these histograms, we consider only velocity perturbations with an absolute value greater than 0.2%.

The discrepancy visibly increases as a function of depth. In Table 4.1 (top) we summarize the analysis of the ratios. In the following analysis, we ignore the small fraction of uncorrelated (i.e., negative) ratios (Figure 4.9). Above the core-mantle boundary, between 2889 km and 2408 km depth, anomalies in the finite-frequency tomographic model are on average 1.6 times larger than in the ray-theoretical model. Smaller anomalies are more affected than larger ones, as expected. The amplitude ratio decreases toward the surface where it is of the order of 1.3. The discrepancy is more evident at depth because wavefront healing effects become more pronounced. The hole of the banana-doughnut kernel becomes wider as a function of depth. Anomalies might partially be hidden in the doughnut hole, where the traveltime sensitivity is zero, and the anomaly amplitudes must compensate for that.

The plot of the model root-mean-square (rms) as a function of depth shows an average discrepancy of about 20% or less (Figure 4.10a) between the average absolute value of the velocity changes in the two models. This average includes smooth areas with small velocity anomalies. The ray-theoretical and finite-frequency solutions are strongly correlated, with correlation coefficients larger than 0.9. However, the correlation diminishes slightly as a function of depth (Figure 4.10b).

If we believe our estimates for the error σ_i in our data to be correct then we can push the system to get to χ^2 exactly equal to n. Figure 4.11 shows the comparison of the ray-theoretical and finite-frequency wave tomography in this $\chi^2 = n$ case. Numerous small-scale heterogeneities appear in the model. Many of these are of major interest since they are perfectly correlated with known structures, such as hotspots. Because of the presence of small-scale anomalies at all depths, finite-frequency velocity changes are now on average about 1.7 times larger than the corresponding ray-theoretical ones between the bottom of the mantle and 2408 km depth and become even more pronounced near the Earth's surface where they are about 2 times larger (Figure 4.12 and Table 4.1 bottom). In this case the fraction of uncorrelated velocity anomalies is more significant, ranging from about 10% near

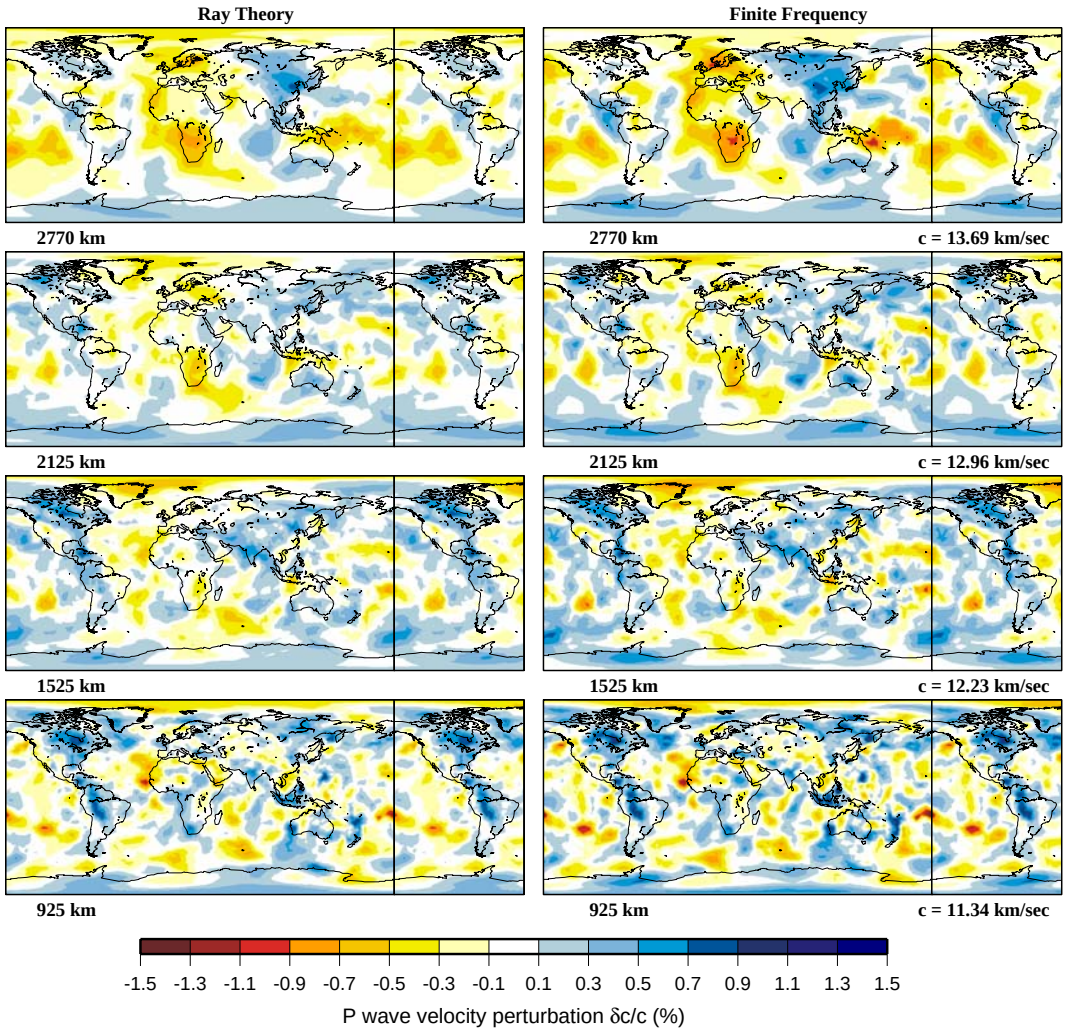


Figure 4.8 Comparison between velocity maps of the smooth, $\chi^2 = 1.2n$ model for ray theory (left) and finite-frequency theory (right) at different depths. The quantity c is the velocity in the reference model shown in Figure 4.2. Maps have been "wrapped around" to aid in visualization of patterns both in the Atlantic and the Pacific Oceans.

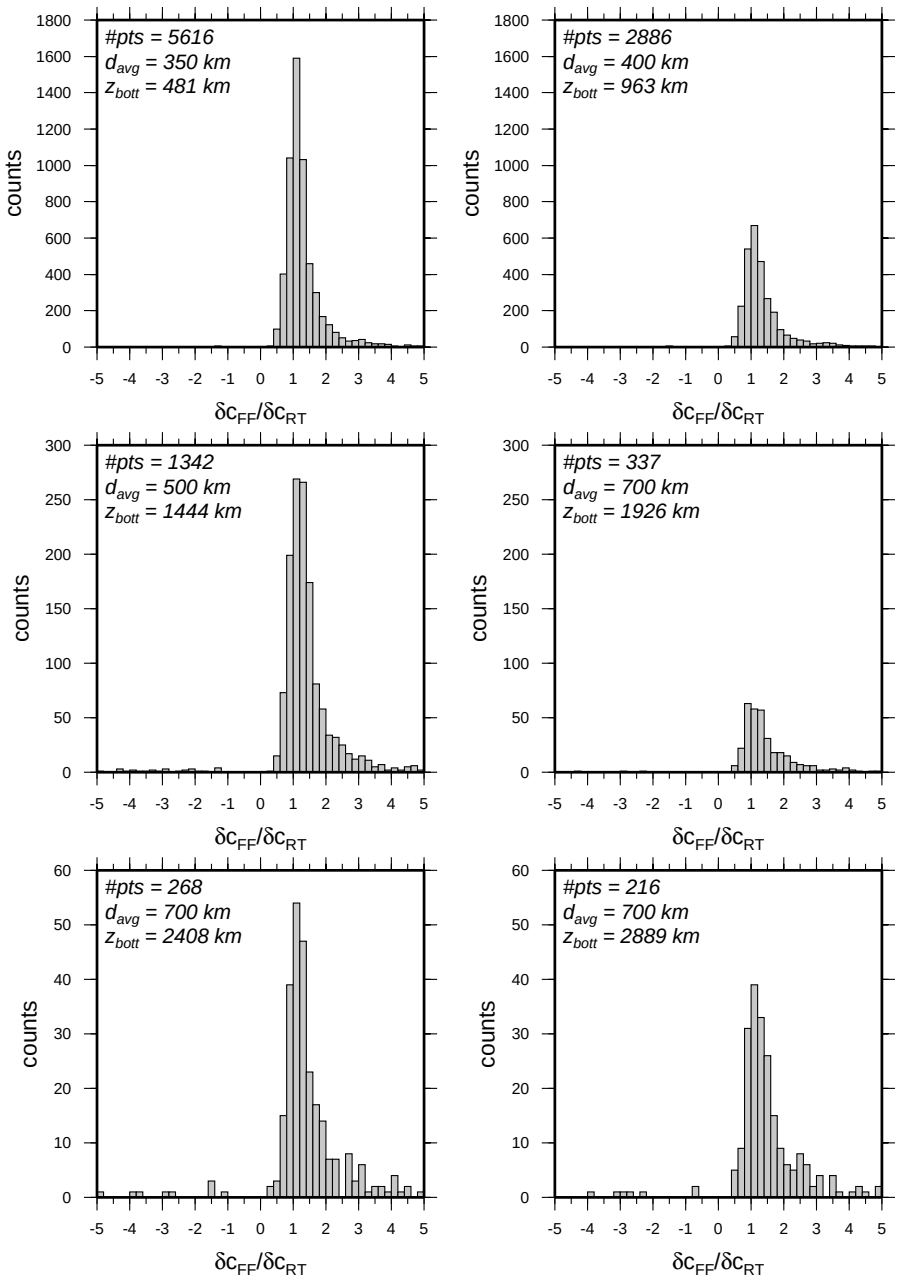


Figure 4.9 Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the model with $\chi^2 = 1.2N$. Only changes with absolute value larger than 0.2% are binned. Depth in the histograms is representative of the depth at the bottom of the shell considered (z_{bott}). There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers. Information about the mean and median for each shell are contained in Table 4.1a. For each layer the total number of points of the grid present in that layer (#pts) and their average distance (d_{avg}) are indicated.

Depth (km)	Pos Mean	Pos Median	Anticorr (%)	Neg Mean	Neg Median
0 to 481	1.29	1.15	0.8	−2.76	−2.68
481 to 963	1.35	1.18	1.4	−2.74	−2.66
963 to 1444	1.44	1.27	2.0	−2.88	−2.91
1444 to 1926	1.49	1.28	0.9	−3.11	−2.80
1926 to 2408	1.53	1.28	3.5	−2.67	−2.78
2408 to 2889	1.57	1.33	3.3	−2.35	−2.73
Depth (km)	Pos Mean	Pos Median	Anticorr (%)	Neg Mean	Neg Median
0 to 481	2.01	1.79	11	−2.19	−1.91
481 to 963	1.77	1.53	16	−2.01	−1.62
963 to 1444	1.78	1.49	19	−2.07	−1.80
1444 to 1926	1.77	1.43	17	−2.11	−1.70
1926 to 2408	1.76	1.51	28	−2.24	−1.97
2408 to 2889	1.71	1.37	32	−2.17	−1.72

Table 4.1 Quantitative analysis of the ratio of finite-frequency versus ray-theoretical model parameters plotted as histograms in Figures 4.9 and 4.12 for the smooth and rough model, respectively. "Anticorr" represents the fraction of model parameters with significant anticorrelation. We also compute the mean and the median. All values are computed for the positive ratios, i.e., larger than 0.2% (Pos – on the left); and negative ratios, i.e., smaller than −0.2% (Neg – on the right). Uppermost table contains values for the smooth, $\chi^2 = 1.2N$ model; lowermost table contains values for the rough, $\chi^2 = N$ model.

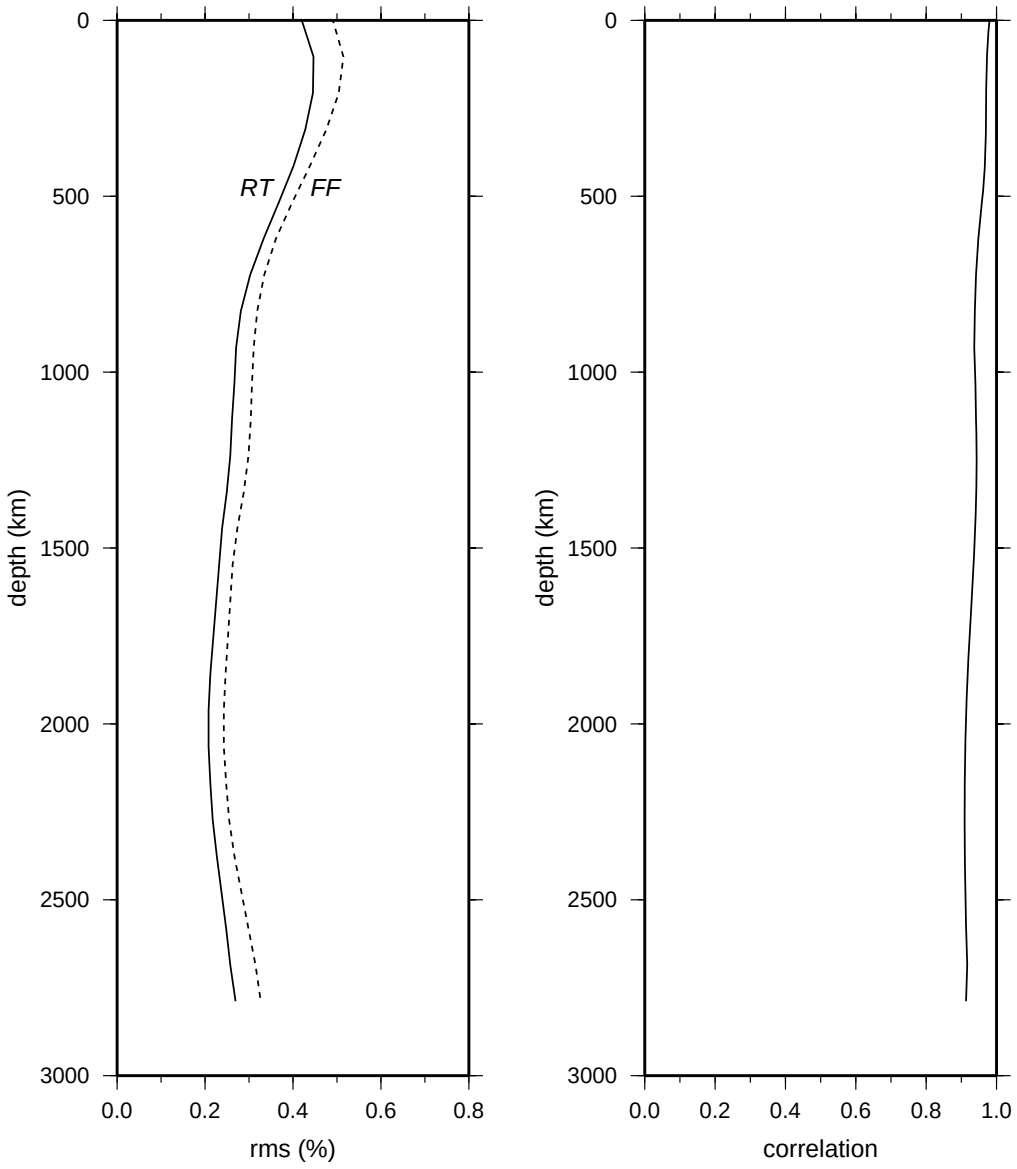


Figure 4.10 (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF – dashed line) and ray-theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2 = 1.2n$.

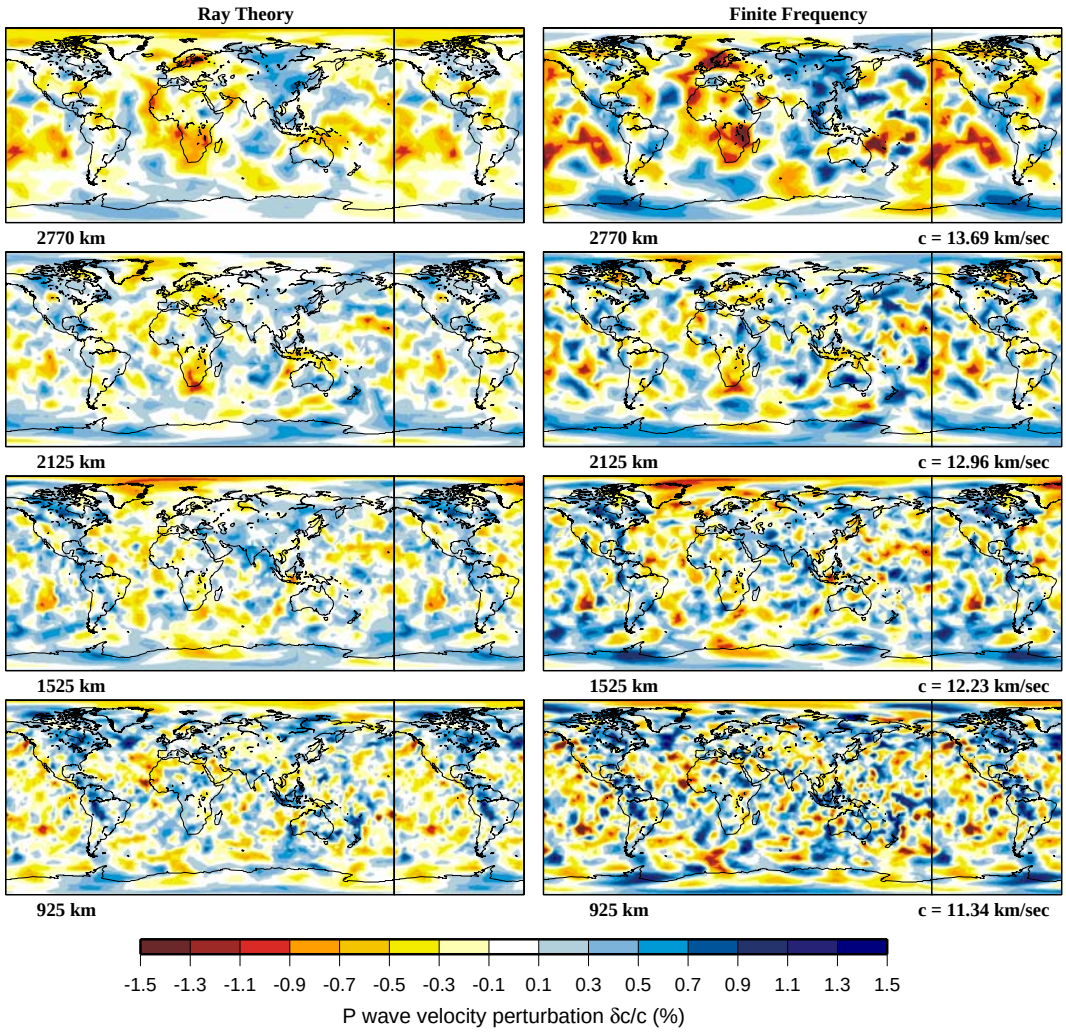


Figure 4.11 Comparison between velocity maps of the rough, $\chi^2 = n$ model for ray theory (left) and finite-frequency theory (right) at different depths. Maps have been "wrapped around" to aid visualization of patterns both in the Atlantic and the Pacific Oceans.

the Earth's surface to 30% near the core-mantle boundary.

Finite-frequency analysis enhances the rms velocity anomaly by more than 30% in the lower mantle (Figure 4.13a). The two models are also much less well correlated. The correlation coefficient reduces to about 0.6 near the core-mantle boundary (Figure 4.13b).

Unfortunately, we have no way to decide objectively between the models in Figure 4.8 and 4.11, unless we put an unrealistic trust in our σ_i estimates. Some of the extra details in Figure 4.11 make sense, such as the splitting of the Pacific superplume into separate anomalies (see discussion in next section). On the other hand, William of Occam's dictum that the simplest hypothesis is the preferred one, leads us to prefer the much simpler images in Figure 4.8, which after all have only 20% larger χ^2 than those in Figure 4.11.

We performed one final test to investigate the influence of the model parameterization on the inversion. To verify that our previous observations are independent of the chosen grid, we performed an inversion with a much finer parameterization of 39,048 points (i.e., approximately twice as many). The distribution of the nodes is roughly proportional to the expected resolving length of our data. The distance among the nodes ranges from about 100 km at the surface to about 600 km near the core-mantle boundary. None of the conclusions reached earlier were affected by this test. We do see small differences in the velocity anomalies, mainly at shallow depths where we lack resolution. The original parameterization with $19,279$ nodes was evidently fine enough to capture the smaller structure that is affected by finite-frequency effects.

We personally do not believe that the variance reduction is a particularly useful parameter to evaluate a tomographic inversion. It is as much a measure of the adequacy of the starting model as it is a measure of goodness of fit of the final solution. However, since other global tomographic models are often specified in terms of their variance reduction, we give our values here. Both inversions with the $19,279$-point grid have a variance reduction of about 84% for $\chi^2 = 1.2n$, and of about 87% for $\chi^2 = n$. In both cases, roughly half of the variance reduction is due to the velocity anomalies and half to the hypocenter corrections.

4.7 Discussion

Even though this tomographic study is primarily meant to study the difference between ray-theoretical and finite-frequency inversions, the results obtained are of enough interest that we cannot avoid offering a few speculations. A more rigorous analysis, including resolution and covariance calculations, will be deferred to a future paper in which we combine the long-period data with short-period data from the ISC.

There is an astonishing agreement between our low-velocity anomalies and the locations of well known hotspots visible both in the maps (Figure 4.8) and in the

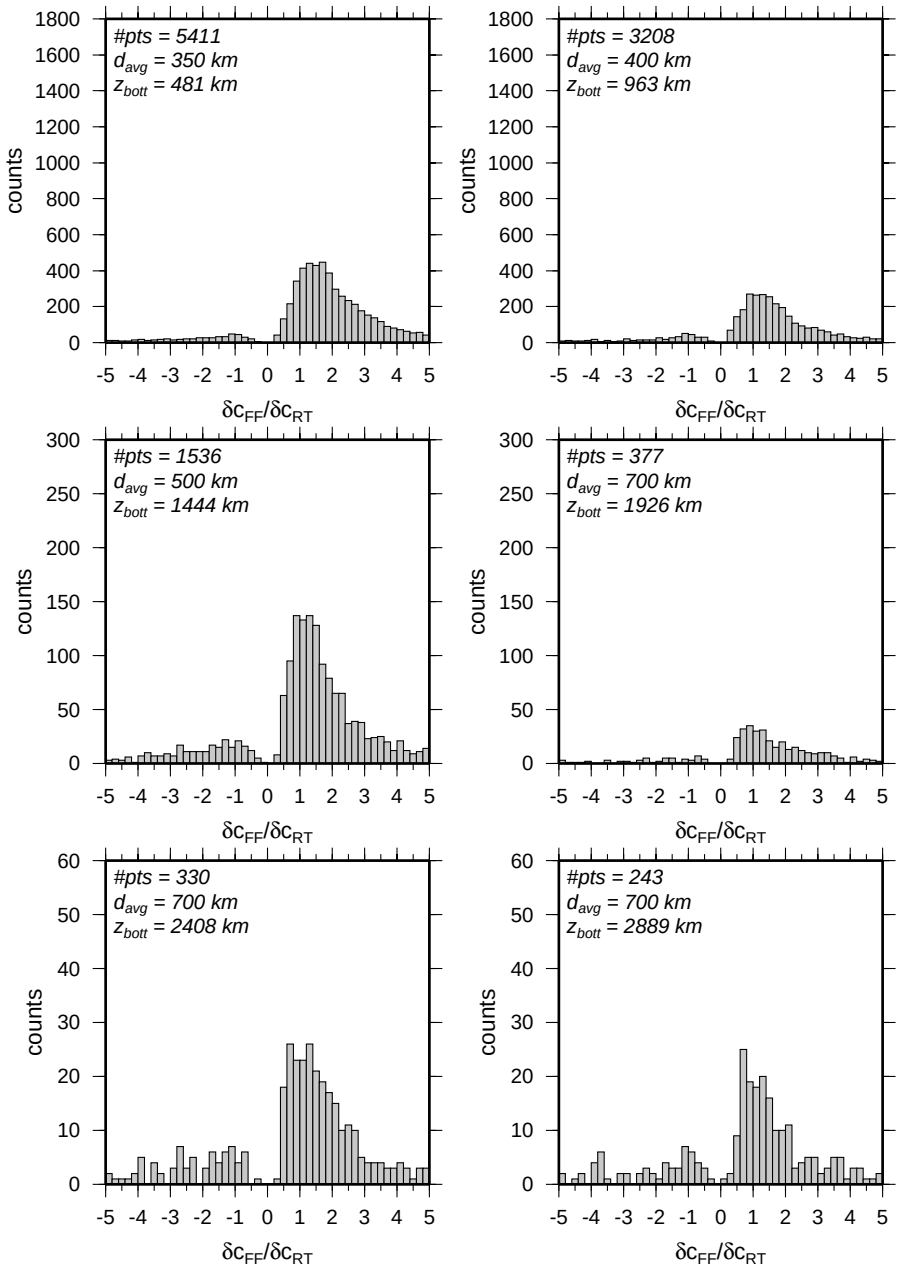


Figure 4.12 Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the models with $\chi^2 = n$. Only changes with absolute value larger than 0.2% are binned. There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers. Information about the mean and median for each shell are contained in Table 4.1b.

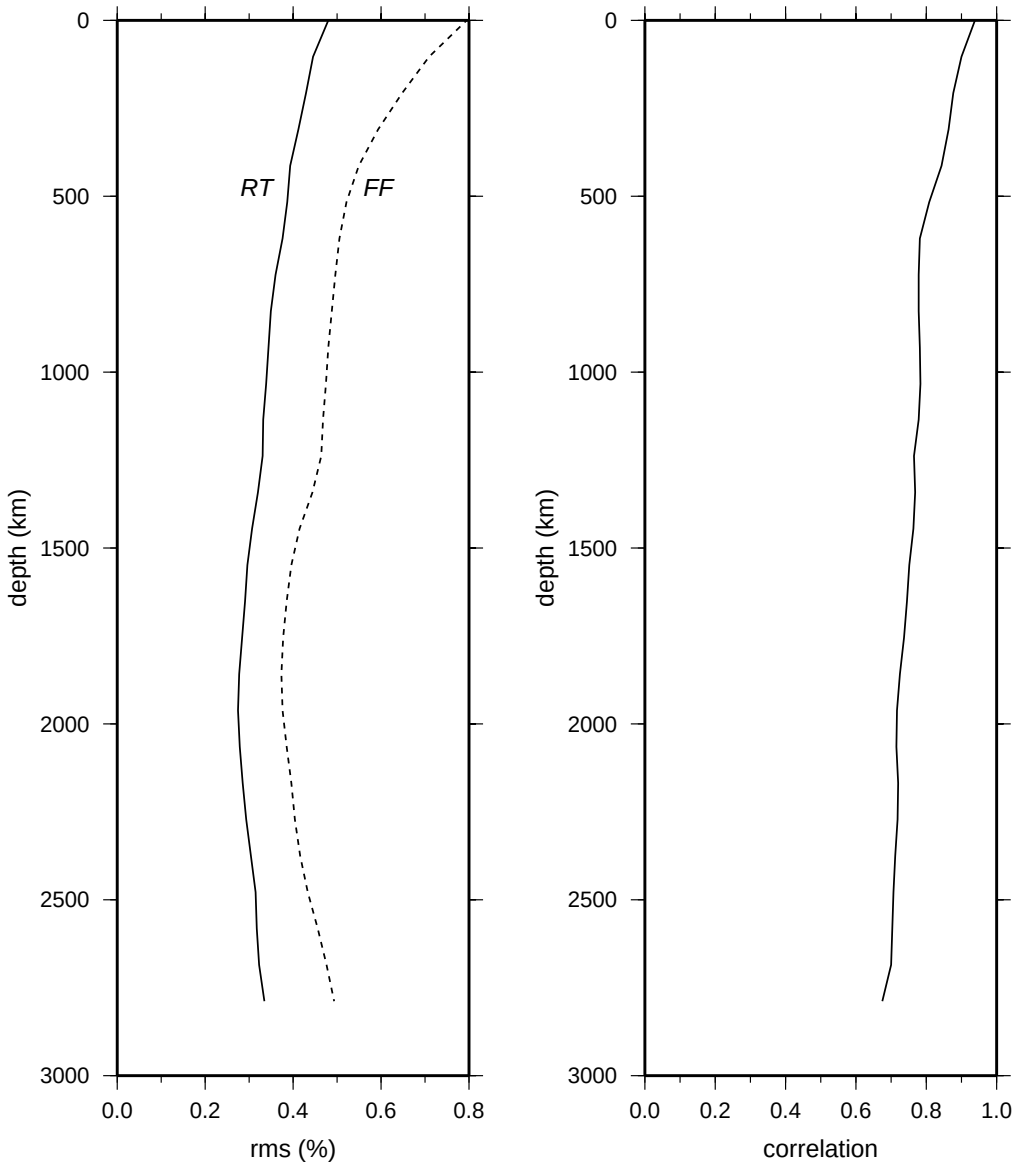


Figure 4.13 (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF–dashed line) and ray theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2 = n$.

Hotspot	Latitude	Longitude	Label
Amsterdam	38.7°S	77.5°E	AM
Ascension	7.9°S	14.3°W	AC
Azores	37.9°N	26°W	AZ
Bouvet	54.4°S	3.4°E	BV
Canary	28.2°N	18°W	CA
Cape Verde	14.9°N	24.3°W	CV
Easter	26.8°S	107.6°W	ES
Kerguelen	49.6°S	69°E	KG
Hawaii	19.1°N	155.1°W	HW
Guadalupe	26.8°N	112.4°W	BC
Iceland	64.4°N	17.3°W	IC
Kilimanjaro	3°S	37.5°E	KL
Madeira	32.7°N	17°W	MA
Reunion	21.2°S	55.7°E	RE
Tahiti	18.1°S	148.3°W	TH
Tasmania	40.8°S	146°E	TA
Tibesti	20.8°N	17.5°E	TI
Yellowstone	44.5°N	110.4°W	YW

Table 4.2 List of major hotspots clearly seen in our tomographic images (locations are from W. Jason Morgan, personal communication 2003)

cross sections (Figure 4.14). A list of major hotspots with their location is given in Table 4.2. We distinctly see Easter Island, Tahiti, Hawaii, Bouvet, Kerguelen, Azores, Canary Island, Cape Verde, Tibesti, Kilimanjaro, Galapagos and Ascension. Only two low-velocity anomalies are not associated with suspected plumes. There is an isolated low-velocity anomaly at (15°N,150°E) at 1200 km depth (not visible in the figures presented here) and another in the northeastern part of Greenland.

Fast anomalies with amplitudes above 1% are observed beneath Tonga-Kermadec, Tasmania, Java and below Asia, the latter presumably identifiable as the Tethys slab (Van der Hilst et al., 1997; Grand et al., 1997; Grand, 1994; Bijwaard et al., 1998; Van der Voo et al., 1999; Gu et al., 2001). Also clear is the familiar signature of the Farallon plate which is migrating eastward with depth (Grand, 1994; Grand et al., 1997; Van der Hilst et al., 1997; Mégnin & Romanowicz, 2000). At 1800 km depth the high-velocity anomalies below North and South America begin to disappear while becoming more pronounced below Central America in the lowermost mantle.

In the lowermost mantle the pattern of heterogeneity is dominated by two large-scale slow velocity anomalies, one in the eastern Atlantic Ocean and one under the South Pacific (Dziewonski et al., 1991, 1993; Grand, 1994; Su et al.,

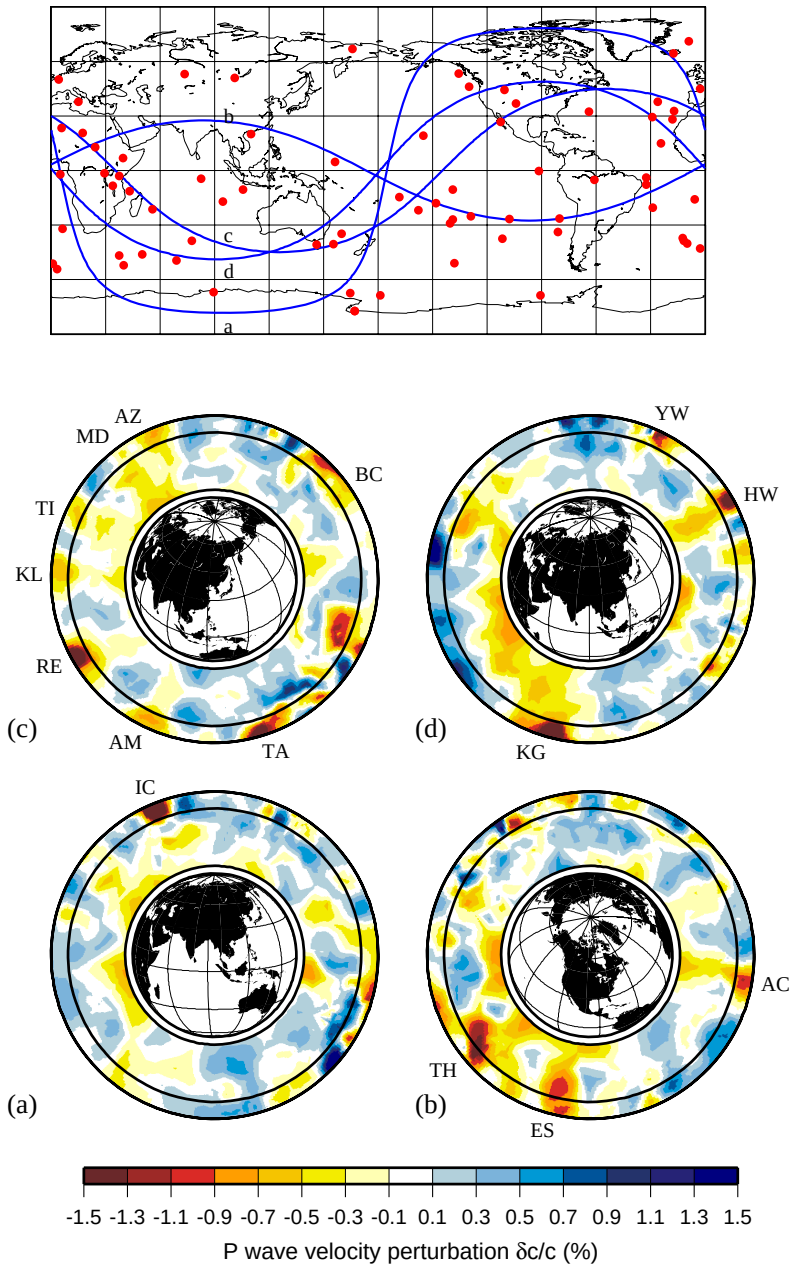


Figure 4.14 Cross sections of the finite-frequency model with $\chi^2 = 1.2n$. The top figure shows the four great circle paths. Letters (a)–(d) on the paths match plots below. (a) cross section across Greenland and Iceland (pole location 94.82°W, 11.57°N), (b) cross section through the Pacific superwell (pole location 96.50°E, 62.44°N), (c) cross section across La Reunion and the African hotspots (pole location 125.39°W, 45.09°N), (d) cross section across the Atlantic superwell and Hawaii (pole location 90.94°W, 41.16°N). Two-letter hotspot identifiers are listed in Table 4.2.

1994; Liu & Dziewonski, 1998; Masters et al., 1996; Ritsema et al., 1999; Mégnin & Romanowicz, 2000; Romanowicz & Gung, 2002), as well as by fast velocity under the circum-Pacific (Dziewonski, 1984; Dziewonski & Woodhouse, 1987; Tanimoto, 1990) common to many other tomographic models.

Both slow regions are loci of major hotspots. The South Pacific "superwell", which is considered to be an exhausted remnant of the Mid-Cretaceous upwelling beneath the Pacific Basin (Larson, 1991; McNutt, 1998), underlies Easter Island, Tahiti and Samoa (Figure 4.14b). The Atlantic "superwell" contains Kerguelen, the African superplume, the African hotspots, Cape Verde, Canary Island, and ends in the Norwegian Sea with a clear signature of the shallow Jan Mayen seamount connected with a deeper anomaly below Greenland (Figures 4.8 and 4.14a,c,d).

Crough & Jurdy (1980) removed subduction-related geoid highs from the observed geoid and found a residual field which has a simple form of two large, elliptical highs surrounded by lows. Broad residual geoid highs are in the central Pacific and the Africa/eastern Atlantic region, in perfect correlation with the regions of highest hotspot concentration. Because hotspots are regions of mantle upwelling, they can contribute significantly to geoid anomalies. The two low-velocity anomalies in our velocity maps seem to be in perfect agreement with the geoid highs of Crough & Jurdy (1980).

Many features are visible both in the ray-theoretical and finite-frequency inversions. However, the continuity of anomalies is generally greater for the finite-frequency images. This is particularly true for Easter Island (not shown) and Hawaii, which clearly shows up as a continuous feature from the surface to the core-mantle boundary (Figure 4.14d). Also, the finite-frequency tomographic images provide compelling evidence that the hotspots named are fed from the lower mantle (Figure 4.14). The Pacific superplume seems to feed the spreading of the South Pacific, whereas the Atlantic megaplume feeds the spreading not only of the North and South Atlantic but also of the Indian Ocean, through a clearly visible conduit "leaning" toward Kerguelen (Figure 4.14d). The interaction of the African superplume with both the Mid-Atlantic Ridge and the Mid-Indian Ridge is present in the shear velocity models obtained previously by Ritsema et al. (1999), Mégnin & Romanowicz (2000) and Romanowicz & Gung (2002).

Major hotspots which do not seem connected to a lower mantle plume include Afar, Ascension, Etna, Galapagos, Iceland, Kilimanjaro, Madeira, Reunion, Tristan. These all seem to originate in the mid mantle. Indications that Iceland is not a deep-seated anomaly were already presented by Ritsema et al. (1999) and a shallow origin was argued from indirect evidence by Foulger & Pearson (2001), Foulger et al. (2001) and Foulger (2003). The result of our inversion confirms these observations and clearly contradict the finding of Bijwaard & Spakman (1999), who proposed a plume extending all the way to the core-mantle boundary.

4.8 Resolution

To determine the reliability of our tomographic images we have performed a limited number of resolution tests. We introduce velocity anomalies having the form of vertical cylinders at location of known hotspots: Iceland, Ascension, Kerguelen, Hawaii, Tahiti and Easter Island. The velocity perturbation decreases with a 3D Gaussian shape from the axes of the hotspot where it is -1.0%. The corresponding velocity perturbation is assigned to each point of the grid lying within the cylinder. Because of the tetrahedral nature of our parameterization, the input velocity model deviates slightly from a smooth cylinder, depending on the distribution of model nodes within and near the synthetic plume.

The diameter of the input cylinders is roughly the same as the quasi-cylindrical anomalies in the smoother tomographic model (Figure 4.8), i.e., the region with the highest amplitude of the perturbation has a diameter of about 500 km. To verify the vertical resolution we performed two kinds of tests. In the first test, the synthetic hotspots reach the core-mantle boundary (Figure 4.15a); in a second test, they stop at about 1400 km depth to simulate plumes originating in the mid-mantle (Figure 4.15b). Synthetic delay times are computed by means of finite-frequency theory, and inverted using both ray theory and finite-frequency theory (Figure 4.15 "exact data"). The same tests are then repeated adding normally distributed random noises to the synthetic residuals (Figure 4.15 "noise added"). The results show that we have adequate resolution in all directions mainly at all synthetic hotspot locations. The fact that we recovered the shape and depth extent of the anomaly, no matter if this was shallow or deep, indicates that we have a reasonable resolution. However, we do not exactly resolve the width of the anomalies. We hope to further improve the resolution with the introduction of the high-frequency data (ISC delays).

Figure 4.16 shows the root-mean-square amplitude of $\delta c/c$ for each synthetic hotspot and the two solutions as a function of depth. The plotted quantity represents an average value of the absolute velocity variations at different depth inside each cylinder. From the noise-free tests, we again see a discrepancy, of about 20%, between finite-frequency and ray theory, as we observed in the tomographic models. Kerguelen and Ascension seem to be particularly poorly resolved. We attribute this poor resolution to a lack of ray path coverage in these particular regions, visible also from Figures 4.5 and 4.6.

Finally, to verify that our velocity anomalies in the lower mantle are not smeared signal from the upper mantle we have performed two more tests. For reasons of space, we do not present the figures here, but only a description of the results. In the penultimate test, the synthetic cylinders stop at 150 km depth with maximum velocity perturbation of -5.0%; in the final test the anomalies stop at the 660 km discontinuity with a maximum velocity perturbation of -1.0%. In both cases we observe very little leakage into the lower mantle. Amplitudes of the recovered anomalies falling below the 660 km discontinuity lie within the interval -0.15% to

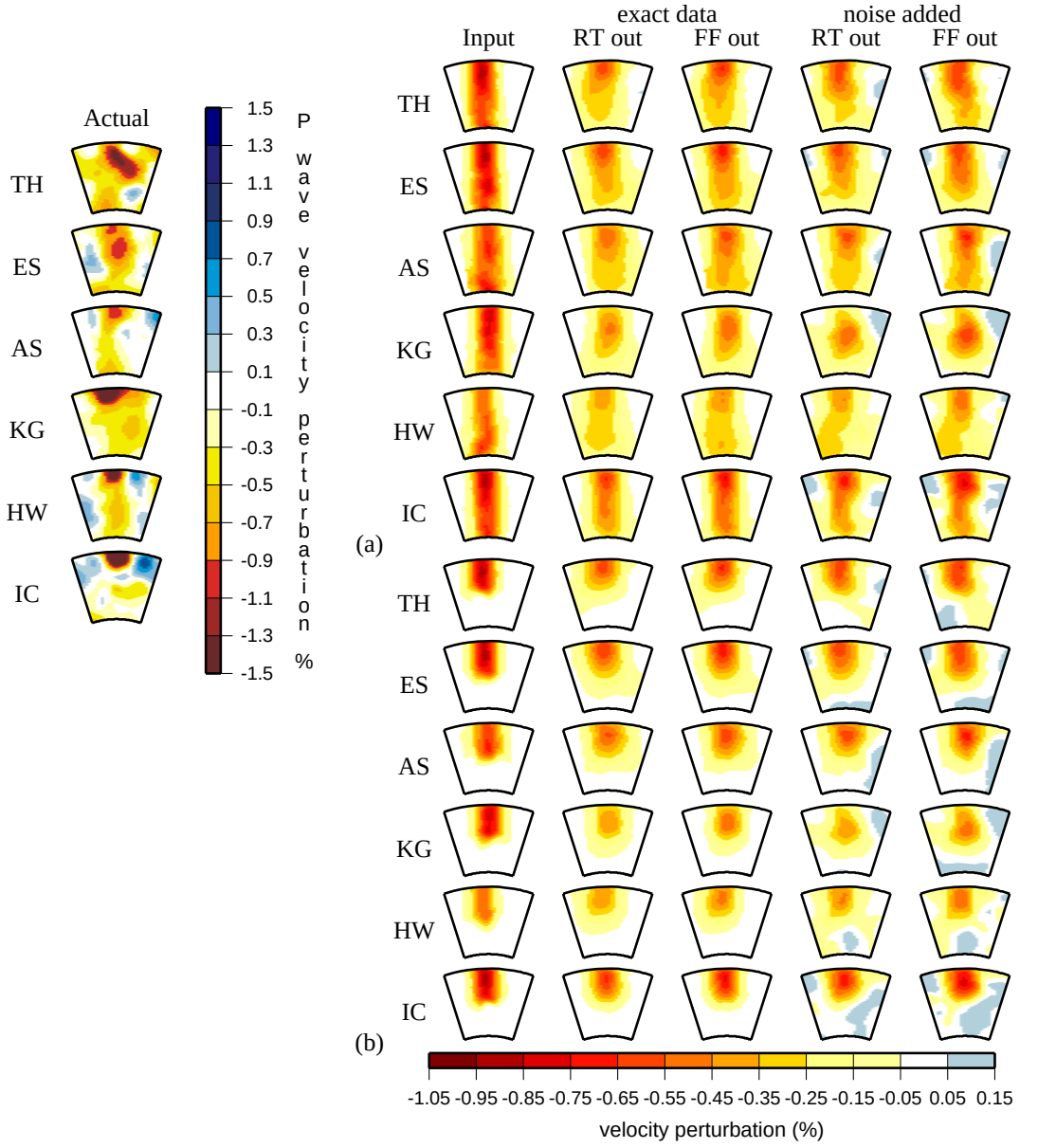


Figure 4.15 Resolution tests for six of the major hotspots observed in the tomographic images. (Left) Recovered velocity model (Actual) for $\chi^2 = 1.2n$; (Right) Resolution tests: from left to right we present the input model (Input), the recovered model obtained by inverting the synthetic delay times δT using the ray-theoretical inverse (RT out) and the finite-frequency inverse (FF out), respectively. The rightmost two columns show the corresponding recovered models in the case we invert the synthetic residuals after the addition of normally distributed random noise. Panel (a) shows the results with the synthetic hotspots reaching the core-mantle boundary; panel (b) shows the results with the hotspots originating in the mid mantle (around 1400 km depth). Two-letter hotspot identifiers are listed in Table 4.2.

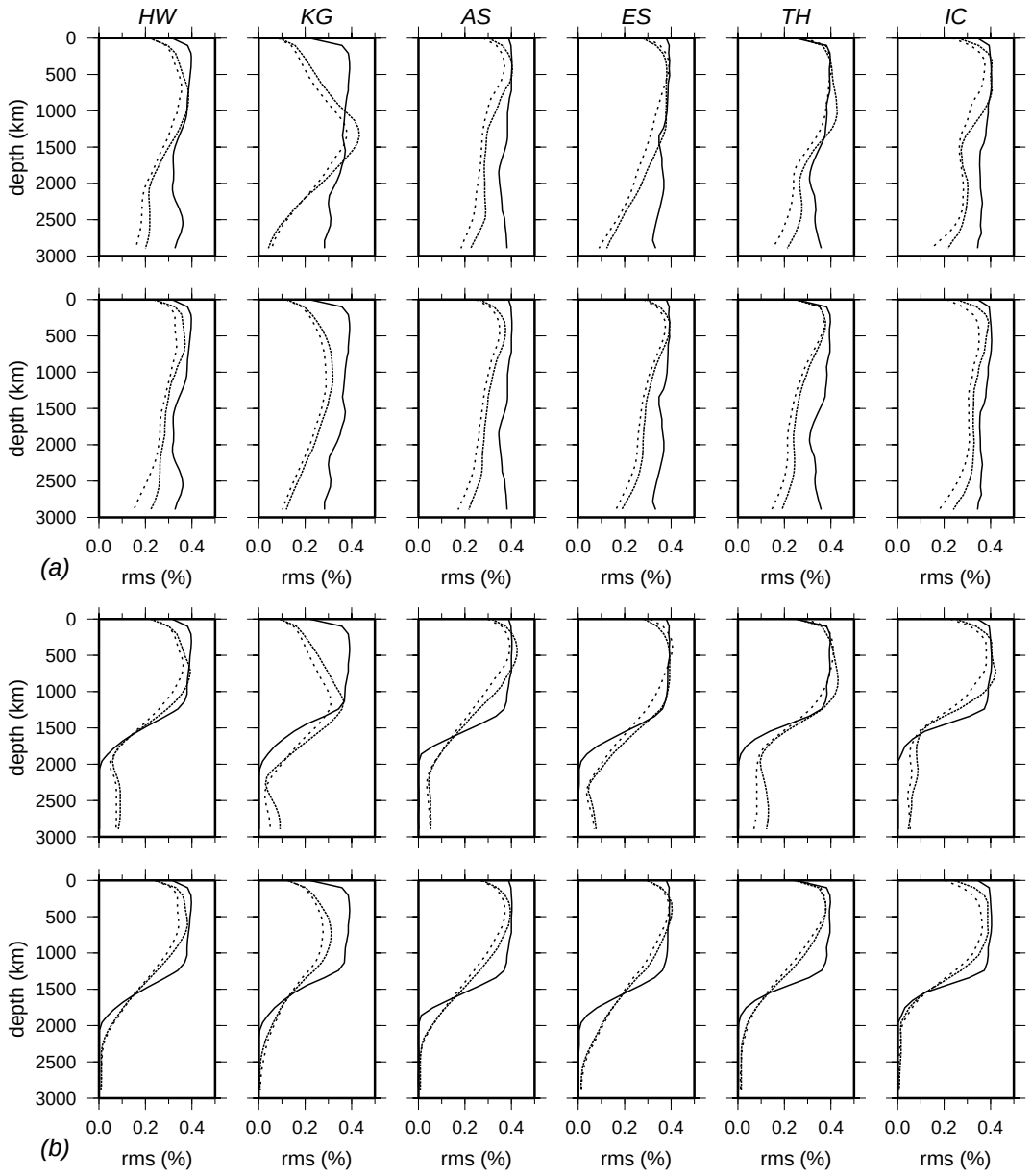


Figure 4.16 Root-mean-square of $\delta c/c$ as a function of depth for each cylindrical anomaly introduced at known hotspot locations. Two-letter hotspot identifiers are listed in Table 4.2. Panel (a) shows the results of the test done with the anomalies reaching the core-mantle boundary, whereas panel (b) shows the results in the case the anomalies stop at the mid-mantle. In each panel we show the results of the inversion of the synthetic residuals without normally distributed error (bottom row) and after adding normally distributed error to the synthetic residuals (top row). Solid line indicates the rms for the input velocity anomalies. Denser and coarser dashed line indicates solution for finite-frequency and ray-theoretical inversion respectively.

