Global multiple-frequency *S*-wave tomography of the Earth's mantle.

Tomographie globale en ondes *S* multi-fréquences du manteau terrestre.

Thèse de doctorat
présentée à l'Université de Strasbourg par

Christophe Zaroli

Ecole et Observatoire des Sciences de la Terre
Institut de Physique du Globe de Strasbourg (UMR 7516)
Laboratoire de Sismologie

en vue de l'obtention du titre de
Docteur de l'Université de Strasbourg
Spécialité: Sciences de la Terre

Soutenue le 2 Septembre 2010
Devant la Commission d'Examen composée de:

Mr. Cara, Michel	Université de Strasbourg	Rapporteur
Mr. Debayle, Eric	Université de Strasbourg	Directeur de thèse
Mr. Nolet, Guust	Université de Nice	Rapporteur
Mr. Ricard, Yanick	Université de Lyon	Rapporteur
Mr. Sambridge, Malcolm	Australian National University	Co-directeur de thèse
Mr. Trampert, Jeannot	Utrecht University	Examinateur

Acknowledgements/Remerciements

Ainsi sonne déjà la fin de ma thèse. J'estime avoir été, pendant ces trois ans et demi, un thésard privilégié. Tout d'abord, j'ai eu la chance d'avoir deux directeurs de thèse, Eric Debayle (à l'IPGS[1], puis à l'université de Lyon) et Malcolm Sambridge (RSES[2], Canberra), dont les qualités humaines et scientifiques m'ont énormément inspiré dans ma recherche. Par ailleurs, mon sujet de thèse - la tomographie sismique globale - m'a passionné, et présente un intérêt majeur pour la communauté des Sciences de la Terre. Que demander de plus? Il ne me reste donc plus qu'à remercier toutes les personnes qui m'ont aidé à devenir docteur en sismologie!

This is already the end of my thesis. During these three and a half years, I have been a very fortunate PhD student. First of all, I have had two great co-supervisors, Eric Debayle (IPGS, now at the université de Lyon) and Malcolm Sambridge (RSES, Canberra). Both of them have incredible scientific and social skills. They have inspired me a lot in the way to conduct efficient research. Moreover, my thesis topic - global seismic tomography - has passionated me, and is of major interest for the Earth Sciences community. What more can I ask? Therefore, I only have to thank all the people who helped me to become doctor of seismology!

Eric Debayle

Merci *Eric* de m'avoir permis de rentrer dans la cour des grands (ou du moins des docteurs en sismologie), en me proposant ce très bon sujet de thèse! Merci de m'avoir toujours soutenu, encouragé, encadré, tout en me laissant la liberté nécessaire pour faire un vrai travail de recherche. J'espère que nos collaborations futures seront aussi fructueuses!

Malcolm Sambridge

Thanks a lot *Malcolm* for having kindly accepted to be the co-supervisor of my PhD. I have learned so much by working with you at RSES. My two stays in Australia were unforgettable experiences for me, from a scientific and personal point of view! Thanks for having taught me so much in science, through all our brainstormings, and also a lot more. I really hope that we will continue to work together !

Michel Cara

Merci *Michel* d'avoir initié mon intérêt pour la sismologie, et en particulier pour la tomographie, lors de vos cours en Master à l'IPGS. J'ai ensuite voulu approfondir mes connaissances en thèse, et je ne l'ai jamais regretté! Merci pour votre soutien quotidien pendant toute ma thèse. J'ai énormément aimé discuter avec vous pendant toutes ces années, et j'espère qu'on aura encore l'occasion de parler sismologie, risque sismique, montagne, etc. Enfin, merci d'avoir accepté d'être rapporteur et président de mon jury de thèse.

[1]Institut de Physique du Globe de Strasbourg
[2]Research School of Earth Sciences

Guust Nolet

Thank you *Guust* for having accepted to be in my jury and to report on my thesis. This was an honor for me. Your "banana-doughnut theory" has always excited me, and I have really been interested in working on this subject. I also wish to thank you for giving me the opportunity to work as a post-doc in your Globalseis project at Geoazur (Nice/Sophia Antipolis university).

Yanick Ricard

Merci *Yanick* pour avoir apporté un regard de grande qualité sur mon travail, en acceptant d'être rapporteur de ma thèse. J'ai ainsi pu pleinement réaliser l'intérêt scientifique d'intégrer mon modèle tomographique globale du manteau avec les résultats de d'autres disciplines en Sciences de la Terre (e.g. la géodynamique, la géochimie, la géologie, etc). Merci aussi pour m'avoir invité à donner un séminaire à l'université de Lyon, et de m'avoir si bien accueilli à plusieurs reprises dans votre labo.

Jeannot Trampert

Merci *Jeannot* d'avoir accepté d'être présent dans mon jury de thèse en tant qu'examinateur. Vos remarques à ma soutenance de thèse vont contribuer à m'aider à approfondir la dernière partie de mon travail de thèse. Merci aussi d'être venu à mon repas de thèse, avec toute la bande de thésards!

Jean-Jacques Lévêque

Merci *Jean-Jacques* pour toutes nos discussions, si passionnantes, qui m'ont tant aidé à avancer dans la dernière ligne droite de ma thèse (notamment pour la partie inversion). Merci aussi d'avoir toujours laissé la porte de votre bureau ouverte à toutes mes questions de thésard en dernière année! Vous avez toujours été de si bon conseil pour moi, que je tiens à vous en remercier encore tout particulièrement.

Luis Rivera & Alessia Maggi

Merci *Luis* et *Alessia* d'avoir partagé mon enthousiasme pour la sismologie pendant toute ma thèse. Nos discussions informelles (notamment sur les synthétiques et la sélection des phases sismiques) m'ont toujours appris énormément. J'espère qu'il y en aura plein d'autres.

Marc Grunberg

Sans ton aide en informatique, je n'aurai pas pu avoir de belles images tomographiques à montrer à la fin de ma thèse! Merci encore *Marc*, et j'espère qu'on continuera à montrer que informaticiens et sismologues sont faits pour s'entendre, au boulot et en dehors!

Jean-Marc Brendle & Xavier Rose

Merci aux deux informaticiens de l'IPGS, *Jean-Marc* et *Xavier*, qui m'ont toujours aidé, dans la bonne humeur, à surmonter certains problèmes informatiques que j'ai rencontrés. Merci aussi à Jean-Marc pour m'avoir apporté de la stabilité mentale pendant ma thèse, en m'enseignant l'art de la danse salsa!

Benoît Tauzin

Merci à mon ancien collègue de bureau, et ami, pour toutes nos discussions passionnées sur nos

sujets de recherches respectifs. J'espère que notre amitié de jeunes sismologues va perdurer pendant encore très longtemps sur l'échelle des temps géologiques!

Même si l'on aime son travail, cela ne peut suffire à être vraiment heureux dans la vie, et en particulier pendant une thèse. Je tiens donc tout particulièrement à remercier, ci-dessous, tous ceux qui ont contribué à faire de moi un thésard heureux!

Je tiens à remercier les équipe de sismologie de l'IPGS et de RSES pour la bonne humeur présente dans les labos! Merci également aux administrations des universités de Strasbourg et Canberra, qui m'ont tant facilité les tâches administratives, afin que je puisse me concentrer sur la recherche. Merci à tous les doctorants et post-doctorants: *Chloé*, *Mélanie*, *Zac*, *Vince*, *Julia*, *Ben*, *Aurore*, *Olivier*, *Antony*, *Maxime*, *Julien(s)*, *Valérie*, *Sophie*, etc (désolé pour ceux que j'oublie, j'ai une mémoire de poisson rouge, mais vous resterez tous dans mon coeur!). Mes années thèse resteront, grâce à vous, de superbes souvenirs pour moi.

Un énorme merci à toute ma *famille*, qui m'a toujours encouragé, soutenu et motivé dans mon parcours scolaire. Finalement, avoir été un "têtard" nous aura souvent fait rire, ce qui s'est avéré être un bon moyen de partager mon travail de thèse avec vous! Je souhaite aussi remercier mon oncle *Philippe*, qui était guide de haute montagne. L'ivresse des cimes que j'ai ressenti grâce à lui, lors de nos itinéraires hors-piste en ski de randonnée, m'a souvent permis de relativiser et de prendre du recul sur ma thèse.

Un grand merci à mes *parents*, qui m'ont toujours encouragé à réaliser un de mes rêves: faire de la recherche mon métier. Grâce à eux, je pense être sur la bonne voie pour y arriver! Je remercie aussi plus particulièrement mon *Papa*, dont les talents d'enseignant en informatique, à l'IUT, m'ont aidé à apprivoiser les subtilités du langage C. Merci aussi d'avoir toujours su écouter pendant des heures, au ski de fond ou en vélo, les dernières avancées de ma thèse!

Enfin, je remercie du fond du coeur celle qui partage ma vie, *Laura*. Après m'avoir déjà rejoint deux semaines à Canberra, merci de m'avoir encore suivi dans cette nouvelle aventure de post-doctorat à Nice. Si j'ai pu finir dans les temps ma thèse, c'est en grande partie grâce à ton soutien.

A mon grand-père...