-0.05%, approximately ten times smaller than the highest amplitude recovered in the upper mantle. In our $\chi^2 = 1.2n$ tomographic images, we believe and interpret only velocity perturbations with amplitudes larger than $\pm 0.2\%$.

4.9 Conclusion

We have introduced 3D finite-frequency Fréchet sensitivity kernels into global P wave tomography. Effects on the resulting tomographic images have been investigated by comparing finite-frequency tomographic maps with velocity images obtained with the more standard technique of ray theory.

We inverted P and PP-P cross-correlation traveltimes of 20 s dominant period. We have demonstrated that finite-frequency analysis of such long-period waves significantly affects the final images. The amplitudes of the velocity perturbations in our finite-frequency model are 30%–60% higher than those obtained with ray theory, depending upon depth and size of the heterogeneity. This demonstrates a major shortcoming of ray theory. It is not possible to neglect wavefront healing effects, as ray theory does. Finally, even though we defer a more conclusive interpretation of our images until we have a clearer idea of resolution including short-period waves, we present the first clear evidence that at least six hotspots originate in the lower mantle. Iceland seems of shallow depth. For other hotspots, not mentioned here, the jury is still out.

Finite frequency analysis makes it possible to combine data of different frequency. In continuing work we are combining the 20 s compressional wave used in this study with ISC delays obtained at ~ 1 s period. We expect this will further constrain the P wave velocity structure in the Earth.

Chapter 5

Joint seismic tomography of short and long period P wave traveltimes

5.1 Résumé

La modélisation aux fréquences finies nous permet d'utiliser différents types de données dans la même inversion. Pour affiner encore plus notre modèle tomographique, j'ajoute aux données à longue période une sélection de données à courte période de très haute qualité. Ce jeu de données, dérivé de la base de données de l'International Seismological Centre (ISC), ont été re-analysé par Prof Bob Engdahl et classée avec une précision de deux décimales sur les temps. Ce chapitre est principalement dédié à l'analyse de compatibilité entre les données à courte période et les données à longue période utilisées dans le chapitre précédent.

Pour définir leur compatibilité, je compare d'abord les données brutes. Pour cela, je sélectionne les événements qui ont le plus de stations en commun. Pour chaque événement, je compare les deux délais obtenus par soustraction du temps théorique calculé dans notre nouveau modèle de référence et les temps mesurés. La comparaison montre que les deux délais sont généralement très bien corrélès. Les temps de retard pour les données à longue période sont bien approximés par une régression lineaire dans laquelle la pente est indicative de l'importance des effets de diffraction, et en particulier de ce que l'on appelle le "wavefront healing". Ce remplissage dépend de l'échelle des anomalies et aussi de la longueur d'onde. La largeur locale du noyau de Fresnel augmente avec la racine au carré de la distance entre la source et le recepteur. Comme l'onde se propage, une petite anomalie est de plus en plus capable de se cacher dans le trou qui se trouve au milieu du noyau: elle n'affecte alors pas significativement le temps de trajet de l'onde. On dit que le front d'onde est capable de guérir et d'oublier les effets de l'anomalie rencontrée dans son parcours vers la station. À cause de cette diffraction, les délais

des données à longue période augmentent moins vite que ceux des données courte période. Plus significatifs sont les effets de la diffraction, moins grandes sont les délais des arrivées à longue periode par rapport aux délais des ondes à courte période et, en conséquence, plus petite est la pente de régression. Environ 80% des événements analysés sont affectés par ce phénomène de diffraction (pente ≤ 1).

La compatibilité des deux familles des donnée peut être aussi estimée à partir des résultats des inversions des données individuellement. Cette comparaison nous permet de voir que les amplitudes des anomalies dans l'inversion de données à courte periode sont 30–50% plus faibles que dans les images obtenues avec les données à longue période suivant la profondeur que l'on considère. Visuellement, la différence est plutôt dans les anomalies courte longueur d'onde. La corrélation est meilleure pour les anomalies de grande longueur d'onde. En particulier, je montre des exemples des panaches: Azores, Canaries et Cape Verde; Hawaii; Iceland etTahiti. La description en detail du modèle final est remis au prochain chapitre. Cette comparaison permet de voir que les panaches sont déjà visibles dans le modèle obtenu avec les données à courte période, mais avec des amplitudes plus faibles. Différentes sont les hypothèses qu'on considère ici pour expliquer cette différence non négligeable entre les amplitudes des deux tomographies. Il est probable que les données courte période sont aussi affectées par les phénomènes de diffraction. Pour ces données, l'estimation des noyaux de Fresnel serait donc aussi nécessaire.

Le nombre de données à courte période est significativement plus large que le nombre de données à longue période. Si on fait l'inversion sans aucun control, la tomographie va être totalement dominée par les données courte période. Mais le but de ce travail est d'exploiter les informations contenues dans chacun des deux familles de données. Pour cela, j'applique un facteur d'échelle aux données à courte periode de façon que les deux jeux de données aient le même poids dans l'inversion. La description de la méthode utilisée pour l'inversion conjointe des deux jeux de donnée est aussi abordée dans ce chapitre.

5.2 Introduction

Aiming to obtain a new high quality P wave velocity model of the Earth, we have performed a global tomographic study in which we jointly invert short and long period traveltime data. We present here an analysis of their compatibility, and we describe the approach we use to invert them. Measures of the consistency between the two data sets can be obtained by studying the correlation between short- and long- period raw delays, and by comparing the velocity models obtained by inverting each data set separately.

Short period data generally are onset times, measured by an analyst from short-period vertical component instruments and sent to national or international agencies such as the National Earthquake Information Centre (NEIC) or the International Seismological Centre (ISC) that distribute them. Teleseismic P arrivals are

assumed to have been read from instrument with a dominant period around 1 s. They represent the arrival of the highest-frequency observable wave, for which ray theory is valid. Because of the mathematical simplicity of ray-theory, they have been extensively used in global seismic tomography.

Broadband digital data are very high-quality measurements whose number increases steadily. Arrival times are often measured by cross-correlation of an observed body-wave phase with the corresponding spherical-earth synthetic phase (Bolton & Masters, 2001). Cross-correlation methods have also been used to measure the differential traveltime of two phases at the same station (Kuo et al., 1987; Woodward & Masters, 1991; Su et al., 1994) or of the same phase in a network (VanDecar & Crosson, 1990). Both absolute and differential travel times have been used in tomographic studies, most recently by Fukao et al. (2003) and Vasco et al. (2003). Long period data are low pass filtered, which facilitates the identification of the different phases.

The measured traveltimes of finite-frequency seismic waves are usually modeled with ray theory. Ray theory assumes that the travel time of a P wave is only influenced by the Earth's properties along an infinitesimally narrow path that follows Snell's law. However, as a result of diffraction effects, the traveltimes of finite-frequency waves are sensitive to velocity structure far away from the geometrical ray, within a volume known as the Fresnel zone (Nolet, 1987, 1992; Woodward, 1992; Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000). Marquering et al. (1999), Dahlen et al. (2000) and Hung et al. (2000) show how the 3D Fréchet kernel expressing this sensitivity is identically zero along the unperturbed ray, and is confined to a banana-shaped volume surrounding the geometrical ray. Objects much smaller than the width of the Fresnel zone will not significantly influence the travel time of the wave (Nolet & Dahlen, 2000); diffraction acts to heal the irregularities and the finite-frequency wavefront continues to propagate unperturbed. Only an infinite-frequency wave always remembers the shift accrued upon passage through an anomaly somewhere along its ray path all the way to the receiver. We account for the diffraction healing of the broad-band body-waves by computing 3D traveltime sensitivity kernels using the formalism of Dahlen et al. (2000). These kernels are based on the paraxial approximation to ray theory and the Born approximation.

Because of their different sensitivity, the two data sets implemented in our tomographic study provide complementary information about the velocity structure.

5.3 Data

For the long period data, we use arrival times of P and PP waves with 20 s dominant period (Bolton & Masters, 2001). Although part of this data set has already been used in an earlier inversion (Bolton, 1996), the original seismograms were re-analyzed and a recently discovered timing error of the IRIS/IDA Global Seis-

mographic Network (GSN) stations equipped with the IDE MK7 data logger was corrected. We invert for 66,210 P absolute traveltimes, as well as 20,147 PP–P and 2,382 pP–P differential traveltimes. Absolute arrival times are measured by cross-correlation of an observed pulse with a synthetic which is constructed by convolving the instrument response of ANMO GSN station, located at Albuquerque (New Mexico), with a t^* attenuation operator. The attenuation time t^* is kept constant at 1 s. This assumption can lead to small but systematic depth- and distance-dependent trends in the data set. To minimize such errors, the fitting procedure concentrates on matching the first swing of the waveforms, which is less affected by interference with depth phases and by effects due to the inaccurate choice of the t^* operator (Bolton & Masters, 2001). Differential PP–P and pP–P times are also obtained by cross-correlation (Woodward & Masters, 1991). Here, the t^* operator accounts for the different attenuation histories of the direct and reflected phase, respectively. Differential travel times are less affected by errors in the source location, and have reduced sensitivity to structure in the vicinity of source and receiver (at least in the ray approximation). Also, they provide useful information about the shallow structure located near the bounce points, thus allowing us to constrain upper mantle regions of the world where there are no sources or receivers.

For the short period data, we use 1,427,114 P and 68,911 pP arrival times extracted from the International Seismological Centre (ISC) and U.S. Geological Survey's National Earthquake Information Center (NEIC) bulletins and re-interpreted by Engdahl et al. (1998), extended to earthquakes measured until the year 2000. We only selected the times listed with two decimal precision and labeled as the highest quality data.

Predicted absolute and differential times are computed using the *iasp91* velocity model (Kennett & Engdahl, 1991). We correct for the delay due to the ellipticity and for the effect of the crust, including topography. Crustal corrections are computed by using the global crustal model CRUST2.0 (model available on the web: *http://mahi.ucsd.edu/Gabi/rem.html*). The delays δT_{LP} of the long-period data, i.e. the difference between the cross-correlated arrival time T_{LP}^{obs} and the theoretical travel time T^{syn} computed in the *iasp91* model, show a baseline shift of about -1 s for PP–P and about +4 s for P. The origin of this shift is not fully understood; it could be due to the imperfection of the 1D reference model, or to systematic error in the source locations and/or origin time. The assumption of a constant t^* operator does not explain this time shift either. Variations in the delays due to attenuation are of the same order as the variations in the t^*, which are of a tenth of a second (see Stewart (1984) and equation (6) in Bolton & Masters (2001)) and therefore far too small to explain the 4 s shift.

A purely technical remedy has been found to remove this offset, a detailed description of which can be found in Montelli et al. (2003) (also chapter 4). Since differential times such as PP–P are quite insensitive to source mislocations, we

assume the -1 s offset to be mainly produced by the inadequacy of the 1D reference velocity model used. We eliminate the offset by making a slight change to the *iasp91* velocity model in the transition zone. This led to an increase of the offset in the P data set to about 5 s, which we remove by subtracting a constant correction to all P traveltimes such that the average is zero. We use this slightly perturbed *iasp91* model, that we indicate as *iasp91mod*, as reference model in our tomographic study to compute the theoretical times for both long and short period. As a result, an offset (of 1 s) is also produced in the short-period delays δT_{SP} (difference between the picked T_{SP}^{obs} and the theoretical travel time T^{syn} computed in the *iasp91mod* model), that again we correct by applying a constant correction to all data such that they average to zero. At first sight, this may seem to introduce a discrepancy in our data set. However, in the next section we argue that we need to correct the origin time for short- and long- period data separately. The effect of the overall time shifts we just discussed is to bring the average of origin time corrections close to 0; it does not introduce discrepancies that cannot be handled by the inversion.

5.4 Technical aspects of the inversion

We simultaneously invert for velocity perturbations and hypocentral parameters (origin time, longitude, latitude and depth). Because the onset of the short period P waves originates from the nucleation point on the fault surface, whereas long period waves average over the rupture process, hypocenters and origin times are treated separately, for the two types of data. We have 5,938 events that produced the long-period data, and the 86,499 that produced the short-period arrival times.

We sample the velocity structure by using a heterogeneous distribution of points to form a Delaunay mesh (Watson, 1981, 1992; Sambridge et al., 1995). The node spacing increases with the expected resolving length of our data and ranges from about 200 km in the upper mantle to about 600 km in the lower mantle.

The system $\mathsf{Ax} = \mathsf{b}$ to invert can be represented as:

$$\begin{pmatrix} \mathsf{A}_{LP} & \mathsf{H}_{LP} & 0 \\ \mathsf{A}_{SP} & 0 & \mathsf{H}_{SP} \end{pmatrix} \begin{pmatrix} \mathsf{x}_c \\ \mathsf{x}_{h_{LP}} \\ \mathsf{x}_{h_{SP}} \end{pmatrix} = \begin{pmatrix} \delta\mathsf{T}_{LP} \\ \delta\mathsf{T}_{SP} \end{pmatrix} \tag{5.1}$$

where the matrix A_{LP} for the long period data is built by using the finite-frequency modeling of Dahlen et al. (2000); the matrix A_{SP} for the short period data is constructed by using standard ray theory; the matrix H_{LP} and the matrix H_{SP} of the partial derivatives with respect to the hypocentral coordinates of the long and short-period data, respectively, are computed with ray-theory (equations 2.11 in Chapter 2). The vector x contains the unknown velocity perturbations x_c and source correction $\mathsf{x}_{h_{LP}}$ and $\mathsf{x}_{h_{SP}}$ for the long- and short period, respectively. b is the vector of the delay times $\delta T = [\delta T_{LP}; \delta T_{SP}]$.

For a detailed description of the mathematical formulation for the construction of the matrices A we refer the reader to Montelli et al. (2003) (also Chapter 4). The difference between ray-theoretical and finite-frequency approaches resides in the elements of the matrix A. In the finite-frequency modeling, each row of the matrix A_{LP} represent one Fréchet kernel connecting the source to the receiver. The elements A_{ij} of the matrix can be seen as the integrated effect of the kernel for datum i contained in all tetrahedrons having node j as common vertex. In the ray-theoretical formulation each row of the matrix A_{SP} represents the geometrical ray connecting the source to the receiver. The elements A_{ij} are interpolation weights integrated along the arc-length of the ray i contained in all tetrahedrons having node j as common vertex. Figure 5.1 shows the column density for the long -, short - period matrices and the full matrix, respectively. We define the density for a given node of the grid as the L_1 norm of the column vector corresponding to that node:

$$d_j^{LP} = \sum_{i=1}^{n_{LP}} \mid A_{ij}^{LP} \mid \tag{5.2}$$

$$d_j^{SP} = \sum_{i=1}^{n_{SP}} \mid A_{ij}^{SP} \mid \tag{5.3}$$

Although the number of long period data is about 17 times smaller than the number of short period data, long-period data provide similar ray path coverage. The maximum amplitude of the matrix density for the long period data are about a factor of 100 smaller than for the short period data. Notable is the difference between the ray path coverage in the northern and southern hemisphere. The last column in the figure shows the density of the joint inversion matrix, which has been weighted to avoid that the larger number of short-period data will dominate the inversion, as described below.

Standard deviations are assigned to the long and short period data to account for the measurement errors. Long period times are divided in three accuracy classes to account for the confidence of the pick. To each class corresponds an error σ. A priori estimates of the σ values were slightly raised from purely subjective estimates (Bolton & Masters, 2001), on the basis of the misfits we obtained in preliminary inversions. Likewise, we adopt standard deviations for the short period data, which are close to the estimates of Morelli & Dziewonski (1987). Values are summarized in Table 5.1. Outliers with a posteriori misfits larger than 3σ are rejected after a first iteration.

To avoid that the image is completely controlled by the much larger short-period data set, we apply a weight ω to the least square system for the simultaneous fit of the two data set. We denote by n_{SP} the number of short-period data and n_{LP} the number of long-period data and by define the reduced chi-square χ^2_{red} as chi-square

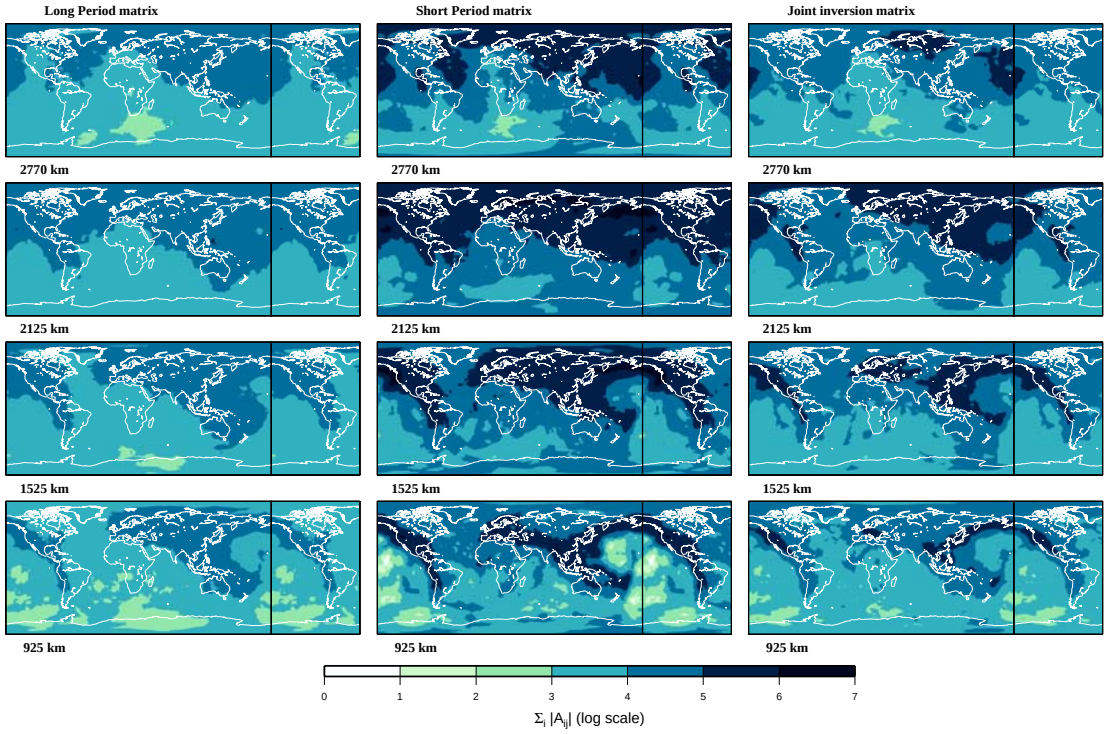


Figure 5.1 Section at different depth of the density of the matrix A for the long period data (left), short period data (center) and joint inversion (right). Note that the maps have been "wrapped around" to facilitate the visualization of patterns both in the Atlantic and the Pacific Ocean.

LP data:

Phase	N. of obs	σ_A	σ_B	σ_C
P	66,210	0.44	0.53	0.79
PP-P	20,147	0.75	0.95	1.15
pP-P	2,382	0.75	0.95	1.15

SP data

Phase	N. of obs	σ
P	1,427,114	0.88
pP	68,911	1.14

Table 5.1 Summary of the data sets used in our tomographic study. Number of measurements and corresponding a priori errors for each of the phases used are provided for the long period - (left table), and short-period (right table) data set, respectively.

divided by the number of data; the system to minimize can be expressed as:

$$\chi^2_{red_{SP}} + \chi^2_{red_{LP}} = \frac{1}{n_{SP}}\chi^2_{SP} + \frac{1}{n_{LP}}\chi^2_{LP} \tag{5.4}$$

the minimum can be found by computing the least square solution to :

$$\begin{pmatrix} \mathsf{A_{LP}} \\ \omega \mathsf{A}_{SP} \end{pmatrix} \mathsf{x}_c = \begin{pmatrix} \delta \mathsf{T}_{LP} \\ \omega \delta \mathsf{T}_{SP} \end{pmatrix} \tag{5.5}$$

with ω given by:

$$\omega^2 = \frac{n_{LP}}{n_{SP}} \tag{5.6}$$

From the number of data used, it follows that $\omega = 0.24$. A different approach, in which the maximum column sum of the matrices $\omega \mathsf{A}_{SP}$ and A_{LP} is equalized, leads to the same scaling factor $\omega = 0.24$.

The influence of the data errors in the tomographic images is kept under control by applying a set of regularizations. We apply norm damping to the velocity perturbation x_c, and to the source corrections $\mathsf{x}_{h_{LP}}$ and $\mathsf{x}_{h_{SP}}$, respectively, which limits the rms (root mean square) of the origin time correction to 0.7 s for the long period and to 0.4 s for the short period data, and the rms of the changes in hypocentral coordinates to about 10 km for the long period and 5 km for the short period in every direction (latitude, longitude and depth). To supplement the norm damping, we also apply an approximate second derivative damping, $\|\epsilon_S \mathsf{S} \mathsf{x}_c\| \to 0$, controlled by the parameter ϵ_S; this biases the solution toward smooth velocity variations in every direction (latitudinal, longitudinal and radial) (Nolet, 1987). Our smoothing operator is not truly a canonical second derivative ∇^2 since it averages on the total number of node's neighbors in an irregular grid. For node i of the grid, which has N near neighbors, the smoothing operator is given by:

$$\epsilon_{\mathsf{S}_i}(\mathsf{x}_{c_i} - \frac{1}{N}\sum_{k=1}^{N} \mathsf{x}_{c_k}) = 0 \tag{5.7}$$

The smoothing parameter ϵ_{S_i} varies for each node of the grid to account for the uneven illumination of the earth by seismic rays. An estimate of the illumination of a node of the grid is given by the column norm (equation 5.2 or 5.3). ϵ_{S_i} varies linearly with the node density between a specified minimum and maximum value. If the column norm is small then the node is poorly illuminated and we obtain a strong smoothing; on the contrary, if the column norm is large, then the node is very well resolved and the smoothing is weak.

The resulting least-square system minimized in the inversion can finally be expressed as:

$$\chi^2_{red_{LP}} + \chi^2_{red_{SP}} + \epsilon_c \|\mathsf{x}_c\|^2 + \epsilon_{h_{LP}} \|\mathsf{x}_{h_{LP}}\|^2 + \epsilon_{h_{SP}} \|\mathsf{x}_{h_{SP}}\|^2 + \|\epsilon_\mathsf{S} \mathsf{S} \mathsf{x}_c\|^2 = \text{minimum}, \tag{5.8}$$

	Long Period data:	Short Period data
(1)	observed travel times	observed travel time
(2)	subtract *iasp91* time	
(3)	modify *iasp91* to center PP–P	
(4)	subtract *iasp91mod* time from (1)	subtract *iasp91mod* time from (1)
(5)	subtract 5 s from all P delay times	subtract 1 s from all P delay times
(6)	invert for $(\delta c/c)_{LP}$ and x_h^{LP}	invert for $(\delta c/c)_{SP}$ and x_h^{SP}
(7)	correct for x_h^{LP}	correct for x_h^{SP}

(8) compare the resulting delays δT_{LP} with the corresponding δT_{SP}
(9) compare the two velocity models obtained by inverting δT_{LP} and δT_{SP}, separately.

Table 5.2 Summary of the procedure followed to analyze the compatibility of the two different data sets.

where ϵ_c is the damping factor for the velocity model; $\epsilon_{h_{LP}}$ and $\epsilon_{h_{SP}}$ the damping parameter for the long- and short-period hypocenters respectively, and ϵ_{S} and S are the diagonal matrix of the smoothing factors and the matrix representing the operator that applies the smoothing.

5.5 Compatibility of the two data sets

To investigate the compatibility of the two data sets we analyze the raw delays $\delta\mathsf{T}_{LP}$ and $\delta\mathsf{T}_{SP}$ and we compare velocity models obtained by independently inverting the two data sets. By 'raw' data, we denote the traveltime anomalies that remain after offset corrections. For origin time and hypocenter we adopt the origin time and hypocentral parameter for the relocated event. The anomaly is w.r.t. model *iasp91mod*. In Table 5.2, we summarize the steps that define our raw delay times.

5.5.1 Analysis of the raw data

We take a closer look at the 58 events that have a large (≥ 5) number of stations in common among the two classes of data. We compare the short period delays δT_{SP} with the long period delay times δT_{LP} for each event. Scatterplots show that the two delay times are generally well correlated. Figure 5.2 shows the scatterplots for events with more than 10 data. The long-period delay times can be approximated by linear regression:

$$\delta T_{LP} = \alpha \delta T_{SP} + \beta \tag{5.9}$$

The slope α is a measure of the importance of diffraction wavefront healing and is controlled by the scale-length of the heterogeneities and the ray-length. The local width of the kernel increases with the square root of the source-receiver distance, so that as the wave propagates more, a small anomaly is increasingly able to hide

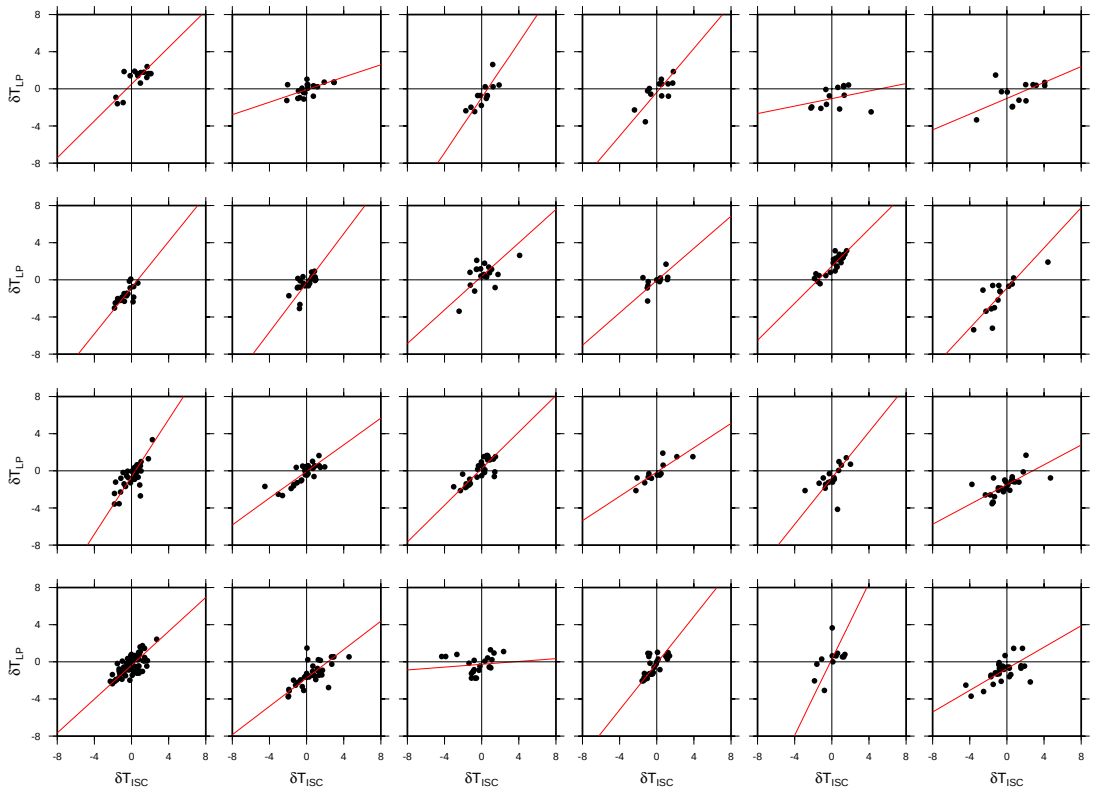


Figure 5.2 Scatterplot comparison of the short-period delays (horizontal axis) with the corresponding long-period residuals (vertical axis) for 24 events. A least-squares criterion has been used to find the best-fitting line $\delta T_{\mathsf{LP}} = \alpha \delta T_{\mathsf{ISC}} + \beta$ (red line).

within the hole of the kernel and does not significantly affect the traveltime of the wave. The delay time is also expected to increase (in a rms sense) as the square root of the ray length, like in the case of the random walk. Because of the healing, the delay δT_{LP} is still growing as the wave continue to propagate, but less than the δT_{SP} for the correspondent ray which is only affected by the rms growth. The larger are the effects of the wavefront healing, the smaller is the long-period residual with respect to the short-period, and as result, the smaller is the slope α. About 80% of the events analyzed are affected by wavefront healing ($\alpha \leq 1$). The other 20% are larger than 1, but are also characterized by larger error (figure 5.3c). The plots show also a shift β in the long period time that is different for each event.

The origin of this shift, and in particular its variability, is not yet fully understood. Scatterplots of origin time shifts as a function of moment magnitude (figure 5.3a), and as a function of depth (figure 5.3b) do not show any characteristic pattern that could explain this shift. The solution might be found by analyzing the waveforms. A collaboration with Guy Masters is planned to study possible causes

of this shift in the measurement procedure. For the present study, we note that the shift is accommodated by the separate origin time corrections $\delta\mathsf{T}_O^{LP}$ and $\delta\mathsf{T}_O^{SP}$ such as not to influence the inversion result.

5.5.2 Comparison of the velocity models

A comparison of velocity models obtained by inverting the two data sets separately also provides a measure of their consistency even though it is clear we must allow for differences in resolving power, and we should never expect complete agreement even if the data are 100% compatible. Figures 5.4 to 5.7 show P wave velocity perturbations with respect to our background model for the long-period data only (a), short period delays only (b) and the joint inversion (c), as a function of depth. All models have a reduced χ^2 close to 1. We will focus our attention here on the two separated inversions, postponing the description of the joint inversion model to the next section.

The only striking difference between the long- and short-period solutions is in the amplitudes of the velocity anomalies. Figure 5.8 (a) shows the model root-mean-square for the three inversions; long-dashed line corresponds to the short-period model, short-dashed line to the long-period only model, and solid line to the final tomographic model, obtained by jointly inverting the two data sets. The norm of the short period model is 32% smaller than the norm of the long period model. Amplitudes of some anomalies in the long-period model are as much as 50% larger (Figure 5.8-a) than in the short-period model. This value changes as a function of depth, and it is even larger in the upper mantle (Montelli et al., 2003). The two models correlate rather badly (long dashed line in Figure 5.8) when a strict numerical measure is used. The correlation coefficient is around 0.5.

Visually, the mismatches are dominated by a lack of correspondence between the two models of several fast velocity anomalies. Two major fast anomalies correlate: the Farallon slab signature and the high velocity anomaly beneath Asia, presumably an imprint of the Tethys slab. But, the fast anomaly located north-east of Hawaii in the long-period model, visible down to 2125 km depth, corresponds to a slow velocity anomaly in the short period model.

Overall, low velocity anomalies tend to correlate better. However, two low velocity anomalies mapped in the long period model are fast anomalies in the short-period images: in the lowermost mantle the super-plume like feature beneath Europe; and in the Indian Ocean the plume-like features traceable all the way down to the core-mantle boundary. In the lowermost mantle, both models are dominated by the two low velocity anomalies common to many other tomographic study and known as "superplumes" (Dziewonski et al., 1991, 1993; Grand, 1994; Su et al., 1994; Liu & Dziewonski, 1998; Masters et al., 1996; Ritsema et al., 1999; Mégnin & Romanowicz, 2000; Romanowicz & Gung, 2002): one in the eastern Atlantic Ocean and one under the South Pacific. Both models show a good agreement between the low-velocity anomalies and the location of known hotspots; many of

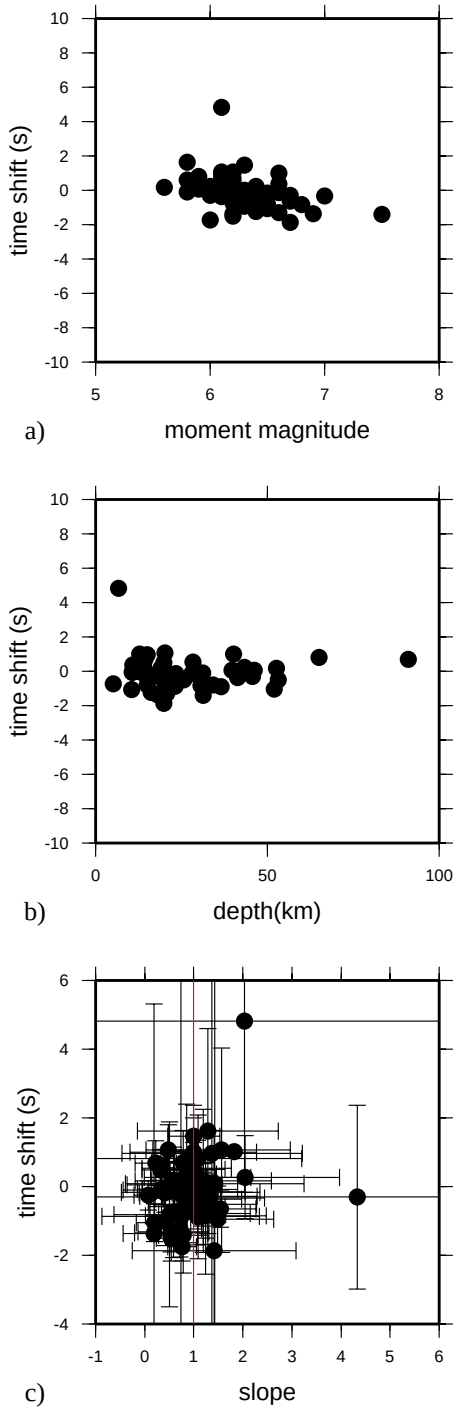
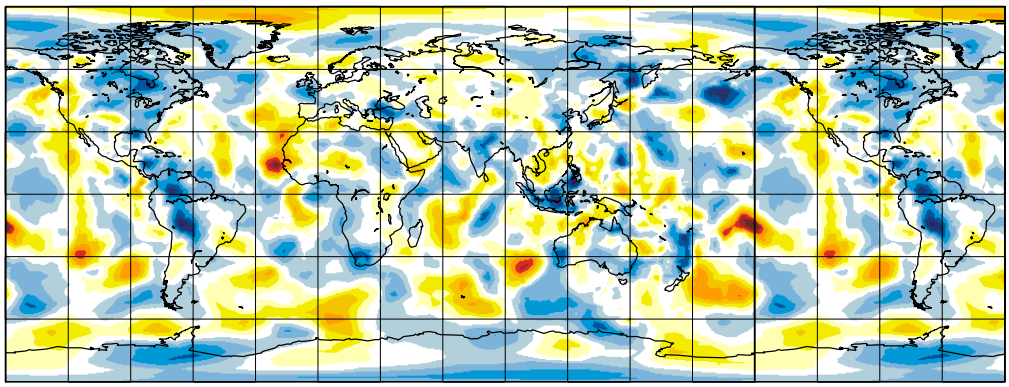
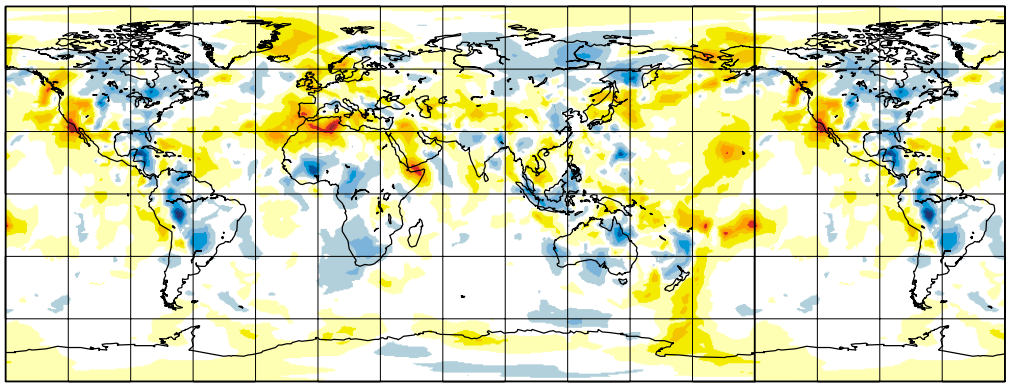


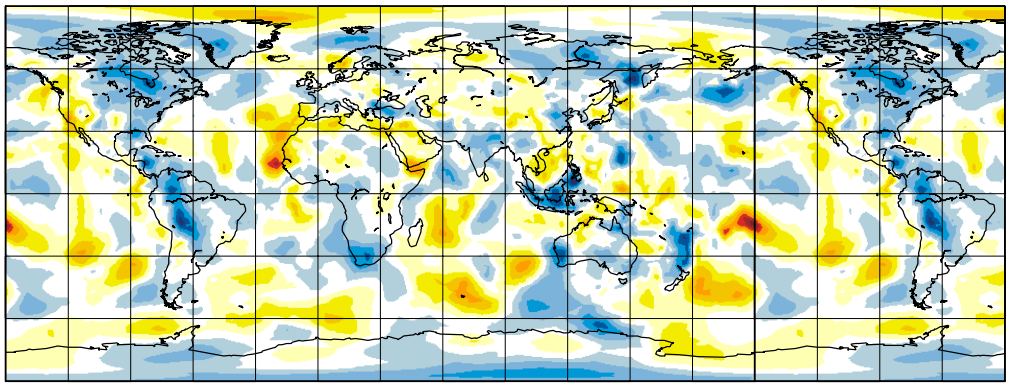
Figure 5.3 Scatterplot comparison of the LP–SP time shifts versus moment magnitude (a), hypocenter depth (b) and slopes (c). Red line in plot (c) indicate slope 1. No pattern is recognizable in any of these plots.



(a) Long Period Only (925 km; c = 11.34 km/sec)



(b) Short Period Only



(c) Joint Inversion

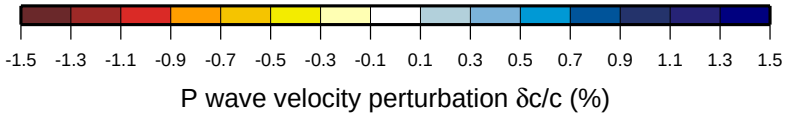
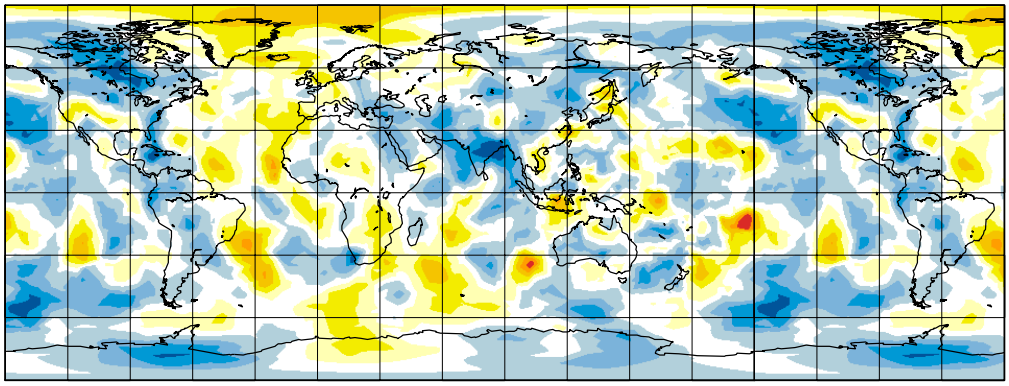
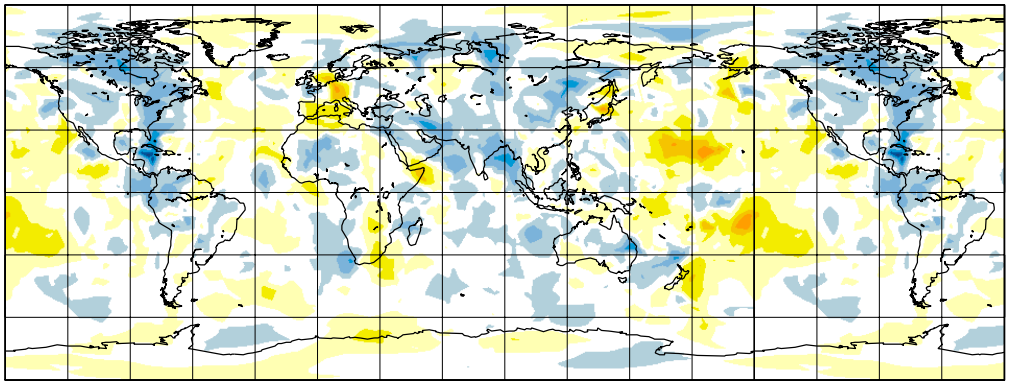


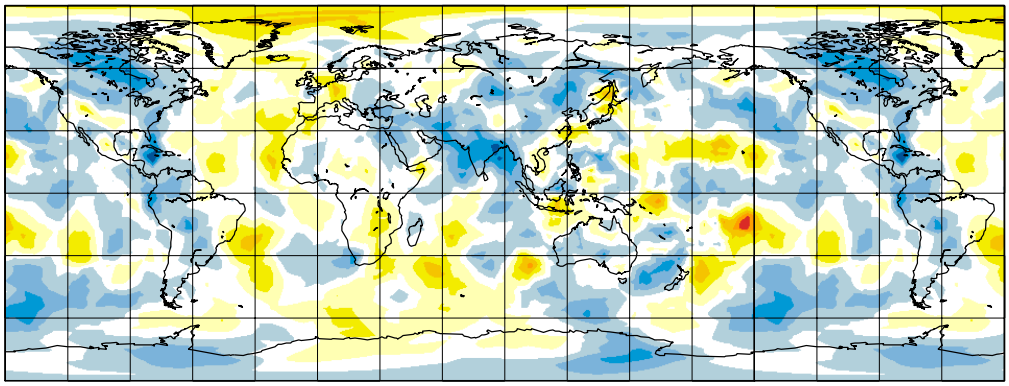
Figure 5.4 Comparison of the a) long period, b) short period and c) joint inversion model at 925 km depth. Note that the maps have been "wrapped around" to facilitate the visualization of patterns both in the Atlantic and the Pacific Ocean.



(a) Long Period Only (1525 km; c = 12.23 km/sec)



(b) Short Period Only



(c) Joint Inversion

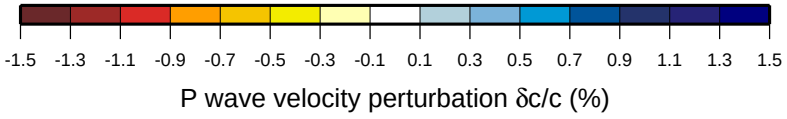
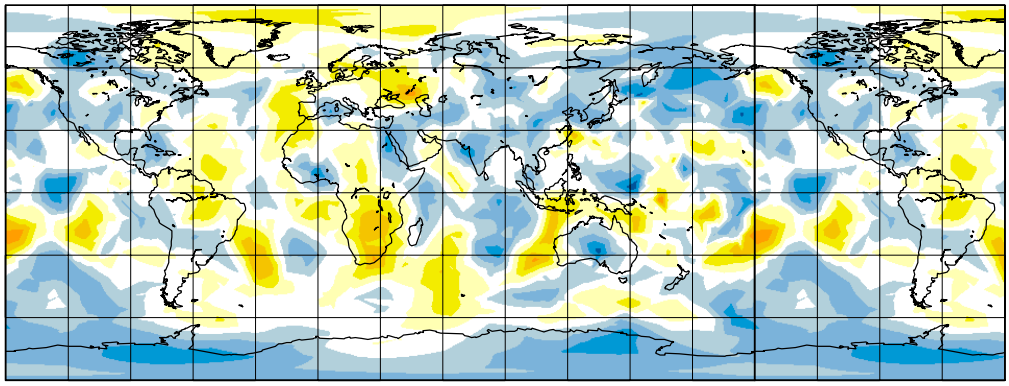
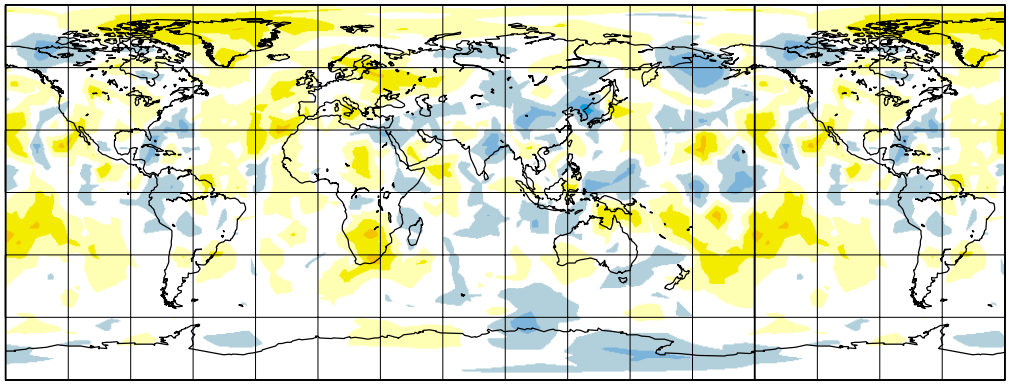


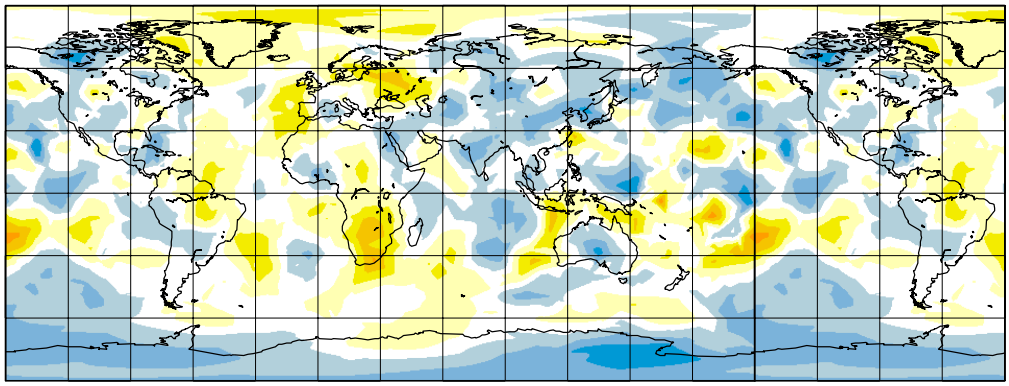
Figure 5.5 Comparison of the a) long period, b) short period and c) joint inversion model at 1525 km depth.



(a) Long Period Only (2125 km; c = 12.96 km/sec)



(b) Short Period Only



(c) Joint Inversion

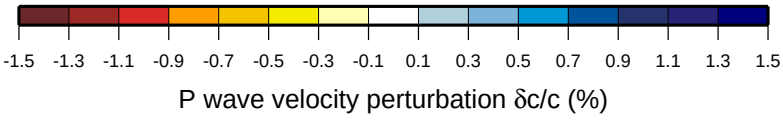
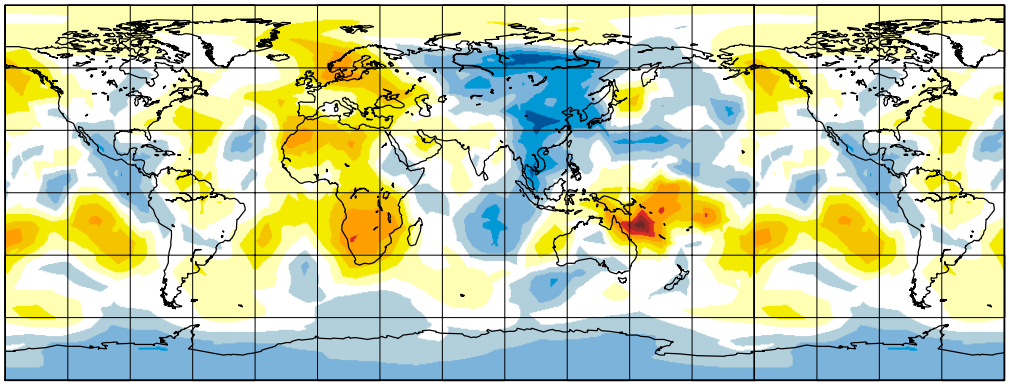
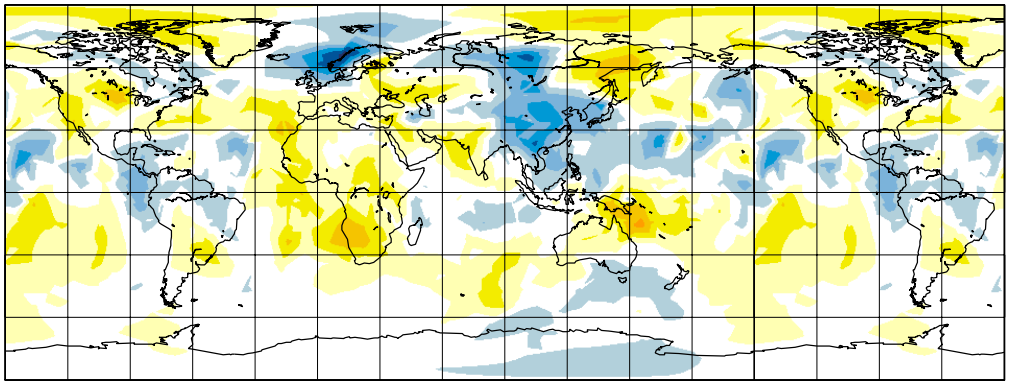


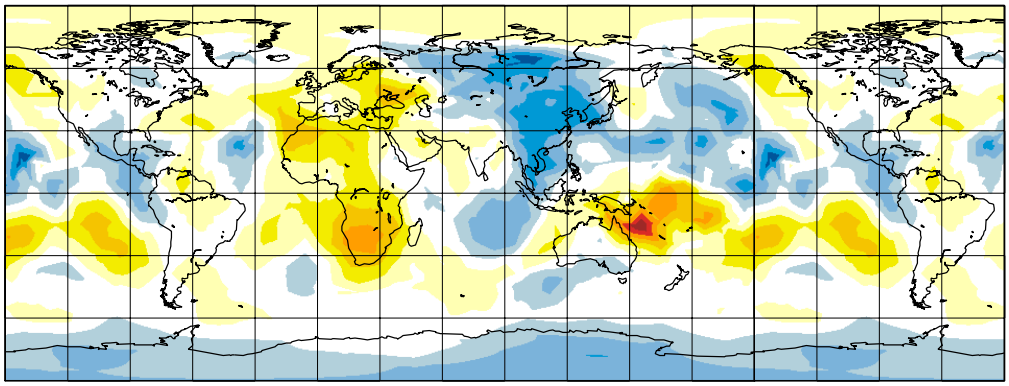
Figure 5.6 Comparison of the a) long period, b) short period and c) joint inversion model at 2125 km depth.



(a) Long Period Only (2770 km; c = 13.69 km/sec)



(b) Short Period Only



(c) Joint Inversion

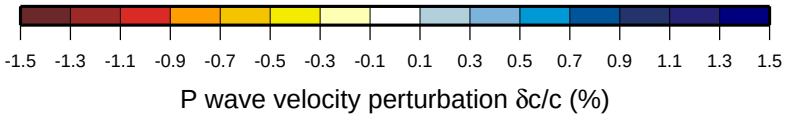


Figure 5.7 Comparison of the a) long period, b) short period and c) joint inversion model at 2770 km depth.

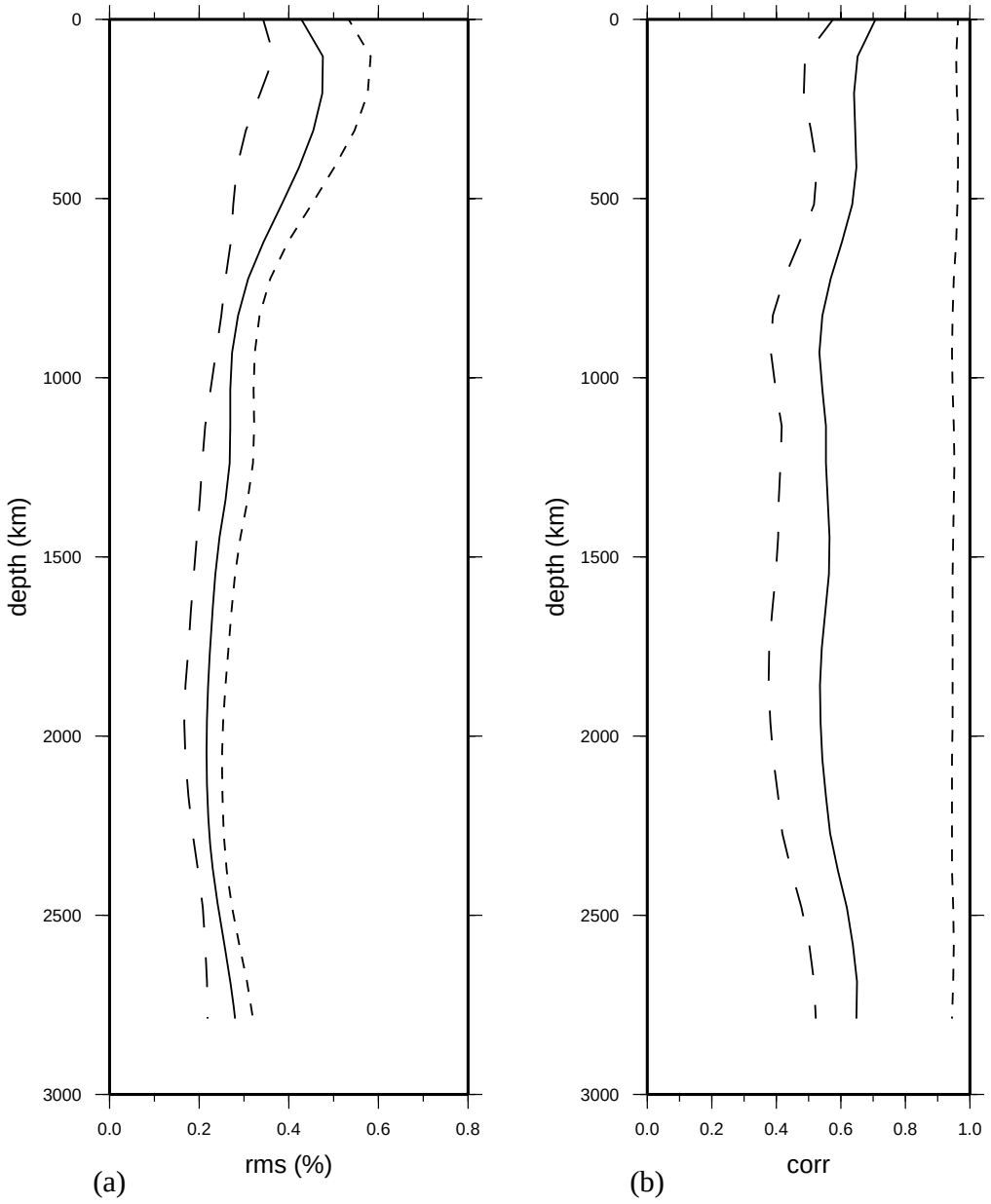


Figure 5.8 a) Model root-mean square as a function of depth for the short-period only inversion (long-dash line), long-period only inversion (short-dash line), and for the joint inversion (solid line). b) correlation of short-period and long-period (long-dashed line), of short period and joint inversion (solid line), and of long-period with joint inversion (short-dash line).

which originate in the "superplume" regions located at the bottom of the mantle. The plumes, clearly visible in the low-frequency inversion, are also recognizable (although they are much weaker) in the short period tomographic images (Figures 5.9, 5.10, 5.11 and 5.12).

Figure 5.9 to 5.12 show four example of plume-like structure detected in our tomographic image. We compare the signature of the plume as detected in the long-period model (left), short-period (center) and in the joint inversion right). We observe a remarkable agreement in the structure, but a strong damping of the amplitude in the short-period images.

Why are the amplitudes in the two individual tomographic models so different? Amplitudes of the velocity anomalies are the most difficult to reconstruct, and are often poorly resolved in tomographic images. They are strongly affected by the regularization.

The comparison of the 'raw' delays has shown that both short and long period seem to see the same Earth.

The ray coverage is very heterogeneous and there might be regions that are badly illuminated (see figure 5.1). The velocity at the points of the grid within these regions can vary without seriously affecting the fit of the linear system. The tomographic system is unstable and the model solution is strongly dependent on the regularization technique. But this does not seems to be our case, since our inversions are quite insensitive to changes to the damping factor.

χ^2 is defined by a priori estimates of the standard deviations in the measurements, the choice of which is in a sense a subjective one. Underestimated σ's would give a solution model less damped; and viceversa, overestimated σ's would lead to overdamping. Since the long period model has larger amplitude than the short period model, we must investigate if standard errors of the long-period arrival times are underestimated, or alternatively, if those associated to the short period delay times are overestimated. The long-period data have been measured from broadband stations by cross-correlation techniques, mainly by a single operator. Various inversions with different values of the regularization show that at this level of chi-square $\chi^2_{red} = 1.1$ there are no signs that the model is dominated by noise. The similarity of the structures among the different resulting models, like the plumes in figures 5.9–5.12, argues against that. Therefore, it seems unlikely that we are underestimating the error in the long period data. Are we then overestimating the errors in the short-period data? Morelli & Dziewonski (1987) estimate the standard error in P waves to be 0.9 s. Conceivable, the re-interpretation strategy of Engdahl et al. (1998) and our rejection of all but the highest quality measurements in this data set could have brought this value of σ down below our adopted value of 0.88 s. The main reason we believe this not to be the case is that a test with an undamped inversion produces a χ^2_{red} of 0.89, only slightly below 1. Since our node spacing is of the order of the width of the Fresnel zone, inadequacy of the model parameterization cannot be a major factor in our inability to bring χ^2_{red} further

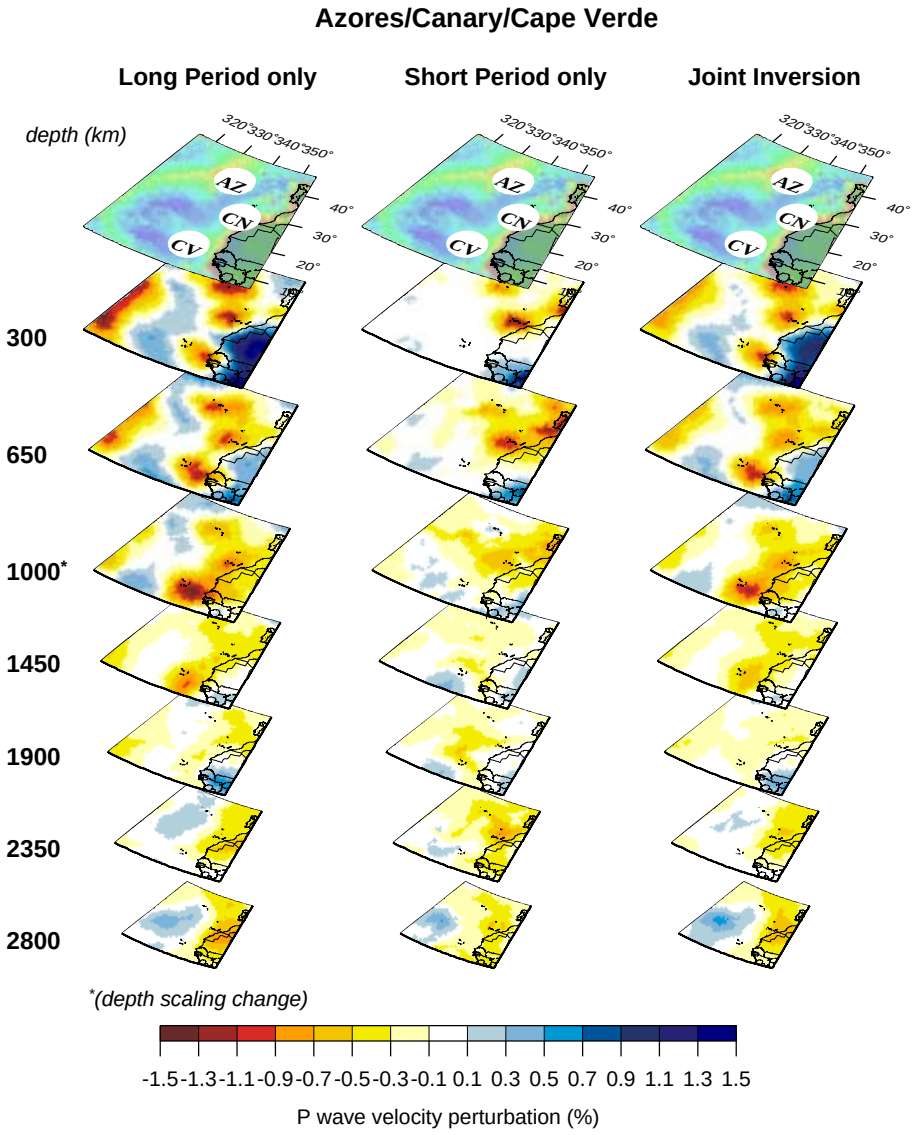


Figure 5.9 Comparison of the plumes like features mapped beneath Azores, Canary and Cape Verde as mapped in the three tomographic inversions: long-period only (left), short-period only (center), joint inversion (right). Short-period tomographic images are characterized by smaller amplitude anomalies, but much sharper resolution.

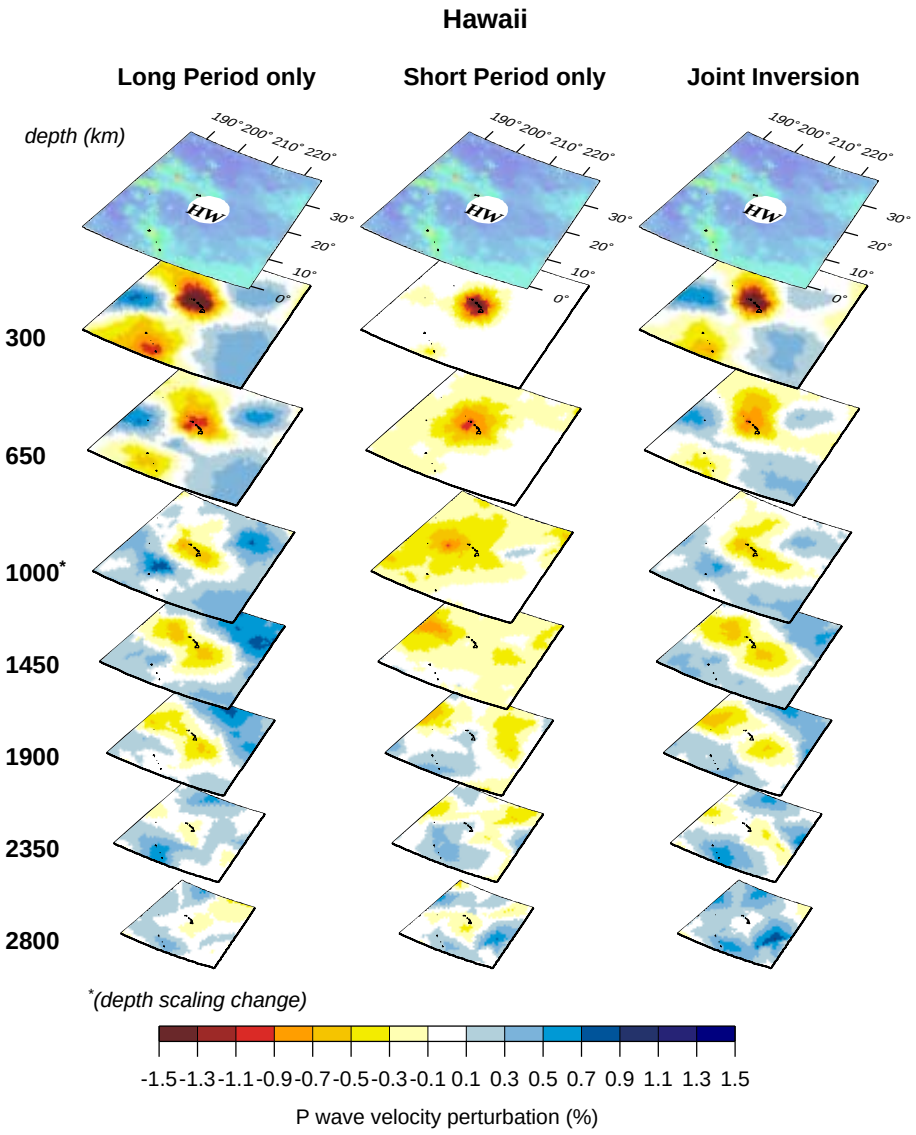


Figure 5.10 Same as figure 5.9 but for the plume beneath Hawaii. Hawaii is also a deep-mantle plume.

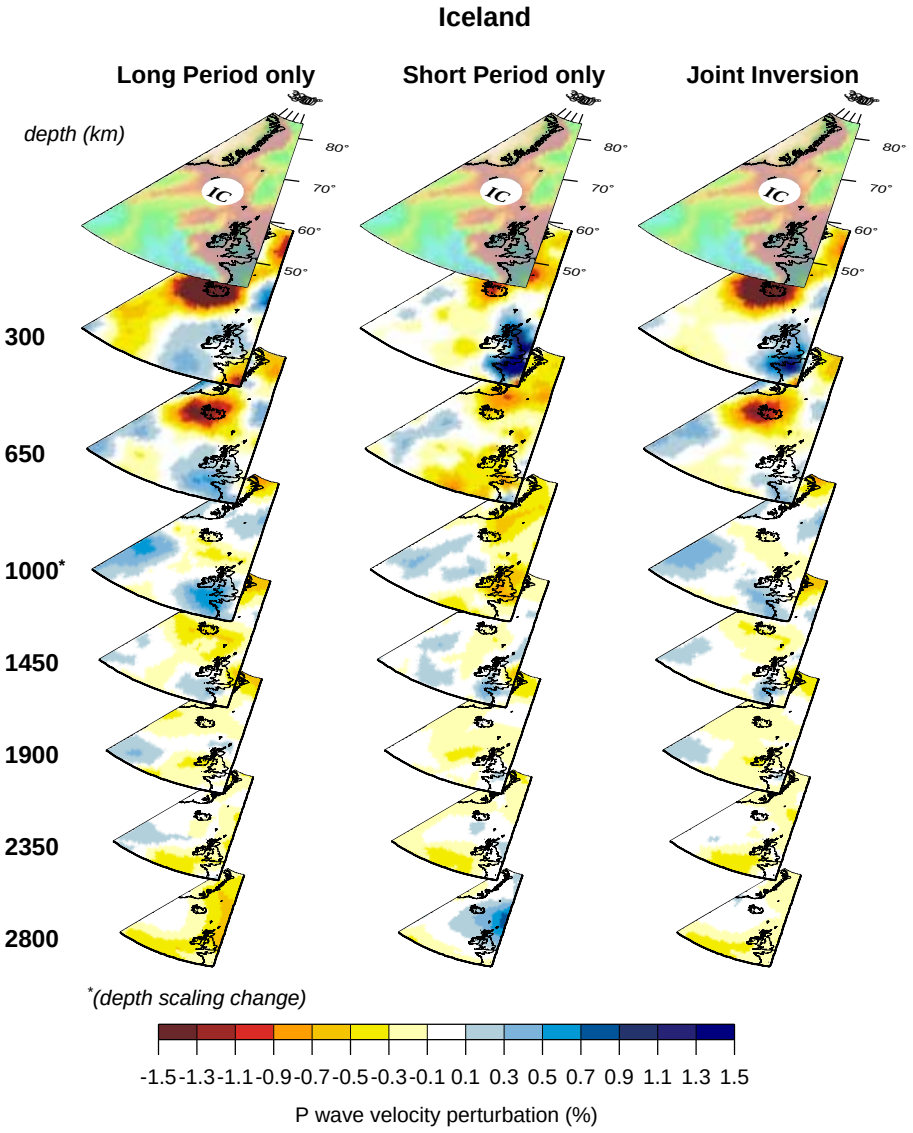


Figure 5.11 Same as figures 5.9 and 5.10, but for the plume beneath Iceland.

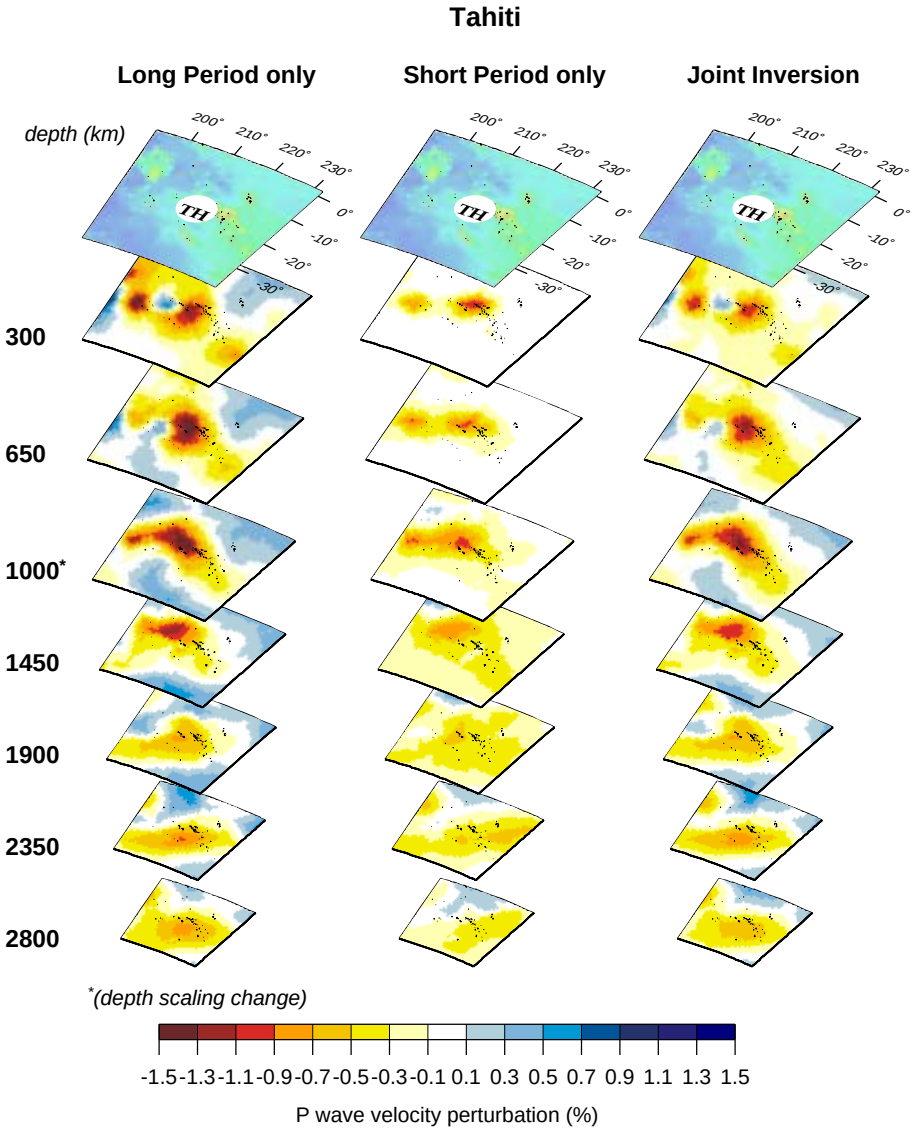


Figure 5.12 Same as figures 5.9–5.11, but for the plume beneath Tahiti.

down. For comparison, the model shown in figures 5.4–5.7 has χ^2_{red}=0.95. No relaxation of regularization allowed us to obtained a reasonable short-period model whose amplitudes were similar to the long-period model. These results suggest that errors adopted for the short period data are not overestimated; if anything, they seem rather underestimated.

An alternative explanation for the lower amplitudes characteristic of the short-period model could be found by invoking wavefront healing for this class of data as well. In Chapter 4, we found that finite-frequency modeling tend to increase the amplitudes of the velocity perturbations with respect to ray theory. For that analysis we were inverting the long-period data set by using the two different techniques and the result indicated that wavefront healing affects the amplitudes in the tomographic image by as much as the difference observed here. Could amplitudes in the short-period model be smaller because they, too, are affected by wavefront healing? In figure 1.1a in chapter 1 we show the Fréchet kernel computed for a P wave with dominant period of 1 s. Sensitivity regions are significantly smaller then those for long-period waves, but still it has a diameter of about 400 km at the turning point of a wave of 80° epicentral distance (Figure 1.1a). Structures smaller than 400 km, as the plumes, will give delays that suffer from healing. Also, because of the assorted ISC instrumentations, it could be possible that many of the onset measurements are representative for a larger dominant period than 1 s, which would induce strong wavefront healing effects in short-period data also for structures larger than 400 km. The solution may be found by doing a test inversion for the short-period data with finite-frequency modeling, but that requires a parallelization of our code which we have not yet accomplished.

5.6 The joint inversion

This section provides a brief overview of the joint inversion model. For a detailed description of the plumes we refer the reader to chapter 6.

The model we obtain satisfies the short period data with a χ^2_{red} of 1.1 and the long period data with a χ^2_{red} of 1.2. The small difference is not judged significant, given the uncertainties in our a priori estimates for σ.

Our model shows several features that have already been identified by earlier investigators (van der Hilst et al., 1998; Grand et al., 1997; Grand, 1994; Bijwaard et al., 1998; Van der Voo et al., 1999; Gu et al., 2001).

Fast anomalies are located beneath Asia, most probably identifiable as the Tethys slab and beneath Java, Tonga Kermadec, and Tasmania. Clear is also the signature of the Farallon plate, sinking eastward beneath North America; and of the subduction of the Nazca plate beneath South America. Again the lowermost mantle is dominated by the two "superplume" located in the Atlantic Ocean and under South Pacific.

There is an astonishing agreement between the low-velocity anomalies and the

location of known hotspots. A detailed analysis of these anomalies (see chapter 6 for a complete description) have shown that plumes do not originate all in the deep mantle, they may as well originate at much shallower depths. Deep plumes are located beneath Ascension, Azores, Canary, Easter Island, Hawaii, Samoa and Tahiti. Others plumes originate at much shallower depths. For instance the Hainan plume, first seen in a tomographic study by Lebedev (Lebedev & Nolet, 2003) is clearly seen down to 1000 km depth. Iceland, Galapagos and Juan de Fuca also originate at much shallower depth, most likely near 660 km. Newly discovered plume-like features have been observed beneath the Mid-Atlantic Ridge (between 15°N and 25°N,45°W), in the Indian Ocean (35°S,100°E) and north of Reunion, beneath Seychelles (5°S,56°W).

5.7 Conclusions

We performed a global tomographic study in which we jointly inverted very high quality short - and long - period data. Despite the different character of the two data sets, a statistical analysis of a subset of delays times observed at the same location shows that the two data sets are reasonably consistent. Also, a direct visual comparison of velocity models obtained by inverting each data set separately shows an overall good agreement between the two inversions. However, amplitudes are significantly different among the two models and are much weaker in the short-period tomographic images. Objectively, we cannot conclude that one data set is better than the other, nor we can confidently argue that the amplitudes in the long-period are better estimated than in the short-period model. With this said, our approach for simultaneously invert the two data sets by minimizing their respective χ^2_{red} seems the most appropriate. By doing this, we guarantee the equal weight of the two data sets on the inversion. Finally, although the origin of this difference in the amplitudes between the two tomographies seems to be encircled by a certain halo of vagueness, we can confidently say that it not due to our assumed standard errors in the data. The comparison of a ray-theoretical tomography with a finite-frequency tomography of the same data set, described in chapter 4, shows that ray theory tends to underestimate the anomalies for measurements that account for diffraction phenomena. We could therefore speculate that wavefront healing also affects the short period data, and explain the lower amplitudes of the short-period anomalies as due to the lack of modeling of the finite-frequency effect by ray theory. We cannot confidently rule this hypothesis out, because we cannot say with certainty that short-period data are all characterized by a very low (≤ 1) dominant period. More tests are required to study possible finite-frequency effects in the very large short period data set.

Chapter 6

Finite-Frequency tomographic reveals a variety of plumes in the mantle

A modified version of this chapter will be submitted for publication in Science as: Montelli R., Nolet. G., Dahlen F.A., Masters G, Engdahl R. E. and Hung S.-H.: Finite-Frequency tomography reveals a variety of plumes in the mantle.

6.1 Résumé

Dans ce chapitre, je fournis une description très détaillée des 32 points chauds trouvés dans notre image du champ de vitesse. D'une manière similaire au test décrit dans le chapitre 4, un test de résolution est fait pour pouvoir définir les limites pour la profondeur et la largeur des panaches qui alimentent les points chauds. Ce test me permet de classifier les panaches découverts en trois familles: les profonds, les superficiels et ceux commençant à profondeur moyenne. Les resultats de cette analyse sont récapitulés dans la table 6.1. Au moins une dizaine des points chauds a une origine profonde: Ascension, St. Helena, Azores, Canaries, Crozet, Kerguelen, Hawaï, Samoa, Cook Island et Tahiti. Une autre dizaine semble avoir une origine dans le manteau au-delà de 1000 km. Les panaches ont besoin d'une couche thermique de frontière. Bien que l'existence d'un tel changement des caractéristiques du manteau à de telles profondeurs a été deja avancée, la preuve n'est pas faite d'une telle discontinuité thermique. Notre travail pourrait donc aussi confirmer cette hypothèse grâce à ces panaches intermédiaires.

Vu l'importance de cette observation, il faut être tr'es critique sur les images tomographiques. J'ai augementé la limite des changements de vitesse qui permet de considérer le panache synthétique comme résolu. Cette analyse me permet de conclure qu'au moins une dizaine des panaches intermédiaires pourrait avoir leur origine à la base du manteau en raison du manque de résolution à cette profondeur.

Seulement deux des panaches analysés passent l'épreuve et semble satisfaire ce test de resolution tres sélectif. Ce sont les panaches au-dessous de Hainan, et de la dorsale medio-Atlantique (à 22° N).

Un nombre de panaches a une origine plus superficielle, tres probablement dans la région de discontinuité des 660 km. Le modèle montre aussi des points chauds qui n'apparaissent dans aucun liste publiée. En même temps, des points chauds comme Yellowstone, Macdonald, Lake Baikal et Bermuda n'apparaissent pas dans le modèle actuel. Pour finir, le modèle montre aussi des structures à panache à la base du manteau qui n'arrivent pas encore à la surface au-dessous de la Mer de Coral et au sud de Java.

6.2 Abstract

Recent improvements in the measurement and interpretation of the small variations in travel times of seismic P waves show that many of the Earth's volcanic hotspots are underlain by deeply reaching thermal plumes. New tomographic images show at least six plumes that extend into the lowermost mantle and are well resolved even at that depth: Ascension, Azores, Canary, Easter, Samoa and Tahiti. Several other plumes, among which is Hawaii may also reach the lowermost mantle, but suffer from lack of resolution in the images. Other plumes originate at much shallower depth. The images show several hitherto unsuspected plumes: two that feed the Atlantic Ridge and the South-east Indian ridge without creating excess in topographic relief, one beneath the Seychelles and two in the lower mantle that do not reach the surface.

6.3 Introduction

Several dozen 'hotspots', characterized by higher temperature, topographic swells, and recent volcanism with isotopic signatures distinct from those that characterize mid-ocean ridge or andesitic basalts are found at the surface of the Earth (Sleep, 1990; Davies, 1988; Courtillot et al., 2003). The best known example is the Hawaii-Emperor chain in the North Pacific. Wilson (1963, 1965) suggested that these volcanoes were formed as the Pacific plate moved over a deep mantle volcanic source. Morgan (1971, 1972) generalized Wilson's idea to all of the global hotspots, and suggested that such intraplate volcanism is the surface manifestation of a deep mantle plume (Richards et al., 1989; Courtillot et al., 1999; Norton, 2000; Courtillot et al., 2003). Plumes are commonly observed in laboratory experiments (Davaille, 1999; Davaille et al., 2002) and numerical simulations (Bunge et al., 1997; Zhong et al., 2000; Cserepes & Yuen, 2000). The plume hypothesis plays an important role in explaining flood basalts, the isotopic signature of ocean island basalts and the topographic profile across swells and plateaus that often accompany hotspots.

While this has led to a coherent theory of much of the geology that accompanies hotspots, though with imperfections, undisputed evidence for the existence of lower mantle plumes in tomographic images of the mantle is so far lacking.

Low velocity anomalies have been found beneath some hotspots in the upper mantle, and broad regions of low velocity are visible in the lowermost mantle beneath the Pacific and Africa (Ritsema & Allen, 2003). Nataf & VanDecar (1993) detected low seismic velocities - presumably due to higher temperatures - in the top of the lower mantle beneath the Bowie hotspot. The upper mantle transition zone close to Iceland is locally thinned (Shen et al., 2001), again a sign of an increase in temperature [1]. Bijwaard & Spakman (1999) image a deep Iceland plume but their finding is disputed by Ritsema et al. (1999) nor supported by the analysis we report in this chapter.

Two 'superplumes', beneath Africa and beneath the South Pacific, are visible in all recent global tomographic images, most clearly in the S-velocity [2]. However, the superplumes do not clearly extend all the way to the surface and only provide a qualitative indication that deep upwelling takes place where the concentration of hotspots is especially dense. Direct tomographic evidence for deep narrow plumes feeding the volcanic activity at the surface has so far not been convincing, presumably because the narrow dimension of the plume conduit makes it hard to see for the long wavelengths of seismic waves that diffract around it.

As a result, thirty years after its formulation, Morgan's deep plume hypothesis has not yet been universally accepted. Anderson and others (Anderson, 1998, 2000; Foulger & Natland, 2003), argue that hotspots could as well be the manifestation of shallow, plate-related stresses that would fracture the lithosphere causing volcanism to occur along these cracks.

6.4 Technical aspects of the inversion

Our tomographic inversion differs in several aspects from earlier attempts to construct detailed images of the P velocity structure in the Earth's mantle: (1) we correct for the effects of wavefront healing in travel times of low frequency P waves, enabling us to combine long- and short-period data sets; (2) we adapt the model parameterization to the lower resolution at depth; (3) we use a re-measured, expanded set of long period data; and (4) we select only the highest quality short period delay times.

Global tomographic models of seismic P wave velocity have so far almost exclusively relied on the approximations of ray theory. Ray theory assumes that the travel time of a P wave is only influenced by the Earth's properties along an in-

[1] Global maps only give correlation with hotspot concentration over large oceanic areas (Flanagan & Shearer, 1998; Gu et al., 2001), but this may be due to lack of resolution in the global images.

[2] see see http://mahi.ucsd.edu/Gabi/rem2.dir/shear-models.html#mods

finitesimally narrow path – the ray – that follows Snell's law. This simplifies the mathematics, but it ignores that the traveltime is sensitive to velocity structure far away from the geometrical ray, within a volume known as the Fresnel zone. The maximum width of the Fresnel zone is $\sqrt{\lambda L}$ for a wave of wavelength λ and ray length L and easily reaches a width of a thousand km at the long periods we consider. Objects much smaller than the width of the Fresnel zone will not appreciably influence the travel time of the wave (Nolet & Dahlen, 2000). The phenomenon is commonly known as wavefront healing. For heterogeneities in size comparable to the Fresnel zone, Dahlen et al. (2000) and Hung et al. (2000) show how one may correct for effects of wavefront healing using a new method of interpretation which we shall refer to as finite-frequency tomography. A preliminary analysis of low-frequency P wave arrival times shows indeed that the velocity anomaly of deep small heterogeneities is underestimated by 30-60% when interpreted by classical ray theory (Montelli et al. (2003), also chapter 4).

Mantle plumes are probably narrow and their images are potentially strongly affected if we ignore the effects of wavefront healing (Nataf & VanDecar, 1993). We present the result of a finite-frequency tomographic inversion of travel times of 66,210 P, and differential times of 20,147 PP-P and 2382 pP-P of waves with a dominant period of 20 s combined with a ray-theoretical interpretation of 1,427,114 short period P and 68,911 pP times extracted from bulletins. The travel times of the long period phases were measured by cross-correlation (Bolton & Masters, 2001). Though part of the data set of long period travel times also served to construct model P16B30 (Bolton, 1996), the original seismograms were re-analyzed and several instrumental timing errors that were only recently discovered have been corrected. Standard deviations were assigned in three accuracy classes and range between 0.5 s and 1.15 s (Montelli et al. (2003), also chapter 4).

The short period times are onset times, picked by analysts and reported to the International Seismological Centre (ISC). These data have been re-interpreted as in Engdahl et al. (1998). By only selecting times listed with two decimal precision, and labeled the highest quality, we are confident the ISC-derived travel times are representative for high frequencies for which ray theory is acceptable. We assigned standard deviations of 0.88 s to P and 1.14 s to pP (Morelli & Dziewonski, 1987) and rejected outliers with deviations larger than 3σ after a first iteration.

We simultaneously invert for perturbation in the compressional velocity $\delta c/c$ and in hypocentral parameters (origin time, longitude, latitude and depth). To avoid that the image is completely dominated by the much larger short-period data set, we weight the least-squares system such that the reduced chi-square χ^2_{red} (χ^2 divided by the number of data) is approximately equal for each data set separately. Because the short period P waves originate from the nucleation point on the fault surface, whereas long period waves average over the rupture process, we use independent hypocentral corrections for the 5,938 events that produced the long-period data, and the 86,499 for the short-period times.

The velocity structure is represented through linear interpolation between a set of flexibly spaced nodes that form a Delaunay mesh (Watson, 1981, 1992; Sambridge et al., 1995). The node spacing increases with the expected resolving length of our data and ranges from about 200 km in the upper mantle to about 600 km in the lower mantle. The grid consists of 19,279 nodes. With the hypocentral corrections we obtain a total number of 389,027 unknowns to resolve with 1,584,764 observations. The influence of data errors in the tomographic image is kept under control by regularization: the source parameters are damped to give a posteriori shifts comparable to their a priori error estimates, the velocity inversion is slightly damped, but mostly regularized by adding a Laplacian smoothing as described in Nolet (1987).

The model we obtain satisfies the short period data with a reduced χ^2_{red} of 1.1, whereas the long period data are fitted with $\chi^2_{red} = 1.2$. In test inversions for the data sets separately, we noted a tendency for the long period data to produce higher anomalies by as much as 50%, though the two models correlate well with a correlation coefficient, which varies with depth, around 0.6. The χ^2 values obtained for the combined inversion show that the amplitude difference is not due to a pronounced difference in internal consistency of the delay times. The difference may indicate that even the ISC data are subject to wavefront healing, that we overestimate the errors in the ISC data set, or that there are as yet unknown biases in the data. This will be subject of a more focused study. But most importantly, the plumes that are visible in the low frequency inversion are also recognizable (though often weaker) in the high frequency inversion, indicating that it is the full combination of all improvements that brings about the elucidation of deep mantle plumes. We accumulate both data sets in one inversion to exploit the difference in sensitivity of each data set.

6.5 The resolution of plumes

The P wave velocity model exhibits a series of deep plumes, most of which are shown in Fig 6.2. The main characteristics for 32 plumes are listed in Table 6.1.

As in every high resolution tomographic study, the fact that we have far more data than unknowns does not guarantee that every model parameter is well resolved. Much of our effort has therefore been spent on determining which plume features are reliable.

Lack of resolution may take different forms: (1) an existing plume may not be imaged with sufficient contrast over its full length or to its source at large depth where resolution often decreases, (2) the regularization may spread a shallow anomaly to larger depth in the image ('leakage'), and (3) the regularization leads to horizontal 'smearing' of an anomaly, resulting in a larger anomaly image with a smaller amplitude than present in the Earth.

To test the resolution, we generated 'synthetic' data sets for Earth models

Label	Name	Lat(°)	Lon(°)	Plate*	Depth (km)†	Radius(km)‡	Remarks
AF	Afar	7	39	af	≥ 1450	200	it could be a deep plume
AR	Atlantic Ridge	15(25)	-45	na	$\simeq 1900$	200	lack of resolution at the base of the mantle gives rise to horizontal smearing
AS	Ascension	-8	-14	sa	$\simeq 2800$	100	robust
AZ	Azores	38	-26	eu	$\simeq 2800$	300	lack of horizontal resolution in the mid-mantle responsible of merging with CN
BV	Bouvet	-54	3	af	≤ 1450	400	connection of BV to African Super plume is not well resolved
BW	Bowie	53	-136	pa	≤ 650	100	plume wider than 100 km would show up if BW is deeper than 300 km
CN	Canary	28	-18	af	$\simeq 2800$	400	robust
CV	CapeVerde	15	-24	af	≥ 1900	300	lack of resolution in the deep mantle
CR	Caroline	3	167	pa	≥ 1000	300	lack of resolution at the bottom of the mantle
CC	Cocos/Keeling	-17	95	au	≤ 1000	200	strong only in the upper mantle
CK	Cook Island	-22	-158	pa	≥ 1450	200	lack of resolution in the lower mantle
CS	Coral Sea	-15	155	au	$\simeq 2800$	300	robust, does not reach the surface, it stops at 1450 km
CZ	Crozet	-46	50	an	$\simeq 2350$	400	vertical leakage present, but confined above 1450 km depth; same origin as KG at 2350 km depth; synthetic plumes are separated down to 1000 km depth
ES	Easter	-27	-108	nz	$\simeq 2800$	400	robust
EA	E.Australian	-41	146	au	$\simeq 650$	100	robust
EF	Eifel	50	4	eu	$\simeq 650$	100	robust
ET	Etna	38	15	eu	≤ 1000		strong only in the upper mantle
SL	East of Solomon	-5	165	pa	$\simeq 1000$		robust
GL	Galapagos	0	-92	nz	≤ 1000	300	strong only in the upper mantle
HN	Hainan	20	110	eu	$\simeq 1000$	200	robust
HW	Hawaii	19	-155	pa	(?)2350		not enough constraints on depth
IC	Iceland	64	-17	na	≤ 1000	100	strong only in the upper mantle (see discussion in text)
IO	IndianOcean	-35	100	au	≥ 1900	400	not well resolved
JF	Juan de Fuca/Cobb	46	-130	pa	≤ 1000	100	strong only in the upper mantle
JZ	Juan Fernandez	-34	-81	nz	≥ 1450	300	lack of resolution below 2350 km
KG	Kerguelen	-50	69	an	$\simeq 2350$	400	robust
LS	Louisville	-54	-141	pa	≥ 1000	300	very weak, newly discovered anomalies located at $(55°S, 150°W)$ and $(60°S, 120°W)$
RN	Reunion	-21	56	af	≥ 1900	200	it could be a D" plume
SM	Samoa	-15	-168	pa	$\simeq 2800$	200	robust, except between 1000 and 1450
SJ	South of Java	-12	112	au	$\simeq 2800$	300	does not reach the surface, it stops at 1450
SY	Seychelles	-5	-56	af	$\simeq 650$		robust
TH	Tahiti	-18	-148	pa	$\simeq 2800$	300	robust, except between 1000 and 1450 km depth

*: af - African plate; an - Antarctic plate; au - Autralian plate; eu - Eurasian plate; nz - Nazca plate; pa - Pacific plate; sa - South American plate;
†: value is the one visible in the tomographic images, sign is constrained from the resolution tests (see text for details)
‡: minimum radius of the plume constrained from the resolution tests (see text for details)

Table 6.1 Summary of the results for the 32 hotspots present in our tomographic images. Depth range and width of the plume have been determined from the resolution analysis.

with cylindrical plume anomalies of (Gaussian) radius w =100, 200, 300 and 400 km: $\delta c(r) = \delta c_0 \exp(-r^2/w^2)$, centered at a large number of observed plumes and extending to depths of 650, 1000, ..., 2800 km (as in Fig 6.2–6.4) [3]. After adding noise, these data were inverted with the same regularization as the real data and the images were inspected for agreement with the input image used to generate them. We consider a plume 'unresolved' if the synthetic image has a contrast $|\delta c/c| < 0.3\%$. In addition to testing the influence of ray coverage, the test also reveals when our interpolation scheme affects the image. We show many of our tests in the Appendix B.

The tests enable us to determine the minimum radius a real plume should have for it to be visible in Figs 6.2–6.4. This plume radius is listed in Table 6.1. The imaged plume radii should be seen as generous upper limits; because of the effects of smearing the actual plume radii may be much smaller.

The depth listed in Table 6.1 is the deepest level at which the contrast exceeds the 0.3% level. We coded these maximum depths in Table 6.1 with a $\geq$ sign if we conclude from the resolution tests that the absence of a plume at greater depth may be due to a lack of resolution, or with a $\leq$ sign if there is a possibility that the image is generated by leakage to this depth. In case we determined that the resolving power is sufficient, we added $\simeq$ to stress that the depths in Table 6.1 are estimates only. Even in the case of well-resolved plumes, they are uncertain to several hundred km at large depth.

This codification of depth values in Table 6.1 is somewhat incomplete in the sense that resolution may depend on the unknown width of the plume. Where relevant, we discuss caveats in the last columns of Table 6.1 and in the next sections.

6.6 Deep plumes (Fig 6.2)

Much to our own surprise (since temperature anomalies such as those in plumes affect S velocities much more than P velocities), our P velocity images show unambiguous evidence that some hotspots cap a plume originating near the core-mantle boundary. In Fig 6.1 we show a vertical average of the P velocity anomaly in the lowest part of the mantle (1800-2800 km depth). This 'filtered' representation of the tomographic model emphasizes features such as plumes that are vertically continuous over all or much of the averaging depth. The 'African superplume' extends high enough to survive the averaging over 1000 km and is visible as the

[3] Synthetics times were computed by using the finite-frequency modeling of Dahlen et al. (2000) for long period data and by using standard ray theory for short period data. The highest velocity perturbation δc_0 in the center of the cylinder is defined by following the pattern for temperature derivatives of P-wave velocities in the mantle as a function of depth given by Karato (Karato, 1993, Fig 1). Assuming a temperature $T = 300°$ K at the center of the plume, the maximum velocity perturbation is -1% below 1000km depth, -1.2% between 600 and 1000 km depth, and -2.4% above 600 km depth (Fig B.1 in the Appendix).

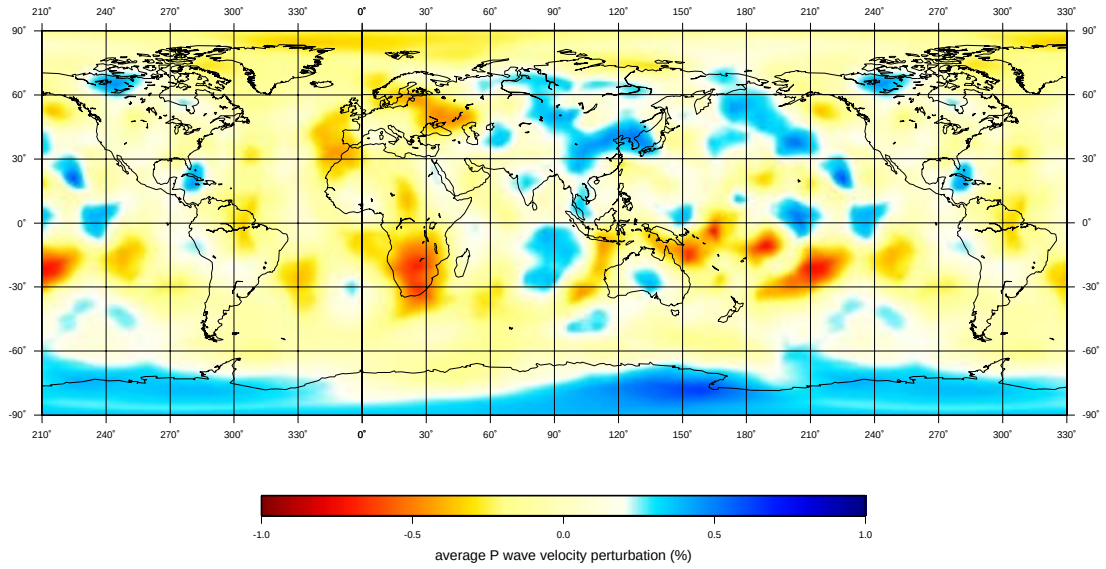


Figure 6.1 Vertical average over 1000 km in the lowest part of the mantle of the relative velocity perturbation which emphasizes features that are continuous with depth. Map has been wrapped around to have complete views of both the Atlantic and the Pacific Oceans.

large low velocity anomaly beneath southern Africa. Inspection of the unfiltered tomographic model shows that the superplume extends locally to depths of 1500 km in accordance with findings by Ritsema et al. (1998) and Ni et al. (2002). This anomaly is part of a broad low velocity region that underlies the Atlantic, the African continent and much of Europe. The broad anomaly in the lower mantle beneath the Pacific south of the equator is different in character: here several maxima are enhanced by the averaging that can be identified as plumes rising from the 'superplume': beneath the Coral Sea, east of the Solomon Islands, Samoa and a broad anomaly centered beneath Tahiti. North of the equator, the Pacific lower mantle is more neutral, with some high velocity anomalies forming a ring that extends beneath East Asia and the eastern part of the Indian Ocean. In the original model, fast anomalies with amplitudes above 1% are observed beneath Tonga, Asia, the latter presumably identifiable as the Tethys slab (Van der Hilst et al., 1997; Grand et al., 1997; Grand, 1994; Bijwaard et al., 1998; Van der Voo et al., 1999; Gu et al., 2001) and in the northern Pacific Ocean. They form a ring of high velocity anomalies around the Pacific that reaches the D". Finally, the polar regions are characterized by lower mantle of opposite velocity anomaly - while the North Pole mantle is hot, it is cold beneath Antarctica.

In Fig 6.2, Ascension, Azores, Canary, Easter, Samoa and Tahiti hotspots all have well resolved deep-rooted origins near the bottom of the mantle.