Abstract

For better constraining the structure of the Earth's interior, new theoretical developments on seismic wave propagation have emerged in recent years, and received increasing attention in tomography. One of these new methods is the "multiple-frequency tomography", which aims at exploiting the frequency-dependency of body wave travel times related to diffraction effects. In this thesis, we have applied this method in order to obtain a "high-resolution" 3-D shear-wave tomographic model of the mantle, that could contribute to a better understanding of the Earth's dynamics.

Firstly, we have built a globally distributed dataset of ∼400,000 frequency-dependent *S*, *ScS* and *SS* travel times, within the 10–51 s period range. After common correction for physical dispersion due to intrinsic anelastic processes (Kanamori & Anderson 1977), we observe a residual dispersion on the order of 1–2 s. This dispersion occurs differently for *S*, *ScS* and *SS*, which is presumably related to their differing paths through the Earth. Our results show that: (1) Wavefront-healing phenomenon, produced by very low velocity anomalies, is observed in our *S* and, to a lesser extent, *SS* travel times; (2) A preferred sampling of high velocity scatterers located at the CMB may explain our observation that *ScS* waves travel faster at low-frequency than at high-frequency; (3) A frequency-dependent attenuation $q(\omega) \propto q_0 \times \omega^{-\alpha}$, with $\alpha \simeq 0.2$, is compatible with the globally averaged dispersion observed for *S* waves. This suggests that the residual dispersion observed in our data is, at least partly, related to seismic heterogeneity and attenuation in the Earth.

Secondly, in order to exploit this *structural* dispersion contained in our global dataset, we have built a "multiple-frequency" *SH*-wave tomographic model (*ZDS-S10*) of the mantle, using the "finite-frequency" formalism of Dahlen *et al.* (2000). Though more investigations are needed, we feel that *ZDS-S10* could provide complementary imaging of seismic structures present in the deep Earth (e.g. parts of slabs sinking in the mantle, hot rising plumes), that are still debated in the community.

Thirdly, we have presented a comparison of single- versus multi-band tomographic models, in order to assess, from a model point of view, the actual benefits of using multi-band rather than single-band data. Our results show that, though not very significant on average, the effect of using multi-band data in the inversion is not negligible. It allows to increase the contrast of some individual velocity anomalies, especially for those with significant amplitude. Expectations are that this could lead to refined tomographic imaging of small-scale objects in the Earth's interior (e.g. plumes), which may play a key role in mantle dynamics.

Résumé

Afin de mieux contraindre la structure interne de la Terre, de nouveaux développements théoriques, sur la propagation des ondes, ont émergé ces dernières années. Une de ces nouvelles méthodes est la "tomographie multi-fréquences", qui vise à exploiter la dépendance en fréquence des temps de parcours des ondes de volume, liée aux effets de diffraction. En utilisant cette méthode, cette thèse a pour objectif d'obtenir un modèle tomographique 3-D du manteau en ondes de cisaillement à "haute-résolution", qui puisse contribuer à améliorer nos connaissances sur la dynamique de la Terre.

Nous avons construit un jeu de données global de ∼400,000 temps de parcours d'ondes *S*, *ScS* et *SS*, mesurés dans la gamme de période 10–51 s. Après avoir corrigé de la dispersion physique liée aux processus anélastiques intrinsèques (Kanamori & Anderson 1977), nous observons une dispersion résiduelle de l'ordre de 1 à 2 s. Cette dispersion est différente pour les ondes *S*, *ScS* et *SS*, ce que nous présumons être lié à des trajets différents dans le manteau terrestre. Nos résultats montrent que: (1) Le phénomène de "guérison du front d'onde", produit par des anomalies à faible vitesse, est observé pour les temps de parcours des ondes *S* et *SS*; (2) Un échantillonage préférentiel de diffracteurs localisés à la CMB, et associés à des anomalies de vitesse élevée, pourrait expliquer pourquoi les ondes *ScS* semblent voyager plus rapidement à basse qu'à haute fréquence; (3) Une atténuation qui dépend de la fréquence, i.e. $q(\omega) \propto q_0 \times \omega^{-\alpha}$ avec $\alpha \simeq 0.2$, est compatible avec la dispersion globale moyenne des ondes *S*. Ceci suggère que la dispersion résiduelle observée dans nos données est, en partie, liée aux hétérogénéités sismiques et à l'atténuation dans la Terre.

Dans le but d'exploiter cette dispersion liée à la structure 3-D, nous avons construit un modèle tomographique multi-fréquences (*ZDS-S10*) de l'ensemble du manteau terrestre, en utilisant le formalisme à "fréquence-finie" de Dahlen *et al.* (2000). Bien qu'encore préliminaire, ce modèle devrait pouvoir apporter des images complémentaires sur certains objets du manteau (e.g. plaques en subduction, panaches mantelliques).

Nous avons essayé de quantifier les bénéfices de nos données multi-fréquences dans le modèle tomographique obtenu. Bien qu'assez faible en moyenne, l'apport des données multi-fréquences, par rapport à des données à une seule fréquence, n'est pas négligeable. En effet, cela semble pouvoir augmenter le contraste de certaines anomalies de vitesse, en particulier de celles avec une forte amplitude. Cela permettra peut-être de détecter des structures de petite taille (e.g. panaches mantelliques), pouvant jouer un rôle clé dans la dynamique du manteau.

Contents

Résumé étendu en français

Mieux comprendre la structure interne du globe terrestre

Depuis l'Antiquité, nombreux sont ceux qui se sont intéressés à la structure interne du globe terrestre: mathématiciens, philosophes, naturalistes, physiciens, géologues, géophysiciens, etc. C'est au cours du XXe siècle que l'analyse des enregistrements des sismographes a bouleversé notre vision de la structure interne de notre planète.

Notre connaissance actuelle des profondeurs de la Terre, et en particulier de la structure du manteau, qui représente 84% du volume terrestre, provient en grande partie de l'analyse des ondes sismiques générées par les tremblements de terre. Ces ondes étant sensibles à certaines propriétés des milieux dans lesquels elles se sont propagées, elles sont porteuses d'informations sur la structure de la Terre, informations qu'il est possible d'extraire des sismogrammes.

Une des motivations de la tomographie sismique du manteau est l'étude de la convection qui s'y produit. Les ondes sismiques étant sensibles à la température des milieux qu'elles traversent, on espère notamment pouvoir imager, par la tomographie sismique, ces cellules de convection.

Ma thèse s'inscrit dans ce contexte: construire un nouveau modèle tomographique de l'ensemble du manteau, avec suffisamment de détail pour contribuer à améliorer notre compréhension des processus géodynamiques du manteau.

Tomographie multi-fréquences: théorie et attentes

Depuis 25 ans, les sismologues utilisent les temps de parcours des ondes télésismiques pour construire des modèles tomographiques de plus en plus détaillés. Ces améliorations ont permis de révéler des comportements différents des plaques de subduction. Certaines restent dans le manteau supérieur (au-dessus de la discontinuité à 660 km de profondeur), tandis que d'autres plongent dans le manteau inférieur (e.g. Grand *et al.*, 1997 ; Albarède & van der Hilst, 1999 ; Fukao *et al.* 2001).

La plupart de ces images tomographiques ont été obtenues en utilisant la "théorie des rais". Cette théorie est basée sur l'hypothèse que les ondes sismiques sont sensibles à la structure de la Terre uniquement le long d'un trajet infiniment étroit (cf. figure 1). Ceci suppose que les ondes aient une fréquence infinie, c'est-à-dire une longueur d'onde nulle. Les ondes sismiques utilisées pour l'étude du manteau terrestre ont pourtant des longueurs d'onde allant de 10 à 1000 km, ou même plus encore. L'approximation "théorie des rais" ne rend donc compte qu'imparfaitement de la réalité. Son utilisation ne se justifie que dans le cas où la longueur d'onde est très inférieure à la taille des hétérogénéités rencontrées par l'onde.

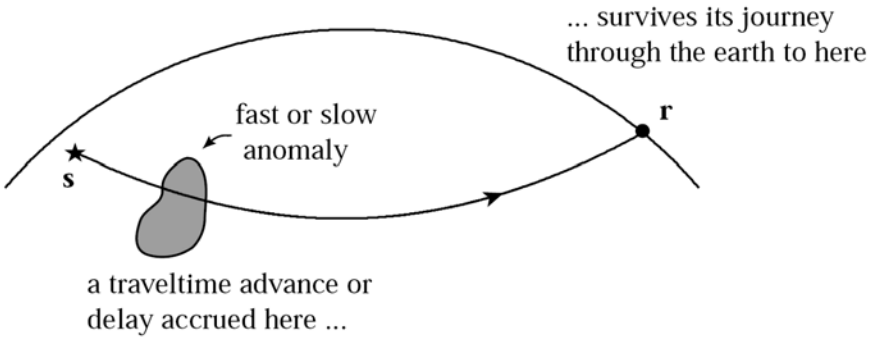


Figure 1: Illustration de la tomographie de temps de trajet d'ondes sismiques en théorie des rais. Une onde, supposée à fréquence infinie, acquiert une avance ou un retard de temps de parcours, qu'on appelle résidu, lors de son passage par une région rapide ou lente (i.e. l'anomalie grisée), le long de son trajet entre la source $\boldsymbol{s}$ *et le récepteur* $\boldsymbol{r}$*. On considère que l'onde sismique préserve ce résidu lors de la suite de son parcours jusqu'au récepteur. Source: Hung et. al (2001)*

La théorie des rais est souvent utilisée en tomographie, car sa simplicité la rend peu coûteuse en temps calcul. On peut ainsi l'appliquer à l'interprétation de gros jeux de données de temps d'arrivée d'ondes de volume, pour des études tomographiques à l'échelle globale. Néanmoins, si l'on souhaite détecter des hétérogénéités de taille équivalente ou inférieure à la longueur d'onde, l'utilisation d'une théorie plus raffinée est pourtant nécessaire. Dans cette thèse, nous avons utilisé une théorie dite à "fréquence finie", par opposition à la théorie des rais à "fréquence infinie".