Hawaii, one of the longest-lived plumes, and by far the strongest in flux as measured by the topographic swell it creates (Sleep, 1990; Davies, 1988), is visible

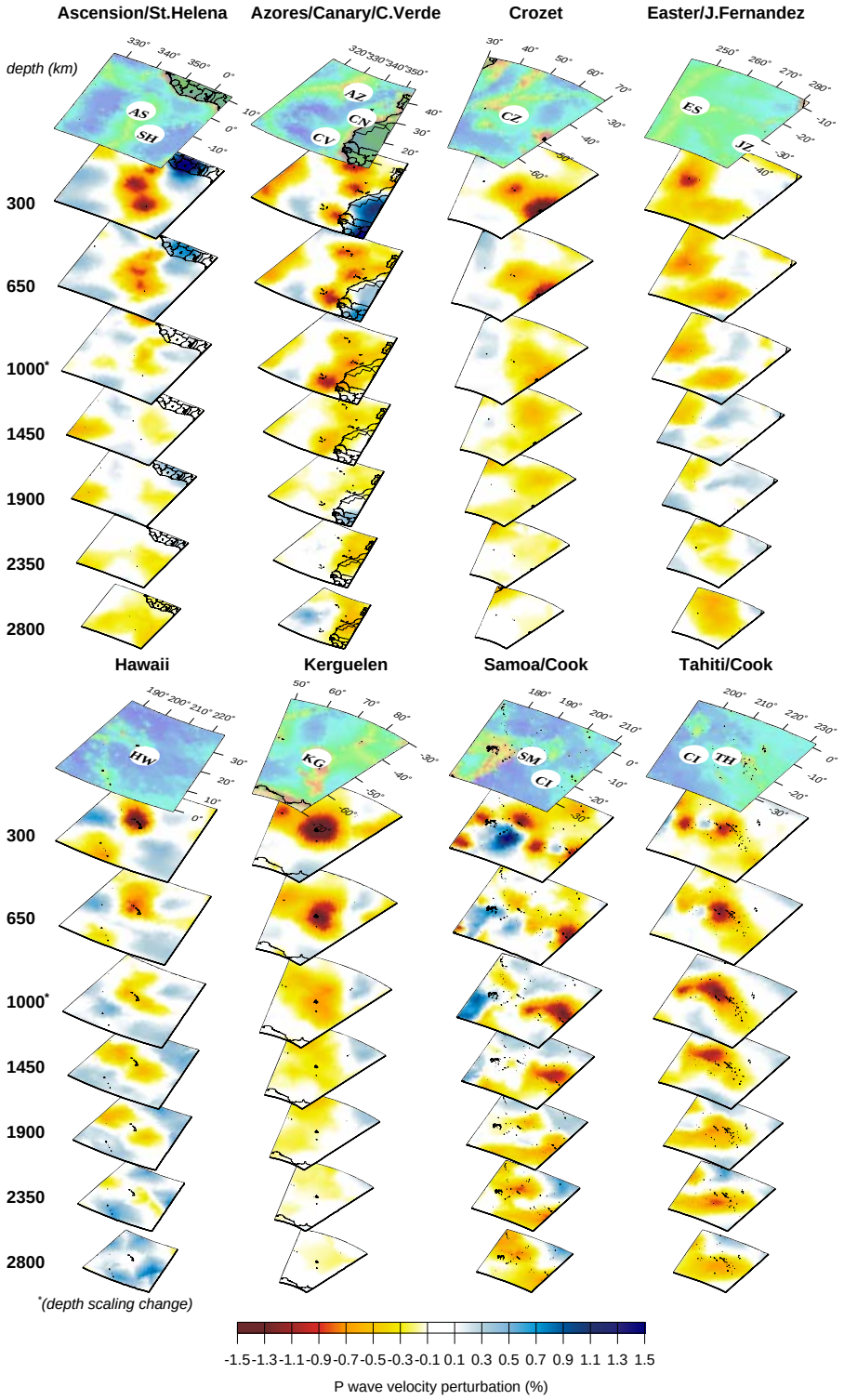


Figure 6.2 Three-dimensional view of deep plumes present in our tomographic model. Maps are 40° by 40° scaled with depth. Depth scaling changes at 1000 km for reasons of space.

as a deeply rooted plume. Resolution is good in the lowermost part of the mantle, although controlled by the very large number of paths running from Tonga subduction zone to North America stations which is such that amplitudes are overestimated especially for wide plumes ($\simeq$ 400 km radius, figure B.3 in the Appendix) and shifted to the south east, where the paths are concentrated. Lack of resolution is present in the lower mantle down to about 2000 km, as already observed in other tomographic studies (Nataf, 2000).

Some plumes reach the lower mantle by confluence with a nearby plume. Kerguelen and Crozet are seen to originate from a common broad anomaly located north of Crozet at 2350 km depth, the width of which is affected by horizontal smearing. St. Helena, next to Ascension, merges with Ascension at about 1000 km depth, a confluence that is well resolved. Azores and Canary originate at the bottom of the mantle. They are distinct plumes down to 1450 km depth, where they merge together and begin to bend eastwards to reach the bottom of the mantle at about (30°N, 10°S). Further south, the Cape Verde plume joins this complex at 1900 km depth. However, their apparent confluence may be the result of smearing of the image. On the Pacific superwell, Tahiti, Cook Island and Samoa appear closely spaced. The images for the Tahiti and Samoa plumes are robust and show independent features to large depth in the mantle. Cook Island merges with Tahiti at about 1450 km depth, but again this may be an effect of lack of resolution.

6.7 Mid-mantle plumes? (Fig 6.3)

Some plumes seems to originate in the mid-mantle, rather than at a recognized thermal boundary layer such as the core-mantle boundary or, perhaps, the phase transition at 660 km depth. Since this would be a finding of considerable impact to geodynamics, we took great pains to determine if this observation is simply due to lack of resolution. In this case, we take a more conservative point of view and only consider a plume 'resolved' if the synthetic image has a contrast $|\delta c/c| > 0.4\%$.

We found indeed that the mid-mantle images of Bouvet, Cape Verde (Fig 6.2) and the newly discovered anomaly in the Indian Ocean, can be explained as the a possible consequence of image 'leakage' from a plume confined to the upper mantle. The source regions of Afar, Juan Fernandez (Fig 6.2), the two newly discovered anomalies nearby Louisville, Cook Island (Fig 6.2), Caroline and Reunion could on the other hand very well be in the D" layer, as our analysis shows that narrow plumes beneath these hotspots suffer a loss of resolution at large depth, making their deep tails invisible.

Two mid-mantle plumes almost survive scrutiny by this strict resolution test: the newly discovered plume beneath the Atlantic Ridge originates at 1900 km depth; the Hainan plume, first seen in a tomographic study by Lebedev (2000), is a robust feature clearly visible in our velocity model down to 1000km. Yet we cannot rule out that they are simply not resolved, though this requires that their

radius is reduced to less than 200 km. We conclude therefore that none of the mid-mantle plumes can be undoubtedly associated to a mid-mantle source region.

6.8 Shallow plumes (Fig 6.4)

A number of plumes originate at much shallower depth, most likely near 670 km: Bowie, Eastern Australia, Eifel, Etna, Iceland, Cocos-Keeling, Galapagos and Juan de Fuca/Cobb.

Eastern Australia and Eifel are both robust features and clearly constrained to shallow depth.

Bowie, Juan de Fuca/Cobb are connected to form a broad low velocity anomaly at 300 km. Bowie is not really identifiable as an isolated plume. In the model it shows only down to 300 km depth. Juan de Fuca/Cobb reaches 1000 km depth. However, the velocity perturbation at this depth is much weaker than in upper mantle. It is hard to believe that the source region of Juan de Fuca/Cobb hotspot could be at 1000 km or deeper. Similar is the case for Cocos/Keeling, Etna, Galapagos and Iceland. The very strong upper mantle plume beneath Iceland has almost disappeared at 1000 km depth. Vertical leakage down and below this depth, could easily explain the anomaly left over of -0.3% beneath Iceland. We suspect that leakage has led an earlier study (Bijwaard & Spakman, 1999), to suggest a deep plume, but it is clear from our image that the strong anomaly observed in the upper mantle is not generated by large flux from the lower mantle. At about 650 km, Etna is connected to a plume-like low velocity anomaly beneath the Gulf of Suez, but lack of resolution beneath Etna suggests us that such a connection may not be real.

One could argue that the large drop in the temperature dependence of V_P in the transition zone (Karato, 1993) would make plumes a much weaker velocity anomaly in the lower mantle, and therefore more difficult to resolve, leading to apparent source regions near 670 km. However, the fact that we do observe numerous plumes to extend to the deep mantle contradicts such a reasoning, at least as a phenomenon affecting the stronger upper mantle plumes.

6.9 Newly discovered plumes (Fig 6.3 and 6.4)

Several plumes are visible in our model for which the existence had not been suggested before.

A deep plume is visible in the lower mantle south of Java (Fig 6.4), which does not reach the surface but stops at about 1450 km depth. The situation here is complicated: an anomaly feeding the South-east Indian ridge (Fig 6.3) is a well defined structure down to 2350 depth where it merges with the anomaly located south of Java. From the resolution tests it follows that all these structures are

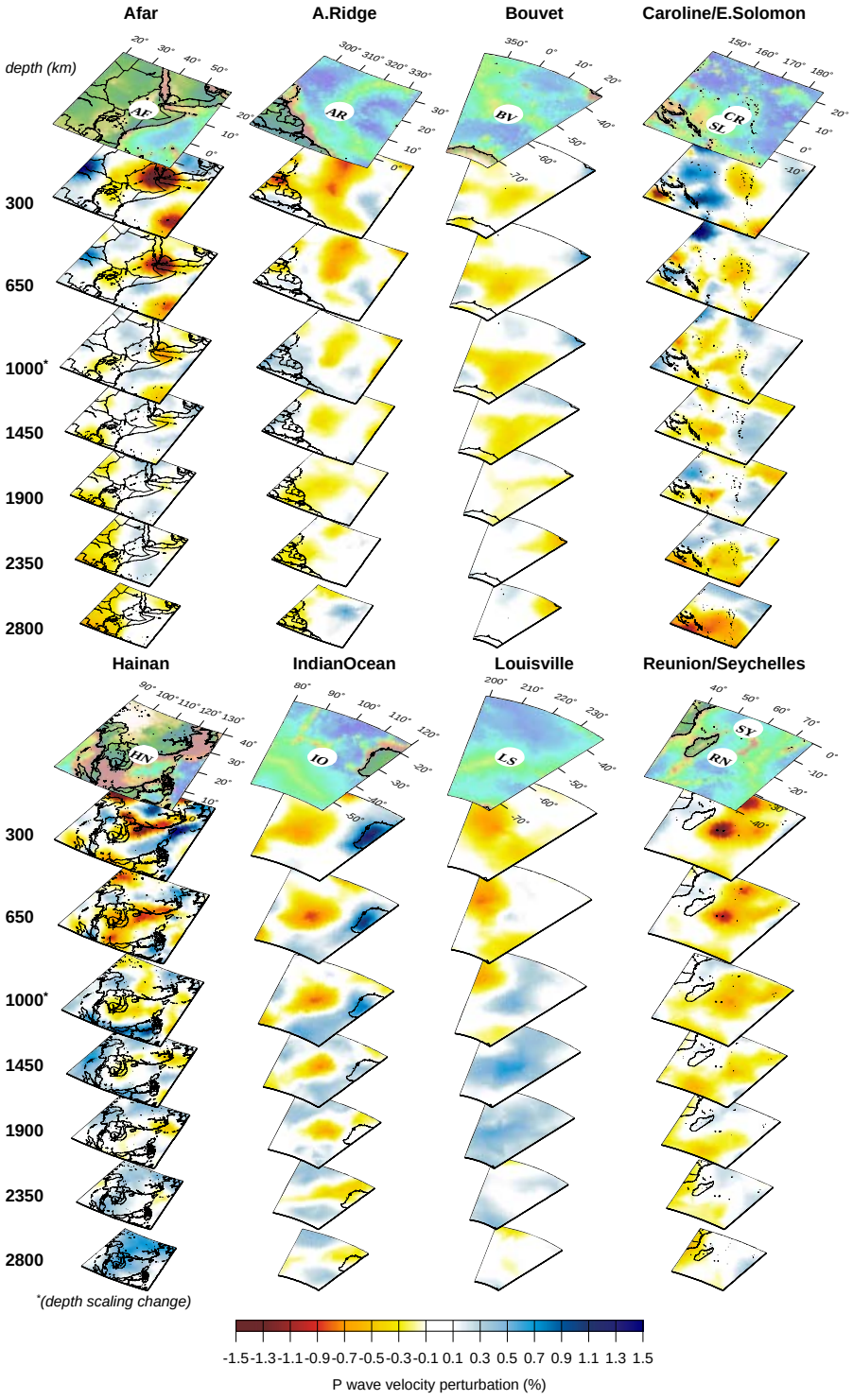


Figure 6.3 Three-dimensional view of the plumes that seems to originate in the mid-mantle.

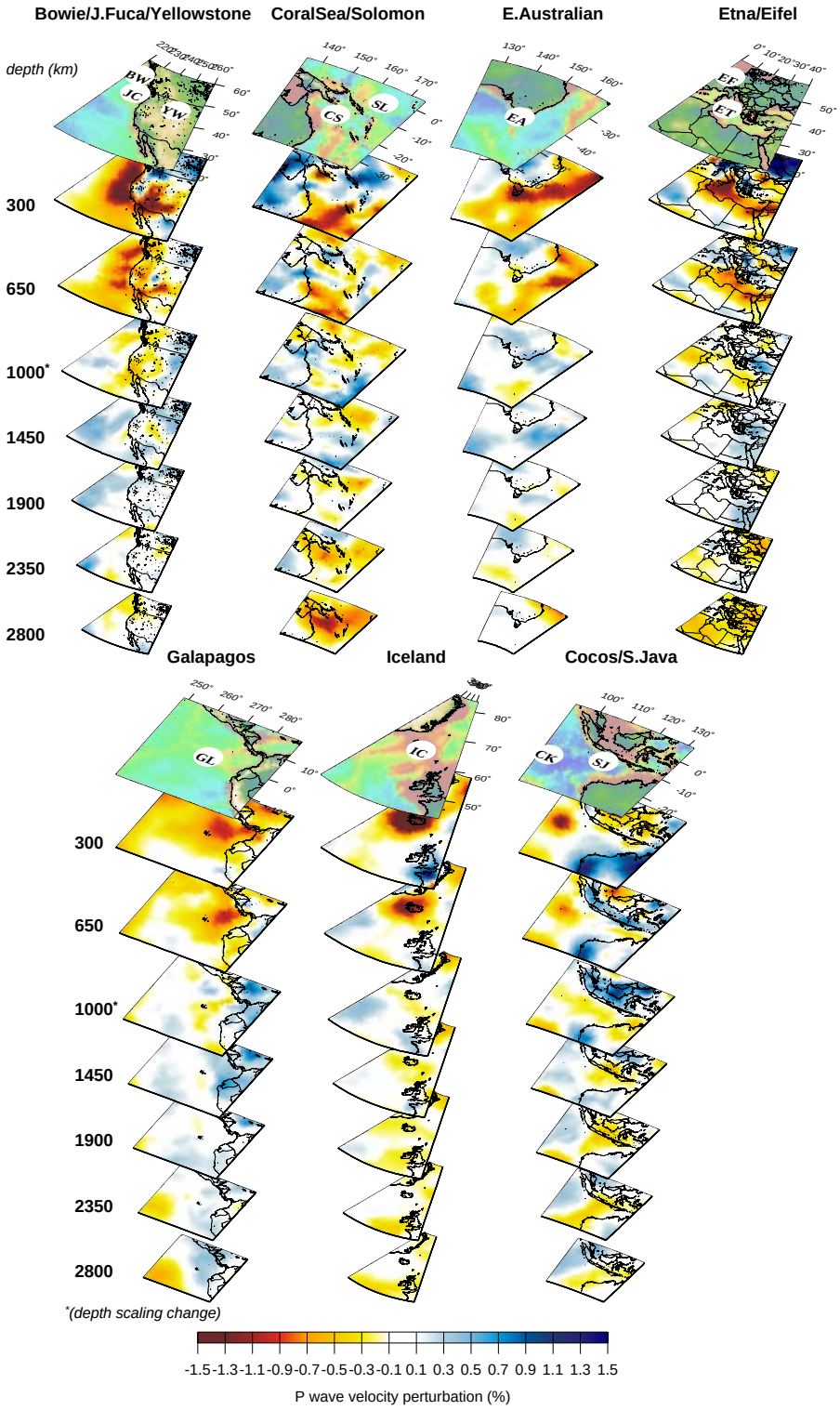


Figure 6.4 View of the shallow plumes and the newly discovered plumes present in our velocity model.

robust. The deep plume beneath the Coral Sea extends upward to 2350 km, with a weak connection to a newly discovered plume beneath east Solomon (5°S,165°E). The east Solomon plume stops at around 1000 km where a weak connection is visible with a plume beneath Caroline Islands (Fig 6.3).

Another hitherto unknown plume is located beneath the Mid-Atlantic ridge bottoming at 1900 km depth. It may be a double plume, since two maximum velocity anomalies are identifiable at about 12°N and 25°N (Fig 6.3).

We recover two low velocity anomalies NW and SE of the Louisville hotspot (Fig 6.3). Depth is not well constrained. Louisville hotspot, characterized by a well delineated narrow island chain, has waned drastically in recent geological times. Its actual location is disputed. For Sleep (1990), Louisville corresponds to a small volcano at 138°W. Morgan [4] places it at 141°W and 54°S. We do not observe a plume directly beneath these attributed locations.

We discovered a new plume north of Reunion, approximately beneath the Seychelles (5°S, 56°E) (Fig 6.3). This plume extends to 650 km and is well resolved. We note that Seychelles is a continental fragment left behind from the break-up of Gondwana (Wegener, 1924) and not a conventional oceanic island.

6.10 Absent plumes

Besides the ones cited earlier (Sleep, 1990; Davies, 1988; Courtillot et al., 1999), numerous lists of as many as 100 or more hotspots have been suggested, often with little agreement amongst each other. It is beyond the scope of this study to test our resolution beneath every one of them. But there are a number of notable absences among our list of plumes, and we shall discuss the most important among them.

No plume underlies Lake Baikal, though a low velocity anomaly is visible down to 1000 km at (48°N,92°W) in western Mongolia which is resolvable for plumes with $w \leq 200$ km.

Guadalupe is not an isolated low velocity anomaly in our tomographic image. It is part of a broad low velocity region connecting Bowie, Juan de Fuca/Cobb and what seems to be more likely a signature of the East Pacific Rise (15°N, 115°W), i.e. south of Guadalupe.

Probably because of the absence of a station on MacDonald we do not resolve a plume in the upper mantle beneath MacDonald island.

Our tests show that even an anomaly of only 100 km radius would be visible beneath Yellowstone leading us to conclude that Yellowstone has no significant plume.

[4] Jason Morgan personal communication, 2003

6.11 Discussion

Our deep-mantle plume images represent the first direct confirmation of Morgan's hypothesis of a "plume model" for a large number of hotspots. Since the D" region is an obvious thermal boundary layer, it is no surprise that this is the source depth for a number of observed plumes. Though the phase transition at 670 km has been a candidate for a second thermal boundary layer in the Earth, recent tomographic evidence for whole-mantle convection had made this less likely. Yet, if the 670 discontinuity can temporarily delay penetration, upwelling material may accumulate and give rise to an apparent source region at or just below 670 (Cserepes & Yuen, 2000).

Our tomographic images show a collection of plumes that apparently originate from the mid-mantle though in each case attempts to obtain a firm constraint on the depth is undermined by doubts about the resolution, certainly if the actual plume is thinner than 200 km. From a geodynamical point of view, plumes that would originate neither at 670 nor near D" would be a problem since no thermal boundary layer has so far been observed in the mid-mantle. The presence of a change in mantle chemistry has been hypothesized from changes in the pattern of subduction (Grand et al., 1997; Van der Hilst et al., 1997; Van der Hilst & Karason, 1999) and from the presence of an anticorrelation between P-wave and S-wave velocity at this depth (Su & Dziewonski, 1997). The presence of a mid-mantle transition has not been confirmed by targeted seismic investigations (Vidale & Schubert, 2001; Vasco et al., 2003). However, Kellogg et al. (1999) developed a geochemical model that has a thermal boundary layer in mid-mantle, which is strongly varying in depth and remains invisible because compositional changes compensate for the higher temperature in keeping seismic velocities approximately the same. Laboratory experiments by Davaille (1999) show that thin plumes arise from such a deep layer. However, this is not what we observe: the minimum radii listed in Table 6.1 are inconsistent with Davaille's results. Tackley's model (Tackley, 2000), identifies the deep layer with the location of the superplumes, which is not where we observe the mid-mantle source regions. The most likely explanation for the plumes originating in mid-mantle is that they are narrower than 200 km in radius in the lowermost mantle, and/or that the velocity contrast is less than 0.4% (which corresponds to a temperature anomaly of about 130°K (Karato, 1993)), making them not or only weakly visible.

Another surprising observation is the lack of correlation between the depth of the source region and the $^3He/^4He$ anomaly. Of the ocean islands with typically high $^3He/^4He$ (Courtillot et al., 2003), Easter, Hawaii, Kerguelen, Samoa and Tahiti are identified as deep plumes in Table 6.1. Afar, Cape Verde, Caroline, Reunion and Jean Fernandez are potential D" plumes and again have correspondent high $^3He/^4He$. Yet others (Galapagos, Iceland) have high $^3He/^4He$ while a deep origin can be ruled out. Conversely, St Helena and Canary Islands have a low $^3He/^4He$ ratio but reach deeply into the mantle.

The maximum anomaly of our observed plumes in the lower mantle is reduced by a factor of about 3 w.r.t. the value found in the upper mantle (which is often in excess of -1.5%); even though the magnitude of each plume anomaly is affected by the resolution, the ratio is in accordance with predictions of dV_P/dT (Karato, 1993).

We are planning a study similar to this one, but targeted to resolve V_S anomalies using the same improvements, that will help us to constrain the physical characteristics of mantle plumes even further.

Chapter 7

Summary and Afterwords

Le travail de cette thése a montré que les phénomènes de diffraction sont importants et doivent être bien pris en compte quand on inverse des données de temps déduites de sismogrammes à longue période. Les améliorations apportées par l'introduction d'un maillage irrégulier, la combinaison d'une selection de très haute qualité de données à courte période et aussi à longue période ont permis, pour la première fois, l'identification des panaches profonds dans le manteau au-dessous d'Ascension, Azores, Canaries, Easter, Hawaï, Samoa et Tahiti. Les images montrent plusieurs points chauds inconnus sous la dorsale medio-Atlantique ou au milieu de l'Océan Indien et au-dessous des Seychelles. Plusieurs autres ont une origine plus superficielle, très probablement à une profondeur d'environ 660 km.

Les resultats obtenus dans cette thèse ont donc un fort impact dans notre compréhension géodynamique. Des évidences de pétrologie, de géochimie et de géodynamique ont soutenu l'existance des panaches profonds dans le manteau depuis les premiers travaux de Wilson (1963, 1965) et Morgan (1971, 1972). Les panaches sont observés dans les expériences de laboratoire (Davaille, 1999; Davaille et al., 2002). Mais la tomographie sismique était simplement incapable de fournir une confirmation complémentaire de la présence de ces structures relativement fines dans le manteau. Certaines études régionales suggéraient l'existence des panaches au-dessous de plusieurs points chauds, mais avec un très large degré d'incertitude. Par exemple, Bijwaard & Spakman (1999) obtient au-dessous du point chaud de l'Islande un panache profond montant de la frontiére entre le manteau et le noyau; cette découverte n'est pas confirmée par le travail tomographique de Ritsema et al. (1999) ou par des évidences indirectes de Foulger & Pearson (2001),Foulger et al. (2001), Foulger (2003). Nos résultats ne confirme pas aussi cette provenance profonde.

À cause de ce manque d'évidences claires fournies par la tomographie sismologique, trente années après sa première formulation par W. J. Morgan, l'hypothèse des panaches profonds n'est pas encore universellement acceptée. Espérons que les résultats de cette thèse mettront un point final à trente années de spéculations et d'hypothèses sur l'origine des points chauds et permettront de fournir une meilleure compréhension de la structure thermique et chimique de la Terre et de

la dynamique de son manteau.

Ce travail n'est pas complet mais beaucoup reste à faire. En particulier, il faut mieux comprendre la provenance du décalage temporel systématique présent dans les données à longue période dont nous avons tenu compte dans notre travail d'une manière ad-hoc mais dont l'origine n'est pas encore comprise.

Des améliorations sont aussi possibles dans la paramétrisation du modèle. Même si le maillage contruit s'est montré performant pour le but assigné, on peut faire encore mieux en introduisant des maillages contruits en tenant compte de la vraie résolution du problème tomographique. Pour cela, il nous faut un expression analytique pour l'estimation de la résolution. Nolet et al. (1999) ont dérivé une expression pour le calcul de l'inverse approximé à utiliser pour l'estimation de la résolution. La description de la méthode est donnée dans l'appendice C. Il faudra utiliser cette méthode pour le calcul de la résolution. Cette résolution sera utilisée ensuite dans une méthode d'optimisation des maillages dans les futures études de tomographie.

Enfin, on sait très bien que le module de cisaillement est plus sensible aux effets de température que le module de rigidité. Il est possible ainsi de répéter l'étude décrite dans cette thèse avec les onde S. À cause de leur plus haute sensibilité à la température et leur région de Fresnel plus petite, on s'attend à obtenir des images des panaches encore plus détaillées.

The work described in this thesis has made clear that diffraction phenomena are very important and must be properly accounted for when inverting lower-frequency waves. The improvements brought about by the introduction of a heterogeneous model parameterization, and the combination of very high quality short - and long-period measurements have allowed, for the first time, the identification of deep-rooted mantle plumes beneath Ascension, Azores, Canary, Easter, Hawaii, Samoa and Tahiti. The images show several hitherto unknown plumes under the Atlantic Ridge, in the South Indian Ridge, and beneath the Seychelles. Many others plumes, Iceland among them, are of more shallow origin, most probably near the 660 km discontinuity.

These results are of tremendous impact in the scientific community. Evidences from petrology, geochemistry, geodynamics have supported, if not required, the existence of deep mantle plumes since the seminal works of Wilson (1963, 1965) and Morgan (1971, 1972). Plumes have been directly observed in laboratory experiments (Davaille, 1999; Davaille et al., 2002). But seismic tomography was so far unable to provide a visual confirmation of the presence of such narrow deep-rooted features in the Earth's mantle. Regional traveltime studies suggest the presence of plumes in the upper mantle beneath several hotspots such as Yellowstone (see Nataf (2000) for a review of these studies) and Iceland, but they are hampered by incertitude. For instance, Iceland appears as a shallow-seated anomaly in the to-

mography of Ritsema et al. (1999). A shallow origin was also argued from indirect evidence by Foulger & Pearson (2001), Foulger et al. (2001), Foulger (2003), and now from delay times by our own results; but Bijwaard & Spakman (1999) revealed a plume-like structure extending from the core-mantle boundary up to the surface.

Because of the absence of undisputed tomographic evidence, thirty years after its formulation Morgan's deep plume hypothesis has not yet been universally accepted, and authors such as Anderson and others (Anderson, 1998, 2000; Foulger & Natland, 2003) have argued that hotspots could as well form along a crack in the lithosphere that allows mantle rock from a few hundred kilometers down to rise to the surface and melt. Our results have the potential to reconciliate 30 years of speculations and observations related to hotspots and finally bring a better understanding of the thermal and chemical structure of the Earth's as well as of the dynamics of its mantle.

Clearly, this work has only just begun. In particular, the analysis of the long-period data have shown a variable offset, characteristic of each event and properly accounted for in our tomography study, but whose origin is not fully understood.

The model parameterization we have used was designed to account for the expected resolving length of our data, with node's spacing of about 200 km in the upper mantle and about 600 km in the lower mantle. Although this grid has shown to be sufficient for our purpose, we can do better by adapting the model parameterization to the true resolution of the tomographic image. This requires an analytical expression for computing the resolution, which provides us with a measure on how our solution is close to the real Earth. For global tomographic inverse problems, attempt at any formal estimate of the resolution is unfeasible. With the exception of Vasco et al. (1999, 2003), no assessment of the resolution of a large global tomographic system involving singular value decomposition has been performed. Nolet et al. (1999) derived an explicit expression for the approximate inverse matrix for use in the estimation of the resolution of very large tomographic systems, that are too large to be solved by singular value decomposition. The description of the method with numerical tests that I performed while at Géosciences Azur in Valbonne (France) are presented in Appendix C. We are currently developing new techniques to re-design grids in term of a given functional characteristic of that grid. We plan to use these techniques with the estimates for the resolution to generate an optimized model parameterization to use in future global tomographic studies.

Finally, it is well known that the shear modulus is more sensitive to the effects of temperature then the bulk modulus. We plan to repeat the study described in this thesis for the shear waves. Because of the higher sensitivity to temperature and their narrower Fresnel zones, we expect this to provide an even better image of the plume-like structures revealed by our P wave velocity model. Also, discrepancies between P and S wave velocity are useful to infer the thermal and chemical structure of the Earth. Shear velocity and compressional velocity are strongly cor-

related throughout the mantle in many tomographic studies, with the exception of the north Pacific region where the P wave anomaly tends to be positive while the S wave anomaly tends to be negative in the lowermost mantle. Joint inversions have been performed to isolate the relative behavior of the bulk sound speed with respect to the shear speed and found an anticorrelation between the two in the lowermost mantle below 1500-2000 km (Su & Dziewonski, 1997; Kennett et al., 1998; Vasco & Johnson, 1998; Masters et al., 2000; Oganov et al., 2001; Saltzer et al., 2001; Ritsema & van Heijst, 2002) for which they invoke the presence of a thermal-chemical boundary layer at the base of the mantle (Lay et al., 2003). Stacey (1998) studies the effect of mixed mineralogy on the pressure and temperature dependences of the bulk modulus and also concluded that a simple thermal interpretation of the velocity anomalies observed at the base of the mantle is not enough and that compositional variations is required as well. These studies tend not to agree on the depth and amplitude of the anticorrelation (Masters et al., 2000; Saltzer et al., 2001; Ritsema & van Heijst, 2002) as well as on the ratio between shear and compressional velocities. In chapter 4 we inverted the same long period P data set with both ray theory and finite frequency, and found that finite-frequency modeling increases the amplitudes of the velocity perturbations. S waves have smaller wavelength that P waves and therefore the banana-doughnut kernel is relatively smaller (1000 km for the S while 1400 km for a P, both of 80°) with the region of insensitivity around the ray-theoretical path also smaller (150 km for the S wave and 300 km for the P wave). This implies that wavefront healing effects tends to be smaller for shear waves than for compressional waves. The comparison of P and S wave anomalies obtained with finite-frequency modeling are expected to provide a better estimate of the anticorrelation found at the bottom of the mantle.

Appendix A

Additional color figures of the plumes

In this appendix, we provide a collection of color figures of the plumes. The sections at different depths are projected on a flat plane. They provide a better view of the location and continuity of these low velocity anomalies. Plotted is the fractional P wave velocity perturbation in % with respect to the *iasp91mod* model (see chapter 5).

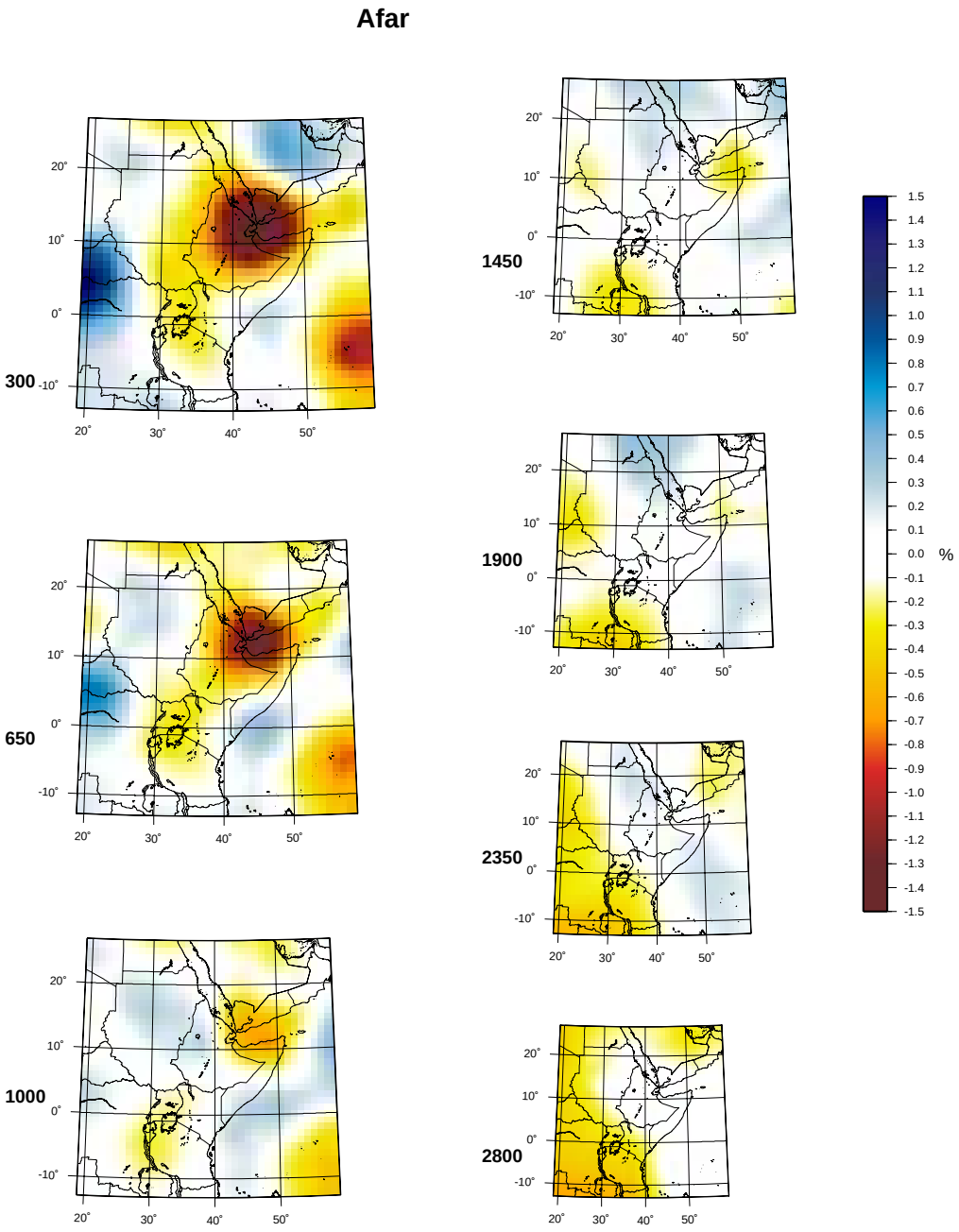


Figure A.1 Sections of the joint inversion velocity model at different depths (300, 650, 1000, 1450, 1900, 2350 and 2800 km) beneath Afar.

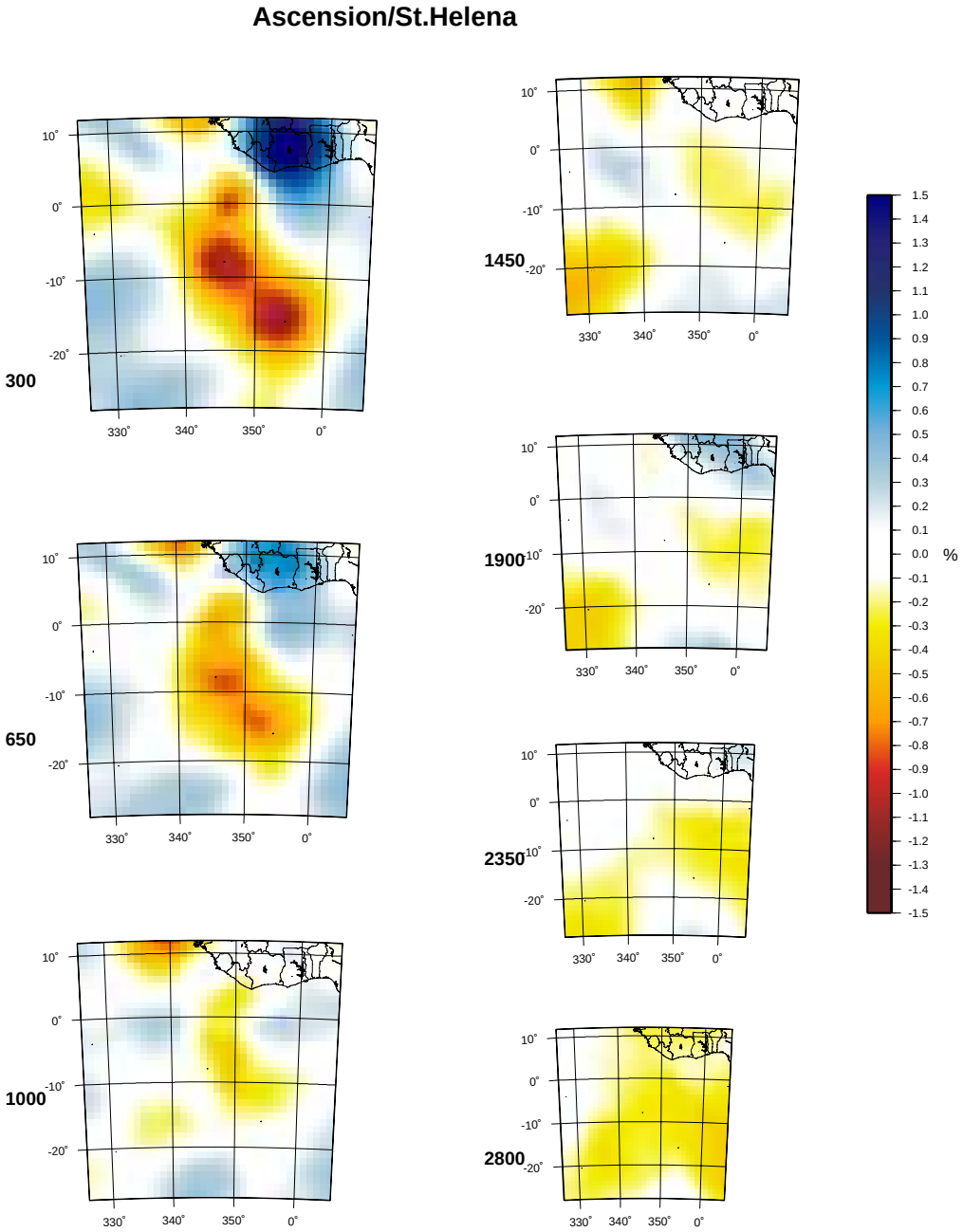


Figure A.2 Sections of the joint inversion velocity model beneath Ascension and St. Helena.

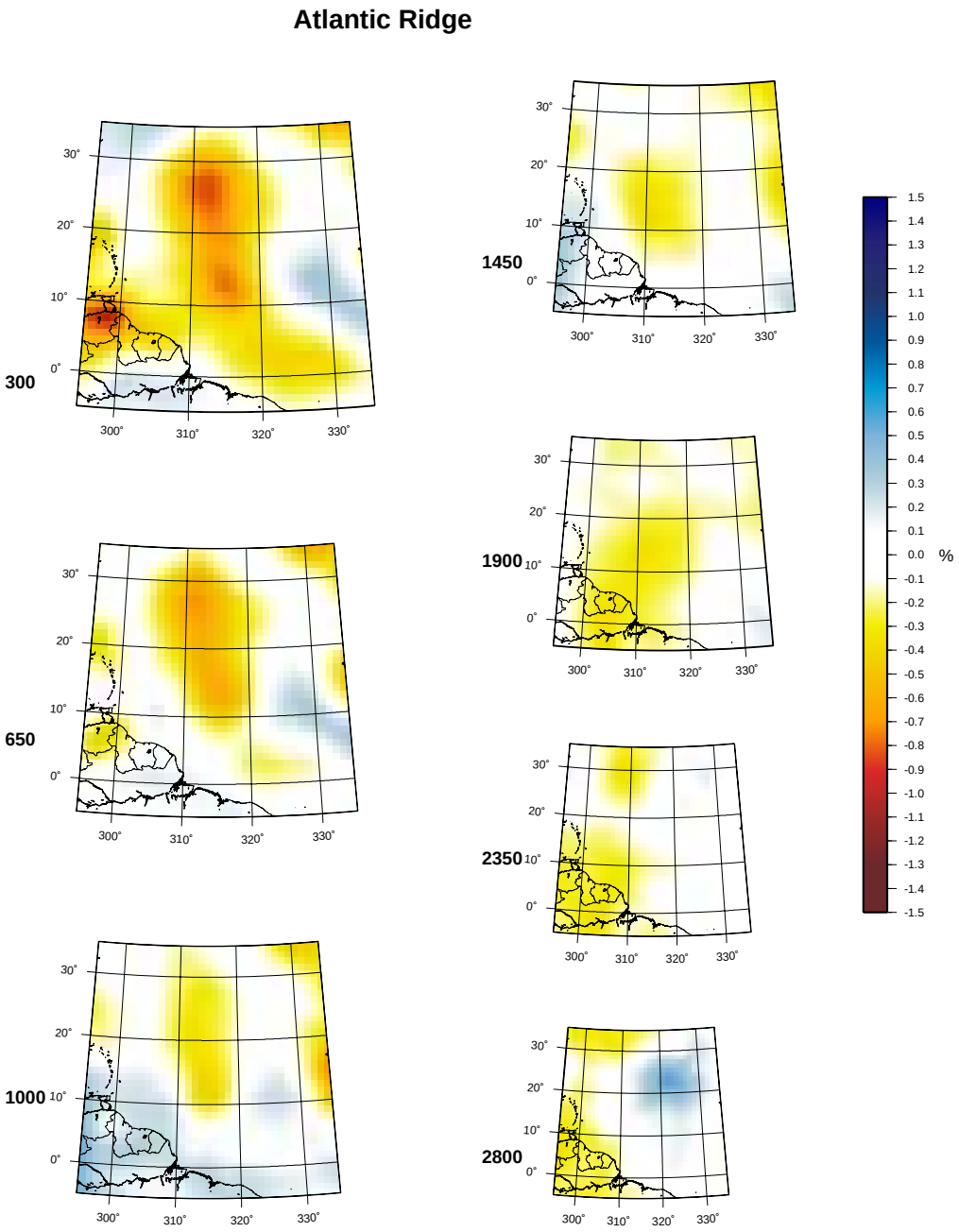


Figure A.3 Sections of the joint inversion velocity model beneath the newly discovered plume along the Atlantic Ridge.

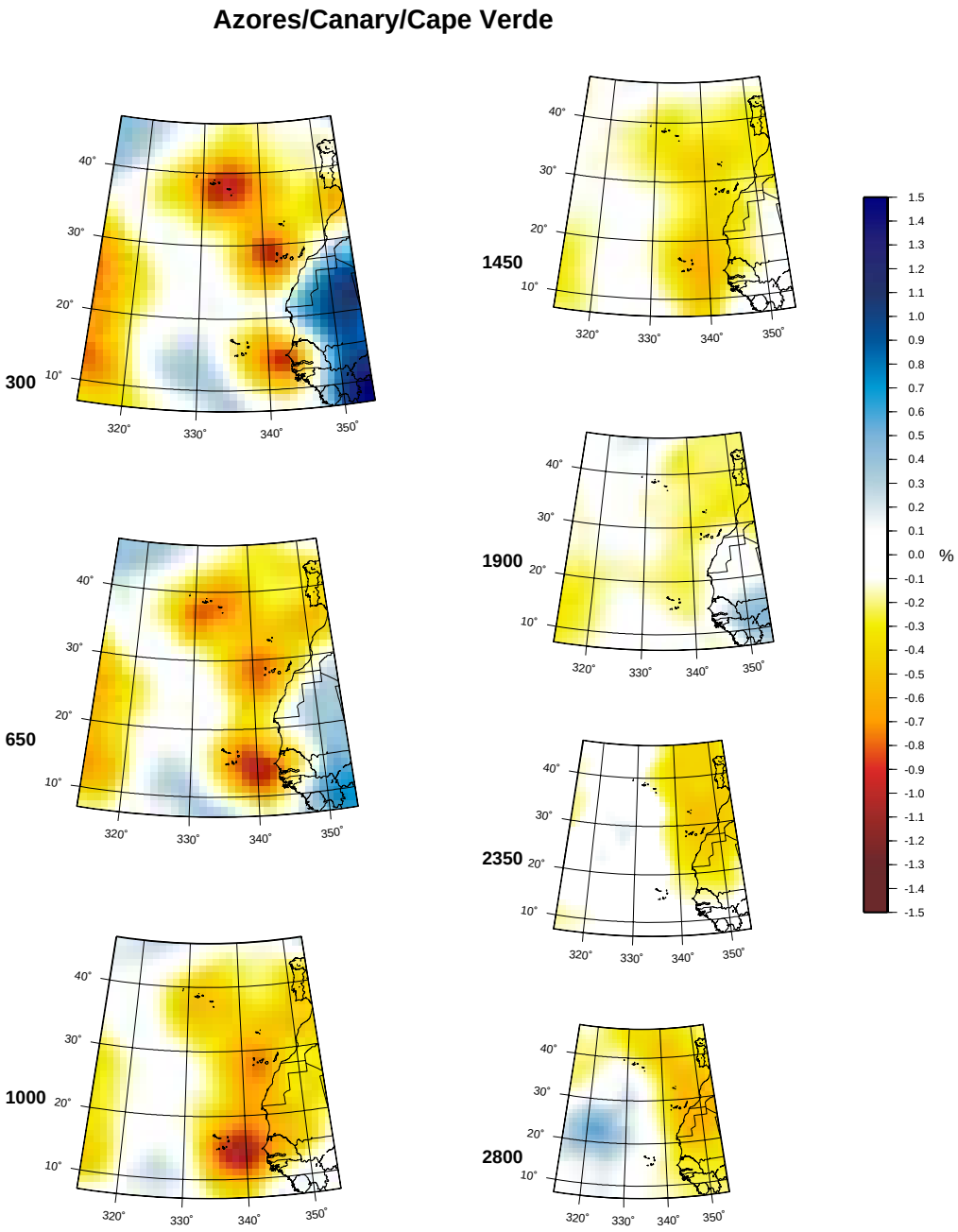


Figure A.4 Sections of the joint inversion velocity model beneath Azores, Canary and Cape Verde.

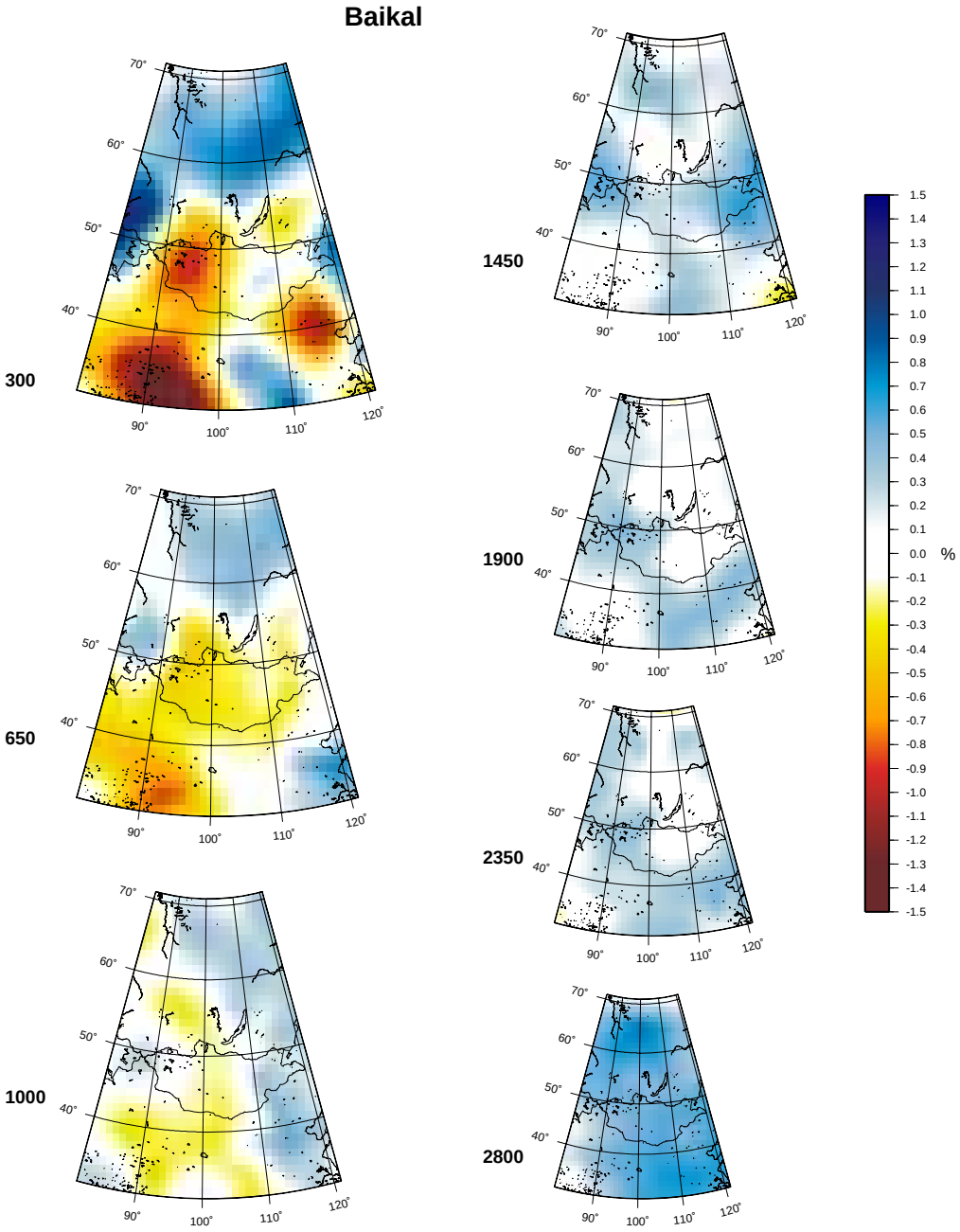


Figure A.5 Sections of the joint inversion velocity model beneath Baikal.

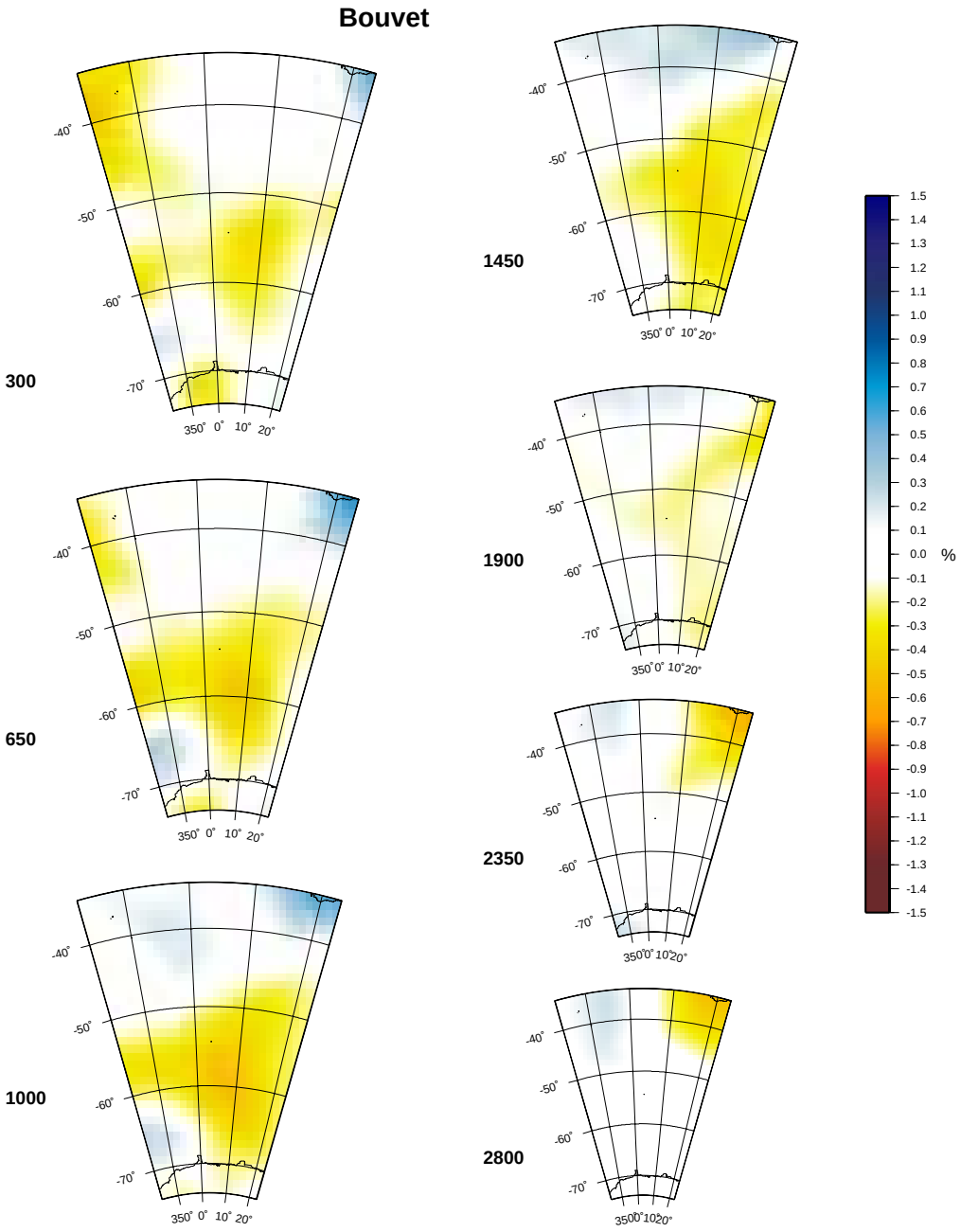


Figure A.6 Sections of the joint inversion velocity model beneath Bouvet.

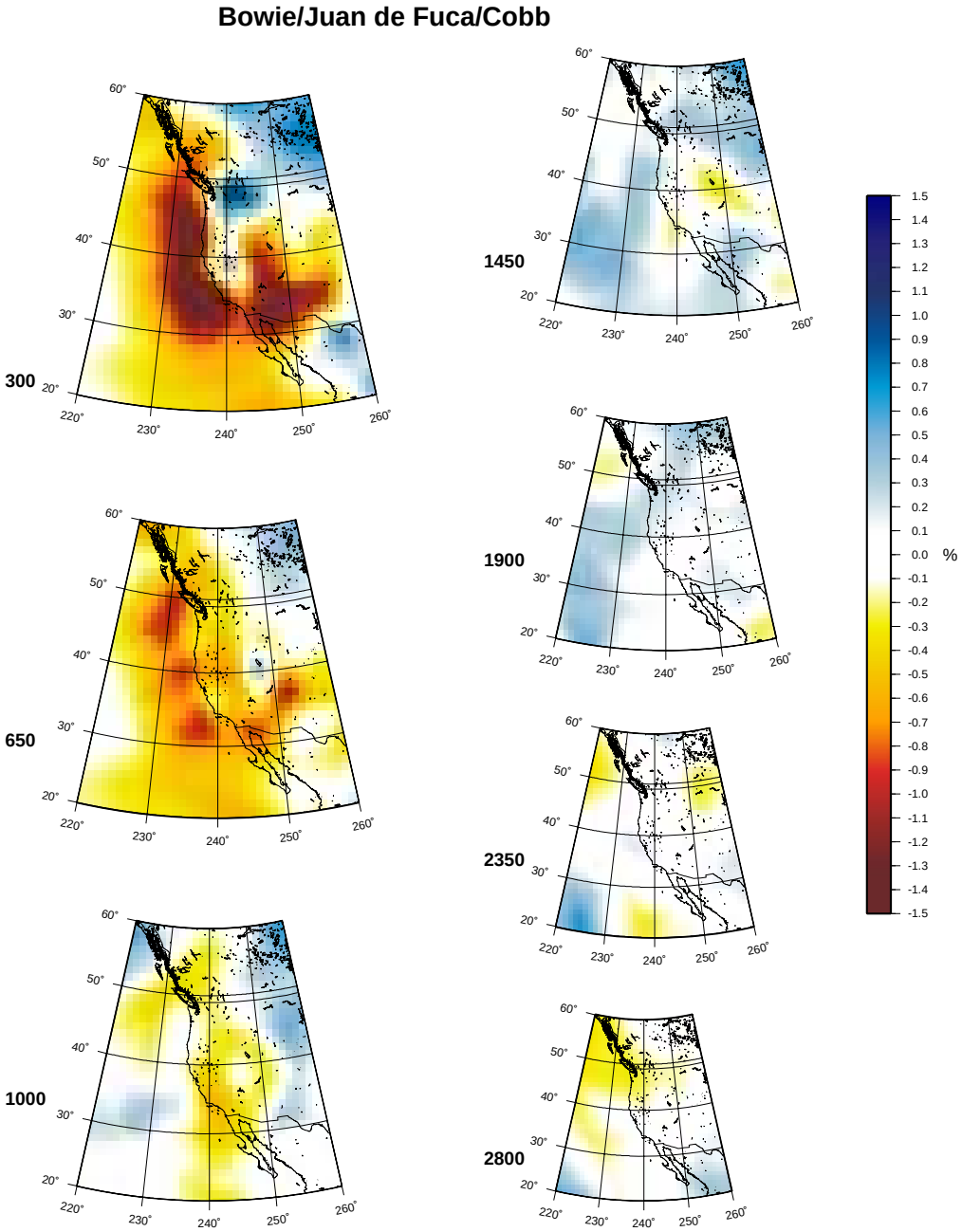


Figure A.7 Sections of the joint inversion velocity model beneath Bowie, Juan de Fuca, Cobb. Yellowstone hotspot is also in the map, but does not seem to have a plume.

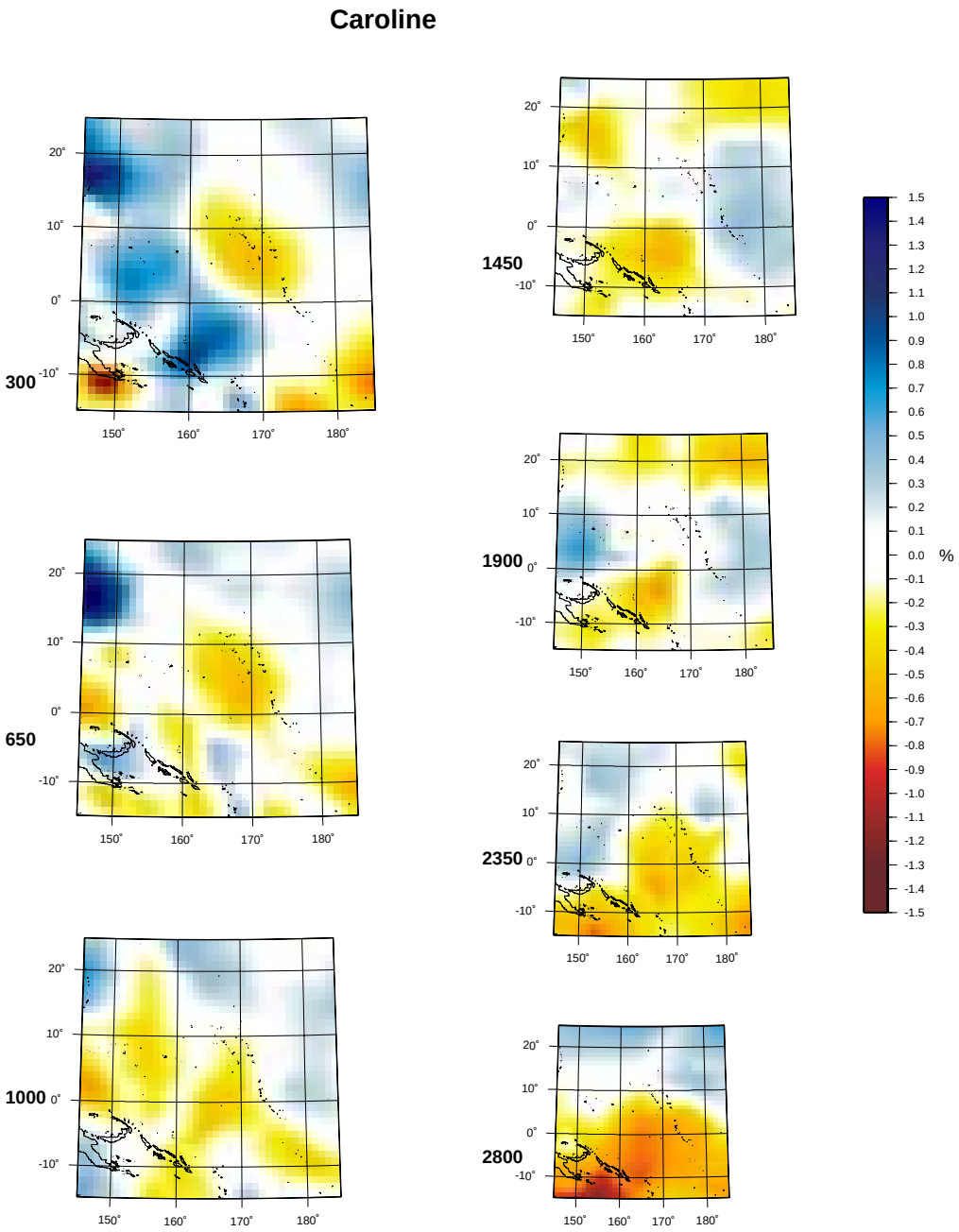


Figure A.8 Sections of the joint inversion velocity model beneath Caroline.

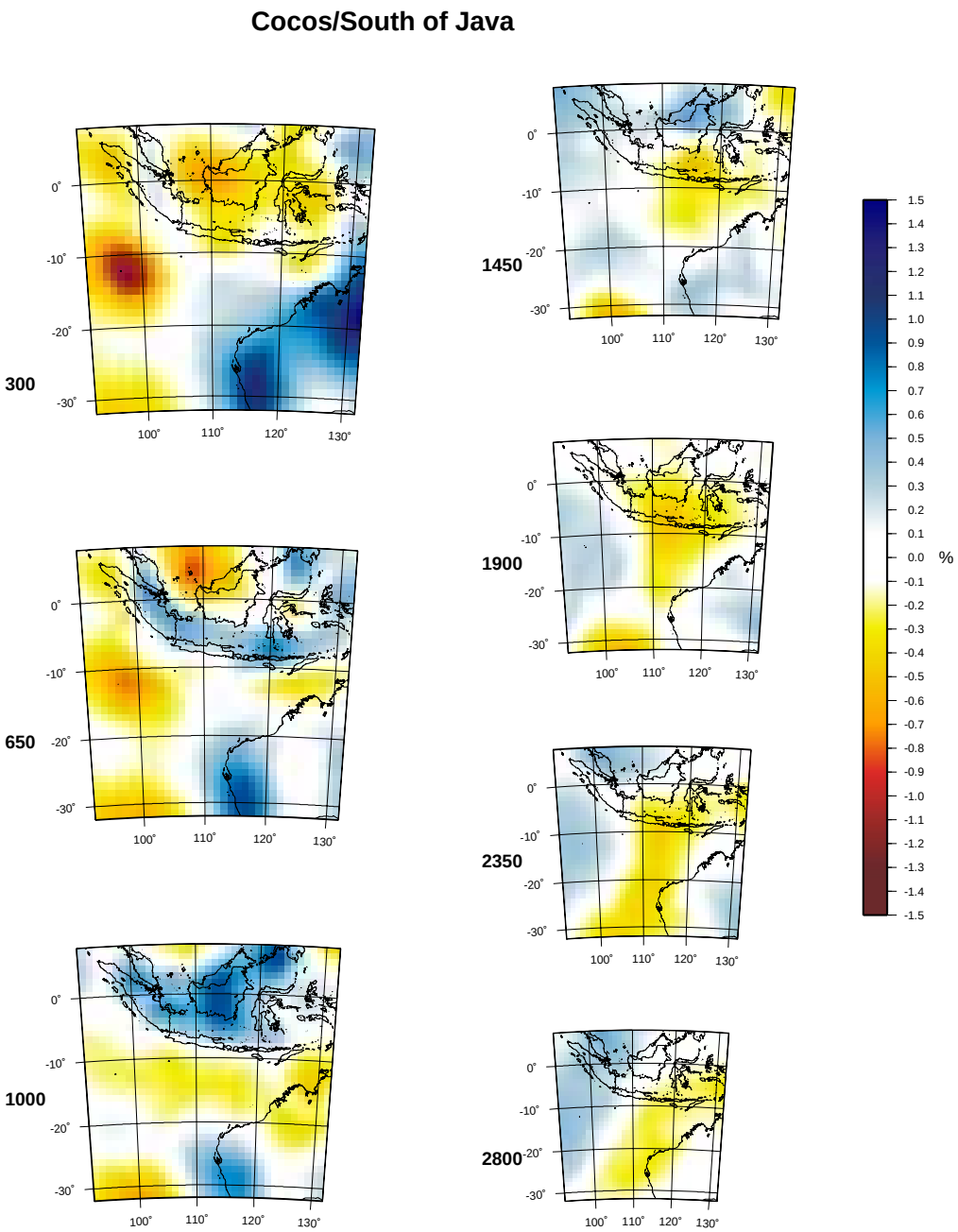


Figure A.9 Sections of the joint inversion velocity model beneath Cocos and South of Java.

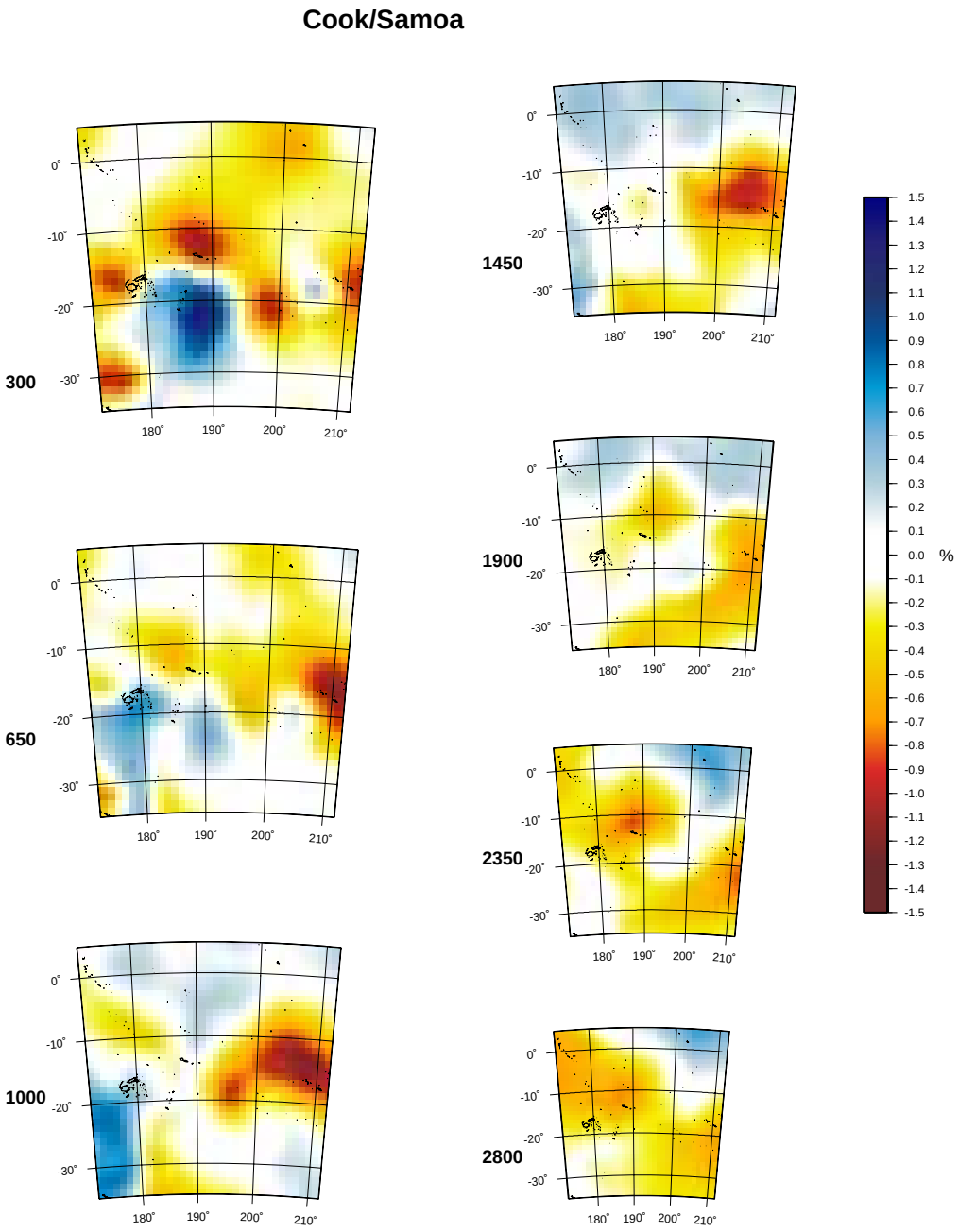


Figure A.10 Sections of the joint inversion velocity model beneath Cook and Samoa.

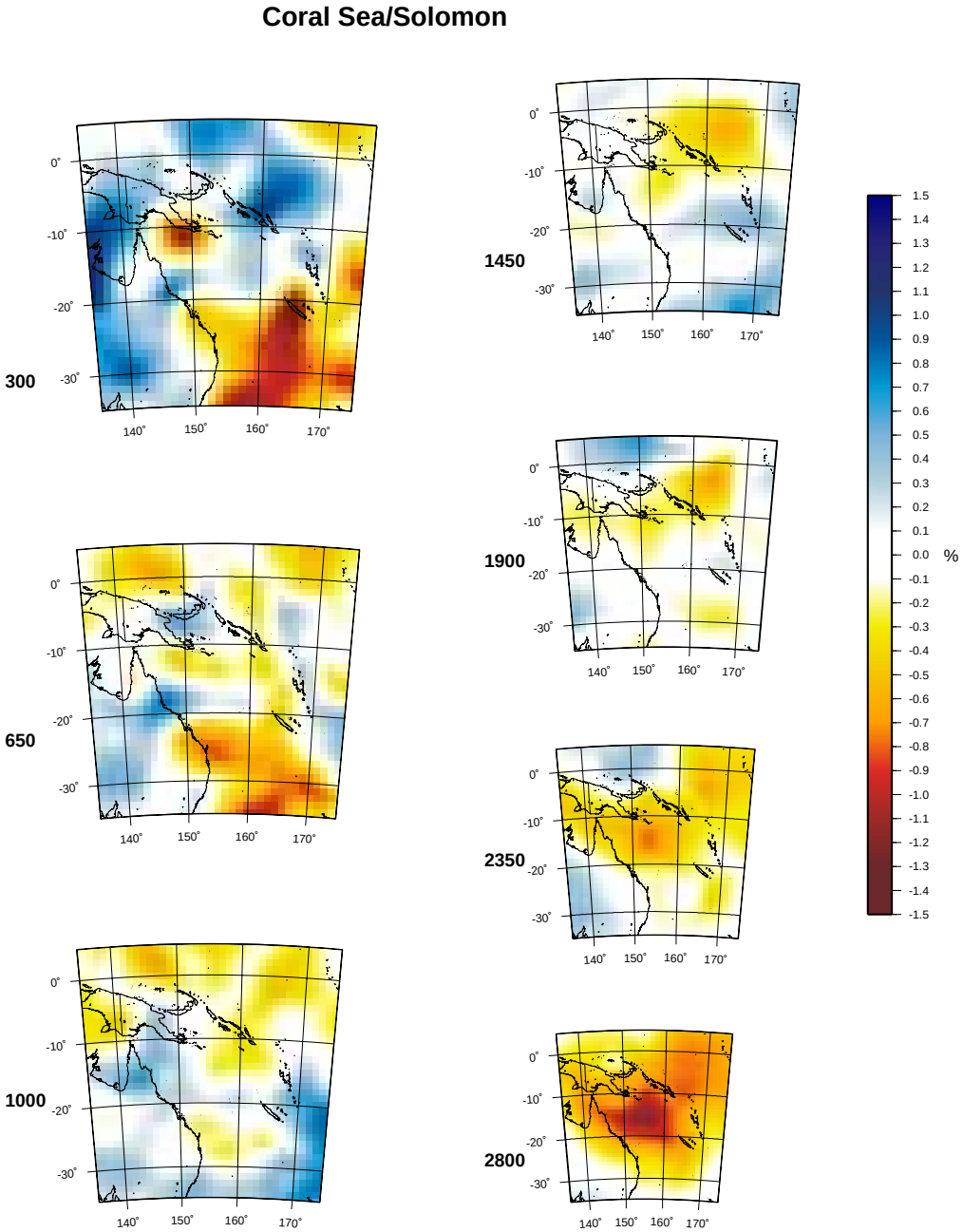


Figure A.11 Sections of the joint inversion velocity model beneath Coral Sea, Solomon area.

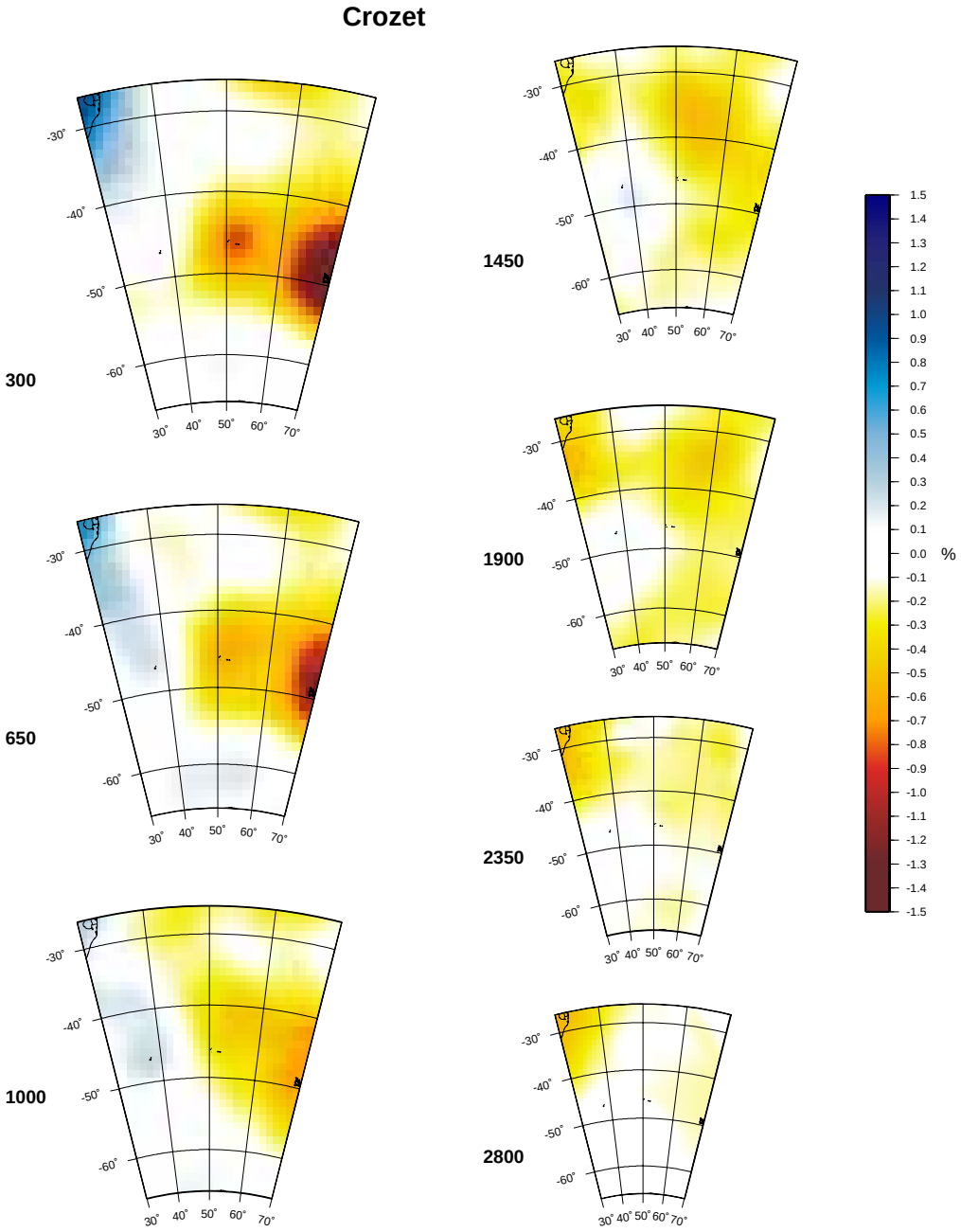


Figure A.12 Sections of the joint inversion velocity model beneath Crozet.

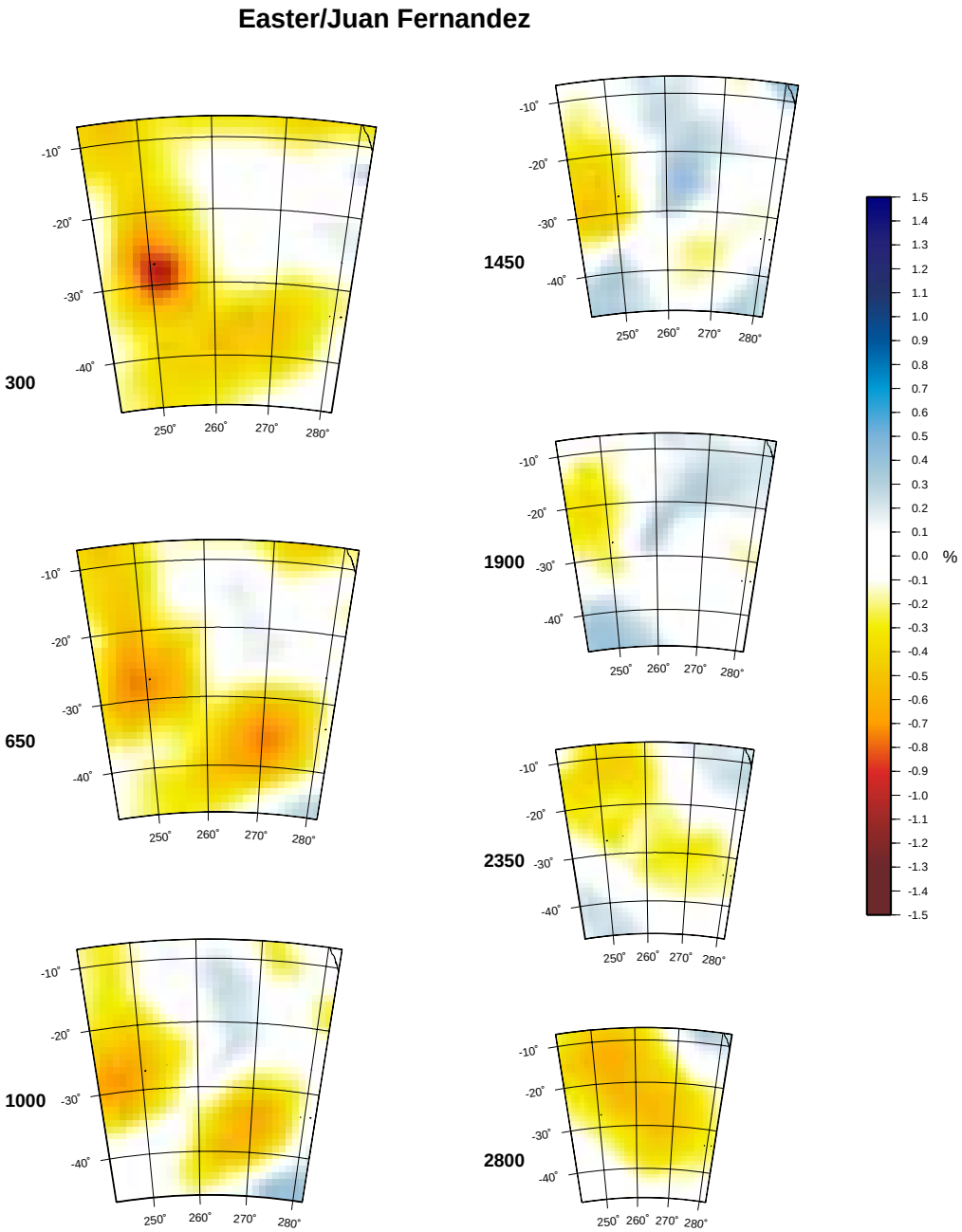


Figure A.13 Sections of the joint inversion velocity model beneath Easter and Juan Fernandez.

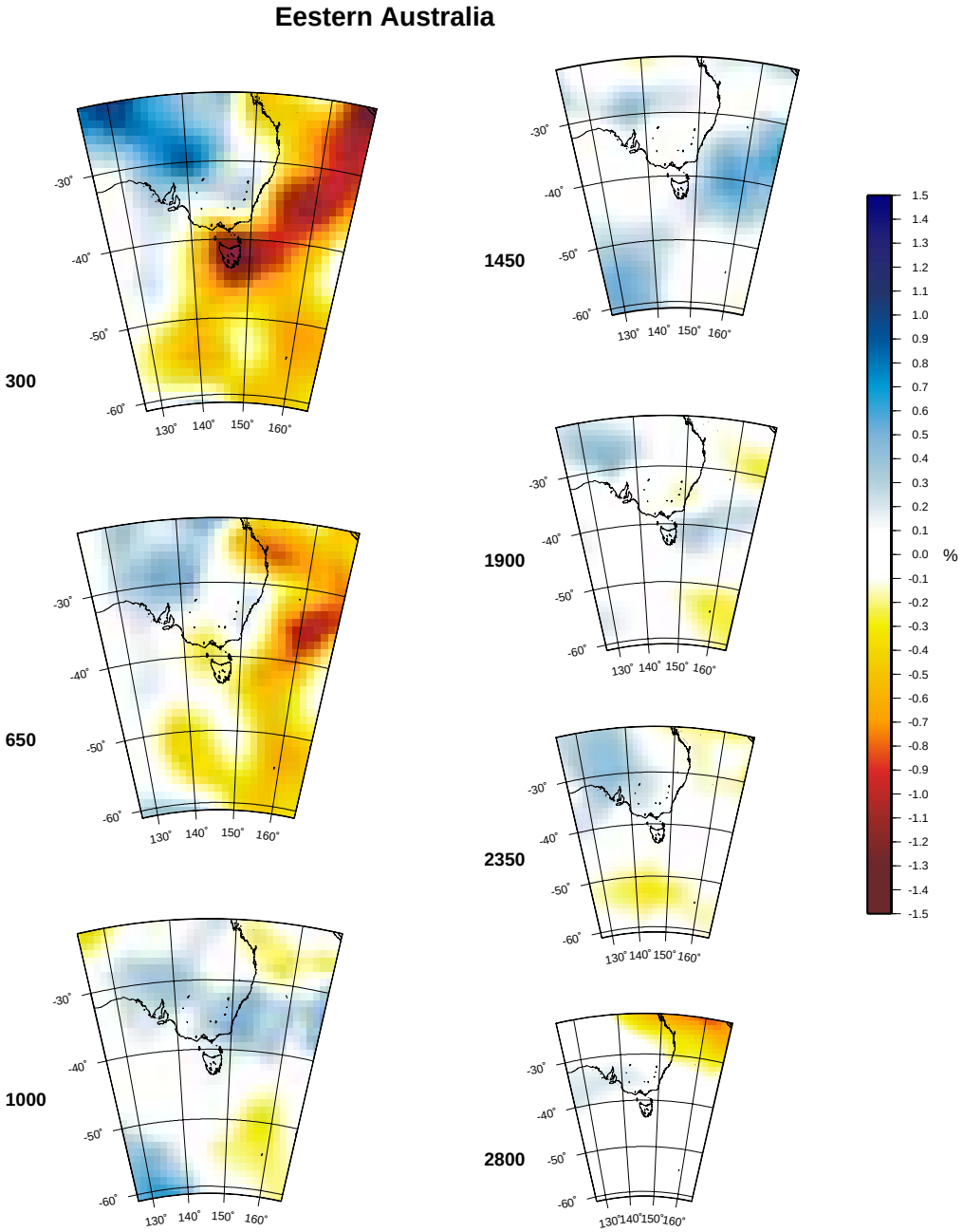


Figure A.14 Sections of the joint inversion velocity model beneath Eastern Australia.

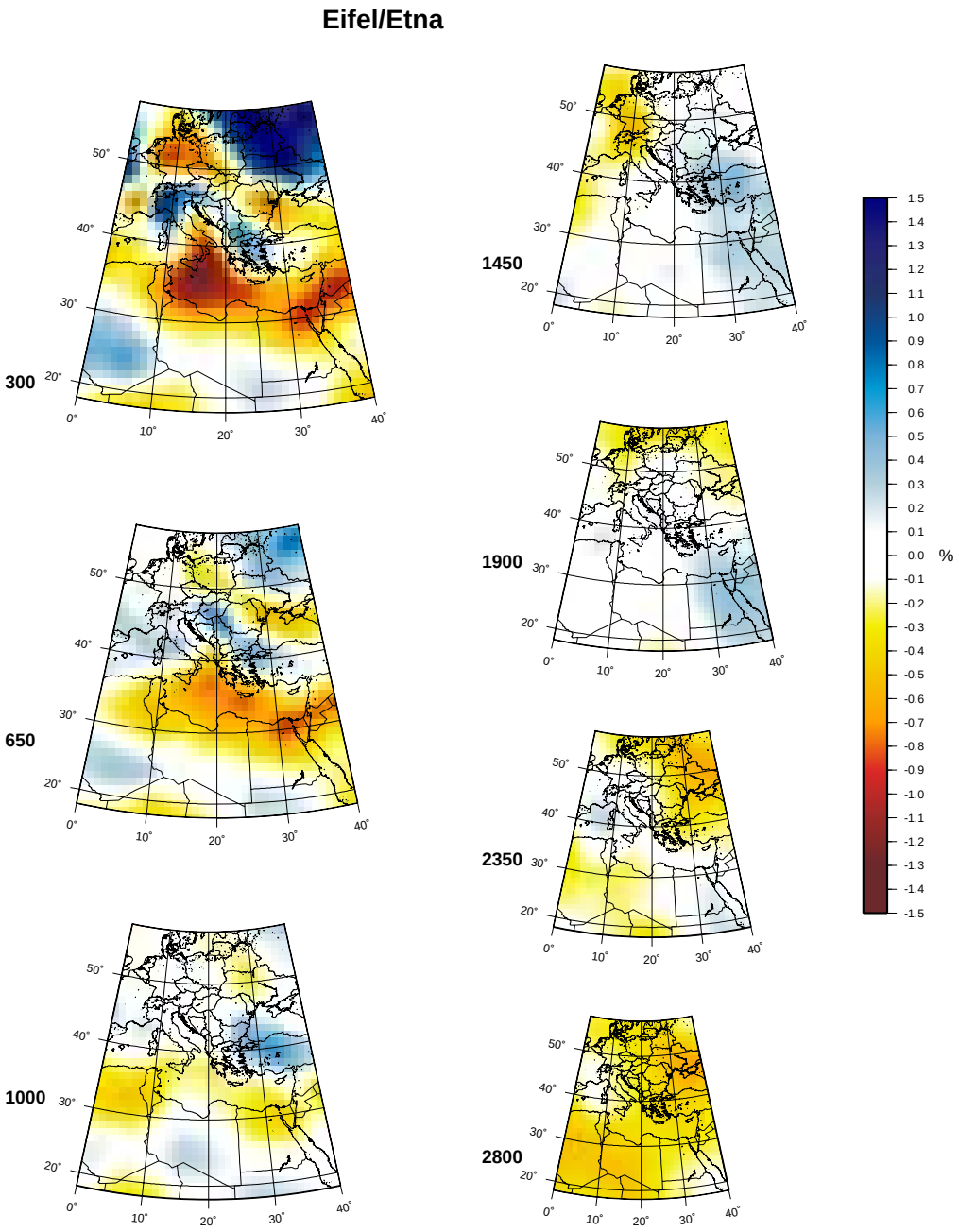


Figure A.15 Sections of the joint inversion velocity model beneath Europe. Visible in the map are Eifel and Etna.

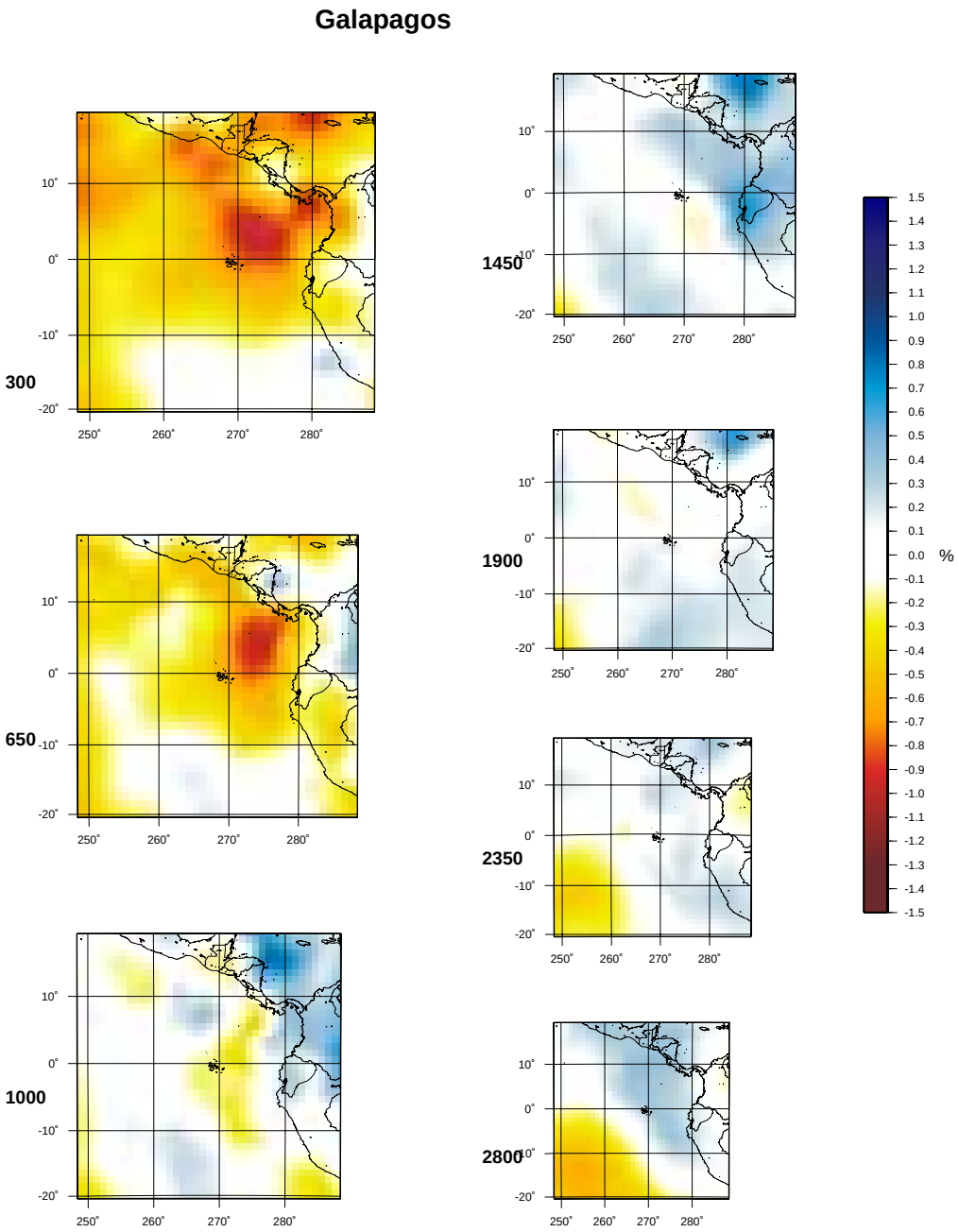


Figure A.16 Sections of the joint inversion velocity model beneath Galapagos.

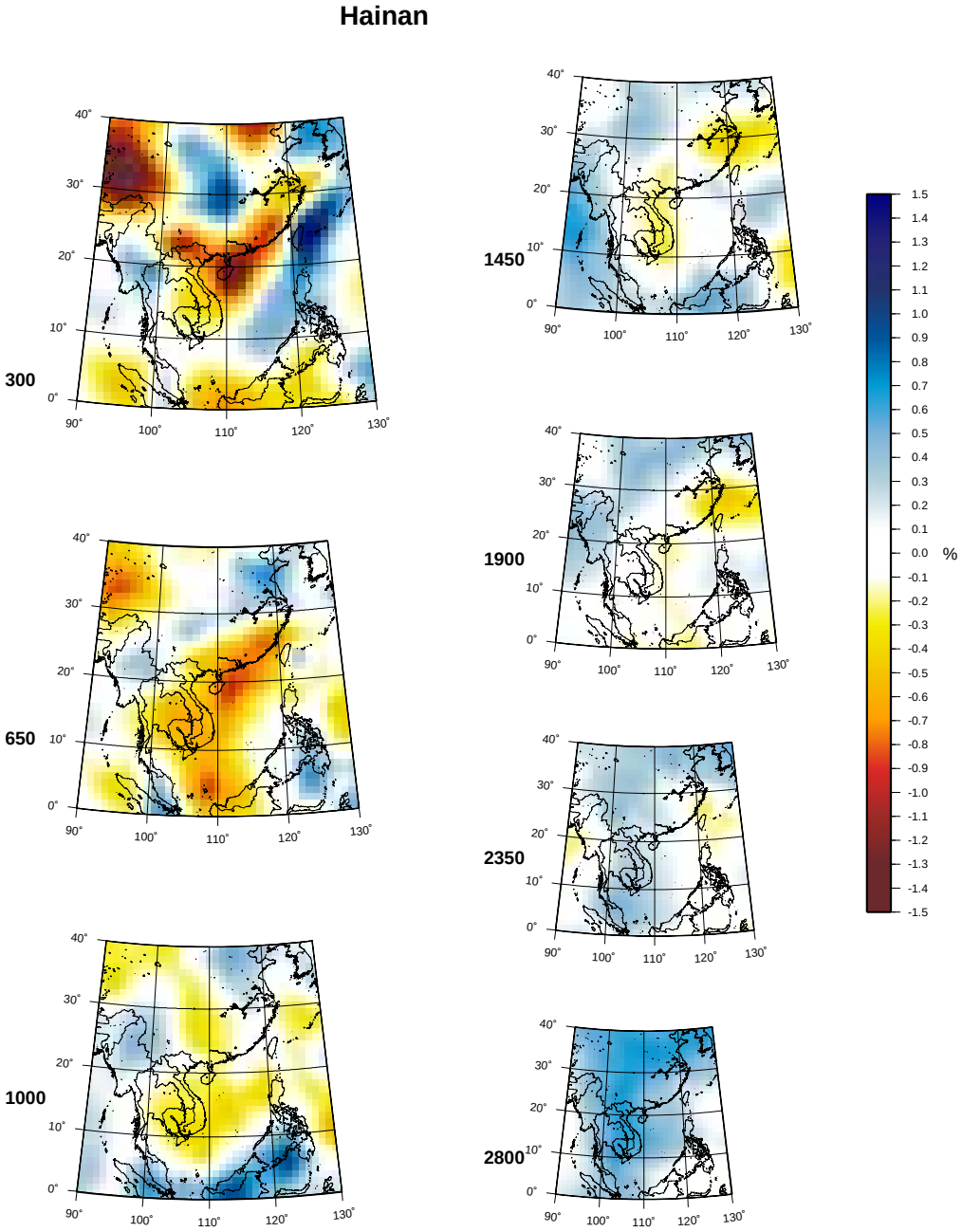


Figure A.17 Sections of the joint inversion velocity model beneath Hainan.

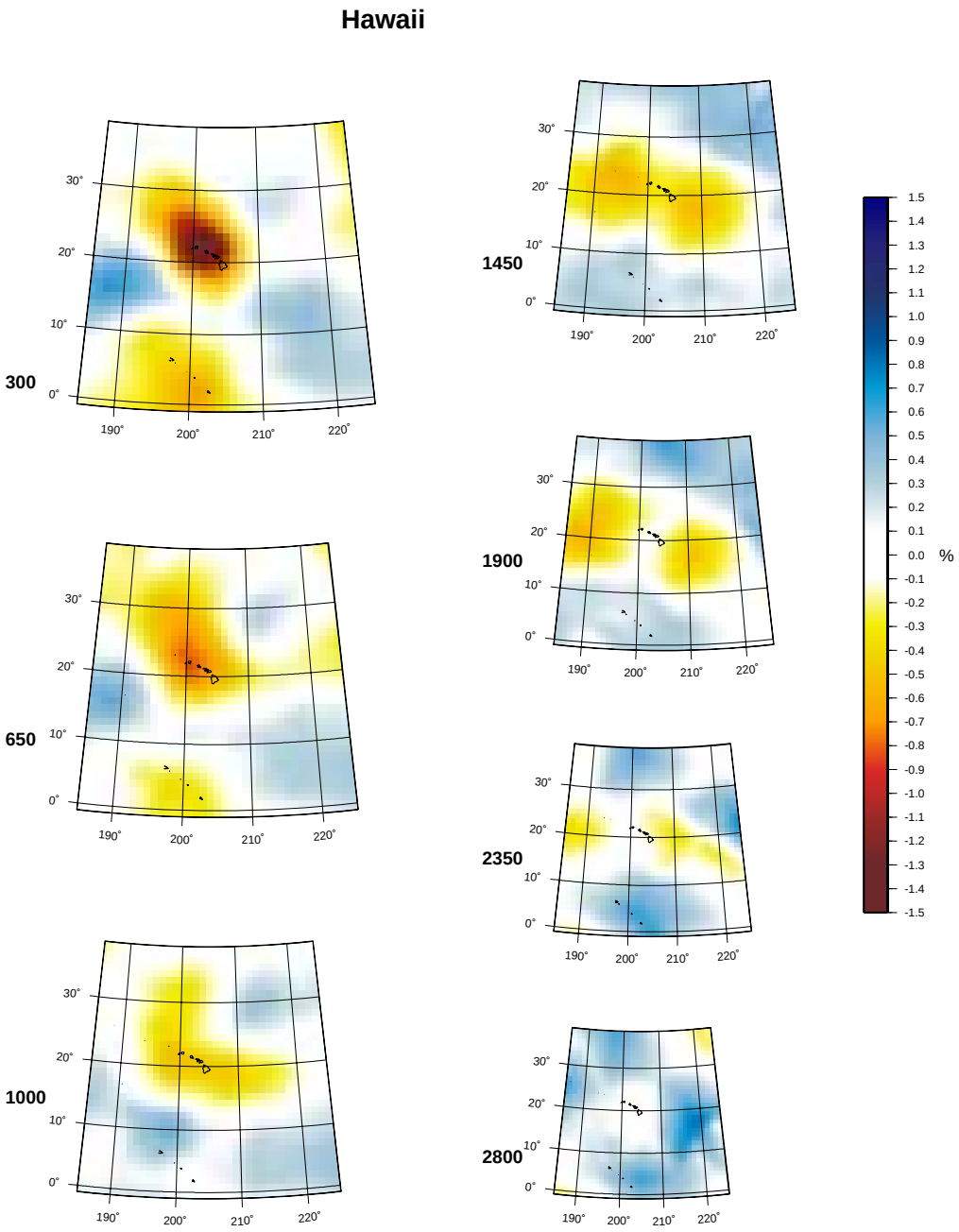


Figure A.18 Sections of the joint inversion velocity model beneath Hawaii.