Les sismologues ont récemment montré que les mesures de temps de trajet d'ondes sismiques, filtrées aux périodes utilisées pour l'étude du manteau, peuvent différer significativement des prédictions de la théorie des rais (e.g. Marquering *et. al* 1999; Dahlen *et. al* 2000; Hung *et. al* 2000; Zhao *et. al* 2000; Nolet & Dahlen 2000; Hung *et al.* 2001). Ces différences seraient dûes aux phénomènes de diffraction par des hétérogénéités (figure 2) qui ne sont pas pris en compte par la théorie des rais.

Les temps d'arrivée des ondes peuvent être mesurés à différentes périodes, en utilisant des techniques de cross-corrélation entre formes d'ondes observées et calculées, filtrées à différentes périodes (cf. figure 3). Du fait de la diffraction, on s'attend à observer une dépendance en fréquence des temps de trajet, qu'il devrait être possible d'exploiter en inversant simultanément des données à différentes périodes. Ce type de tomographie à "multi-fréquences" a le potentiel d'améliorer la "résolution" des modèles tomographiques, en utilisant le fait que les temps d'arrivées mesurés à différentes périodes contiennent une

information sur la taille des hétérogénéités. En effet, à chaque période la forme d'onde est influencée par un volume différent de la structure interne terrestre, à travers la zone de sensibilité (i.e. noyau de Fréchet) 3-D correspondante. En mesurant le temps de trajet d'une phase sismique à plusieurs périodes, on peut ainsi espérer augmenter le nombre d'informations indépendantes dans le problème inverse.

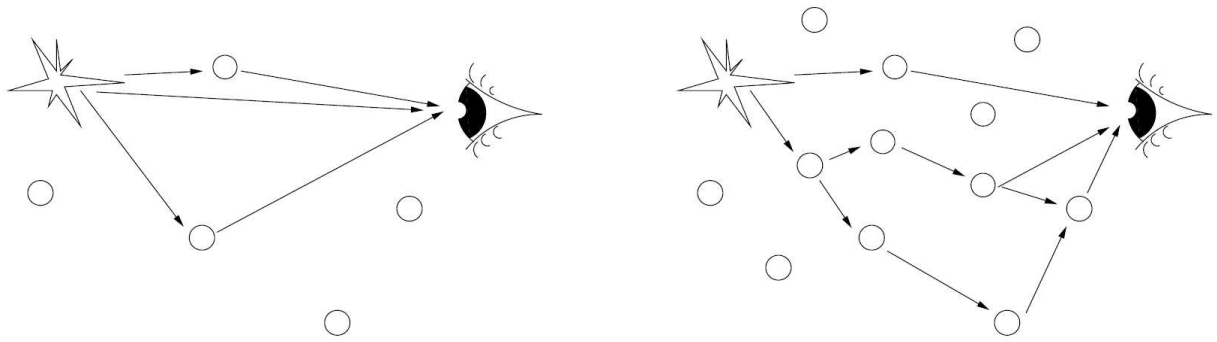


Figure 2: Illustration de la diffraction simple (à gauche) et de la diffraction multiple (à droite). Source: Anache (2008)

Le calcul efficace de ces noyaux de sensibilité, notamment en termes de temps de calcul pour des études tomographiques à l'échelle globale, a été rendu possible par le développement d'une théorie à "fréquence-finie" par Dahlen *et. al* (2000). Ces noyaux sont néanmoins calculés en utilisant une approximation paraxiale, qui néglige la diffraction multiple, et ne prend en compte que la diffraction simple (cf. figure 2 pour une illustration de ces notions). D'autres techniques plus complètes pourraient permettre une modélisation encore plus exacte des zone de sensibilité des ondes sismiques, comme les méthodes des Eléments Spectraux (e.g. Komatitsch *et. al* 2002; Tromp *et al.* 2005). Actuellement, ces approches sont néanmoins encore trop coûteuses en temps calcul, à l'échelle globale, dans la gamme de périodes de cette étude (entre 10 et 51 s).

Dans cette étude, nous avons donc utilisé le formalisme de Dahlen *et. al* (2000) pour le calcul des noyaux de sensibilité[3] à "fréquence finie".

Une base de données globale d'ondes S à différentes fréquences

Des études tomographiques récentes (e.g. Montelli *et al.* 2004a, 2004b, 2006b) suggèrent que l'utilisation d'une approche à "fréquence-finie" - via le formalisme de Dahlen *et. al* (2000) - pour interpréter les temps d'arrivée d'ondes télésismiques, peut permettre

[3]N.B. Le "noyau de sensibilité" est un concept utilisé depuis très longtemps (notamment en inversion), et pas du tout lié au concept "fréquence finie". Ainsi, le rai géométrique (cf. figure 1) est un noyau particulier.

de mieux imager des hétérogénéités de petite taille dans le manteau. Ces auteurs ont par exemple suggéré qu'elle permettrait de détecter des "panaches mantelliques". Ces panaches sont des structures très étroites, assimilées à des remontées de matière chaude, dont le diamètre est seulement de l'ordre de 200 km. Leur existence a été suggérée par Morgan (1971). L'observation de telles structures apporterait des informations importantes sur la dynamique du manteau terrestre.

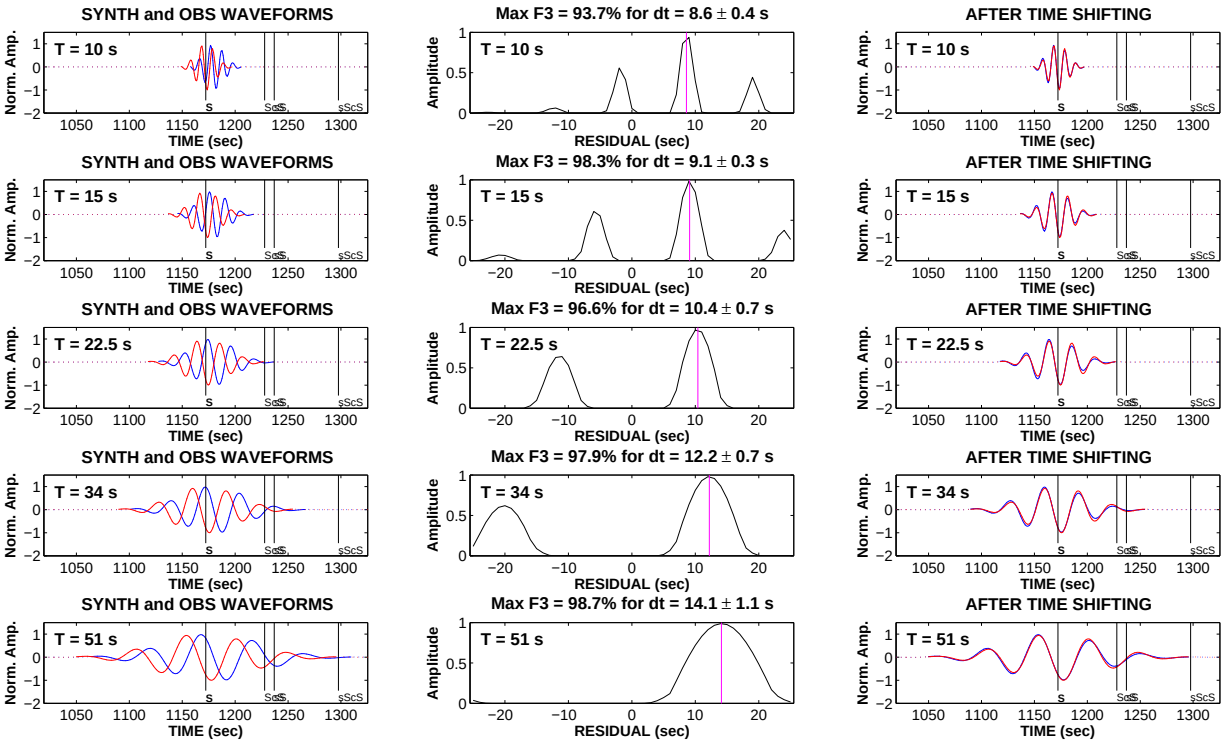


Figure 3: Procédure de mesure des résidus à différentes périodes dans le cas d'une onde S (chaque ligne correspond à une période de filtrage). Colonne de gauche: formes d'onde observée (en bleu) et synthétique (en rouge). Colonne du milieu: pour chaque période T, le résidu $\tau_m(T)$ correspond au délai τ pour lequel la fonction $F_3(\tau)$ (similaire à la fonction de cross-corrélation) atteint son maximum. Colonne de droite: formes d'onde observée et synthétique alignées après correction de $\tau_m(T)$.

L'observation de panaches mantelliques par les sismologues, grâce à l'utilisation d'une approche à "fréquence-finie", reste néanmoins encore controversé dans la communauté scientifique (e.g. Sieminski *et al.* 2004; de Hoop & van der Hilst 2005a; Dahlen & Nolet 2005; de Hoop & van der Hilst 2005b; Julian 2005; Trampert & Spetzler 2006; Montelli *et al.* 2006a; van der Hilst & de Hoop 2006; Boschi *et al.* 2006). Tout d'abord, les bénéfices réels des noyaux de sensibilité, calculés d'après le formalisme de Dahlen *et. al* (2000), sont

suspectés d'être moins important, dans les images finales de tomographie, que d'autres facteurs entrant en jeu dans l'inversion des données (régularisation, fit des données, etc). D'autre part, les premières tomographies globale à "fréquence finie" (Montelli *et. al* 2004a, 2004b 2006b) ont été construites à partir de jeux de données préexistants (Bolton & Masters 2001), dont les mesures avaient été obtenues, à une seule période, par des méthodes ne permettant pas d'exploiter complètement le formalisme de Dahlen *et. al* (2000).

Pour cette thèse, nous avions besoin d'une base de données globale de temps de trajet d'ondes de volume, mesurés à différentes périodes (cf. figure 3). Cette base de données n'existait pas. Dans un premier temps, nous l'avons donc construite, pour des ondes de cisaillement *S*, *ScS* et *SS*, dans la gamme de période 10–51 s. A chaque période, les temps d'arrivée sont mesurés par cross-corrélation entre formes d'ondes observées et calculées. Les ondes de cisaillement peuvent être directement combinées avec les ondes de surface dans une inversion jointe, ce qui à terme nous permettra d'augmenter la résolution de notre modèle tomographique dans le manteau supérieur. Notre base de donnée globale comprend ainsi environ 400 000 données.

Nous avons donc construit la première base de données globale pour des ondes de cisaillement mesurées à différentes périodes. A partir de là, nous avons pu:

1. Vérifier que la dépendance en fréquence, observée dans les temps d'arrivée des ondes télésismiques, peut effectivement être associée à des phénomènes de diffraction (e.g. wavefront-healing) liés aux hétérogénéités sismiques du manteau. Ceci justifie, du point de vue des données, l'utilisation des noyaux de sensibilité à fréquence finie.

2. Effectuer la première tomographie multi-fréquence de l'ensemble du manteau terrestre, en inversant "simultanément" tous les temps d'arrivée mesurés à différentes périodes. Le modèle tomographique obtenu est compatible avec d'autres modèles sur les grandes structures du manteau. Une analyse plus approfondie, notamment des structures de petite taille présentes dans notre modèle, devrait contribuer à mieux comprendre certains processus géodynamiques du manteau (cellules de convection, etc).

3. Quantifier l'apport réel, à l'échelle globale, d'une tomographie "multi-fréquences" (inversion de données à plusieurs périodes) par rapport à une tomographie basée sur des données mesurées à une seule période.

Observation d'une dispersion créée par la structure 3-D: enjeux pour la tomographie

Une des questions clés qui s'est posée dans ce travail de thèse peut se formuler de la manière suivante: "Existe-t-il une dispersion (i.e. dépendance des temps de parcours des ondes avec la fréquence) produite par les hétérogénéités 3-D du manteau dans notre base de donnée globale?".

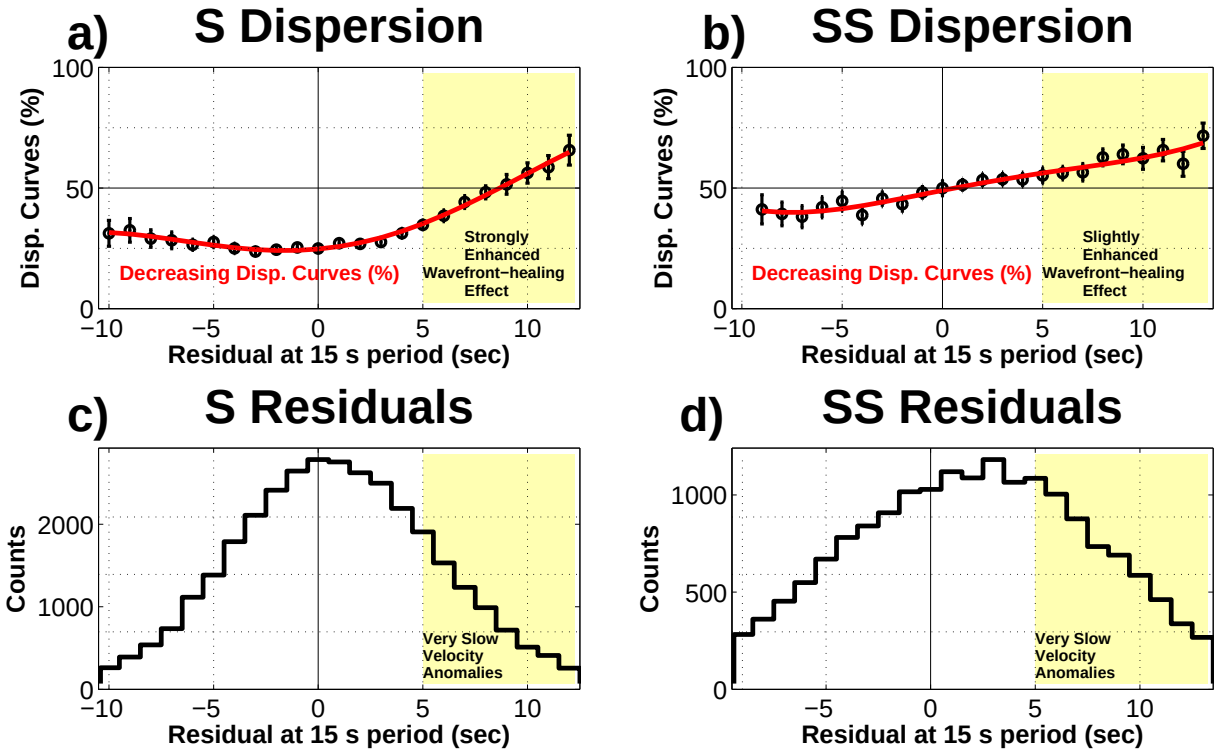


Figure 4: Nous considérons ∼32,000 courbes de dispersion d'ondes S et ∼17,500 courbes de dispersion d'ondes SS, pour lesquelles les résidus en temps ont été mesurés aux périodes 15, 22.5 et 34 s. ***a)*** *et* ***b)****: Une forte (faible) augmentation du pourcentage des courbes de dispersion décroissantes est observée pour les ondes S (SS) ayant traversé des anomalies de vitesse très faibles. Cette observation suggère que l'effet causé par le "wavefront-healing" est présent à l'échelle globale.* ***c)*** *et* ***d)****: Histogrammes des résidus en temps des ondes S et SS à 15 s de période, montrant les anomalies de vitesse très faibles (en jaune) accentuant le phénomène du wavefront-healing.*

Cette question est fondamentale puisque les sismologues supposent, depuis plusieurs décennies, que les ondes de volumes ne sont pas dispersives (à l'opposé des ondes de surface); cette approximation est valide lorsque la théorie des rais s'applique, c'est-à-dire à haute fréquence. Les phénomènes de diffraction, qui rendent les ondes de volume dispersives,

sont théoriquement connus. En pratique, ils ont commencé à être pris en compte dans les tomographies à grande échelles seulement depuis l'émergence du formalisme à "fréquence-finie" de Dahlen *et. al* (2000).

Cette thèse aura notamment confirmé que la dispersion des ondes de volume (de cisaillement) n'est pas négligeable à l'échelle globale. Nous avons montré qu'elle est de l'ordre de 1 à 2 s pour les ondes de type S, dans notre intervalle de période (entre 10 et 51 s). Pour autant, les scientifiques n'avaient jamais, jusqu'à présent, réussi à mettre en évidence, à l'échelle globale, une dispersion liée à la structure, c'est-à-dire aux hétérogénéités sismiques du manteau. Ce ne fut le cas que pour des études locales (e.g. Sigloch & Nolet 2006).

Dans un article[4] publié dans *Geophysical Journal International*, nous avons réussi à montrer que la dispersion observée dans notre base de donnée globale est, au moins en partie, liée aux hétérogénéités sismiques présentes dans le manteau terrestre. Par exemple, nos résultats montrent que le phénomène de "wavefront-healing", induit par des anomalies de vitesse très faibles, est observé dans nos temps de trajet des ondes S et SS (cf. figure 4).

Tomographie multi-fréquences: problème inverse et résultats

Nous avons construit une tomographie multi-fréquences de l'ensemble du manteau, afin d'exploiter l'information structurale que nous avons mise en évidence dans notre base de donnée globale.

Dans le contexte de la tomographie multi-fréquences, le problème inverse, linéaire, peut s'écrire: $\mathbf{d} = G\mathbf{m}$, avec $\mathbf{d}$ et $\mathbf{m}$ les vecteurs des données et des paramètres du modèle, respectivement, et G la matrice représentant la projection des noyaux de sensibilité sur la paramétrisation du modèle.

Dans cette étude, l'ensemble du manteau est paramétrisé par:

1. 18 couches concentriques, depuis la surface jusqu'à la CMB (Core Mantle Boundary, limite noyau/manteau).

2. Dans chaque couche, on définit un maillage qui est constitué de cellules. Ces cellules ont la forme de prisme dont la base et le sommet sont des triangles sphériques.