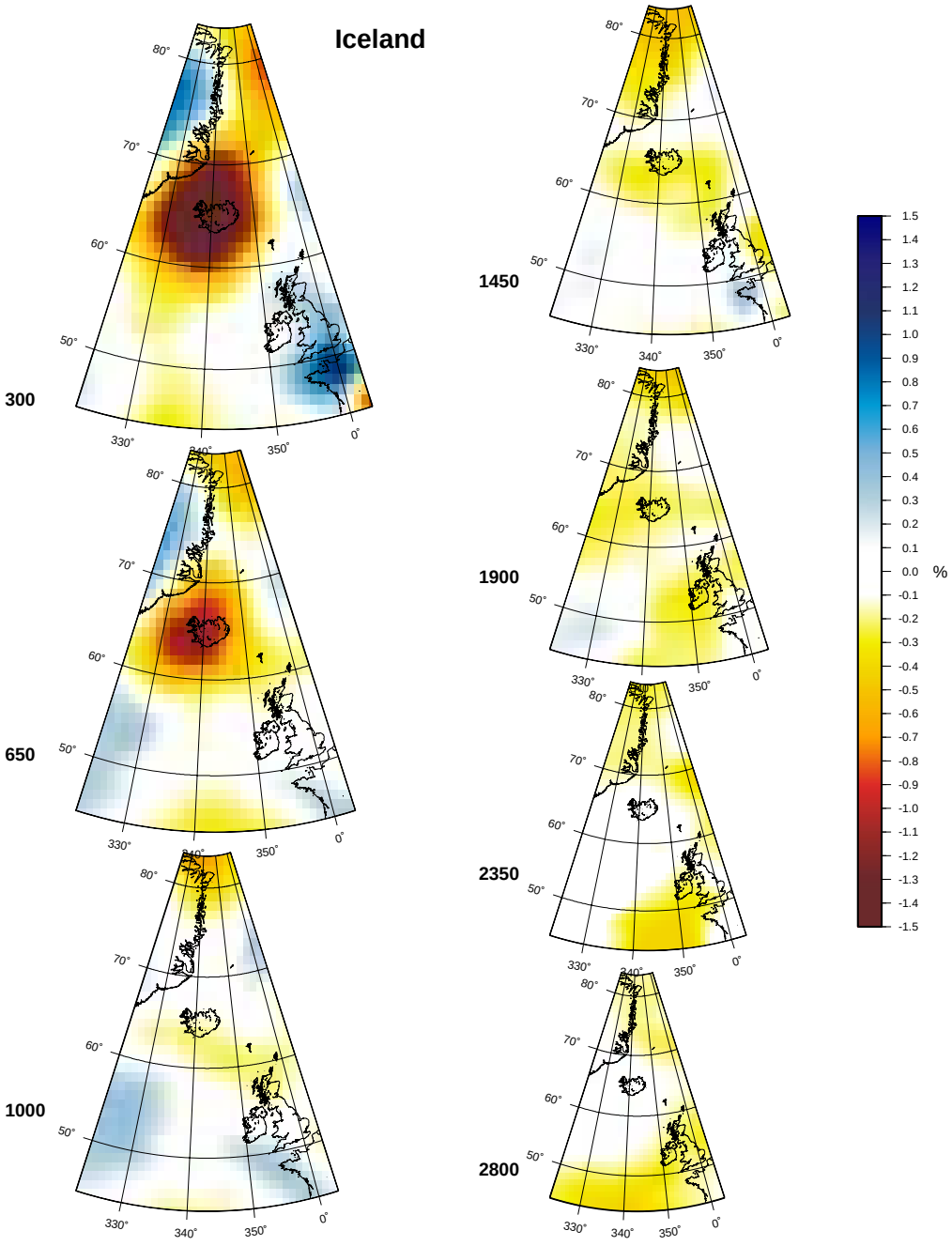


Figure A.19 Sections of the joint inversion velocity model beneath Iceland.

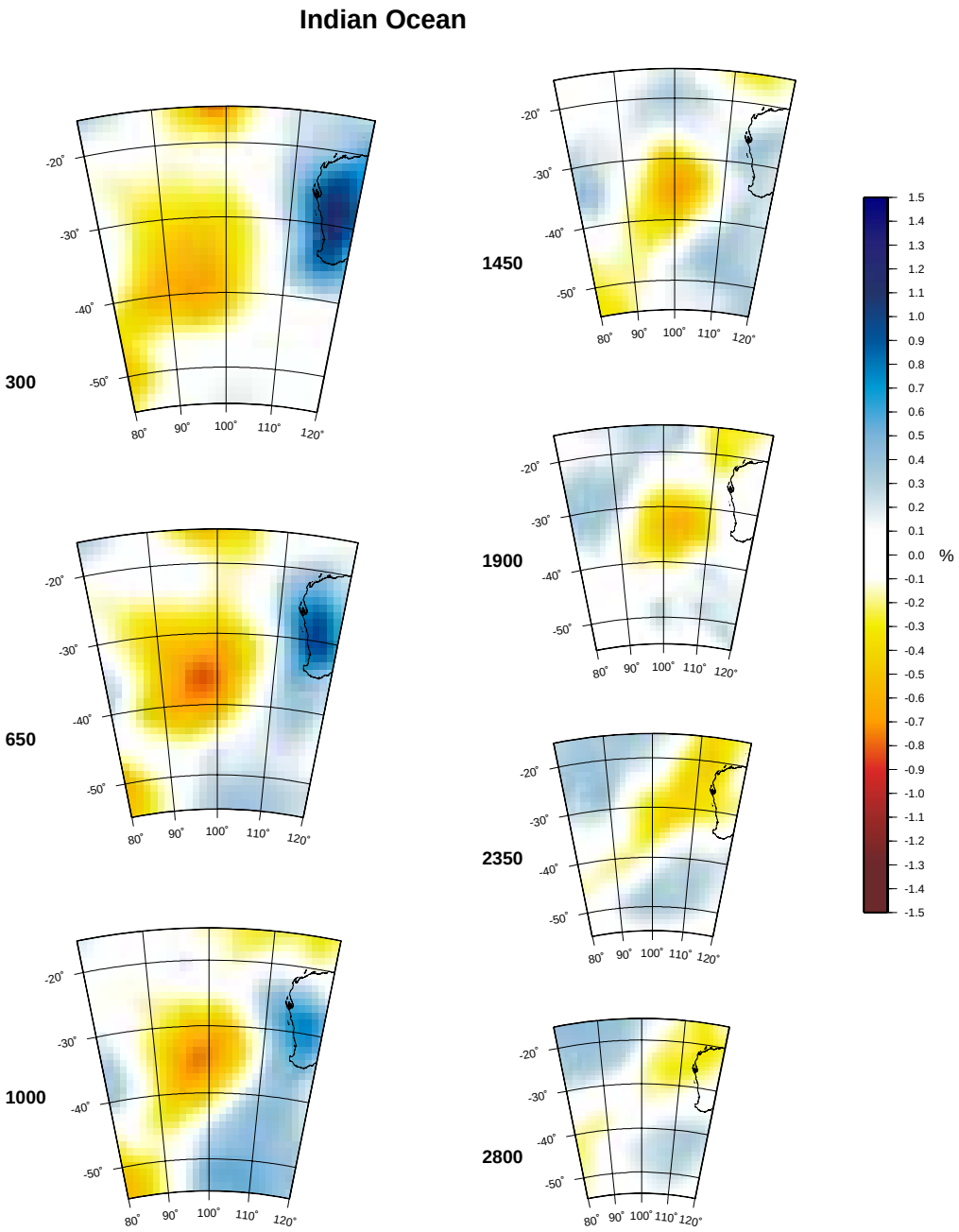


Figure A.20 Sections of the joint inversion velocity model beneath Indian Ocean.

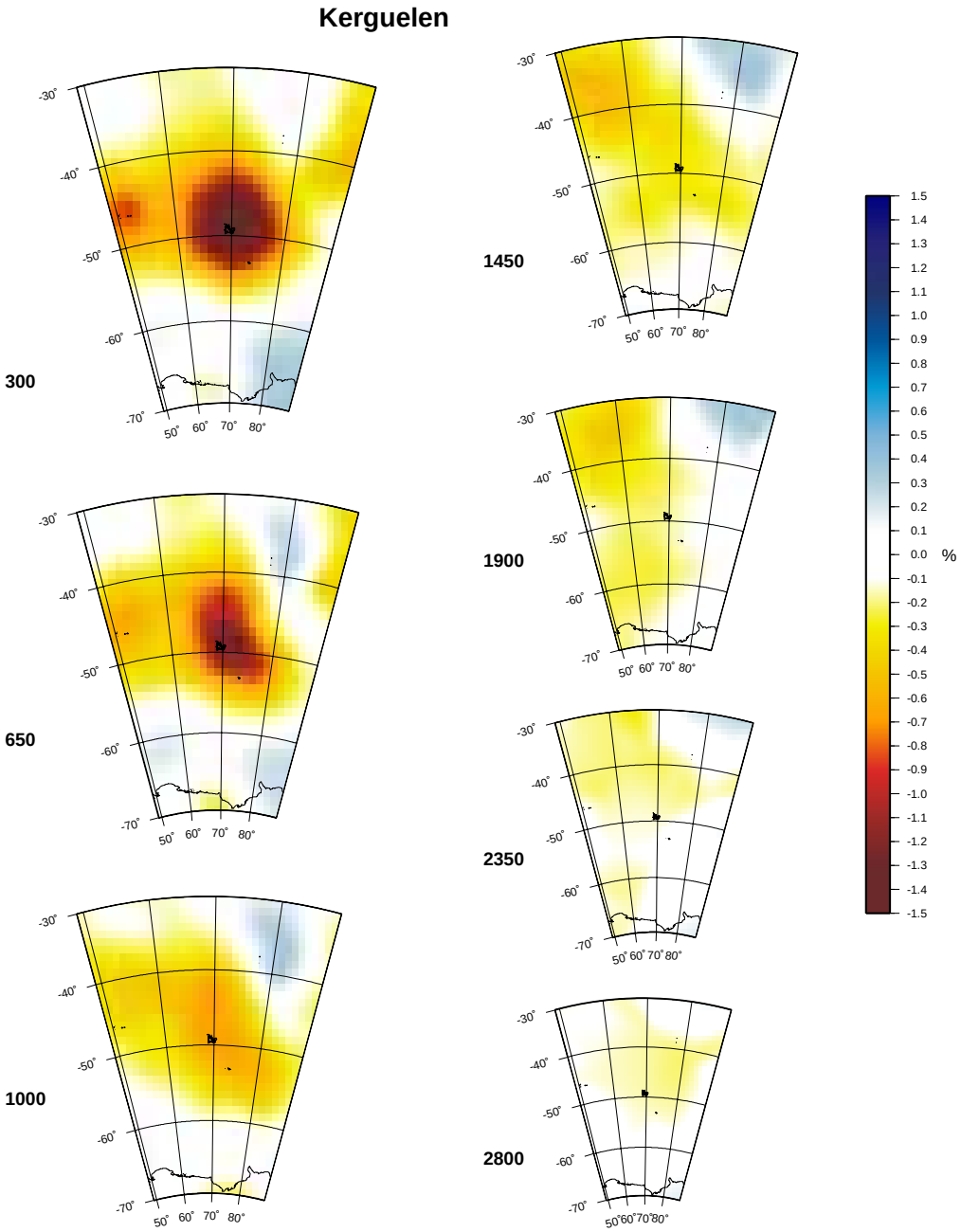


Figure A.21 Sections of the joint inversion velocity model beneath Kerguelen.

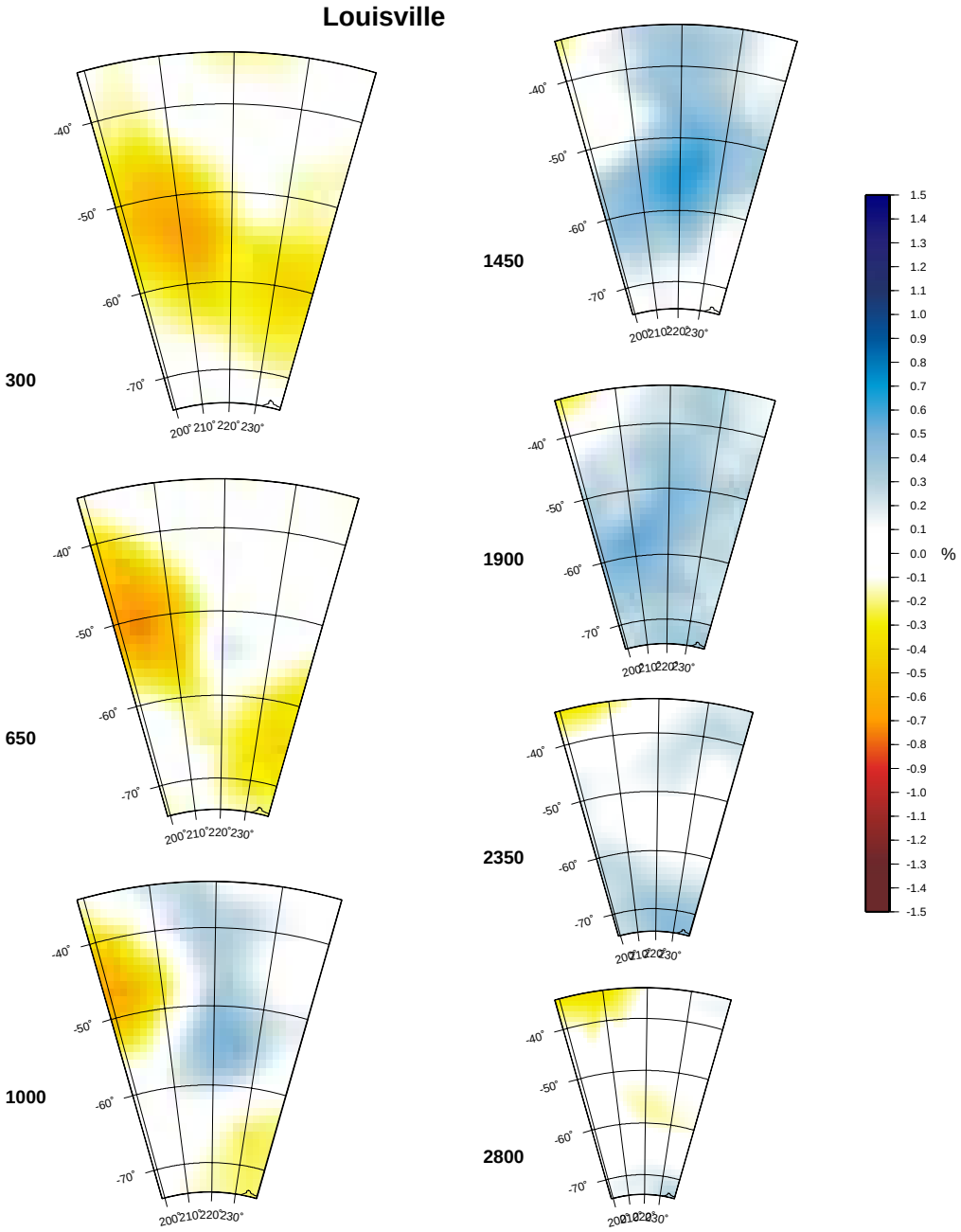


Figure A.22 Sections of the joint inversion velocity model beneath Louisville.

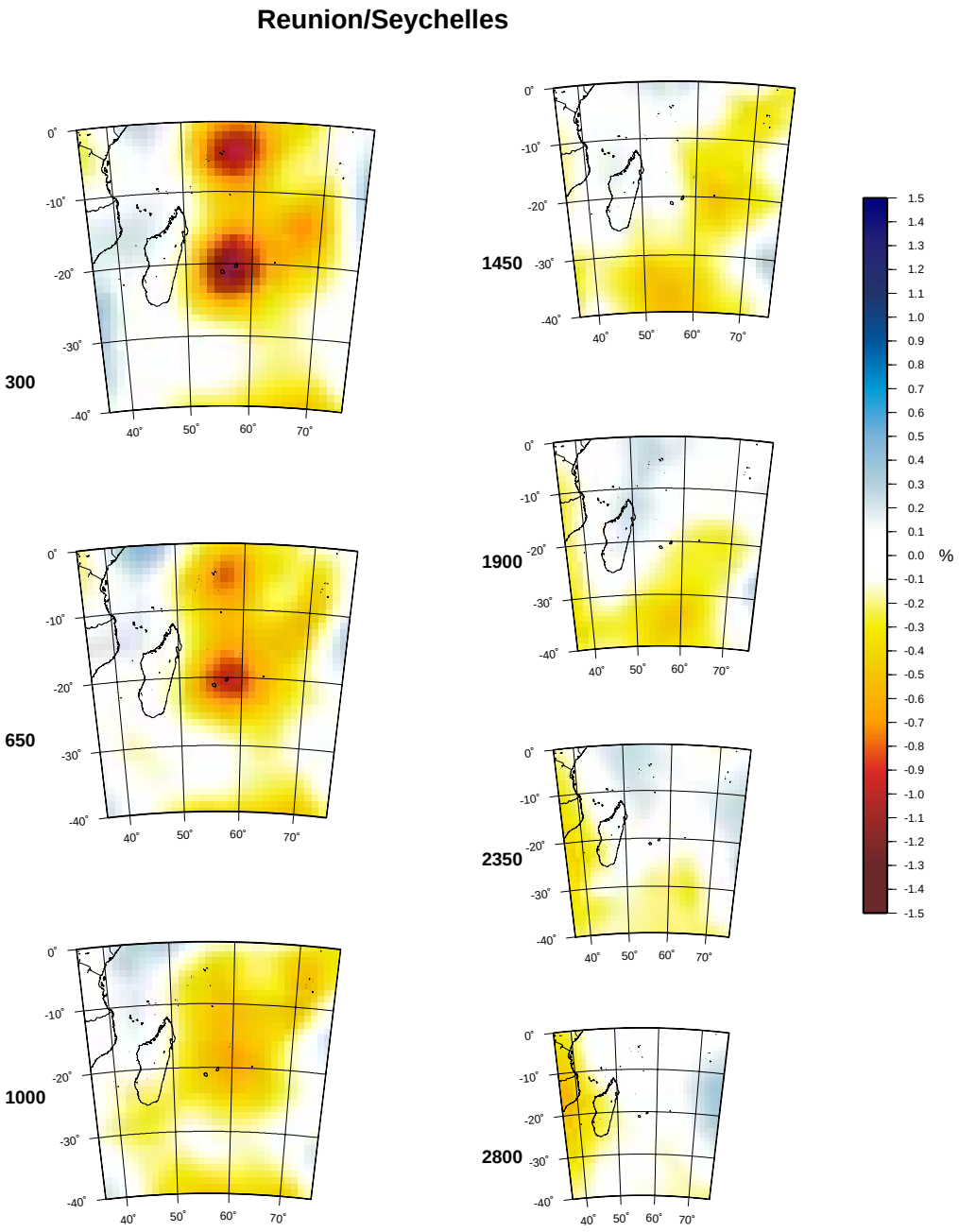


Figure A.23 Sections of the joint inversion velocity model beneath Reunion. A plume like structure is also visible beneath Seychelles, north of Reunion.

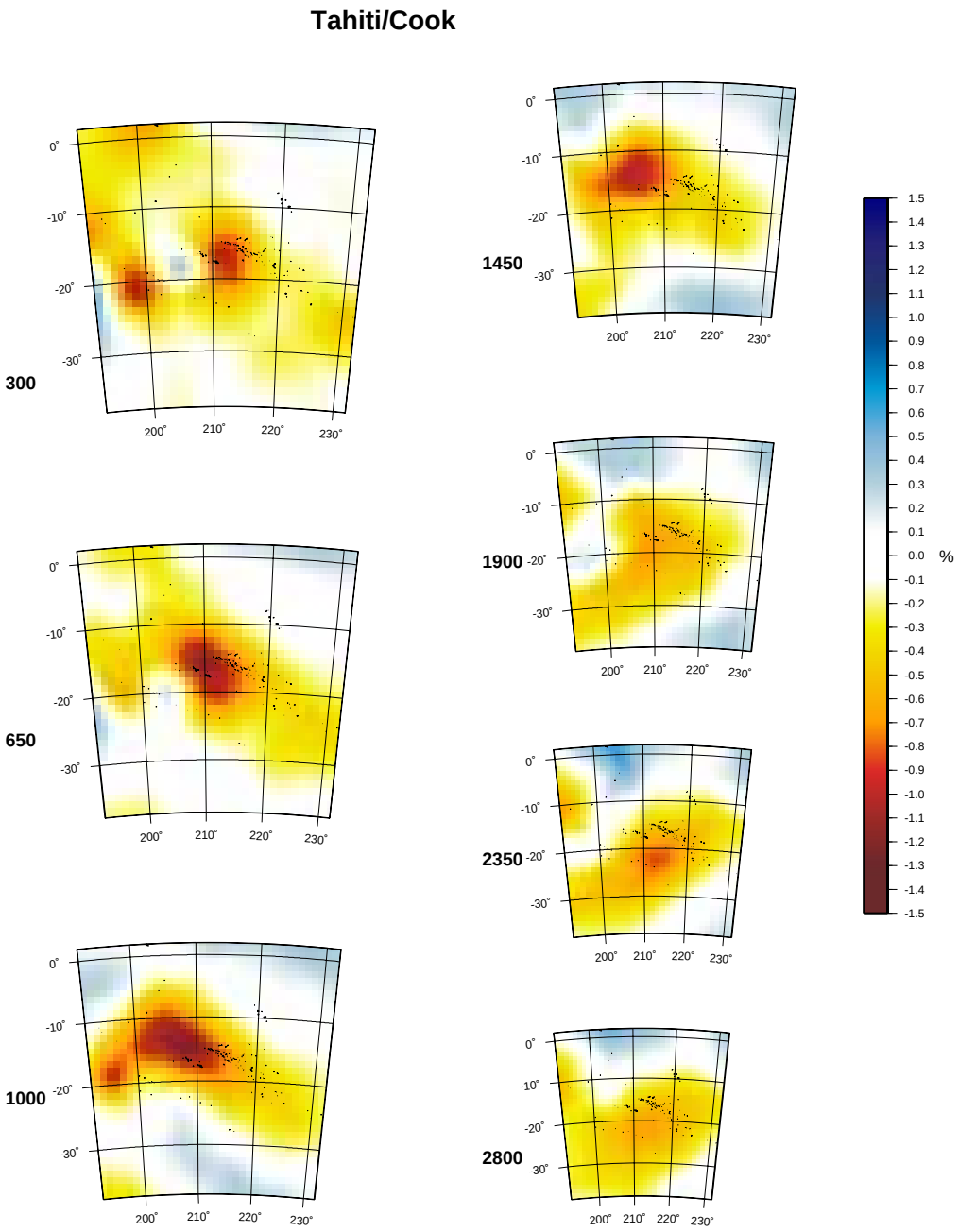


Figure A.24 Sections of the joint inversion velocity model beneath Cook and Tahiti.

Appendix B

Resolution analysis of the plumes.

In this appendix, we show the result of an extensive resolution analysis aimed to estimate the reliability of the plumes visible in our joint tomographic model. We test : a) if the shallow and mid-mantle plumes are the result of lack of resolution at depth; and b) if the mid-mantle and deep plume are due to leakage from the upper mantle.
Figure B.1 to B.5 are the supporting material of the paper which is about to be submitted to Science (Chapter 6). We show all reconstructed synthetic plumes originating at 650 km depth with 400 km radius; and rising from the bottom of the mantle with 200, 300 and 400 km radius respectively.

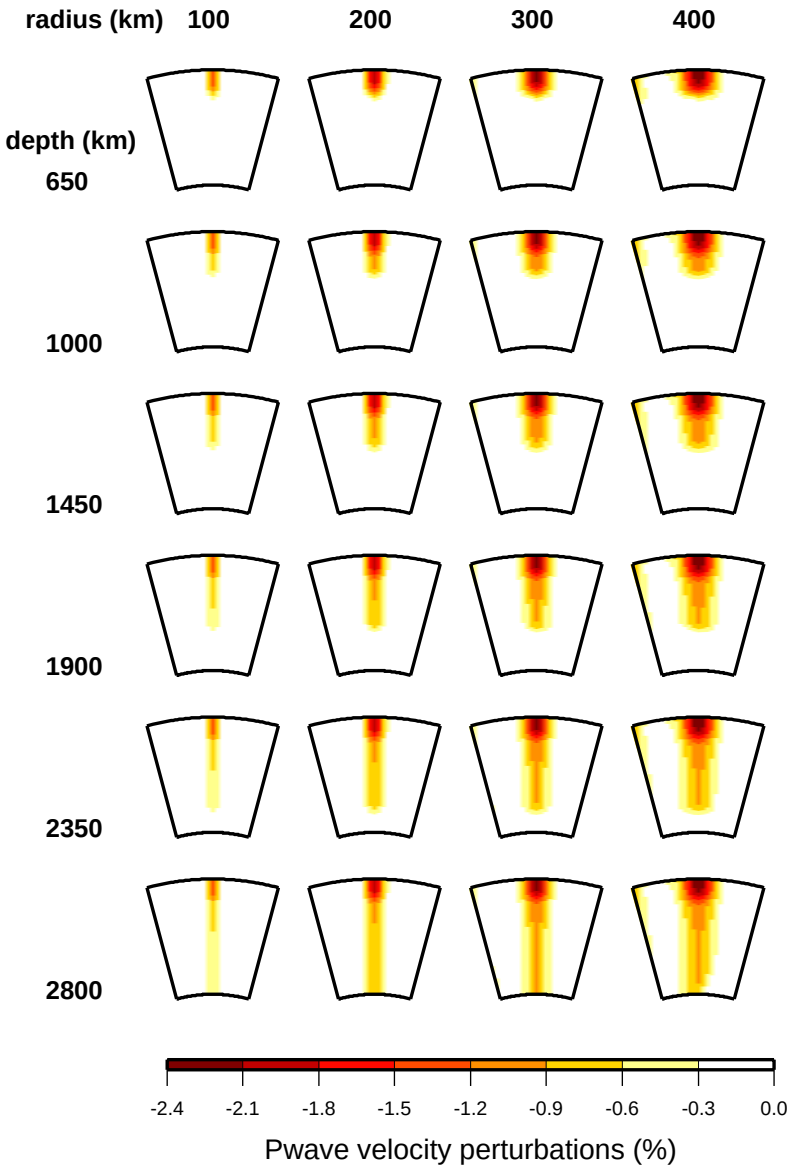


Figure B.1 Cross-sections through the synthetic plumes used in the resolution tests. The one shown are beneath Iceland. To estimate the width and depth reliability of the low velocity anomalies found in the velocity model, we use different plume widths (radii of 100, 200, 300 and 400 km, respcetively) for plumes originating at different depths in the mantle: 650, 1000, 1450, 1900, 2350 and 2800 km. The velocity perturbation in the synthetic plume follows a three-dimensional gaussian centered on the axis of the plume and changes as a function of depth as predicted by Karato (1993) for a temperature constrast of +300°K at the center.

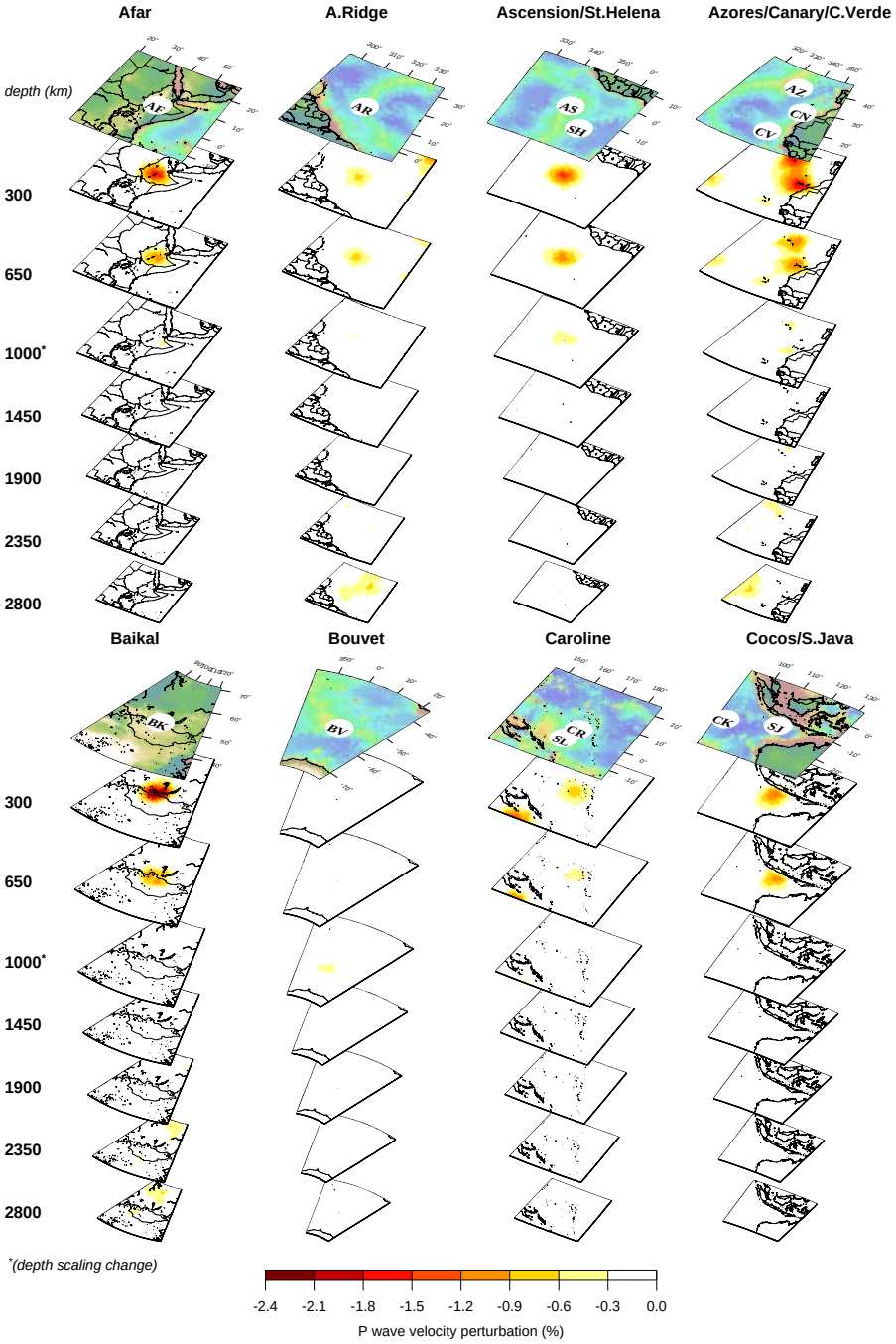


Figure B.2 Reconstructed synthetic plumes whose input depth was 650 km. Plumes are listed alphabetically. Labels on the surface are listed in Table 1 and indicate published locations of the hotspots.

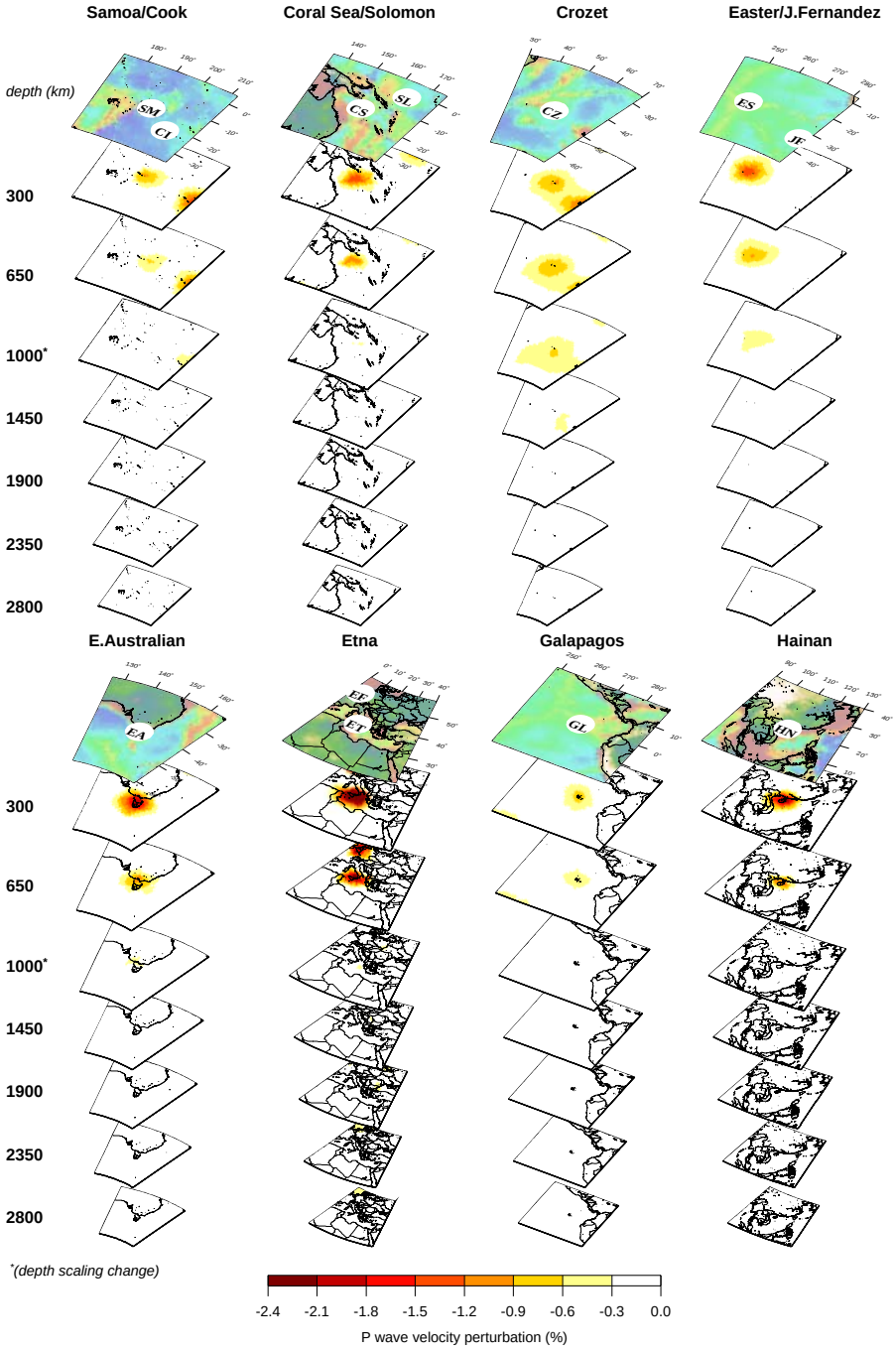


Figure B.2 continue

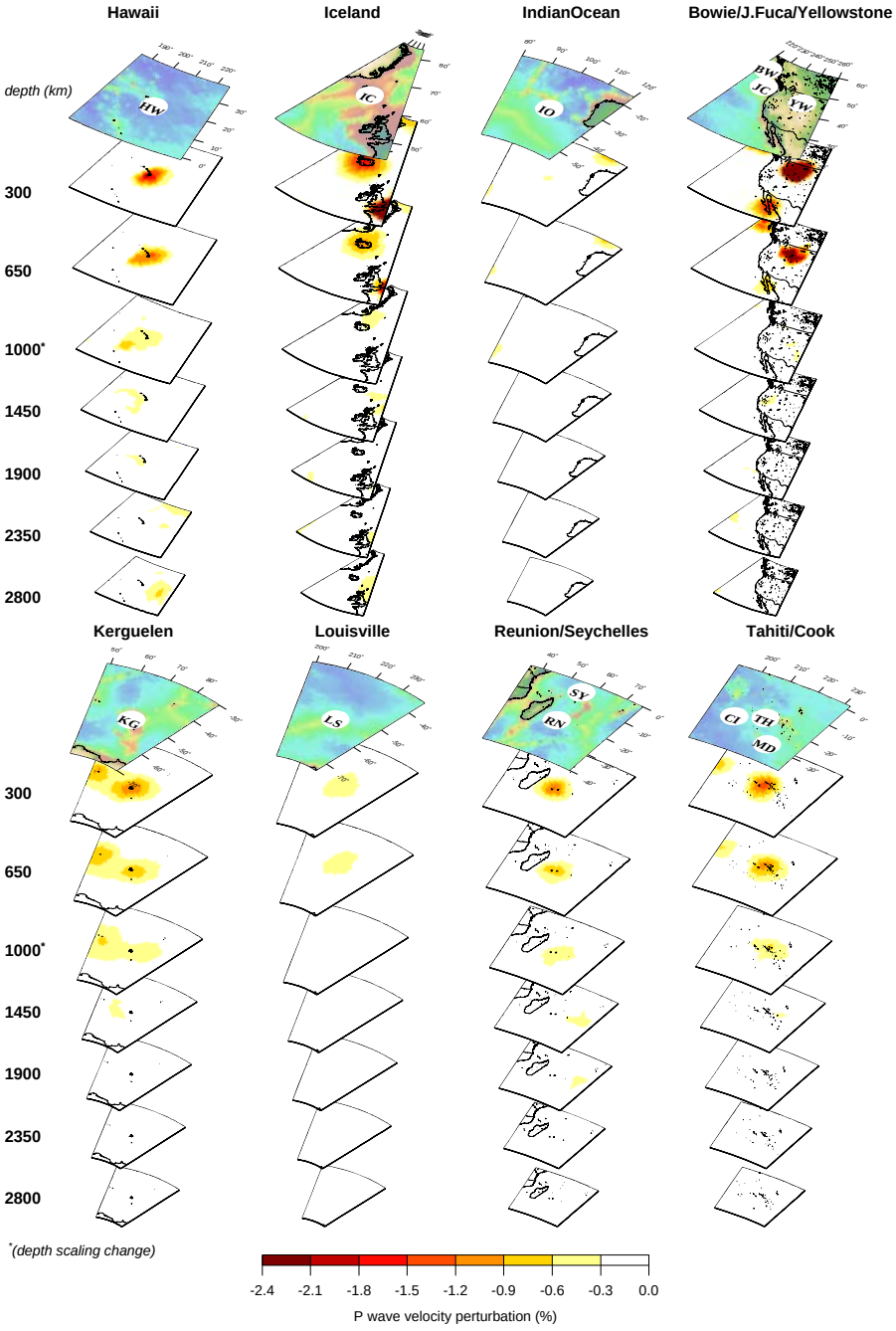


Figure B.2 continue

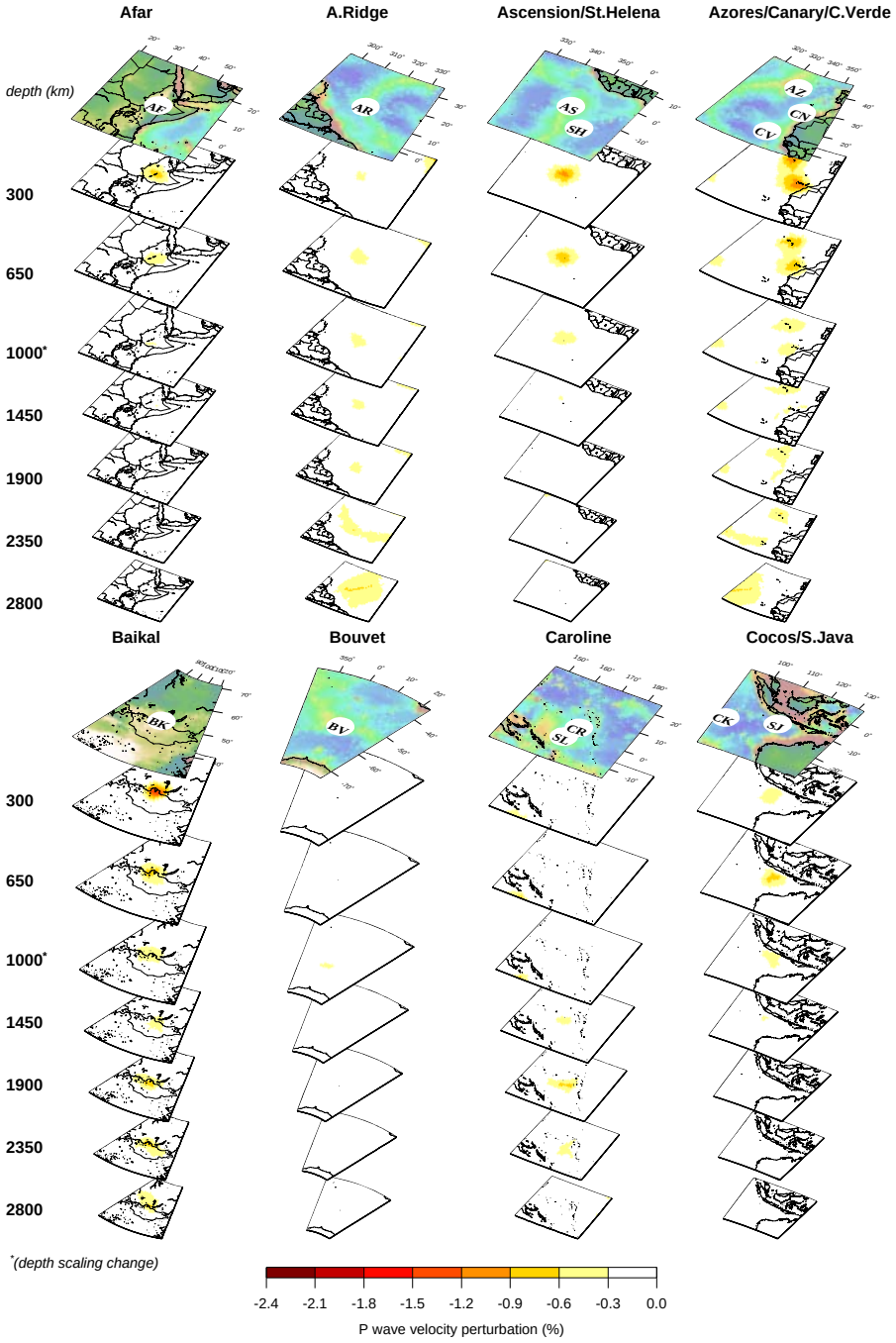


Figure B.3 Reconstructed synthetic plumes whose input depth and radius were 2800 km and 200 km, respectively.