3. Le maillage des régions bien échantillonnées par les ondes sismiques consiste en des cellules de petite taille, afin d'extraire le maximum d'information des données. A

[4]Zaroli, C., Debayle E. & Sambridge, M., 2010. Frequency-dependent effects on global S-wave travel times: wavefront-healing, scattering and attenuation, *Geophys. J. Int.*, **182**, 1025–1042.

l'opposé, le maillage des régions mal échantillonnées consiste en des cellules de grande taille. La paramétrisation obtenue est donc "irrégulière" (cf. figure 5).

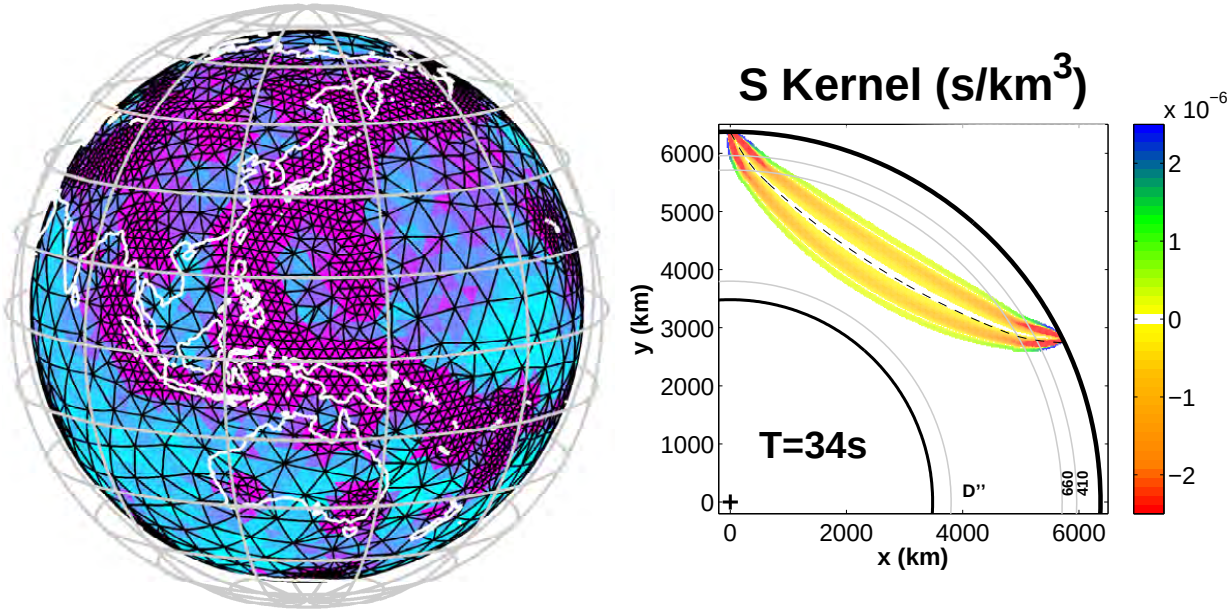


Figure 5: ***Figure de gauche:*** *Illustration de la paramétrisation dans le cas de la sixième couche (530–660 km de profondeur). Maillage optimal (triangles noirs), superposé à la fonction de résolution, qui reflète essentiellement la densité de rais (l'échelle de couleur est linéaire: de 210 km, en cyan, jusqu'à 850 km, en magenta).* ***Figure de droite:*** *noyau de sensibilité d'une onde S à 34 s de période (vue en 2-D).*

Les noyaux 3-D à "fréquence finie" (Dahlen *et al.* 2000) ont été calculés afin d'associer une zone de sensibilité pour chaque onde sismique, en fonction de sa période dominante (cf. figure 5). Nous avons utilisé une fonction source temporelle de type Gaussien, afin d'obtenir des expressions analytiques pour les noyaux, ce qui permet de calculer plus rapidement des centaines de milliers de noyaux (nécessaires dans cette thèse).

Finalement, nous avons développé un programme qui permet de projeter tous les noyaux sur la paramétrisation du modèle, et de calculer la matrice G en quelques jours seulement. Nous avons donc été en mesure d'inverser, simultanément, toutes nos données multi-fréquences. L'algorithme d'inversion utilisé est LSQR (Paige & Saunders 1982). Nous avons nommé le modèle tomographique "multi-fréquences" résultant: *ZDS-S10*.

Ce modèle reste préliminaire: nous pensons pouvoir encore l'améliorer significativement, par exemple en relocalisant les séismes. Au premier ordre, il est similaire à d'autres modèles globaux, provenant d'études antérieures (e.g. Ritsema *et. al* 1999; Montelli *et. al*

2006b; Houser *et. al* 2008). La convection à très grande échelle, dans le manteau inférieur, est par exemple clairement observée (figure 6). De plus, notre modèle *ZDS-S10* semble contenir beaucoup de détails nouveaux (cf. annexe E), qui pourraient contenir une information nouvelle sur les hétérogénéités de tailles intermédiaires (quelques centaines de km). Les résultats sont donc extrêmement prometteurs.

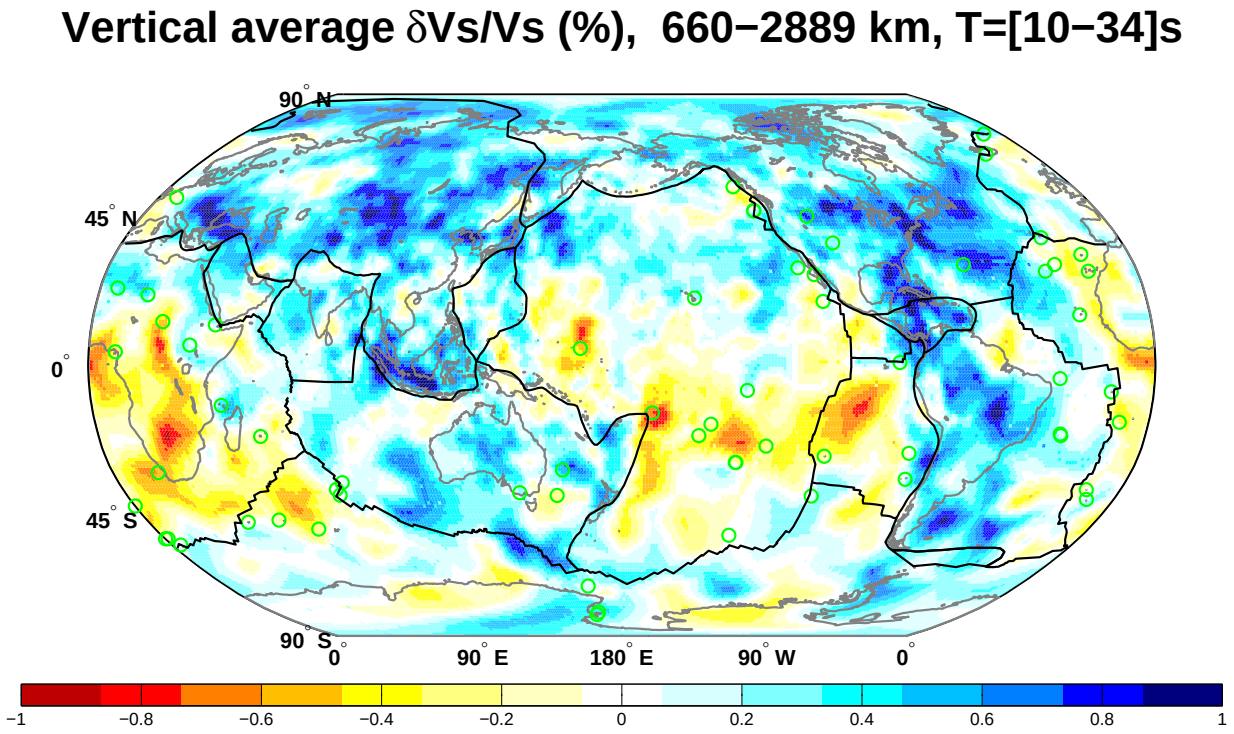


Figure 6: Moyenne verticale de l'ensemble du manteau inférieur (entre 660 km et 2889 km de profondeur) de la perturbation de vitesse $\delta V_S/V_S$ (en %) de notre modèle tomographique "multi-fréquences" (ZDS-S10). Il est ainsi plus facile de distinguer les structures verticales continues. Couleurs chaudes: zones lentes; couleurs froides: zones rapides. Cercles verts: points chauds; ligne continue grise:continents; ligne continue noire: plaques tectoniques.

Puisque ce modèle tomographique est encore préliminaire, nous avons préféré, dans un premier temps, tester et discuter sa robustesse, et quantifier le bénéfice des données multi-fréquences. Nous prévoyons de faire l'interprétation du modèle *ZDS-S10*, d'un point de vue géodynamique, dans un deuxième temps.

Tomographie multi-fréquence: raffiner les images du manteau terrestre

Nous avons explicitement montré que notre base de donnée globale contient une dispersion structurale. Néanmoins, les effets liés aux phénomènes de diffraction sont très

subtiles et n'ont pas été faciles à mettre directement en évidence dans nos données. On peut donc se demander si, après inversion des données, le modèle tomographique multi-fréquences contient, significativement, plus d'information sur la structure interne du manteau, qu'un modèle obtenu avec des données à une seule fréquence?