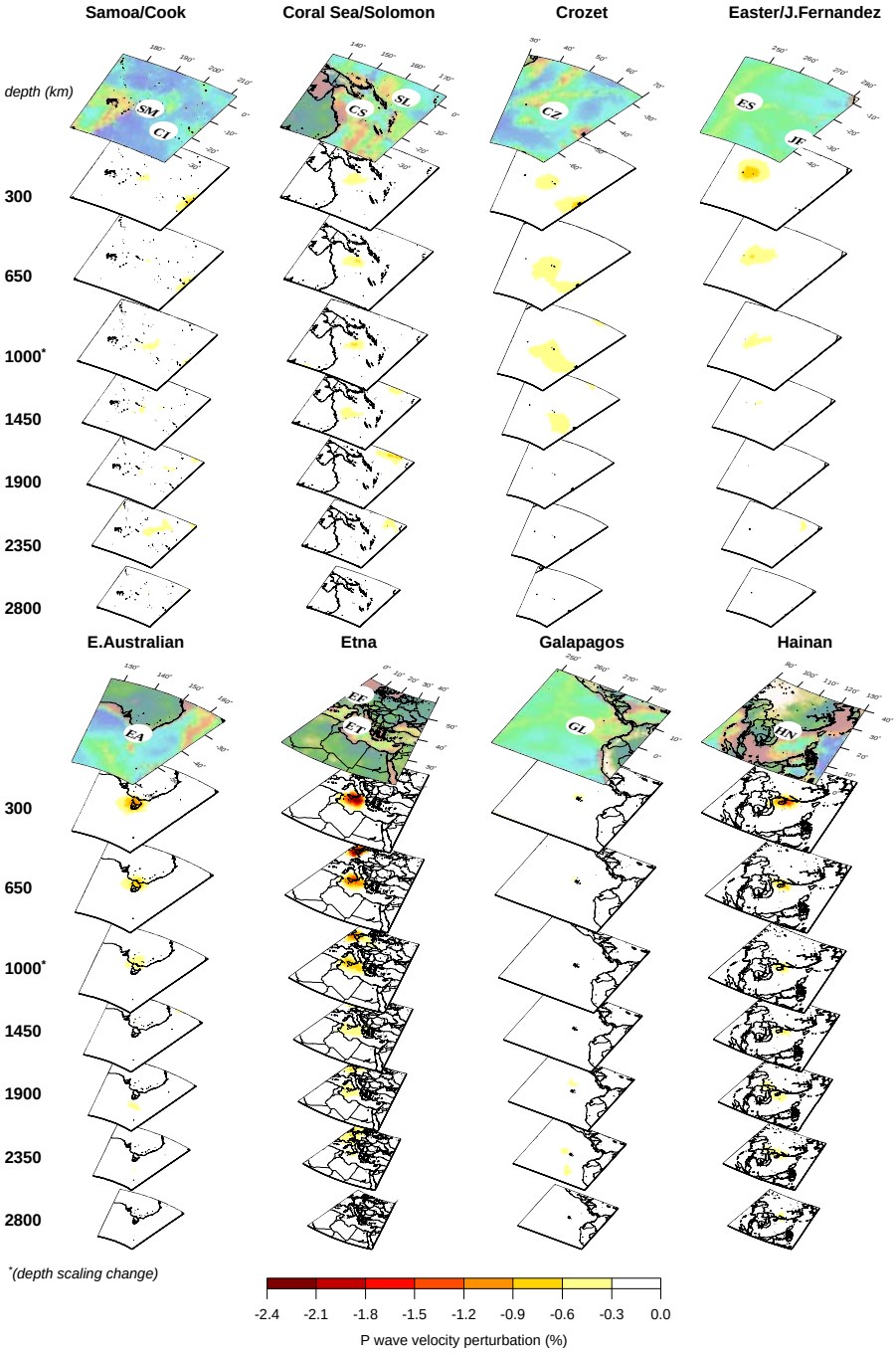


Figure B.3 continue

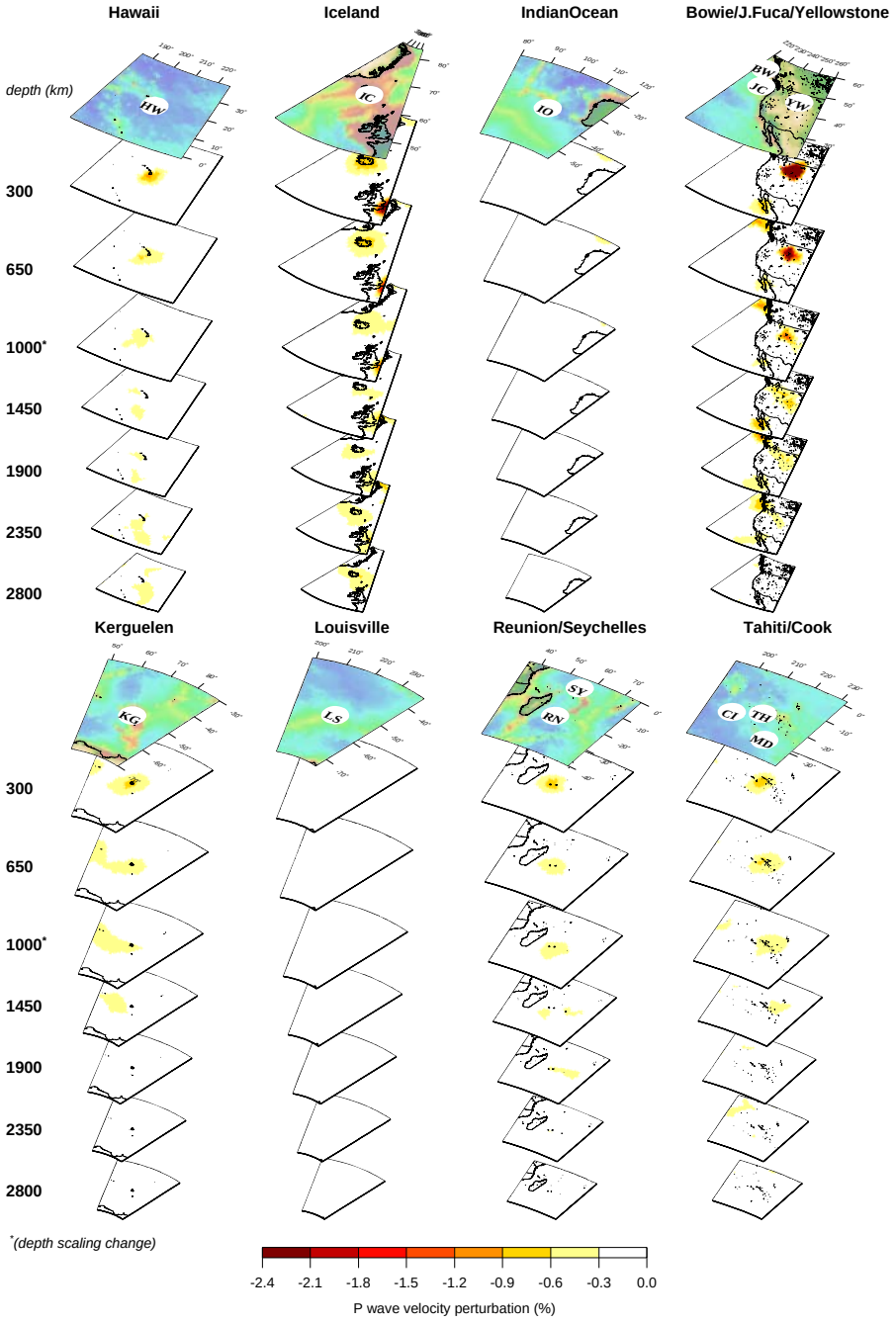


Figure B.3 continue

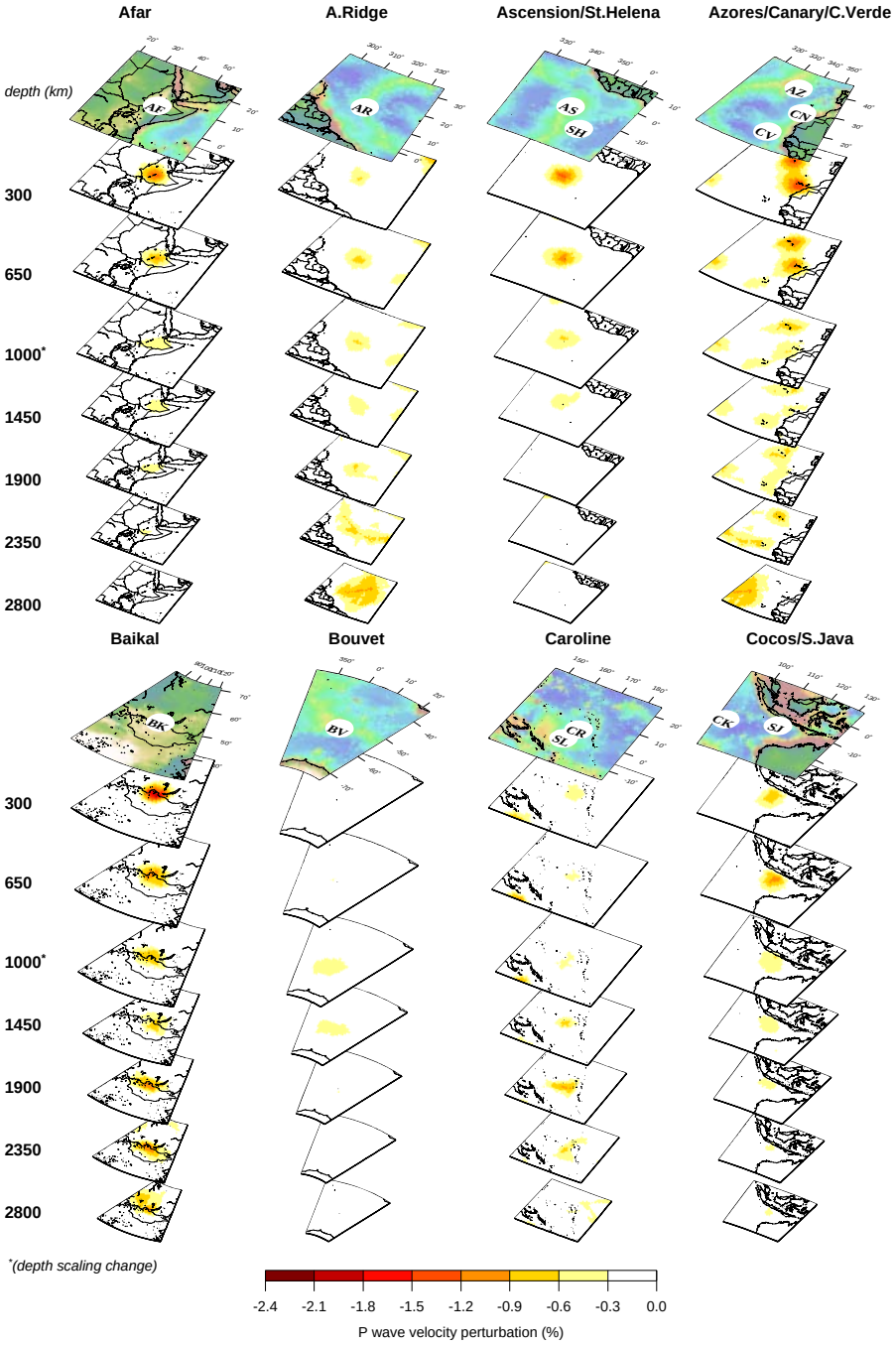


Figure B.4 Same as Fig. B.3, but for synthetic plumes with radius of 300 km

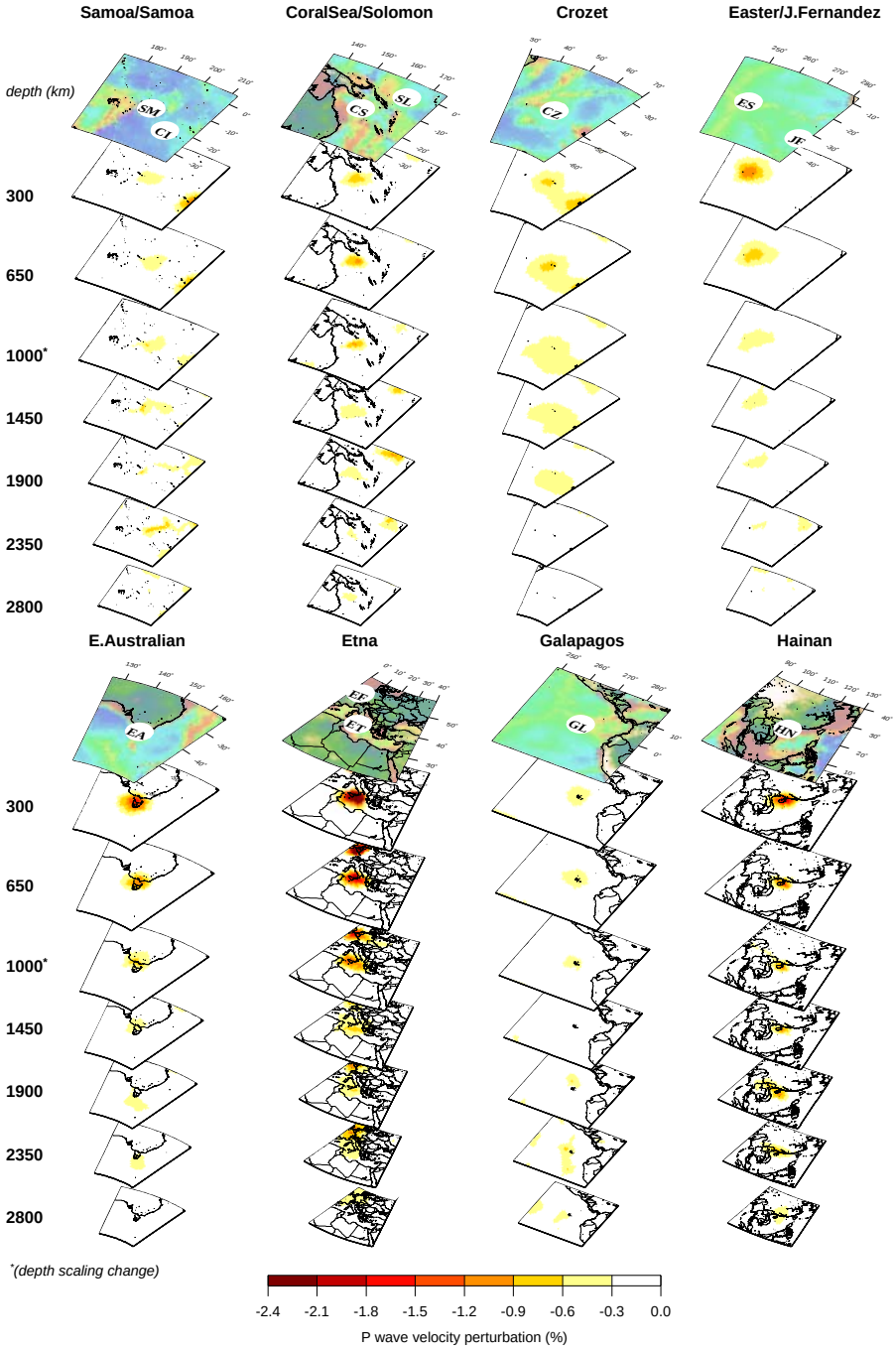


Figure B.4 continue

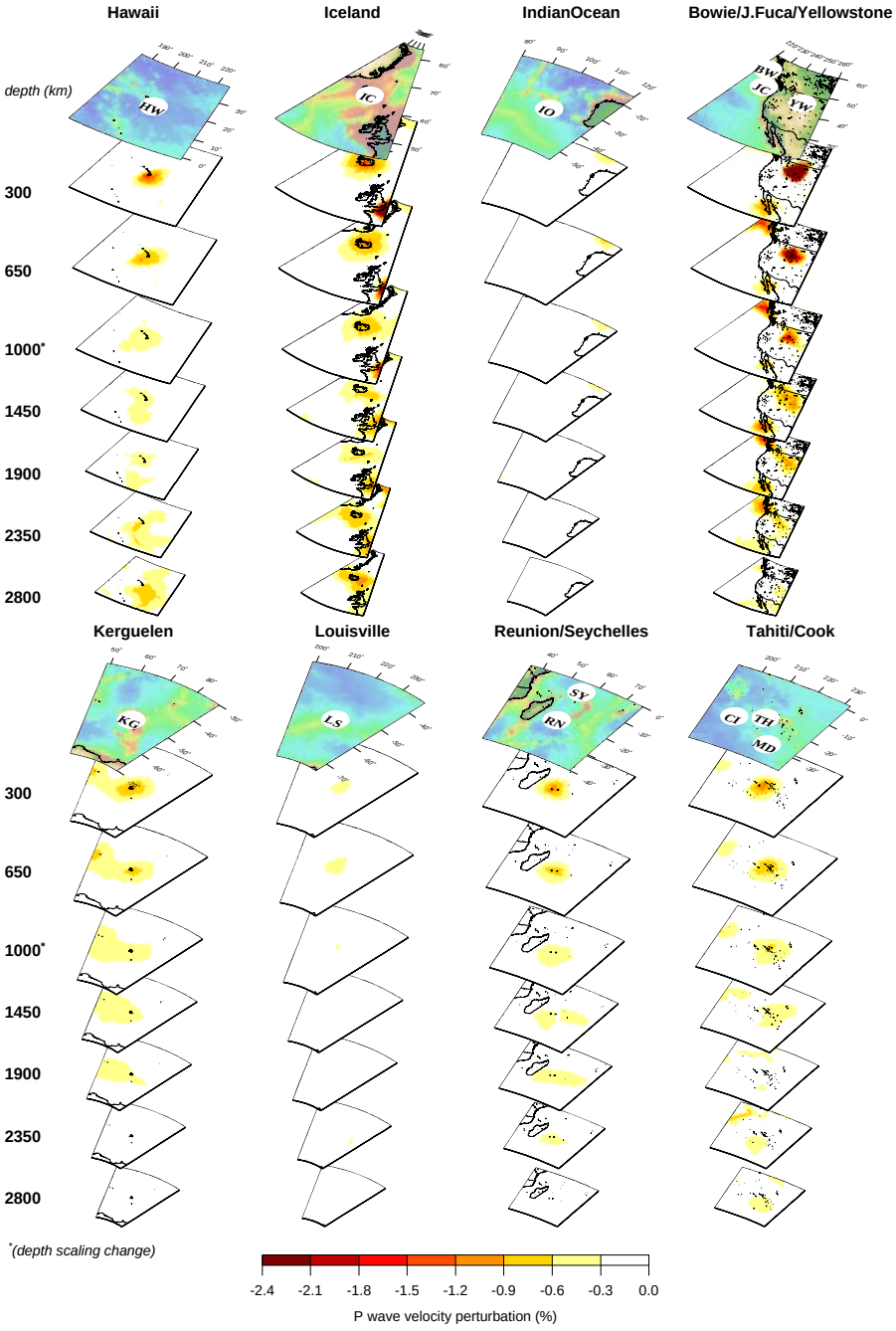


Figure B.4 continue

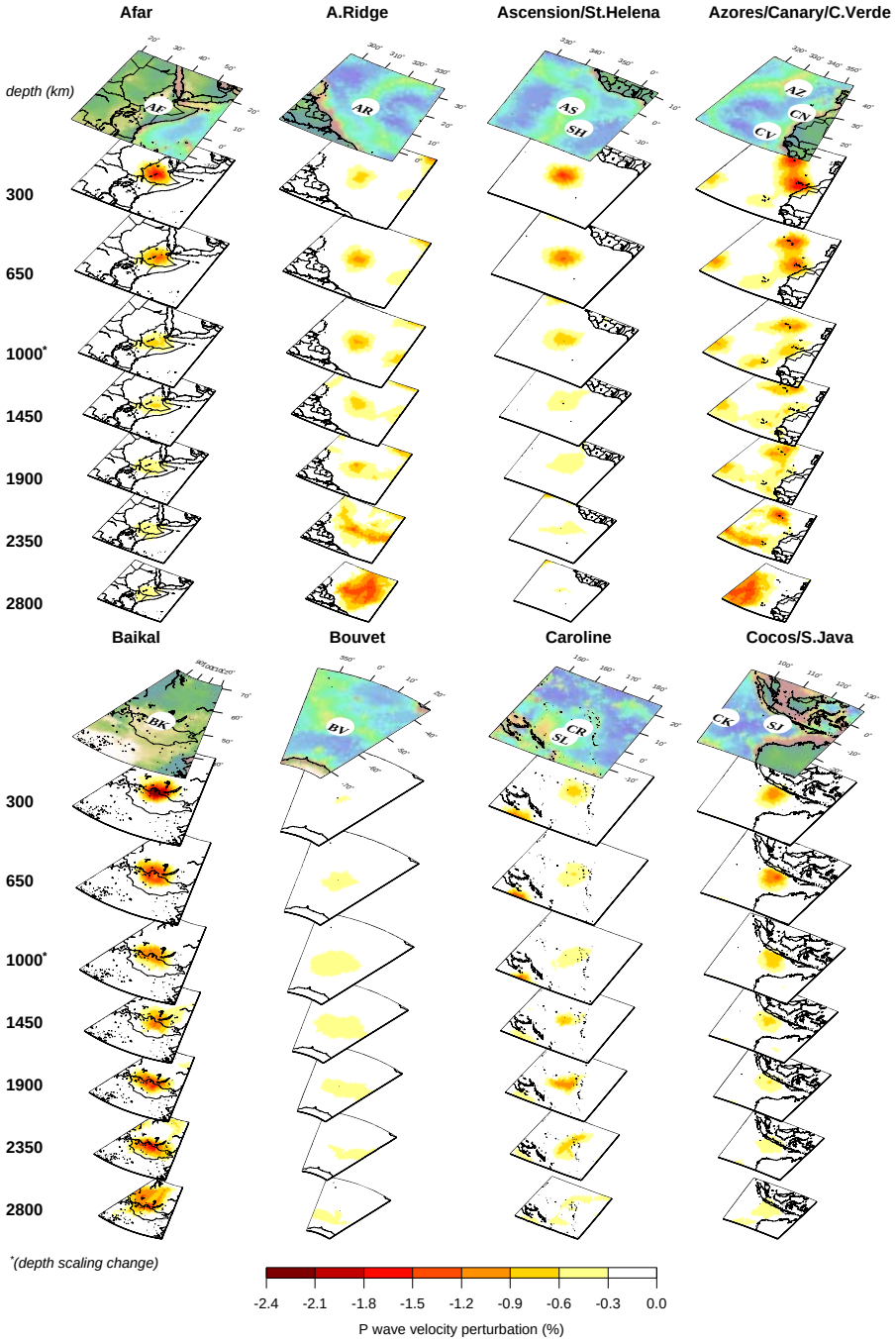


Figure B.5 Same as Fig. B.3 and B.4, but for synthetic plumes with 400 km radius.

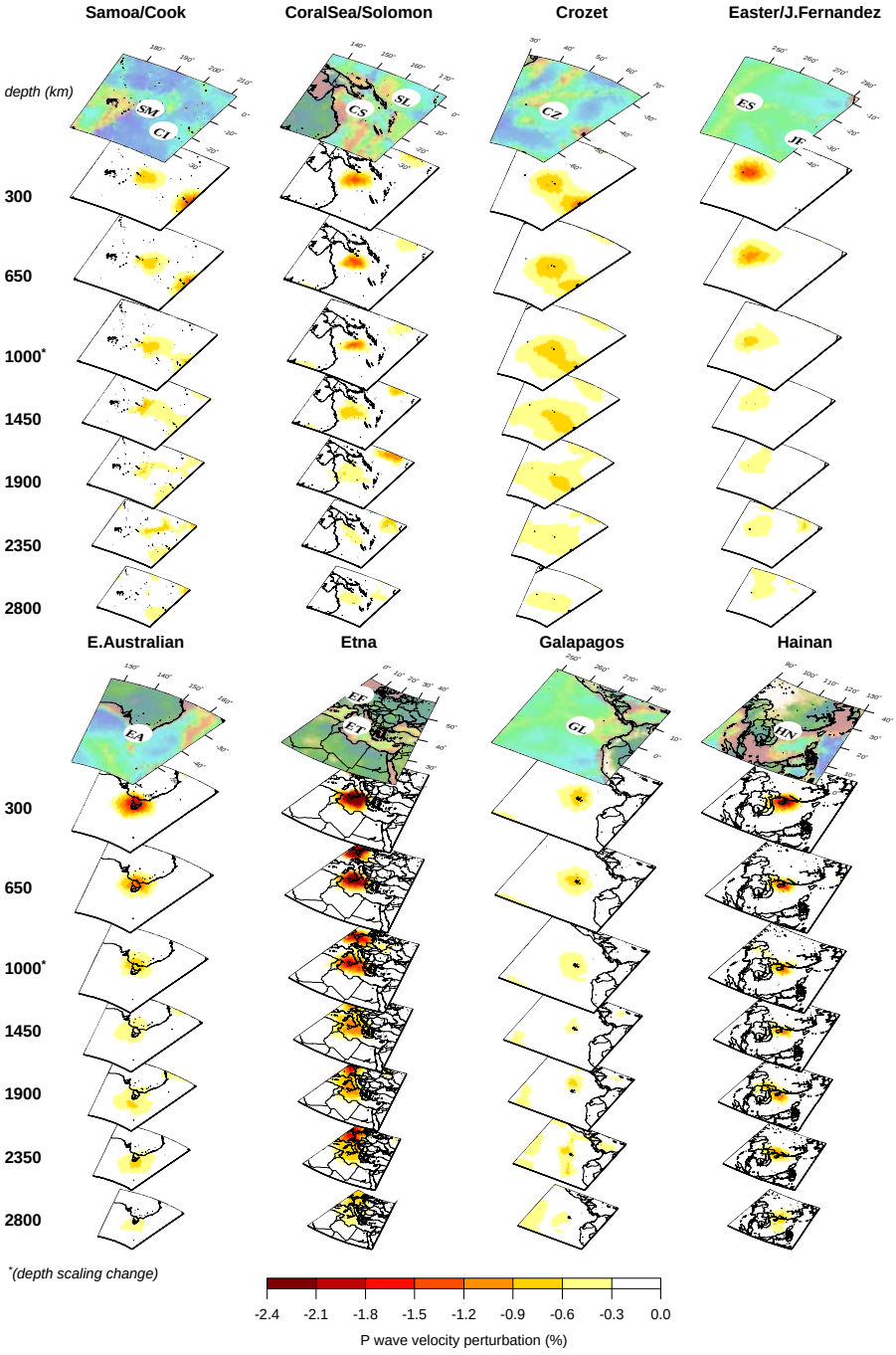


Figure B.5 continue

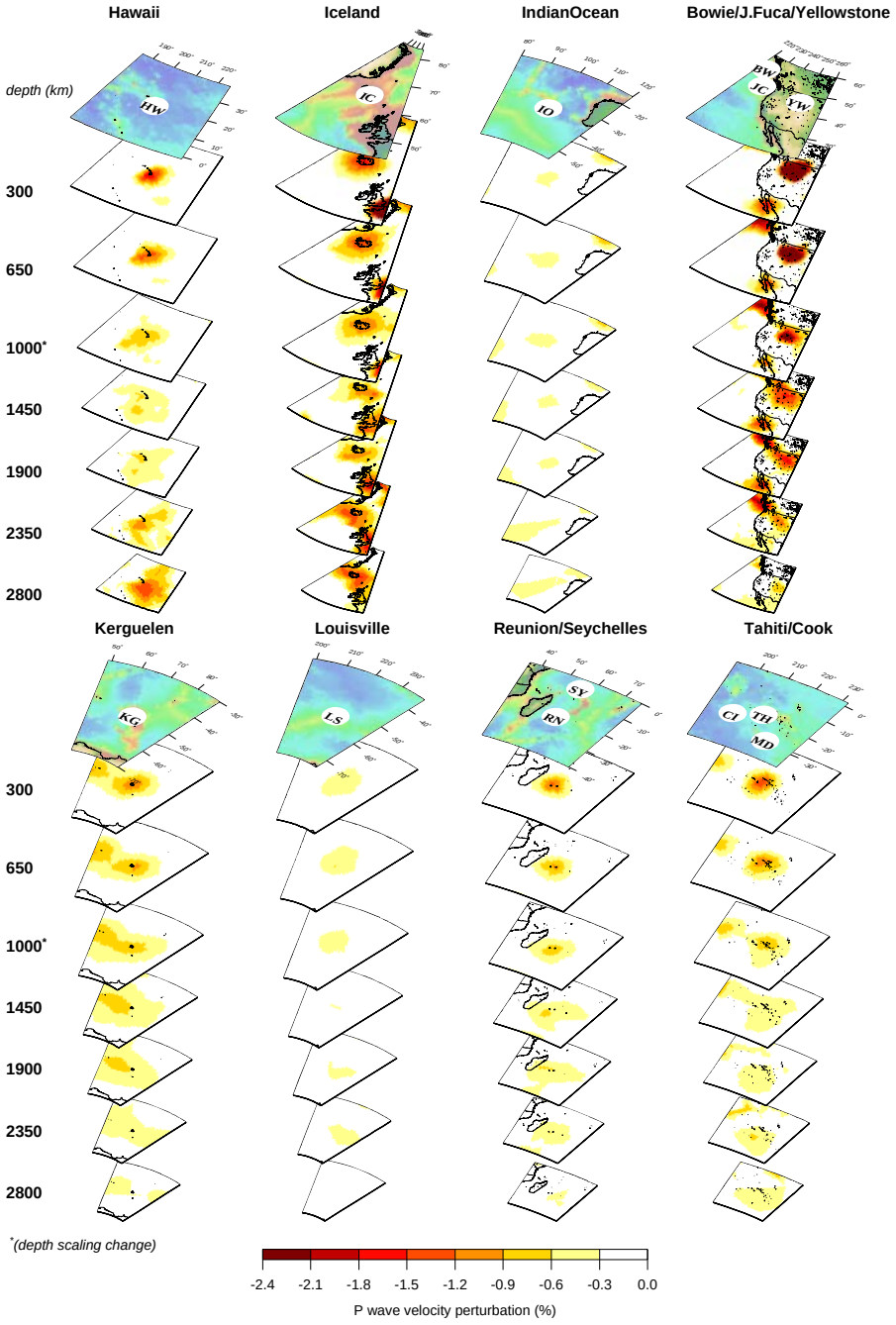


Figure B.5 continue

Appendix C

Resolution and a posteriori covariance of massive tomographic systems

This appendix has been published as: Nolet G., Montelli R. Virieux J., 1999: Explicit approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, Geophysical Journal International, Vol 138, 36-44

C.1 Abstract

We present an approximate method to estimate the resolution, covariance and correlation matrix for linear tomographic systems $Ax = b$ that are too large to be solved by singular value decomposition. An explicit expression for the approximate inverse matrix A^- is found using one-step backprojections on the Penrose condition $AA^- \approx I$, from which we calculate the statistical properties of the solution. The computation of A^- can easily be parallelized, each column being constructed independently.

The method is validated on small systems for which the exact covariance can still be computed with singular value decomposition. Though A^- is not accurate enough to actually compute the solution x, the qualitative agreement obtained for resolution and covariance is sufficient for many purposes, such as rough assessment of model precision or the reparametrization of the model by grouping of correlating parameters. We present an example for the computation of the complete covariance matrix of a very large (69043×9610) system with 5.9×10^6 nonzero elements in A. Computation time is proportional to the number of nonzero elements in A. If the correlation matrix is computed for the purpose of reparametrization by combining highly correlating unknowns x_i, a further gain in efficiency can be obtained by neglecting the small elements in A, but a more accurate estimation of the correlation requires a full treatment of even the smaller A_{ij}. We finally develop

a formalism to compute a damped version of A^-.

C.2 Introduction

Seismic tomography is playing an increasingly large role in the study of the Earth and its dynamic behaviour. Tomographic images now assist us in understanding, amongst others, the deep structure of continents, the details of the subduction process, and magma upwelling under ocean ridges and volcanoes. As the relevance of these seismological investigations grows for other Earth Science disciplines, it becomes important to deal with a fundamental shortcoming of all tomographic imaging: the non-uniqueness of the solution. The model resulting from an inversion is just one member of a subspace of models that satisfy the data equally well or better. Since the choice of the 'preferred' model in the subspace invariably involves a damping of ill-resolved aspects of the model, whereas well-resolved characteristics are more or less fixed, such damping usually reveals a strong influence of the ray path coverage in tomographic images.

The non-uniqueness of the solution can be characterized by its *resolution* and its *variance*, usually represented by the resolution matrix and the (*a posteriori*) covariance matrix. For small scale problems, these matrices can be calculated conveniently using a singular value decomposition of the problem (Wiggins, 1972; Jackson, 1972). For larger problems this becomes impractical or downright impossible. The resolution can still be investigated using sensitivity tests (e.g. Spakman & Nolet (1988)). Such tests have shortcomings (Leveque et al., 1993), but an even greater disadvantage is that such tests measure the sensitivity only with respect to a fixed pattern of cells (e.g. a checkerboard test), and the estimation of the resolution of single cells requires the repetition of many sensitivity tests. Furthermore, no satisfactory method exists to find the *a posteriori* covariance of the solution, other than adding random errors to the sensitivity tests and estimating the covariance matrix from the results of many such tests (Kennett & Nolet, 1978), a practice too laborious to have found general acceptance. Techniques like 'jacknifing' or 'bootstrapping' (Tichelaar & Ruff, 1989) rely on the overdetermined nature of an inverse problem and should never be applied to an underdetermined system of equations. Their use on large mixed over/underdetermined problems such as found in tomography is not only highly questionable but also computationally very expensive.

Recently, the estimation of the resolution matrix from the first few Ritz vectors (approximate eigenvectors) resulting from a Lanczos-type iteration on the linear system has been proposed (Zhang & McMechan, 1995, 1996). Such schemes are seriously flawed unless the number of Ritz vectors approaches the effective rank of the matrix, a goal which is impractical for inversions with, say, the number of data and model parameters exceeding 10^5 (Deal & Nolet, 1996). We can summarize the situation as follows:

1. for models with many degrees of freedom, it becomes impossible to calculate the *a posteriori* covariance matrix of the result;

2. resolution calculations by means of a limited number of sensitivity tests have serious shortcomings;

3. there is no satisfactory way to suppress the influence of the uneven distribution of raypaths in the final result.

In this paper we shall develop a simple approximate algorithm to estimate the resolution *and* the *a posteriori* covariance of a tomographic solution which avoids the calculation of eigenvectors or Ritz vectors. Whilst we leave an investigation of the third problem to a future paper, we believe the influence of the ray path distribution should be reduced by an adaptive reparametrization of the model, for which an estimation of the model covariance is a necessary prerequisite.

C.3 Statement of the problem

We consider the $n \times m$ linear inversion problem for a model x, given (exact) data b with errors ϵ:

$$Ax = b + \epsilon = \hat{b}, \tag{C.1}$$

scaled such that covariance matrix of the data error ϵ is the $n \times n$ unit matrix:

$$C_\epsilon = I_n. \tag{C.2}$$

Without loss of generality, we assume that the expected value of the data errors as well as the model parameters is zero:

$$E[\epsilon_i] = 0 \quad i = 1, ..., n \tag{C.3}$$

$$E[x_i] = 0 \quad i = 1, ..., m. \tag{C.4}$$

Let A^- denote the inverse of A in a generalized sense; for example $A^-\hat{b}$ might be the minimum norm solution of the least-squares system belonging to (C.1). While there is considerable freedom in the choice of A^-, a generalized inverse must satisfy $AA^-A = A$, or, as paraphrased by Jackson (1972):

$$AA^- \approx I_n, \tag{C.5}$$

$$A^-A \approx I_m. \tag{C.6}$$

which we shall refer to as the two 'Penrose conditions'. We can express the error of the solution $\hat{x}$ in terms of A^- (Nolet 1987):

$$\hat{x} - x^{true} = A^-(b + \epsilon) - x^{true} = (A^-A - I_m)x^{true} + A^-\epsilon \tag{C.7}$$

which expresses the well-known result that the error in the solution has two causes: the inadequacy of the generalized inverse to satisfy the second Penrose condition (C.6) exactly, and the propagation of error terms through multiplication with A^-. Using the terminology of statistics, the first term constitutes the *bias* of the solutions, the second term the statistical fluctuations (for different realizations of the observational errors) around the biased solution.

Similarly, for the data misfit we find:

$$\mu \equiv A\hat{x} - b = (AA^- - I_n)b + AA^-\epsilon \tag{C.8}$$

from which we see that the data misfit $\chi^2 = |\mu|^2$ also has a bias and a variance. If we succeed in satisfying the first Penrose condition (C.5) we reduce the bias.

Setting $\epsilon = 0$ in (C.7), we find an expression for the resolution matrix:

$$\hat{x} = Rx^{true}, \tag{C.9}$$

where

$$R = A^-A. \tag{C.10}$$

If, as is usually the case, (C.1) is a linear approximation to a non-linear problem, we may define ϵ to include also the errors due to linearizations, or other approximations in the forward problem (Tarantola, 1987). This will undoubtedly introduce some correlations between the components of the error vector ϵ, which in principle could be removed through a linear transformation. To make a reasonable *a priori* estimate of the covariance matrix of ϵ is, however, a task so daunting that we are not aware of any successful efforts to do so for the seismic tomography problem. The unscaled C_ϵ is therefore generally assumed to be diagonal, so the transformation to satisfy (C.2) reduces to a trivial multiplication. The *a posteriori* covariance matrix of the solution $\hat{x}$ is then given by

$$C_{\hat{x}} = A^-C_\epsilon(A^-)^T = A^-(A^-)^T. \tag{C.11}$$

As usual, this is the covariance in the 'minimum norm' solution, which may be small either because a parameter is well constrained by the data, or because it is strongly damped toward 0. For the latter, the 'bias' is large but with little uncertainty. A true indication of the model accuracy can only be obtained by inspecting both the resolution matrix R and the covariance matrix $C_{\hat{x}}$.

From (C.11) we can easily compute the elements of the correlation matrix, defined as:

$$\rho_{ij} = \frac{C_{ij}}{[C_{ii}C_{jj}]^{\frac{1}{2}}} \tag{C.12}$$

where we suppressed the subscript $\hat{x}$. Fully unresolved parameters (for which the column in A is empty) require a special treatment: their variance, while infinite in reality, is numerically zero because the nullspace of A is excluded from the solution space, and the correlation is undefined. We set such $\rho_{ij} \equiv 0$.

Intuitively, one understands from (C.7) that the statistical error term will grow when A^- has large components. Forcing the elements of A^- to remain small will reduce the variance, but increase the bias, since we also reduce our ability to satisfy (C.6). The early literature on geophysical inverse problems is exhaustive in its analysis of this trade-off between bias and error, or variance, of the solution, either in discrete systems such as considered here (Wiggins, 1972; Jackson, 1972), or in systems where models are not discretized *a priori* Backus & Gilbert (1970); Tarantola (1987). However, it invariably requires the inversion of large matrices, which is generally done by the application of singular value decomposition (SVD). While the increasing capacities of large computers now allow us to apply SVD to matrices where m and $n \approx 10^3$, large-scale traveltime inversions commonly deal with $n = 10^4$ to 10^6 or more data, and require 10^3-10^5 or more model elements.

Of course, the computation of the exact generalized inverse of A with SVD is not feasible for such large tomographic problems. Therefore, we can only attempt an approximate solution to our problem. We note that the computation of the solution itself does not require the computation of the inverse A^-, since we can use iterative techniques to do so. We do need the inverse, however, to characterize the resolution by means of R and $C_{\hat{x}}$.

Since we define our solution as $\hat{x} = A^-\hat{b}$, (C.1) implies the condition $AA^-\hat{b} = \hat{b}$, and it is obvious that the first Penrose condition (C.5) is the equation that we shall wish A^- to satisfy as best as we can. We shall see later that this is not an optimum solution to (C.6) in our approximate analysis of the problem; that is, it does not minimize the model bias.

Define c^k as the vector equal to the $k-$th column of A^-, and e^k as the n-dimensional unit vector in direction k. (C.5) implies the following:

$$Ac^k = e^k \quad (k = 1, ..., n). \tag{C.13}$$

Naganishi & Suetsugu (1986) have proposed solving (C.13) exactly for all k, a strategy which is only possible for small n. We derive a fast, approximate solution using backprojection. The backprojection direction is found by taking the negative gradient $-\nabla_c$ at location c_0^k in model space of the misfit $|Ac^k - e^k|^2$, which is equal to $-A^T(Ac_0^k - e^k)$. In our case $c_0^k = 0$, from which we find:

$$v^k \equiv A^T e^k, \tag{C.14}$$

where v^k is a vector of dimension n. (C.14) gives simply

$$v_i^k = A_{ik}^T = A_{ki} \quad (i = 1, ..., m). \tag{C.15}$$

We seek an approximate solution to (C.13) by imposing the condition that c^k is in the direction of v^k: $c^k = \alpha_k v^k$. If we impose the condition that the misfit is minimized, this implies orthogonality of the misfit vector: $(\alpha_k Av^k - e^k, \alpha_k Av^k) = 0$. Hence the coefficient

$$\alpha_k = \frac{(e^k, Av^k)}{(Av^k, Av^k)}, \tag{C.16}$$

or, when written out explicitly,

$$\alpha_k = \frac{\sum_{j=1}^{m} A_{kj}^2}{\sum_{i=1}^{n}(\sum_{j=1}^{m} A_{ij}A_{kj})(\sum_{q=1}^{m} A_{iq}A_{kq})}. \tag{C.17}$$

Since

$$A_{ik}^{-} = c_i^k = \alpha_k v_i^k = \alpha_k A_{ki} \tag{C.18}$$

the generalized inverse therefore can be written as

$$A^{-} = A^T D \tag{C.19}$$

where D is a diagonal matrix, its diagonal elements equal to α_k:

$$D_{kk} = \frac{(AA^T)_{kk}}{\sum_{i=1}^{n}(AA^T)_{ik}^2} \quad (k = 1, ..., n). \tag{C.20}$$

Unfortunately, Penrose's second condition (C.6) leads to a different approximate solution. Following the same backprojection method, we find

$$A^{-} = D'A^T \tag{C.21}$$

with the elements of D' defined by

$$D'_{kk} = \frac{(A^TA)_{kk}}{\sum_{i=1}^{m}(A^TA)_{ik}^2} \quad (k = 1, ..., m). \tag{C.22}$$

Finally, we notice a difference between the last equation and the approximate inverse we would obtain by simply reducing A^TA to its diagonal. In that case we would have an inverse similar to (C.21): $A^{-} = D''A^T$ with $D''_{kk} = (A^TA)_{kk}^{-1}$. We investigated this third possibility briefly and abandoned it as quickly because of its complete lack of fit to either (C.5) or (C.6).

It is well known that one iteration of a backprojection step will converge to the correct solution in the case where all singular values of A are equal. This is not even remotely the case for tomographic systems. However, backprojections often give very reasonable data fits. The reason must be sought in the sparse nature of the matrices. If there is little overlap between rays, the products involved in AA^T will involve multiplications with zeros, unless two rays sample the same model cell. Therefore, AA^T is likely to be diagonally dominant. One can easily checks that (C.5) is satisfied as long as $(AA^T)_{ij}$ $(i \neq j)$ can be neglected with respect to $(AA^T)_{ii}$. Since cells always correlate with neighbouring cells, the diagonal of A^TA is probably less dominant, which would explain the inferior performance of the diagonal approximation $D''_{kk} = (A^TA)_{kk}^{-1}$. This approximation may work better for systems in which $E[A_{ij}] \approx 0$, such as in diffraction tomography, but is obviously bad for systems from body wave tomography where $E[A_{ij}] \gg 0$. Such considerations

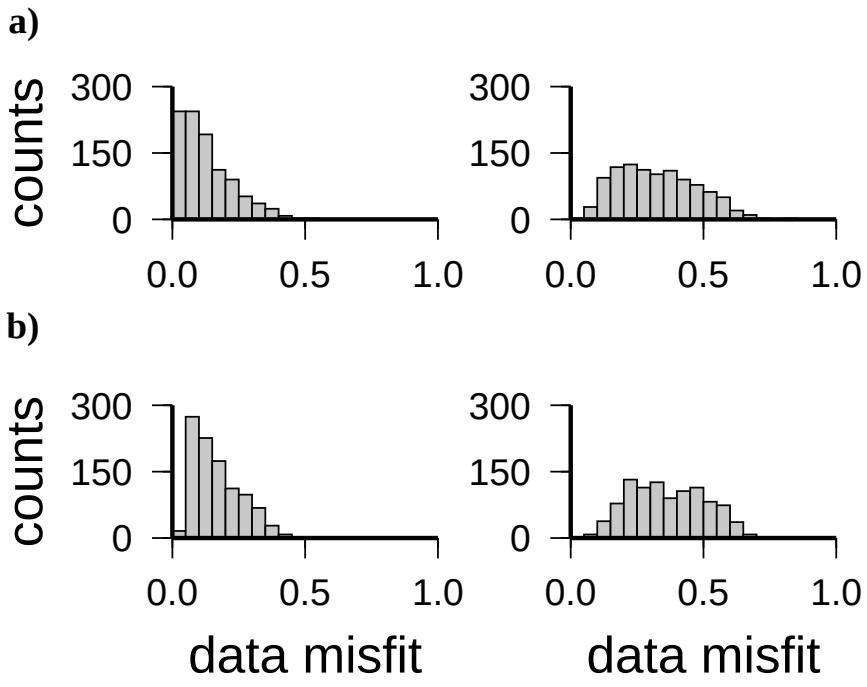


Figure C.1 Histograms showing the data misfit $|Ax - b|^2/|b|^2$ for a Monte Carlo simulation using 1000 random data vectors b in the range of (left) A^{East} and (right) A^{West}. a: A^- defined with the first Penrose condition (C.5), b: A^- defined using (C.6).

are , however, far from conclusive, and in the next subsection we shall rely on a numerical test to justify our approach.

We can use (C.19) in (C.10) and (C.11) to obtain estimates of the resolution and the covariance matrix, respectively. Note that these expressions have an added advantage over the expressions for R and $C_{\hat{x}}$ as computed by SVD, apart from the saving on computer memory and CPU: they allow us to compute only part of these matrices, which is useful if our parameters are 'local' (for example, spline supports, rather than non-local parameters such as spherical harmonic coefficients). Thus, we can isolate velocity or slowness parameters from parameters designating source or station corrections, or even isolate a particular geographic region of interest. The parameter transformations inherent to SVD prohibit this with the exact computations.

Another advantage is that the computation of A^- lends itself naturally to parallel computations, since each of the columns of A^- is computed independently from the others.

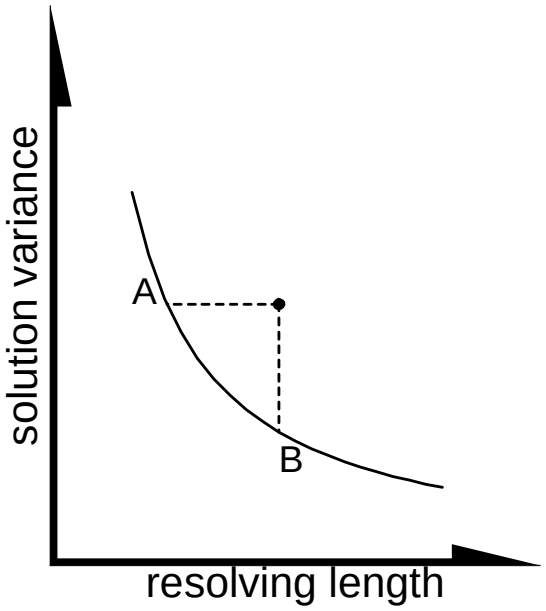


Figure C.2 Schematic diagram showing the trade-off of variance versus resolving length for the SVD solution. The approximate inverse yields estimates for these quantities which are off this curve (black dot), and which may be compared either with the SVD solution with similar variance (point A) or with similar resolving power (point B), or in between.

C.4 Validation on a small linear system

The validity of our approach depends on how well (C.13) is solved with only one backprojection step. Earlier experience with the iterative inversion of sparse matrices suggests that the first backprojection step almost always provides the bulk of the variance reduction, often reducing the data misfit by more than 50% of the total (converged) reduction. In this subsection we investigate the validity of our approach.

Although SVD is always to be preferred for matrices of small dimensions, since it allows for an exact computation of the solution statistics, we use our method on such small system to make a comparison with the exact solution possible. We choose realistic examples. Two matrices A are taken from the Sn tomography study of Nolet et al. (1998), and are denoted by 'east' and 'west', respectively. A^{East} is a 121×115 system with a rather sparse coverage of ray paths (see Fig. 2 in Nolet et al. (1998)). In contrast, A^{West}, 839×429, has a dense coverage with many, often overlapping ray paths. Both systems include source and station corrections in addition to unknown slowness anomalies in the vector x.

In a first test we randomly generate synthetic data vectors b that satisfy (C.1) exactly, and test how well $x = A^{-}b$ satisfies the data. This is a direct test of the first Penrose condition (C.5). Fig. C.1 shows histograms of the fits (defined as

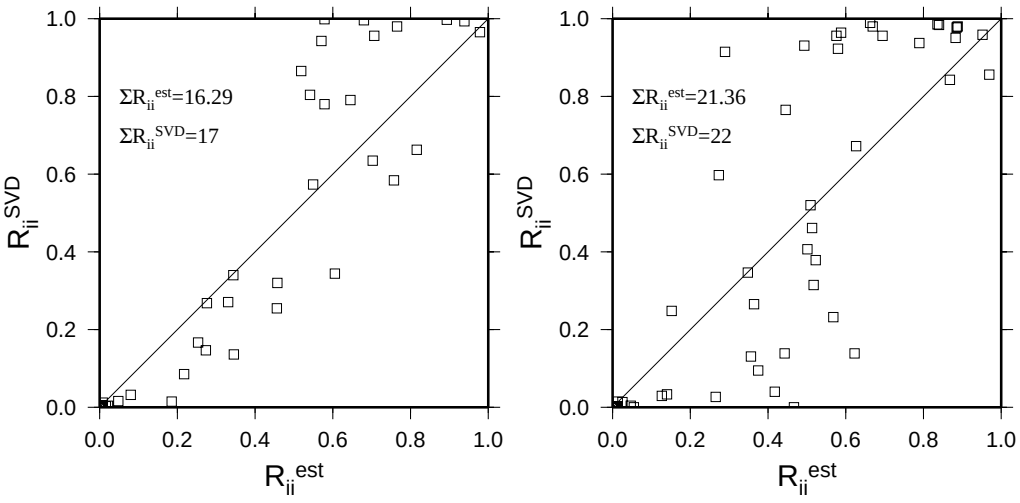


Figure C.3 Comparison between the estimated diagonal elements of the resolution matrix R^{est} and the correct values R_{ii}^{SVD} for (left) A^{East} and (right) A^{West}. The SVD results were computed with 17 and 22 eigenvectors, respectively. See text for discussion.

$|Ax - b|^2/|b|^2$) for A^- computed with (C.19) and (C.21). For both east (left) and west (right) it is clear that (C.19) (a) yields a superior data fit, as is to be expected since (C.19) was constructed to satisfy (C.5), but the difference with (C.21) (b) is not large.

Although it is clear that the variance reduction is not complete, it is obvious that A^-b reduces the variance by at least 50 per cent, and often by much more than that. Since this is not a small variance reduction for many tomographic inversions, and since our aim is to estimate the statistics, not to construct $\hat{x}$, we judge this outcome highly encouraging. Since backprojections work most efficiently for non-overlapping ray paths (and would result in the optimal fit if every cell was visited only once), we conjecture that the differences between east and west are due to the difference in raypath overlap, with the estimate becoming less accurate as the ray paths overlap more. This would imply that the more accurate estimate of A^- is obtained by assembling overlapping raypaths into 'summary rays' (Morelli & Dziewonski, 1987).

For the actual computation of the *solution* of $Ax = b$ the application of repeated backprojections in a conjugate gradient algorithm such as LSQR is not only more accurate but also faster (Paige & Saunders, 1982; Nolet, 1983).

When comparing estimates of $C_{\hat{x}}$ and R with their exact counterparts we face a problem related to the damping of the SVD solution. The situation is schematically sketched in Fig.C.2.

In this figure, our estimated variance and resolving length (or correlation dis-

tance) is shown by the dot. The curve represents the trade-off between variance and resolving length for a truncated SVD solution as we vary the number of eigenvalues. Since our estimate A^- is not exact, our solution is not exactly *on* the curve that describes the trade-off between variance and resolving power.

We can damp the SVD solution such that we obtain the same resolving length as with the approximate inverse A^-, or the same resolving power, or make a choice in between. We shall compare variance estimates for equally resolved models (that is, point B in Fig. C.2). We use the effective rank of the inverse matrix as a measure of the overall resolving power. Wiggins (1972) showed that the effective rank of the truncated SVD inverse (the number of eigenvectors used to construct the generalized inverse) is equal to the sum of the diagonal elements of R:

$$k_{eff} = \sum_{i=1}^{m} R_{ii}. \tag{C.23}$$

For R^{West} we find $k_{eff} = 21.4$. In Fig. C.3 we compare the estimated and the true values of R_{ii} for 22 eigenvectors, and similarly for R^{East} for which $k_{eff} = 16.3$ we choose 17 eigenvectors. Clearly, for well-resolved parameters with $R_{ii} > 0.5$ there is broad agreement; although R_{ii} may be in error by as much as 50 per cent, only a few 'unresolved' parameters are plotted as resolved. The few that have $R_{ii}^{SVD} < 0.1$ but for which our estimate exceeds 0.1 are all event or station corrections, not slowness parameters. This conclusion does not seem to depend on the exact choice of k_{eff}, since adding or subtracting an eigenvector affects only the ill-resolved parameters.

In Fig. C.4 we compare the part of the covariance matrix relating to the slowness parameters with their exact counterparts for east (again calculated with 17 eigenvectors) and west (calculated with 22 eigenvectors). The colour scale is chosen to highlight parameters with a large (co)variance; that is, for which the tomographic image might be suspect.

An eyeball comparison again shows broad agreement between the estimated $C_{\hat{x}}$, denoted by 'EST', and the exact ones ('SVD'). Variances in the east, where the ray density is less than the west, are generally higher than in the west. On the diagonal, many gaps correspond to unresolved parameters for which the variance is 'numerically' zero due to the minimum norm character of the solution. Generally, the order of magnitude of the variances is well reproduced by the estimations, as are groups of covariances around the diagonal. In the off-diagonal bands corresponding to nearest- neighbour cells, the estimations seem to be biased towards somewhat larger values, but far off the diagonal (that is, for parameters located further apart) the estimated covariance is lower than the true value. A^{West} is an order of magnitude larger in size than A^{East} but no strong dependence of accuracy on matrix size is evident. If anything, the estimations for $C_{\hat{x}}^{West}$ seem to be slightly better than for $C_{\hat{x}}^{East}$. Since there is also no reason *a priori* to assume that the accuracy degrades with the size of the matrix, we are confident that, even for very

large systems, the order of magnitude of the variance is estimated correctly.

C.5 Application to a large system

We have also tested the algorithm on a much larger system. While we have no ground truth to compare the results, we investigated the efficiency of the algorithm as well as the effect of neglecting small matrix elements. For this purpose we created a matrix A simulating a P wave tomography experiment covering central and eastern Asia, including the subduction in the northwest Pacific, using one year of seismicity (1993). The system - formulated without source/station correction terms - has 69043 rows and 9610 columns, and could not be handled with SVD even on a large computer. Using a linear spline parametrization (Thurber, 1983), with pivots roughly 200 km apart, the matrix has 5.9×10^6 nonzero elements (0.9 per cent of the total). 44 per cent of these are smaller than 1 per cent of the largest element, 21 per cent less than 0.1 per cent.