En d'autres termes, d'autres facteurs jouant un rôle dans l'inversion, tel que la régularisation, ne peuvent-ils pas masquer l'information structurale contenue dans nos données multi-fréquences? Nous avons tenté d'apporter un début de réponse à cette question, en proposant une comparaison entre notre modèle "multi-fréquences", obtenu avec des données mesurées à des périodes de 10, 15, 22.5 et 34 s, et un modèle "mono-fréquence", obtenu avec uniquement les données mesurées à la période de 22.5 s. Les deux modèles ont été choisis de telle sorte qu'ils expliquent, statistiquement, de la mème manière leurs données communes, c'est-à-dire les mesures à 22.5 s de période.

Nos résultats montrent que prendre en compte des mesures à différentes fréquences semble pouvoir raffiner les images tomographiques, en fournissant des contraintes supplémentaires sur la structure 3-D du manteau (cf. figure 7), mais de manière modérée en moyenne. Ce type de "raffinement" est potentiellement très important dans le cas d'anomalies d'amplitude significative, mais de petite taille. En effet, l'augmentation du contraste de telles anomalies dans les images tomographiques pourrait, à terme, convaincre le sismologue d'interpréter l'anomalie comme étant structurale, et non pas un artefact de méthode ou de mesure.

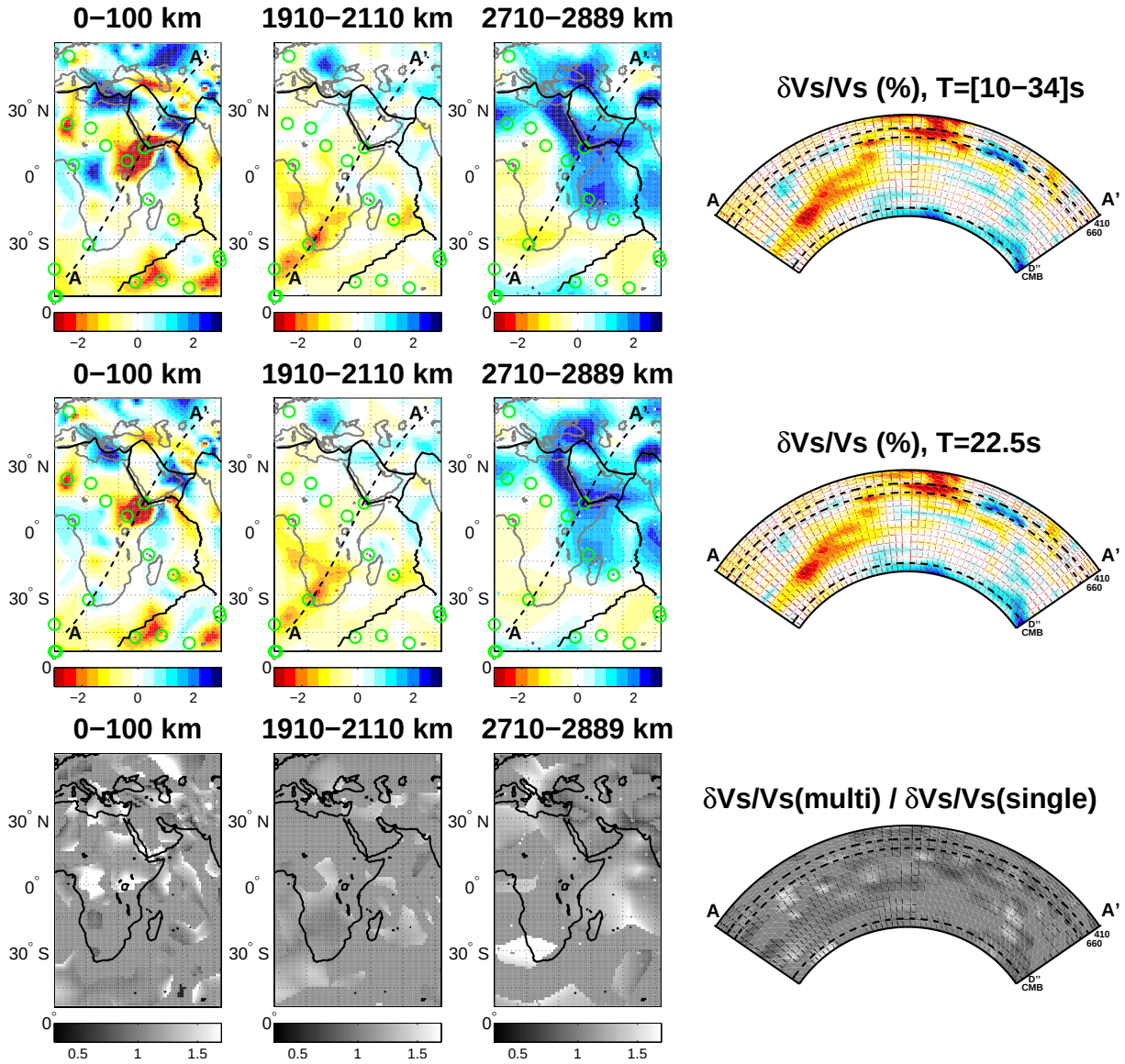


Figure 7: Vers une augmentation du contraste des anomalies de vitesse, $(\delta V_S/V_S)$, grâce aux données multi-fréquences. Exemple de l'anomalie à faible vitesse sous l'Afrique. ***Ligne du haut:*** *modèle tomographique "multi-fréquences" $(\delta V_S/V_S)_{multi}$.* ***Ligne du milieu:*** *modèle tomographique "mono-fréquence" $(\delta V_S/V_S)_{single}$.* ***Ligne du bas:*** *rapport $(\delta V_S/V_S)_{multi}/(\delta V_S/V_S)_{single}$, calculé pour les anomalies supérieures à 0.4% (en valeur absolue). Les valeurs du rapport supérieures à 1 (en blanc) signifient une augmentation de contraste dans le modèle "multi-fréquences", par rapport au modèle "mono-fréquence".*

Conclusion et perspectives

Dans l'objectif d'améliorer l'imagerie de la structure interne du manteau terrestre, de nouveaux développements méthodologiques reçoivent de plus en plus d'attention en tomographie. Une de ces méthodes est la tomographie multi-fréquences, que nous avons appliquée dans cette étude. Notre but étant de construire un modèle tomographique, en ondes de cisaillement, à "haute-résolution" de l'ensemble du manteau terrestre.

Pour ce faire, nous avons construit une base de donnée globale d'environ 400 000 temps de trajets d'ondes S dépendant de la fréquence. Une technique automatique a été utilisée pour mesurer les temps de trajet des ondes télésismiques S, ScS et SS, à 10, 15, 22.5, 34 et 51 s de période.

Après correction de la dispersion physique liée aux processus anélastiques intrinsèques de la Terre, nous avons observé une dispersion résiduelle de l'ordre de 1–2 s, dans l'intervalle de période considéré. Cette dispersion semble différente pour les ondes S, ScS et SS, ce que nous présumons être liée à leurs trajets différents dans le manteau terrestre. Nos observations suggèrent qu'une dispersion résiduelle "structurale" est explicitement présente dans nos données, que nous avons donc incorporée dans une tomographie multi-fréquences du manteau.

Le modèle tomographique multi-fréquences obtenu (*ZDS-S10*) s'est avéré compatible avec d'autres études antérieures. Enfin, nous avons essayé de quantifier les changements, dans les images tomographiques, liés à l'utilisation de données multi-fréquences - par rapport à l'utilisation de données à une seule période. Nos résultats montrent que les différences sont, en moyenne, faibles, mais qu'elles peuvent aussi, pour certaines parties du modèle, être très significatives. Ceci mènera peut-être à des images tomographiques plus fines, notamment pour des structures de petite taille mais d'une grande importance dans la dynamique du manteau (e.g. panaches mantelliques).

En conclusion, cette thèse montre, de manière explicite, que le traitement des données (enregistrement des sismogrammes, méthode de mesure, etc) est assez précis pour que la dispersion résiduelle observée des temps d'arrivée d'ondes de cisaillement contienne une information *structurale*, au moins en partie. Ceci justifie donc, du point de vue des données, l'utilisation d'une théorie "à fréquence-finie" qui puisse, à l'opposé de la théorie des rais, prendre en compte les effets des phénomènes de diffraction (e.g. wavefront-healing, scattering) que subissent les ondes télésismiques. Enfin, il semble que du point de vue des images tomographiques, l'inversion de données multi-fréquences soit bénéfique, en permettant de raffiner localement les contrastes des anomalies.

Dans le futur, différentes améliorations peuvent encore être apportées à cette étude

tomographique. On pourra par exemple prendre en compte la relocalisation des séismes dans l'inversion. En effet, une meilleure localisation des séismes devrait supprimer une partie des erreurs présente dans les données, et donc, *a fortiori*, nous permettre de mieux exploiter l'information, subtile, contenue dans les données multi-fréquences. L'enjeu principal est, à terme, d'être capable de détecter grâce aux données "multi-fréquences" des structures de petite taille, qui ne ressortent pas clairement dans des images tomographiques basées sur des mesures "mono-fréquence".

Introduction

Better understanding the Earth's interior with seismic tomography.

A large contribution to our knowledge of the structure and evolution of our planet has been obtained from the analysis of seismic waves triggered by earthquakes. As these waves are sensitive to some properties of the media through which they propagate, they contain information on the Earth's structure. Global seismic tomography aims at extracting this information from seismograms.

The dynamics of our planet largely occurs in the mantle, which represents 84 per cent of the total volume of the Earth, and extends from the base of the crust (a shallow layer that does not exceed 70 km thickness), to the Core Mantle Boundary (CMB), at 2889 km depth. A detailed knowledge of the geometry and amplitude of seismic anomalies within the mantle is crucial for better constraining the physical parameters and the forms of mantle convection. This requires high resolution 3-D seismic images of the entire mantle, that can be compared directly with geodynamical and geochemical findings.

The TOMOGLOB project.

This thesis is part of the TOMOGLOB[5] project, which aims at building the first French 3-D tomographic model of the entire Earth's mantle, from the crust to the CMB. According to the project:

1. this model will be mainly constrained in the upper mantle with the analysis of surface waves (Debayle *et al.* 2005);

2. the topography of the mantle transition zone seismic discontinuities will be included, with the analysis of *Ps* receiver functions and *SS* precursors (Tauzin *et al.* 2008);

3. this model will be constrained in the lower mantle with the analysis of long-period body waves (*S*, *ScS*, *SS*) using a "finite-frequency" approach - this is the contribution from this thesis;

4. the final phase of the TOMOGLOB project will consist in jointly invert for the complete data sets, including surface waves, converted/reflected waves and long-period body waves.

[5]The TOMOGLOB project has been supported by the young researcher ANR (Agence Nationale de la Recherche) no ANR-06-JCJC-0060. This project, leaded by E. Debayle, started in January 2007 for a four years period, involving a young and motivated scientific team (Debayle, E., Sambridge, M., Vergne, J., Maggi, A., Tauzin, B., and Zaroli, C.).

The resulting 3-D model is expected to have an important impact on the community, by contributing to answer several crucial geodynamical questions (e.g. size of convection cells, vertical circulation within the mantle, material exchanges between upper and lower mantle).

In the context of the TOMOGLOB project, the aim of this thesis is then to obtain a "high-resolution" 3-D shear-wave tomographic model of the Earth's (lower) mantle, that could contribute to a better understanding of mantle dynamics.

Multiple-frequency tomography: towards higher-resolution seismic imaging.

The "resolution" of global body wave 3-D seismic tomographic models has significantly improved in the last 25 years, thanks to growing international seismic networks, improved computational facilities, and development of new seismological tools which extract more information from seismograms. Until recently, ray theory (RT) formed the backbone of global seismic tomography, mainly because of its simplicity and short computing time, and has largely contributed to better resolve the 3-D seismic structure of the deep Earth (e.g. Grand *et al.* 1997; Albarède & van der Hilst, 1999; Fukao *et al.* 2001; Romanowicz 2003). RT is based on the approximation that the travel time of a body wave only depends upon the 3-D seismic structure along the (infinitesimally narrow) geometrical ray path (figure 1.1, of chapter 1). This assumes that seismic waves have an infinite-frequency, or a zero wavelength. In reality, seismic waves have wavelengths ranging from 10 to 1,000 km, or even more. RT is then only applicable if the heterogeneities are much larger than the Fresnel zone. Therefore, it breaks down when used for imaging small scale heterogeneities, because diffraction effects make travel time (and amplitude) anomalies dependent on Earth structure in a 3-D region around the geometrical ray path. Since wave diffraction phenomena are not taken into account in RT based seismic tomography, it seems that progress towards "higher-resolution" imaging of small-scale features present in the mantle requires a movement away from RT.

Recently, "finite-frequency" approaches (in contrast to the "infinite-frequency" RT) have emerged in seismic tomography (e.g. Marquering *et al.* 1998; Dahlen *et al.* 2000; Zhao *et al.* 2000; Komatitsch *et. al* 2002; Calvet & Chevrot 2005; Tromp *et al.* 2005; Nissen-Meyer & Dahlen 2007), in order to take wave diffraction effects into account, and make the imaging of smaller objects possible. In "finite-frequency" tomography, body wave travel time (and amplitude) anomalies are "frequency-dependent", and the geometrical ray paths are replaced by volumetric sensitivity (Fréchet) kernels. Therefore, measuring travel times at several periods increases the amount of independent informations, since at each

period the seismic waveform is influenced by a different weighted average of the Earth's structure, through the corresponding 3-D sensitivity kernel. One can exploit this frequency-dependency, by simultaneously inverting body wave travel times from different frequency bands, which is called "multiple-frequency tomography".

As the main aim of this thesis was to build a "high-resolution" 3-D shear-wave model of the deep Earth, we have performed a global multiple-frequency S-wave tomography of the mantle. We have chosen to use the "finite-frequency" formalism of Dahlen *et al.* (2000) to compute the needed sensitivity kernels (chapter 1). This formalism, which is based on the Born approximation, takes into account first-order and linear wave diffraction effects (e.g. scattering, wavefront-healing). Alternative "finite-frequency" approaches, that may overcome some of the limitations raised in the Born theory (linearity and single-scattering), are available, but are still too much time consuming at present time (e.g. Marquering *et al.* 1998; Zhao *et al.* 2000; Komatitsch *et. al* 2002; Calvet & Chevrot 2005; Tromp *et al.* 2005; Nissen-Meyer & Dahlen 2007). Though approximate, the formalism of Dahlen *et al.* (2000) has allowed us to compute hundreds of thousands of 3-D kernels needed for large scale multiple-frequency tomography.

The need for a global dataset of frequency-dependent *S*-wave travel times.

Recently, Montelli *et al.* (2004a; 2004b; 2006b) published P- and S- wave global "finite-frequency" tomographic models, claiming to confirm the existence of deep mantle plumes below a large number of postulated hotspots. They mainly attributed this to an improvement in the resolving power of their "finite-frequency" approach, based on the formalism of Dahlen *et al.* (2000), which is still debated in the community. That is, a number of recent studies suggest that the effect of such a "finite-frequency" approach could be smaller than that of practical considerations, such as the regularization or the data weighting (e.g. Sieminski *et al.* 2004; de Hoop & van der Hilst 2005a; Dahlen & Nolet 2005; de Hoop & van der Hilst 2005b; Julian 2005; Trampert & Spetzler 2006; Montelli *et al.* 2006a; van der Hilst & de Hoop 2006; Boschi *et al.* 2006).

The global S-wave "finite-frequency" tomographic model (hereafter referred to as *PRI-S05*) obtained by Montelli *et al.* (2006b) is based on (S, ScS-S and SS-S) travel times measured in a "single" frequency band ($\sim$20 s). Hence, it does not benefit from the increased spatial resolution afforded by sensitivity kernels for a range of frequencies. Moreover, their travel times were measured by matching the "first swing" of the observed waveform with a synthetic (Bolton & Masters 2001). As noticed by Montelli *et al.* (2004b), such a measurement scheme presents a possible bias in dominant frequency, by emphasizing

the early part of the waveform.

A first step towards improving the *PRI-S05* model is to use travel times measured in different frequency bands, and in a way which is fully consistent with the kernels. To our knowledge, there is no global dataset of S-wave travel times measured by cross-correlation between observed and synthetic "full" waveforms, as needed in the formalism of Dahlen *et al.* (2000), over a broad frequency range (10–51 s period). We have then focussed on building such a frequency-dependent S-wave dataset (chapter 2).

Assessing the relevance of multiple-frequency tomography

This new frequency-dependent database has enabled us to assess several fundamental questions on the relevance of "multiple-frequency tomography", from both a data point of view, and a model point of view. That is:

1. Is there an explicit *structural* dispersion, related to mantle heterogeneities, in our global dataset of frequency-dependent S-wave travel times?

2. Does a multi-band tomographic model contain significantly more information, about the Earth's mantle seismic structure, than a single-band model?

The analysis of our global dataset suggests that the residual dispersion observed in our data is, at least partly, related to seismic heterogeneity and attenuation in the Earth's interior (chapter 2). In order to exploit this *structural* dispersion contained in our global dataset, we shall use a "finite-frequency" approach to incorporate this new observable in tomography, as RT cannot take it into account.

We have then built a multiple-frequency SH-wave tomographic model of the Earth's mantle. The inverse problem has been posed using the "finite-frequency" formalism of Dahlen *et al.* (2000), and a data driven parameterization (chapter 3). We show that the "multi-band" model (*ZDS-S10*), resulting from the inversion of ∼300,000 S, ScS and SS travel times measured at 10, 15, 22.5, and 34 s periods, is consistent with other studies (chapter 4).

Finally, we present a comparison of single- versus multi-band tomographic models, in order to assess, from a model point of view, the actual benefits of inverting multi-band rather than single-band data (chapter 4). Expectations are that multi-band data could provide additional constraints on the 3-D elastic structure of the mantle, and lead to refined tomographic imaging of small scale features (e.g. plumes), which may play a key role in the Earth's dynamics.

CHAPTER 1

Multiple-frequency tomography: theory

"He was always late on principle, his principle being that punctuality is the thief of time."
Oscar Wilde

1.1 Introduction

During the last decades, global seismic travel time tomography has largely contributed to better constraining the structure of the Earth's interior. Until recently, ray theory formed the backbone of all body wave tomographic studies (e.g. Inoue *et. al* 1990; Su & Dziewonski 1992; Masters *et. al* 1996; Van der Hilst *et. al* 1997; Grand *et. al* 1997), mainly because of its simplicity and its small computing time.