The inspection of several rows of $C_{\hat{x}}$, plotted as correlations to facilitate the colour scaling, gives further confidence in the results. Fig. C.5 gives these correlation coefficients for three locations, plotted in cross-subsections as a function of latitude, longitude and depth. We compare the resolution in three different geographical locations plotted in Fig. C.6. On the left in Fig. C.5, , the solution at point a, located near the surface is clearly well constrained horizontally, but suffers from lack of resolution in the depth direction. In the center of Fig. C.5, the solution in point b, located at 500 km depth just NE of lake Baikal, correlates with points as far as 1000 km away. Finally, on the right of Fig. C.5 one sees the effect of ray bundles for point c in the Japan slab, where the N-S cross-section evidently samples ray paths towards Australian stations, and where the lack of crossing ray paths at depth causes the elongated shape of the correlating structure.

On the Sun UltraSparc processor the computations of R and $C_{\hat{x}}$, including some overhead to calculate matrix statistics, take about 5 hr for 10^6 non-zero elements. When we neglect the smallest elements in A, computations of A^- are faster and we find that the computation time depends linearly on the number of non-zero elements of A (Fig. C.7). However, the accuracy is clearly affected by truncation. We tested this by counting the number of correlation coefficients ρ_{ij} larger than a certain threshold. When we truncate A_{ij} at a level as large as 10 per cent of the maximum, we greatly increase the speed of computation (by a factor of 6), but we lose about 60 per cent of the $\rho_{ij} > 0.8$, which are now underestimated in magnitude; this is even worse for smaller ρ_{ij} (Fig.C.8). Inspection of the actual covariances shows that it is mostly the smaller covariances that are affected. Since these probably belong to the ill-resolved parameters (the well resolved parameters are associated with large elements in A), the situation shown in Fig.C.7 may give a view that is too pessimistic.

A modest truncation level of 1 per cent may be acceptable if we only use the

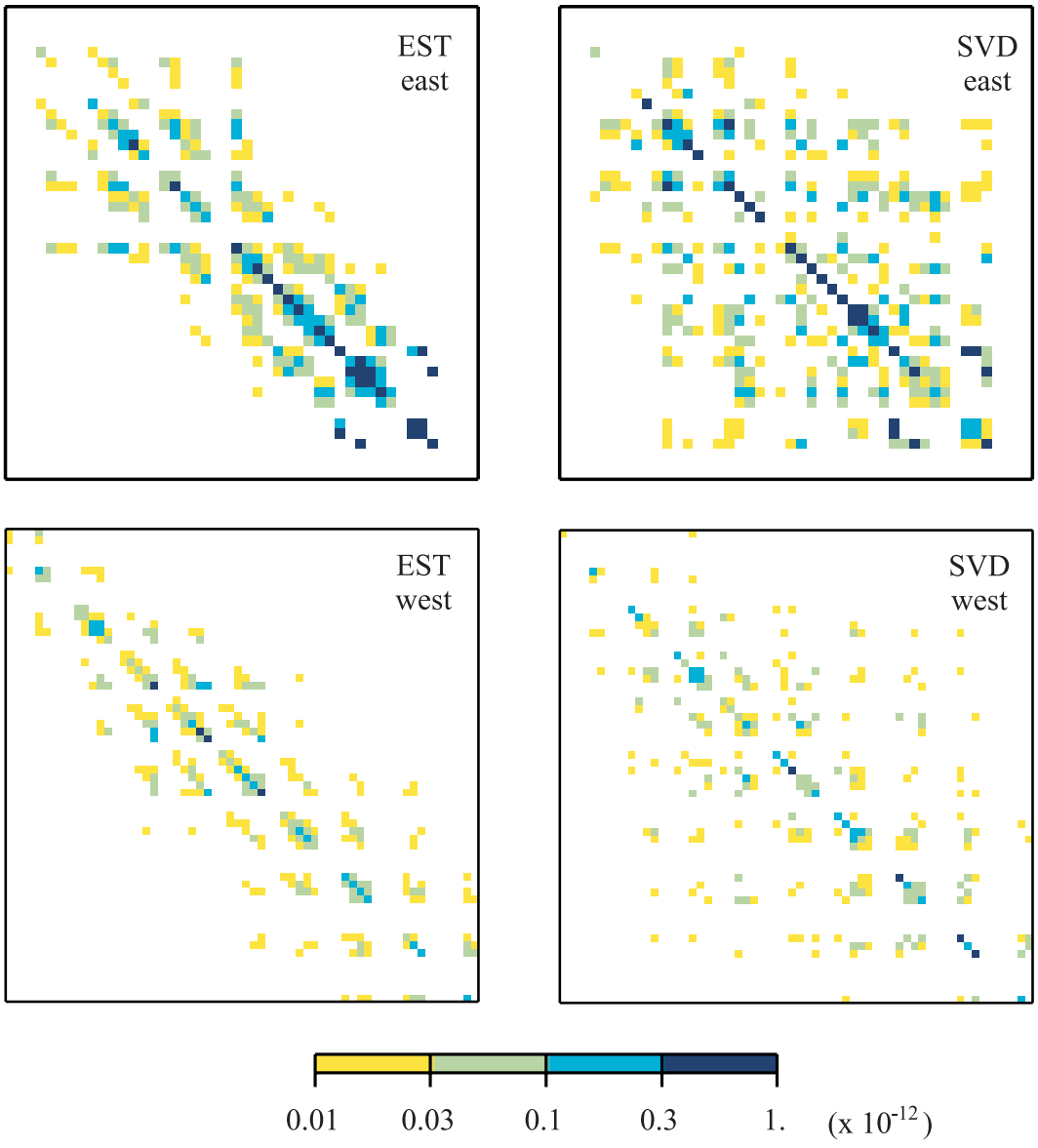


Figure C.4 Comparison between the estimated covariance matrix (EST) and the correct covariance matrix (SVD) as computed for A^{East} (top) and A^{West} (bottom). Only the covariances of the slowness parameters are plotted. The scale is in 10^{-12} s^2/m^2 for an assumed variance in delay times of 1.0 s^2.

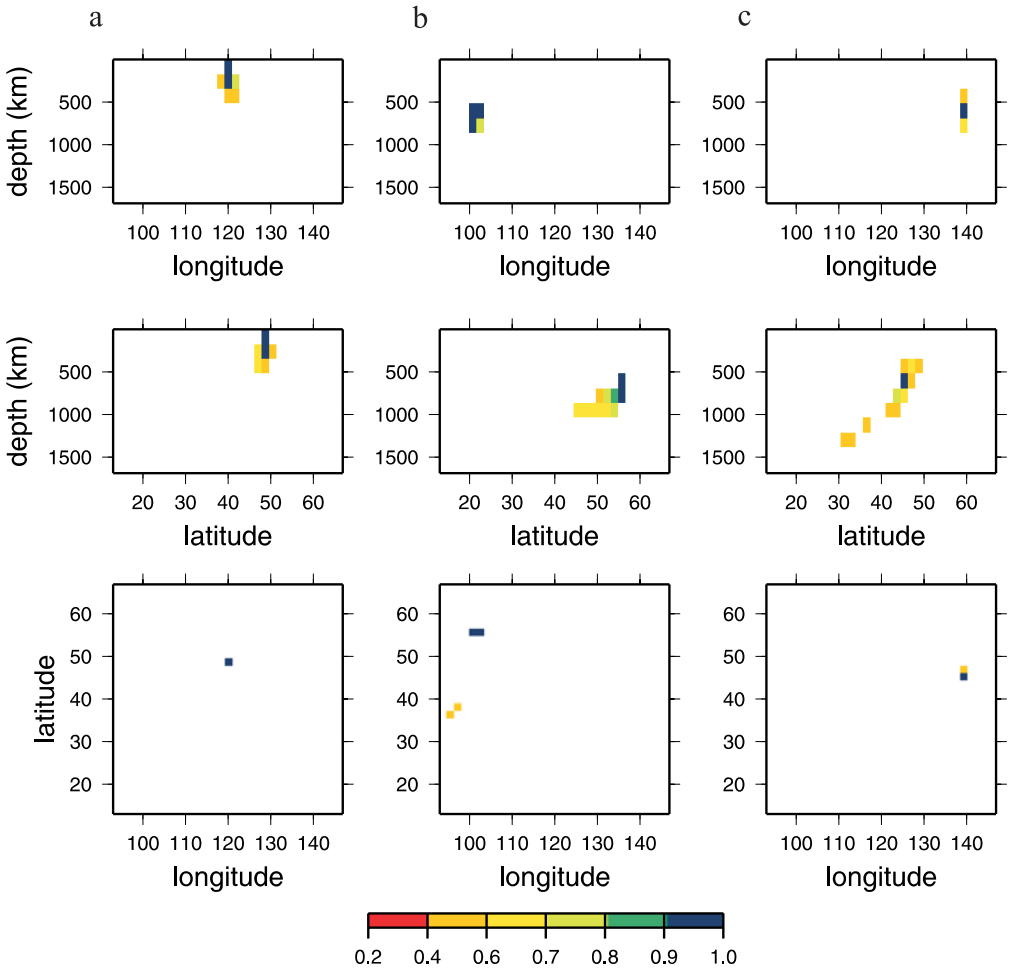


Figure C.5 This figure shows three examples of rows of the correlation matrix for the large (69043×9610) problem, plotted by way of cross subsections with fixed (from top to bottom) latitude, longitude, and depth, respectively. a: the P velocity near the surface below the Grand Khingan mountains in Mongolia, slowness variance 3.1×10^{-4} s^2/km^2, b: at 500 km depth NE of Lake Baikal, variance 6.1×10^{-4} s^2/km^2 c: at 500 km depth in the Japan subduction zone, variance 2.1×10^{-3} s^2/km^2. The variances quoted are for an assumed variance in the delay time observations of 1.0 s^2.

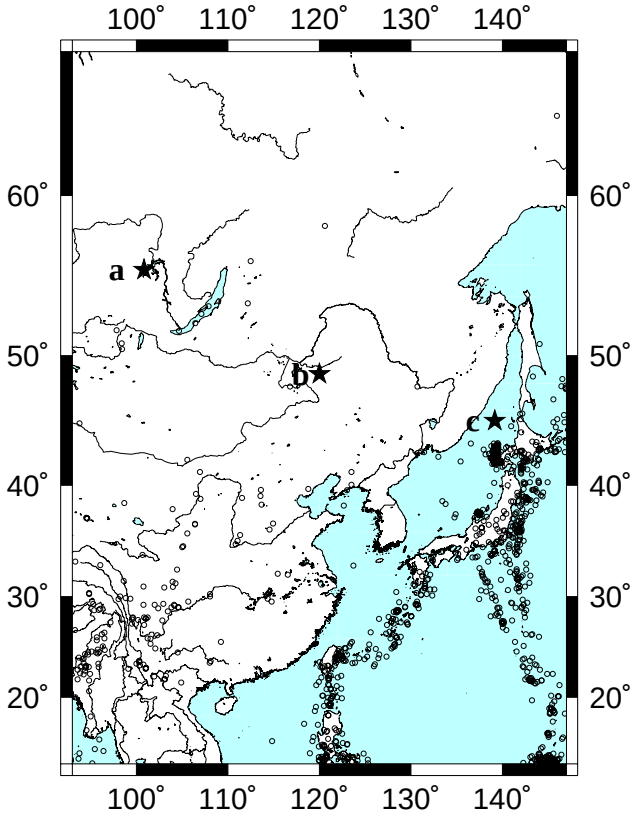


Figure C.6 Geographical locations of the cells shown in Figure C.5.

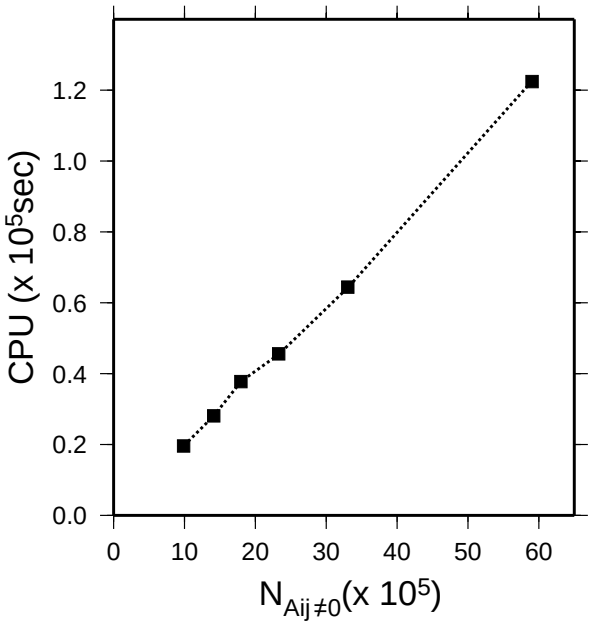


Figure C.7 CPU time needed to compute R and $C_{\hat{x}}$ on a Sun UltraSparc processor as a function of the number of nonzero elements in A.

ρ_{ij} for the purpose of reparametrization; this would result in a reduction of CPU time by a factor of about 2.

C.6 Discussion

Programming considerations The efficiency of the code depends strongly on some elementary programming considerations. The most commonly used scheme for storage of non-zero elements of the matrix A is row-wise. This involves storage overhead more then double the memory required to store just the values of non-zero A_{ij}, since one has to store the column number for each element as well as the number of nonzeros in each row of A. The scheme allows for fast computation of the product of both A and A^T with a vector, by looping through the elements of A in the same order as they are stored. Whilst this strategy can still be followed for the matrix product AA^T by repeatedly multiplying A with one of its own rows, it fails for A^TA, when A^T is multiplied with *columns* of A. Although the computation of D in (C.20) only requires the product AA^T, reverse products occur in the computation of $R = A^TDA$ and $C_{\hat{x}} = A^TD^2A$. We have found it most efficient to store A twice: once in row and once in column order.

We also note that neither A^TA nor AA^T can be expected to be truly sparse matrices and that one should avoid storage of these products. Fortunately, for the computation of D with (C.20) one needs only one row of AA^T at the time. For the

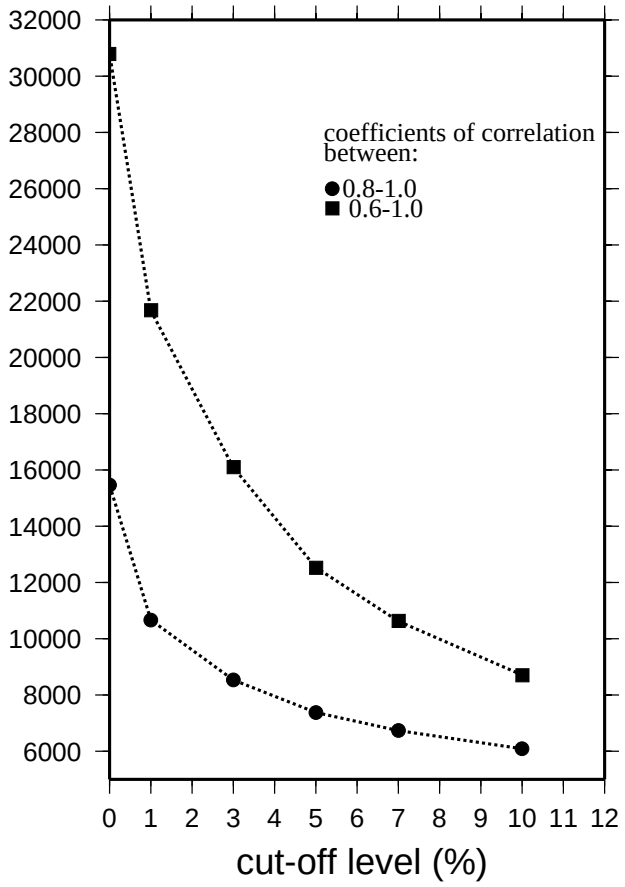


Figure C.8 Test of the effect of neglecting small elements of A by measuring the number of correlation coefficients larger than 0.6 and 0.8, respectively, as a function of the cut-off threshold, defined in per cent of the largest matrix element.

end-products R and $C_{\hat{x}}$ one may use mass storage to store these, generally row-wise in the form of 2D or 3D 'images', and for many applications a heavy truncation of smaller elements is allowed. Since the correlation matrix can be computed from $C_{\hat{x}}$ no separate storage of this is needed. If the correlation matrix is only computed to construct a sensible reparametrization of the model, an advisable strategy is to compute the diagonal elements of R first, then work from the smallest diagonal elements to compute the correlations within that row and regroup parameters. This will quickly eliminate the parameters with the worst resolution and avoid unnecessary calculations.

Sensitivity tests: Our method is similar to that of sensitivity tests (Spakman & Nolet, 1988; Leveque et al., 1993) but on n data vectors in which only one datum is equal to 1 and all others set to zero, rather than setting one model parameter to 1 to construct a right-hand side. We also restrict the matrix solver to just one iteration. One could in fact try to forego an analytical treatment as given here, and simply solve (C.13) using more iterations with a matrix solver such as LSQR. However, for large systems this will quickly saturate the available computer time. Since A^- will lose its sparse nature, this strategy may also invite storage problems, whilst truncating small elements of A^- may result in a loss of the extra precision gained by the extra iterations.

In comparison with sensitivity tests, our method gives a rough global estimate of both covariance and resolution, whereas sensitivity tests with spikes give an accurate image of the resolution, but for a few selected model parameters only, and no information on the covariances. Which is preferred depends on the application, and sometimes one may wish to use both methods, since they nicely complement each other. The main application we have in mind for our method is the reparametrization of the model by grouping of highly correlating parameters.

Lanczos iteration: Using the Ritz vectors (approxiamte eigenvectors) resulting from a Lanczos or conjugate gradient iteration to compute the resolution of large systems (Zhang & McMechan, 1995) as an alternative to explicit computation of the full eigensystem as in SVD. However, as pointed out by Deal & Nolet (1996), it quickly becomes infeasible to compute all the Ritz vectors needed to span the solution space as the size of A and its effective rank grows, due to a prolific growth of duplicate vectors in the conjugate gradient scheme. For example, the effective rank k_{eff} of the large matrix used in the previous section is estimated with (C.23) to be 574, and to compute that many eigenvectors is very costly, and for somewhat larger problems probably even beyond the reach of iterative algorithms. We certainly do not agree with Zhang & McMechan (1996) that it is sufficient to compute the uncertainty in $\hat{x}$ by considering only a subset of Ritz vectors constructed from the data vector b: for a correct estimation of model statistics one has to allow perturbations of the model in *all* directions. Nor could one assume that the selection of a subspace spanned by an incomplete set of Ritz vectors constitutes a good basis for reparametrization (smooth models for which the statistics could then be

computed). The reason is that the set of Ritz vectors is dependent on the data vector used to generate them and will ignore other directions in model space, even those that are associated with quite large eigenvalues (Deal & Nolet, 1996).

This leaves the method described in this paper as the only one to estimate the resolution and covariance matrix for large systems.

Damping: Strictly speaking, the estimated covariance and resolution is valid only for an inversion with the same damping properties as A^-. However, if the first iteration of a backprojection method such as LSQR defines the major characteristics of the solution, R and $C_{\hat{x}}$ should be useful as order of magnitude estimates. Since the variance in the data is not precisely known to begin with, attempts to increase the precision of R and $C_{\hat{x}}$ may seem futile. In principle, one could apply Newton iteration to obtain more precise versions of the inverse of A (the first correction would be $A^-(I-R)$) but the added computational effort will soon become prohibitive for really large systems. We may, however, investigate the case that (C.1) needs to be damped strongly to keep the propagation of data errors under control.

Since A^- already involves a minimum degree of damping, we are limited in controlling the damping of our approximate inverse. However, in many cases the signal-to-noise ratio of the data vector $\hat{b}$ may be very small. For example, Morelli & Dziewonski (1987) estimated the variance of teleseismic P-delays at 1 s^2, which implies a signal-to-noise ratio of the order of 1. For S waves, tomographic systems are even less accurate than that. In such cases it may actually be advisable to damp the solution even further. This can be done by means of a simple adapation of A^-. We may damp (C.1) in the same way as done in ridge regression, adding to the system (C.1) m equations of the form $\lambda x_i = 0$, where λ serves to weigh these equations against the 'true' constraints:

$$\begin{pmatrix} A \\ \lambda I_m \end{pmatrix} x \equiv Bx = \begin{pmatrix} b \\ 0 \end{pmatrix} \tag{C.24}$$

We then define the inverse as

$$B^- = B^T \hat{D} = (A^T D^{(1)} \quad \lambda D^{(2)}) \tag{C.25}$$

where

$$D^{(1)}_{kk} = \frac{(AA^T)_{kk}}{(\sum_{i=1}^{n}(AA^T)^2_{ik} + \lambda^2 \sum_{i=1}^{m} A^2_{ki})} \qquad (k = 1, ..., n) \tag{C.26}$$

$$D^{(2)}_{kk} = \frac{1}{\sum_{i=1}^{n} A^2_{ik} + \lambda^2} \qquad (k = 1, ..., m) \tag{C.27}$$

The definition of R now depends on a subtle interpretation of the damping. If we consider the added m equations $\lambda x = 0$ as true information on the model, that is, if we have reason to assume that the true earth model x^{true} is really 0, we would define R as before as B^-B and find:

$$R = A^T D^{(1)} A + \lambda^2 D^{(2)}. \tag{C.28}$$

More probably, the damping equations are not reflecting true information, but are introduced to bias the model towards 0 and reduce its variance. In that case we can only say that $Ax^{true} = b$, so that

$$x = (A^T D^{(1)} \quad \lambda D^{(2)}) \begin{pmatrix} Ax^{true} \\ 0 \end{pmatrix} = A^T D^{(1)} A x^{true}, \tag{C.29}$$

which implies

$$R = A^T D^{(1)} A \tag{C.30}$$

For the covariance we find, using the same interpretation of the damping:

$$C_{\hat{x}} = A^T D^{(1)\ 2} A. \tag{C.31}$$

For $\lambda = 0$, this reduces to (C.11), and the variances behave asymptotically as λ^{-2} for $\lambda \to \infty$, as we should expect.

C.7 Conclusions

We have developed an approximate but explicit expression for the covariance and resolution of the solution of tomographic systems. In contrast to schemes based on SVD or Lanczos iteration, this can be applied to very large matrices. The CPU time required varies linearly with the number of non-zero elements in the matrix. The accuracy has been investigated with small systems and was shown to be sufficient for most purposes. Work on the application of these results in a strategy for automatic reparametrization of the model is currently in progress.

Bibliography

Aki, K. & Lee, W.-H.-K. (1976). Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes; 1, A homogeneous initial model. *J. Geophys. Res.*, **81**, 4381–4399.

Aki, K., Christoffersson, A., & Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Geophys. J. Roy. Astron. Soc.*, **82**, 277–296.

Anderson, D. L. (1998). The edges of the mantle. In *The core-mantle boundary region*, pages 255–271. American Geophysical Union, Washington DC.

Anderson, D. L. (2000). The thermal state of the upper mantle: no role for mantle plumes. *Geophys. Res. Lett.*, **27**, 3623–3626.

Backus, G. & Gilbert, J. F. (1970). Uniqueness on the inversion of inaccurate gross Earth data. *Phil. Trans. Roy. Soc. Lond.*, **A266**, 123.

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. T. (1996). The Quickhull algorithm for convex hulls. *ACM Trans. on Mathematical Software*, **22**, 469–483.

Baumgardner, J. (1983). *A three-dimensional finite element model for mantle convection*. Ph.D. thesis, University of California, Los Angeles.

Bijwaard, H. & Spakman, W. (1999). Tomographic evidence for a narrow whole mantle plume below Iceland. *Earth Planet. Sci. Lett.*, **166**, 121–126.

Bijwaard, H., Spakman, W., & Engdahl, E. R. (1998). Closing the gap between regional and global travel time tomography. *J. Geophys. Res.*, **103**, 30055–30078.

Böhm, G. & Vesnaver, A. L. (1999). In quest of the grid. *Geophysics*, **64**(4), 1116–1125.

Böhm, G., Rossi, G., & Vesnaver, A. (1997). Adaptive regridding in 3D reflection tomography. *Annali di Geofisica*, **XL**(1), 69–83.

Böhm, G., Galuppo, P., & Vesnaver, A. (2000). 3d adaptive tomography using Delaunay triangles and Voronoi polygons. *Geophys. Prospecting*, **48**, 723–744.

Bolton, H. (1996). *Long period travel times and the structure of the mantle.* Ph.D. thesis, IGPP, University of California San Diego.

Bolton, H. & Masters, G. (2001). Travel times of P and S from global digital seismic networks: Implication for the relative variation of P and S velocity in the mantle. *J. Geophys. Res.*, **106**, 13527–13540.

Boschi, L. & Dziewonski, A. M. (1999). High- and low-resolution images of the Earth's mantle: Implications of different approaches to tomographic modeling. *J. Geophys. Res.*, **104**, 25567–25594.

Bunge, H.-P., Richards, M. A., & Baumgardner, J. R. (1997). A sensitivity study of the three-dimensional spherical mantle convection at 10 (super 8) Rayleigh number; effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *J. Geophys. Res.*, **102**, 11991–12007.

Cardimora, S. & Garmany, J. (1993). Smoothing operators for waveform tomographic imaging. *Geophysics*, **58**, 1646–1654.

Castle, J. C., Creager, K. C., Winchester, J. P., & Van der Hilst, R. D. (2000). Shear wave speeds at the base of the mantle. *J. Geophys. Res.*, **105**, 21543–21557.

Chiao, L.-Y. & Kuo, B.-Y. (2001). Multiscale seismic tomography. *Geophys. J. Int.*, **145**, 517–527.

Christensen, U. (1998). Fixed hotspots gone with wind. *Nature*, **391**, 739–740.

Courtillot, V., Jaupart, C., Manighetti, I., Tapponier, P., & Besse, J. (1999). On causal between flood basalts and continental breakup. *Earth Planet. Sci. Lett.*, **166**, 177–195.

Courtillot, V., Davaille, A., Besse, J., & Stock, J. (2003). Three distinct types of hotspots in the earth's mantle. *Earth Planet. Sci. Lett.*, **205**, 295–308.

Crough, S. T. & Jurdy, D. M. (1980). Subducted lithosphere, hot-spots and the geoid. *Earth Planet. Sci. Lett.*, **48**, 15–22.

Cserepes, L. & Yuen, D. A. (2000). On the possibility of a second kind of mantle plume. *Earth Planet. Sci. Lett.*, **183**, 61–71.

Curtis, A. (1999). Optimal design of focused experiments and surveys. *Geophys. J. Int.*, **139**, 205–215.

Dahlen, F. A., Hung, S.-H., & Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes — I. Theory. *Geophys. J. Int.*, **141**, 157–174.

Davaille, A. (1999). Simultaneous generation of hot spots and superswells by convection in a heterogeneous planetary mantle. *Nature*, **402**, 756–760.

Davaille, A., Girard, F., & Bars, M. L. (2002). How to anchor hotspots in a convecting mantle? *Earth Planet. Sci. Lett.*, **203**, 621–634.

Davies, G. F. (1988). Ocean bathymetry and mantle convection 1. large-scale flow and hotspots. *J. Geophys. Res.*, **93**, 10467–10480.

Deal, M. & Nolet, G. (1996). Comments on estimation of resolution and covariance for large matrix inversions by Zihang and McMechan. *Geophys. J. Int.*, **127**, 245–250.

Dziewonski, A. M. (1984). Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6. *J. Geophys. Res.*, **89**, 5929–5952.

Dziewonski, A. M. & Woodhouse, J. H. (1987). Global images of the Earth's Interior. *Science*, **236**, 37–48.

Dziewonski, A. M., Hager, B. H., & O'Connell, R. J. (1977). Large-scale heterogeneties in the lower mantle. *J. Geophys. Res.*, **82**, 239–255.

Dziewonski, A. M., Su, W.-J., & Woodward, R. L. (1991). Grand structures of the Earth's Interior. *Eos Trans. AGU*, **72**, 451.

Dziewonski, A. M., Forte, A. M., Su, W.-J., & Woodward, R. (1993). Seismic tomography and geodynamics. In K. Aki and R. Dmowska, editors, *Relating Geophysical Structures and Processes: The Jeffreys Volume*, volume 76, pages 67–105. AGU Geophys. Monograph.

Eberhart-Phillips, D. (1986). Three-dimensional velocity structure in nothern California Coast Ranges from inversion of local earthquake arrival times. *Bull. Seis. Soc. Am.*, **76**, 1025–1052.

Engdahl, E. R., Van der Hilst, R. D., & Buland, R. (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seis. Soc. Am.*, **88**, 722–743.

Evans, J. R. & Achauer, U. (1993). Teleseismic velocity tomography using ACH method: theory and application to continental studies. In H. M. Iyer and K. Hirahara, editors, *Seismic tomography*, pages 657–763. Chapman and Hall, London.

Flanagan, M. P. & Shearer, P. M. (1998). Global mapping of topography on transition zone velocity discontinuities by stacking ss precursors. *JGR*, **103**, 2673.

Foulger, G. R. (2003). Plumes, or plate tectonic processes? *Astronomy and Geophysics*, **43**, 6.19–6.23.

Foulger, G. R. & Natland, J. H. (2003). Is "hotspot" volcanism a consequence of plate tectonics? *Science*, **300**, 921–922.

Foulger, G. R. & Pearson, D. G. (2001). Is Iceland underlain by a plume in the lower mantle? Seismology and helium isotopes. *Geophys. J. Int.*, **145**, F1–F5.

Foulger, G. R., Pritchard, M. J., Julian, B. R., Evans, J. R., Allen, R. M., Nolet, G., Morgan, W. J., Bergsson, B. H., Erlendsson, P., Jakobsdottir, S., Ragnarsson, S., Stefansson, R., & Vogfjörd, K. (2001). Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle. *Geophys. J. Int.*, **146**, 504–530.

Fukao, Y., To, A., & Obayashi, M. (2003). Whole mantle P-wave tomography using P and PP-P data. *J. Geophys. Res.*, **108**.

Gelchinsky, B. (1985). The formulae for the calculation of the Fresnel zones or volumes. *J. Geophys.*, **57**, 33–41.

Grand, S. P. (1994). Mantle shear structure beneath the americas and surrounding oceans. *J. Geophys. Res.*, **99**, 11591–11621.

Grand, S. P., Van der Hilst, R. D., & Widiyantoro, S. (1997). Global Seismic Tomography: a snapshot of convection in the Earth. *GSA Today*, **7**, 1–3.

Gu, Y. J., Dziewonski, A. M., Su, W., & Ekström, G. (2001). Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities. *J. Geophys. Res.*, **106**, 11169–11199.

Hagendoorn, O. (1954). A process of seismic reflection interpretation. *Geophys. Prospecting*, **2**, 85–127.

Hounsfield, G. N. (1973). Computerized transverse axial scanning (tomography). part 1: Description of system. *Br J Radiol.*, **46**, 1016–1022.

Hung, S.-H., Dahlen, F. A., & Nolet, G. (2000). Fréchet kernels for finite-frequency travel times — II. Examples. *Geophys. J. Int.*, **141**, 175–203.

Husen, S. & Kissling, E. (2001). Local earthquake tomography between rays and waves: fat ray tomography. *Phys. Earth Planet. Inter.*, **123**, 129–149.

Inoue, H., Fukao, Y., Tanabe, K., & Ogata, Y. (1990). Whole mantle P-wave travel time tomography. *Phys. Earth Planet. Inter.*, **59**, 294–328.

Jackson, D. D. (1972). Interpretation of inaccurate, insufficient and inconsistent data. *Geophys. J. Roy. Astron. Soc.*, **28**, 97–109.

Karato, S.-I. (1993). Importance of anelasticity in the interpretation of seismic tomography. *GRL*, **20**, 1623–1626.

Kellogg, L. H., Hager, B. H., & Van der Hilst, R. D. (1999). Compositional stratification in the deep mantle. *Science*, **283**, 1881–1884.

Kennett, B. L. N. & Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophys. J. Int.*, **105**, 429–465.

Kennett, B. L. N. & Nolet, G. (1978). Resolution analysis for discrete systems. *Geophys. J. Roy. Astron. Soc.*, **53**, 413–425.

Kennett, B. L. N., Widiyantoro, S., & van der Hilst, R. D. (1998). Joint seismic tomography for bulk-sound and shear wavespeed in the earth's mantle. *J. Geophys. Res.*, **103**, 12,469–12,493.

Kissling, E., Husen, S., & Haslinger, F. (2001). Model parameterization in seismic tomography: a choice of consequence for the solution quality. *Phys. Earth Planet. Inter.*, **123**, 89–101.

Kravtsov, Y. A. (1988). Rays and caustics as physical objects. In E. Wolf, editor, *Progress in optics XXVI*, pages 229–348. Elsevier Science Publisher.

Kuo, B.-Y., Forsyth, D. W., & Wysession, M. (1987). Lateral heterogeneity and azimuthal anisotropy in the North Atlantic determined from SS-S differential travel times. *J. Geophys. Res.*, **92**, 6421–6436.

Larson, R. L. (1991). Latest pulse of Earth: evidence for a mid-Cretaceous superplume. *Geology*, **19**, 547–550.

Lay, T., Garnero, E. J., & Williams, Q. (2003). Partial melting in a thermochemical boundary layer at the base of the mantle. *Phys. Earth Planet. Inter.* submitted.

Lebedev, S. (2000). *The upper mantle beneath the Western Pacific and Southern Asia*. Ph.D. thesis, Princeton University.

Lebedev, S. & Nolet, G. (2003). Upper mantle beneath Southeast asia from S velocity tomography. *J. Geophys. Res.*, **108**.

Lee, W. H. K. & Stewart, S. W. (1981). *Principles and applications of microearthquake networks*, volume Supp. 2, page 293. NY Academic Press.

Leveque, J. J., Rivera, L., & Wittlinger, G. (1993). On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophys. J. Int.*, **115**, 313–318.

Li, X.-D. & Tanimoto, T. (1993). Waveforms of long-period body waves in a slightly aspherical Earth model. *Geophys. J. Int.*, **121**, 695–709.

Liu, X.-F. & Dziewonski, A. M. (1998). Global analysis of shear wave velocity anomalies in the lower-most mantle. In M. Gurnis, M. Wysession, E. Knittle, and B. Buffett, editors, *The Core-Mantle Boundary Region, Geodynamics Serie*, volume 28, pages 21–36. American Geophysical Union.

Marquering, H., Nolet, G., & Dahlen, F. A. (1998). Three-dimensional waveform sensitivity kernels. *Geophys. J. Int.*, **132**, 521–534.

Marquering, H., Dahlen, F. A., & Nolet, G. (1999). Three-dimensional sensitivity kernels for finite-frequency travel times: the banana-doughnut paradox. *Geophys. J. Int.*, **137**, 805–815.

Masters, G., Johnson, S., Laske, G., & Bolton, H. (1996). A shear-velocity model of the mantle. *Phil. Trans. Roy. Soc. Lond.*, **354**, 1385–1411.

Masters, G., Laske, G., Bolton, H., & Dziewonski, A. (2000). The relative behaviour of shear velocity, bulk sound speed and compressional velocity in the mantle: implications for chemical and thermal structure. In *Earth's deep interior: mineral physics and tomography from the atomic to the global scale*, volume 117, pages 63–87. AGU Geophys. Monograph.

McNutt, M. K. (1998). Superswells. *Rev. Geophysics*, **36**, 211–244.

Mégnin, C. & Romanowicz, B. (2000). The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms. *Geophys. J. Int.*, **143**, 709–728.

Menke, W. (1989). *Geophysical data analysis: discrete inverse theory*, volume 45. Academic Press, INC. International Geophysics Series.

Michelena, R. J. & Harris, J. M. (1991). Tomographic traveltime inversion using natural pixels. *Geophysics*, **56**, 635–644.

Michelini, A. (1995). An adaptive-grid formalism for traveltime tomography. *Geophys. J. Int.*, **121**, 489–510.

Montelli, R., Nolet, G., Masters, G., Dahlen, F. A., & Hung, S.-H. (2003). Global P and PP traveltime tomography : rays versus waves. *submitted to GJI.*

Morelli, A. & Dziewonski, A. M. (1987). The harmonic expansion approach to the retrieval of deep earth structure. In G. Nolet, editor, *Seismic Tomography*, pages 251–274. Reidel, Dordrecht.

Morgan, W. J. (1971). Convection plumes in the lower mantle. *Nature*, **230**, 42–43.

Morgan, W. J. (1972). Plate motions and deep mantle convection. *The Geolog. Soc. of America, Inc.*, **132**, 7–22.

Müller, G., Roth, M., & Korn, M. (1992). Seismic-wave traveltimes in random media. *Geophys. J. Int.*, **110**, 29–41.

Naganishi, I. & Suetsugu, D. (1986). Resolution matrix calculated by a tomographic inversion method. *J. Phys. Earth*, **34**, 95–99.

Nataf, H.-C. (2000). Seismic imaging of mantle plumes. *Ann. Rev. Earth Planet. Sci.*, **28**, 391–417.

Nataf, H.-C. & VanDecar, J. (1993). Seismological detection of a mantle plume? *Nature*, **364**, 115–120.

Ni, S., Tan, E., Gurnis, M., & Helmberger, D. (2002). Sharp sides to the african superplume. *Science*, **296**, 1850–1852.

Nolet, G. (1983). Inversio and resolution of linear tomographic systems. *Eos Trans. AGU*, **64**, 775–776.

Nolet, G. (1985). Solving or resolving inadequate and noisy tomographic systems. *J. Comp.Phys.*, **61**, 463–482.

Nolet, G. (1987). Seismic wave propagation and seismic tomography. In G. Nolet, editor, *Seismic Tomography*, pages 1–23. D. Reidel Publishing Company.

Nolet, G. (1990). Partitioned wave-form inversion and 2D structure under the NARS array. *J. Geophys. Res.*, **95**, 8513–8526.

Nolet, G. (1992). Imaging the deep earth: technical possibilities and theoretical limitations. In A. Roca, editor, *Proc. XXIIth Assembly ESC, Barcelona 1990*, pages 107–115.

Nolet, G. & Dahlen, F. A. (2000). Wavefront healing and the evolution of seismic delay times. *J. Geophys. Res.*, **105**, 19043–19054.

Nolet, G., Coutlee, C., & Clouser, R. (1998). Sn velocities in the western and eastern North America. *Geophys. Res. Lett.*, **25**, 1557–1560.

Nolet, G., Montelli, R., & Virieux, J. (1999). Explicit, approximate expressions for the resolution and *a posteriori* covariance of massive tomographic systems. *Geophys. J. Int.*, **138**, 36–44.

Norton, I. O. (1995). Plate motions in the North Pacific - the 43 Ma nonevent. *Tectonics*, **14**, 1080–1094.

Norton, I. O. (2000). Global hotspot reference frames and plate motion. In A. G. Union, editor, *The history and dynamics of global plate motions, Geophysical Monograph*, volume 121, pages 339–357.

Oganov, A. R., Brodholt, J. P., & Price, G. D. (2001). The elastic constants of $MgSiO_3$ perovskite at pressures and temperatures of the earth's mantle. *Nature*, **411**, 934–937.

Paige, C. C. & Saunders, M. A. (1982). LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans.Math.Software*, **8**, 43–71.

Parker, R. L. (1994). *Geophysical Inverse Theory*. Princeton University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77, Second Edition, The Art of Scientific Computing*, volume 1. Cambridge University Press.

Richards, M. A. & Lithgow-Bertelloni, C. (1996). Plate motion changes, the hawaiian-emperor bend, and the apparent success and failure of geodynamics models. *Earth Planet. Sci. Lett.*, **137**, 19–27.

Richards, M. A., Duncan, R. A., & Courtillot, V. E. (1989). Flood basalts and hot-spot tracks: plume heads and tails. *Science*, **246**, 103–107.

Ritsema, J. & Allen, R. M. (2003). The elusive mantle plume. *Earth Planet. Sci. Lett.*, **207**, 1–12.

Ritsema, J. & van Heijst, H. J. (2002). Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P,PP,PPP and PKPab traveltimes. *Geophys. J. Int.*, **149**, 482–489.

Ritsema, J., Ni, S., V., D. V. H. D., & Crotwell, H. P. (1998). Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa. *Geophys. Res. Lett.*, **25**, 4245–4248.

Ritsema, J., Van Heijst, H. J., & Woodhouse, J. H. (1999). Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science*, **286**, 1925–1928.

Romanowicz, B. & Gung, Y. C. (2002). Superplumes from the core-mantle boundary to the lithosphere: implications for heat flux. *Science*, **296**, 513–516.

Saltzer, R. L., van der Hilst, R. D., & Karason, H. (2001). Comparing P and S wave heterogeneity in the mantle. *Geophys. Res. Lett.*, **28**, 1335–1338.

Sambridge, M. & Gudmundsson, O. (1998). Tomographic systems of equations with irregular grids. *Geophys. J. Int.*, **103**, 773–781.

Sambridge, M., Braun, J., & McQueen, H. (1995). Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophys. J. Int.*, **122**, 837–857.

Shen, Y., Solomon, S., Bjarnason, I. T., Nolet, G., Morgan, W. J., Allen, R., Vogfjord, K., Jakobsdottir, S., Stefansson, R., Julian, B., & Foulger, G. (2001). Seismic evidence for a tilted mantle plume and North-South mantle flow beneath Iceland. *Earth Planet. Sci. Lett.*, **197**, 261–272.

Sleep, N. H. (1990). Hotspots and mantle plumes: some phenomenology. *J. Geophys. Res.*, **95**, 6715–6736.

Smith, W. H. F. & Sandwell, D. T. (1997). Global sea floor topography from satellite and ship depth soudings. *Science*, **277**, 1956–1962.

Spakman, W. (1991). Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia Minor. *Geophys. J. Int.*, **107**, 309–332.

Spakman, W. & Bijwaard, H. (2001). Optimization of cell parameterization for tomographic inverse problems. *Pure Appl. Geophys.*, **158**, 1401–1423.

Spakman, W. & Nolet, G. (1988). Imaging algorithms, accuracy and resolution in delay time tomography. In N. J. Vlaar, G. Nolet, M. J. R. Wortel, and S. A. P. L. Cloetingh, editors, *Mathematical Geophysics*, pages 155–188. Reidel.

Stacey, F. D. (1998). Thermoelasticity of a mineral composite and a reconsideration of lower mantle properties. *Phys. Earth Planet. Inter.*, **106**, 219–236.

Stark, P. B. & Nikolayev, D. I. (1993). Toward tubular tomography. *J. Geophys. Res.*, **98**, 8095–8106.

Steinberger, B. & O'Connell, R. J. (1998). Advection of plumes in mantle flow: implications for hotspot motion, mantle viscosity and plume distribution. *Geophys. J. Int.*, **132**, 412–434.

Stewart, R. C. (1984). Q and the rise and fall of a seismic pulse. *Geophys. J. Roy. Astron. Soc.*, **76**, 793–805.

Su, W.-J. & Dziewonski, A. M. (1997). Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. *Phys. Earth Planet. Inter.*, **100**, 135–156.

Su, W.-J., Woodward, R. L., & Dziewonski, A. M. (1994). Degree 12 model of shear velocity heterogeneity in the mantle. *J. Geophys. Res.*, **99**, 6945–6980.

Tackley, P. J. (2000). Mantle convection and plate tectonics: toward an integrated physical and chemical theory. *Science*, **288**, 2002–2007.

Tanimoto, T. (1990). Long-wavelength S-wave velocity structure throughout the mantle. *Geophys. J. Int.*, **100**, 327–336.

Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier, Amsterdam.

Tarantola, A. & Valette, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophysics*, **20**, 219–232.

Tarduno, J. A. & Cottrell, R. D. (1997). Paleomagnetic evidence for motion of the hawaiian hotspot during formation of the emperor seamounts. *Earth Planet. Sci. Lett.*, **153**, 171–180.

Thurber, C. H. (1983). Earthquake locations and three-dimensional structure in the Coyote Lake area, central California. *J. Geophys. Res.*, **88**, 8226–8236.

Tichelaar, B. W. & Ruff, L. R. (1989). How good are our best models? *Eos Trans. AGU*, **70**, 593–606.

Van der Hilst, R. D. & Karason, H. (1999). Compositional heterogeneity in the bottom 1000 kilometers of earth's mantle: toward a hybrid convection model. *Science*, **283**, 1885–1888.

Van der Hilst, R. D., Engdahl, E. R., Spakman, W., & Nolet, G. (1991). Tomographic imaging of subducted lithosphere below northwest pacific island arcs. *Nature*, **353**, 37–42.

Van der Hilst, R. D., Widiyantoro, S., & Engdahl, E. R. (1997). Evidence for deep mantle circulation from global tomography. *Nature*, **386**, 578–584.

van der Hilst, R. D., Widiyantoro, S., Creager, K. C., & McSweeney, T. J. (1998). Deep subduction and aspherical variations in P-wavespeed at the base of earth's mantle. In *The Core-Mantle Boundary Region*, volume Geodynamics 28, pages 5–20. American Geophysical Union, Washington DC.

Van der Voo, R., Spakman, W., & Bijwaard, H. (1999). Mesozoic subducted slabs under Siberia. *Nature*, **397**, 246–249.

VanDecar, J. C. & Crosson, R. S. (1990). Determination of teleseismic relative phase arrival times using multi-channel cross correlation and least squares. *Bull. Seis. Soc. Am.*, **80**, 150–169.

Vasco, D. & Johnson, L. (1998). Whole earth structure estimated from seismic arrival times. *J. Geophys. Res.*, **103**, 2633–2671.

Vasco, D. W. & Majer, E. L. (1993). Wavepath travel time tomography. *Geophys. J. Int.*, **115**, 1055–1069.

Vasco, D. W., Johnson, L. R., & Marques, O. (1999). Global earth structure: inference and assessment. *Geophys. J. Int.*, **137**, 381–407.

Vasco, D. W., Johnson, L. R., & Marques, O. (2003). Resolution, uncertainty, and whole earth tomography. *J. Geophys. Res.*, **108**, 2022–2048.

Vesnaver, A. (1994). Towards the uniqueness of tomographic inversion solutions. *J. of Seismic Explor.*, **3**, 323–334.

Vidale, J. E. & Schubert, G. (2001). Unsuccessful initial search for a midmantle chemical boundary with seismic arrays. *Geophys. Res. Lett.*, **28**, 859–862.

Watson, D. F. (1981). Computing the n-dimensional Delaunay tesselation with application to Voronoi polytopes. *Comput. J.*, **24**, 167–172.

Watson, D. F. (1992). *Contouring: A guide to the analysis and display of spatial data*. Pergamon, Oxford.

Wéber, Z. (2001). Optimizing model parameterization in 2d linearized seismic traveltime tomography. *Phys. Earth Planet. Inter.*, **124**, 33–43.

Wegener, A. (1924). *The Origin of Continents and Oceans*. Methuen, London.

Wielandt, E. (1987). On the validity of the ray approximation for interpreting delay times. In G. Nolet, editor, *Seismic Tomography*, pages 85–98. D. Reidel Publishing Company.

Wiggins, R. A. (1972). General linear inverse problem - Implication of surface waves and free oscillations for Earth structure. *Rev. Geophys. Space Phys.*, **10**, 251–285.

Wilson, J. T. (1963). A possible origin of the hawaiian islands. *Can. J. Physics*, **41**, 863–870.

Wilson, J. T. (1965). Convective currents and continental drifts. *Phil. Trans. R. Soc. London*, **258A**.

Woodhouse, J. H. & Dziewonski, A. M. (1984). Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms. *J. Geophys. Res.*, **89**, 5953–5986.

Woodward, M. J. (1992). Wave equation tomography. *Geophysics*, **57**, 15–26.

Woodward, R. L. & Masters, G. (1991). Global upper mantle structure from long-period differential travel times. *J. Geophys. Res.*, **96**, 6351–6377.

Yomogida, K. (1992). Fresnel zone inversion for lateral heterogeneities in the Earth. *Pageoph*, **138**, 391–406.

Yoshizawa, K. (2002). *Development and application of new techniques for surface wave tomography.* Ph.D. thesis, The Australian National University.

Zhang, J. & McMechan, G. A. (1995). Estimation of resolution and covariance of large matrix inversion. *Geophys. J. Int.*, **121**, 409–426.

Zhang, J. & McMechan, G. A. (1996). Reply to comment by M. M. Deal and G. Nolet on 'Estimation of resolution and covariance of large matrix inversions'. *Geophys. J. Int.*, **127**, 251–252.

Zhao, L., Jordan, T. H., & Chapman, C. H. (2000). Three-dimensional Fréchet differential kernels for seismic delay times. *Geophys. J. Int.*, **141**, 558–576.

Zhao, L., Chen, L., & Jordan, T. H. (2001). Full Three-dimensional tomography experiments in the Western Pacific Region. *Eos Trans. AGU*, **82**(S52F-0698), 47.

Zhong, S., Zuber, M. T., Moresi, L., & Gurnis, M. (2000). Role of temperature-dependent viscosity and surface plates in spherical shell modesl of mantle convection. *J. Geophys. Res.*, **105**, 11063–11082.