Ray theory is based on the approximation that the travel time of a body wave only depends upon the 3D perturbation in wave speed along the unperturbed spherical-earth ray path (see figure 1.1). This assumes that seismic waves have an infinite-frequency, or a zero wavelength. However, teleseismic waves have wavelengths ranging from 10 to 1000 km, or even more.

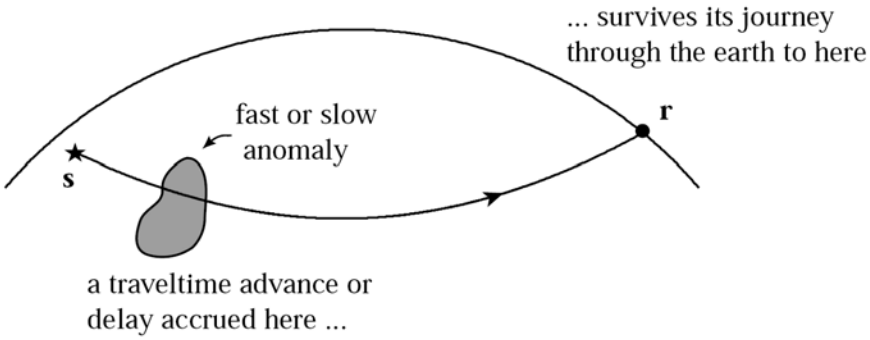


Figure 1.1: Cartoon illustration of the main ideas underlying conventional "ray theoretical" travel time tomography. First, ray theory assumes that seismic waves have an infinite-frequency, or a zero wavelength. Therefore, the travel time of a body wave only depends upon the 3D perturbation in wave speed along the infinitesimally narrow geometrical ray path (black solid line). Second, a seismic wave that accrues a travel time advance, or delay, upon passage through a fast, or slow, anomaly, "remembers" that time shift as it continues to propagate towards the receiver. From Hung et. al (2001).

At the macroscopic scale, the validity of the approximation of ray theory depends on the length of the ray path, L, compared to the size of the heterogeneity, a. Seismologists (e.g. Aki & Richards 1980) often use the 'wave parameter'

$$D = \frac{4L}{ka^2} \tag{1.1}$$

which represents the ratio between L/a and ka ($k = 2\pi/\lambda$ is the wave number), with the factor 4 chosen to differentiate different wave propagation regimes around $D \simeq 1$. The

'wave parameter' is related to the Fresnel zone as

$$D \propto (\frac{r}{a})^2 \tag{1.2}$$

where r is the radius of the Fresnel zone ($r = 1/2 \cdot \sqrt{\lambda L}$).

Ray theory is only applicable if the heterogeneities are much larger than the Fresnel zone, i.e. if $D \leq 1$, because diffraction effect around them is small. However, ray theory breaks down when used for imaging heterogeneities which are smaller than the Fresnel zones, i.e. if $D \geq 1$, because diffraction effects (e.g. scattering and wavefront-healing) make travel time (and amplitude) anomalies dependent on Earth structure in the 3-D region around the geometrical ray path.

It is interesting to assess the validity of ray theory for teleseismic S-waves in the period range of interest of this PhD study (i.e. 10–51 s), if one aims to image small objects (e.g. slabs, plumes), that are likely to be rather limited in size (∼200 km horizontally). Let us consider a typical 20 s period S wave, with a velocity of almost 5 km/s, and a small heterogeneity with a size of 200 km (horizontally). The ray-theoretical regime is only valid as long as $L \leq ka^2/4 \simeq 630$ km. As the shortest propagation distance of an S wave, in this study, is around 3,300 km (i.e. epicentral distance of 30°), we are never in the strict ray-theoretical regime in the case of such small heterogeneities - which are of considerable interest in understanding the Earth dynamics.

On the other hand, for a large heterogeneity of 1,000 km (horizontally), the ray-theoretical regime is valid as long as $L \leq 15,700$ km. As the largest propagation distance of an S wave is around 10,000 km (i.e. epicentral distance of 90°), ray theory may be used for imaging such very long-wavelength heterogeneity - as expected.

Since wave diffraction phenomena are not taken into account in ray theory based seismic tomography studies, it seems that progress towards higher-resolution images of small heterogeneities in the mantle requires a movement away from ray theory. In an effort to improve upon the infinite-frequency approximation of ray theory, that is only applicable to the time of the wave onset (considered as the very high-frequency content of the wave), "finite-frequency" approaches (in contrast to the "infinite-frequency" ray theory) have recently emerged in seismic tomography (e.g. Marquering *et al.* 1998; Dahlen *et al.* 2000; Zhao *et al.* 2000; Komatitsch *et. al* 2002; Calvet & Chevrot 2005; Tromp *et al.* 2005; Nissen-Meyer & Dahlen 2007). For example, they may take the effects of wave diffraction into account, which makes the imaging of smaller objects possible.

In "finite-frequency" tomography, travel time (and amplitude) anomalies are "frequency-dependent", and geometrical ray paths are replaced by volumetric sensitivity (Fréchet)

kernels. Delay-times (or time-residuals) observed in different frequency bands contain information on the size of seismic heterogeneities - provided that we can detect a structural dispersion in the data (cf. chapter 2). Measuring travel times at several periods increases the amount of independent informations, since at each period the waveform is influenced by a different weighted average of the Earth's structure, through the corresponding 3-D sensitivity kernel. One can exploit this frequency-dependency, by inverting data from different frequency bands simultaneously - which is called "multiple-frequency tomography" (cf. section 1.6).

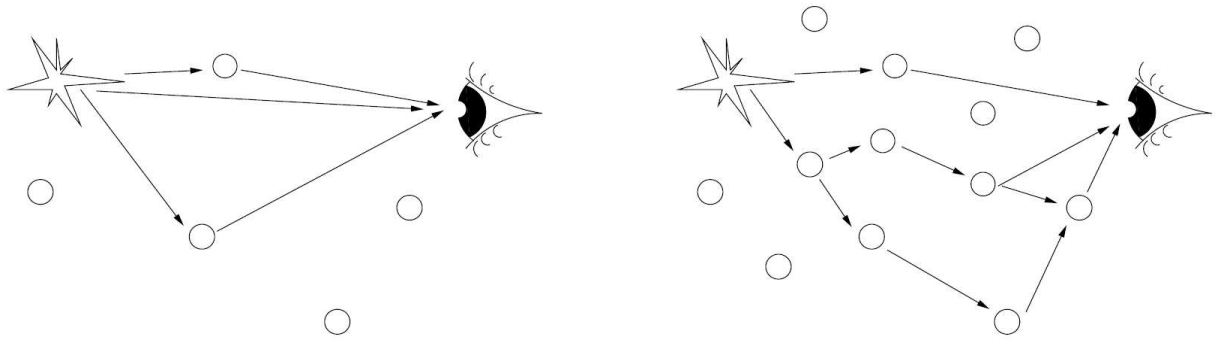


Figure 1.2: The "Born approximation" assumes that the scattered wave energy is small, such that one may ignore the fact that the scattered waves themselves induce scattered energy (i.e. multiple scattering). The "finite-frequency" formalism of Dahlen et. al (2000) assumes that the Born approximation is valid, and therefore only takes single-scattering into account. Left: single scattering. Right: multiple scattering. From Anache (2008).

In this thesis, we aim to build a global multiple-frequency S-wave tomography of the Earth's mantle, for better constraining mantle dynamics. We have chosen to use the "finite-frequency" formalism of Dahlen *et al.* (2000) to compute sensitivity kernels. This formalism only takes the first-order effects of wave diffraction into account. It assumes that the scattered part of wave energy is small, so that the energy coming from multiple-scattering (figure 1.2) rapidly decreases and can be neglected - which is known as the "Born approximation". The assumption of infinite-frequency waves (ray theory) is then replaced by the assumption of single-scattering (figure 1.2). The main advantage of this formalism, which is based on the paraxial approximation[1], is the use of dynamic ray tracing (e.g. Tian *et al.* 2007b) to compute extremely fast the kernels. It makes manageable the computation time of hundreds of thousands of 3-D kernels, needed for large scale tomography.

However, the "finite-frequency theory" of Dahlen *et al.* (2000) has some limitations - which are detailed in section 1.5. It is a first-order and linear theory: multiple-scattering

[1]The paraxial approximation assumes scatterers are not too far from the direct ray.

is neglected, and the non-linear part of diffraction effects (e.g. wavefront-healing) cannot be modeled. Moreover, the paraxial approximation breaks down for long distance waves (e.g. *SS*) near the antipode, near-field effects are not taken into account, and headwaves or diffracted waves (e.g. *Sdiff*) cannot be handled.

Alternative "finite-frequency" approaches, which may overcome some limitations raised in the Born theory (e.g. linearity and single-scattering), are available (e.g. Marquering *et al.* 1998; Zhao *et al.* 2000; Komatitsch *et. al* 2002; Calvet & Chevrot 2005; Tromp *et al.* 2005; Nissen-Meyer & Dahlen 2007). For instance, a more exact modeling of the propagation of teleseismic waves, in a 3-D Earth model, can be achieved with the Spectral Element Methods (e.g. Komatitsch *et. al* 2002). Unfortunately, all these promising approaches are still too much time consuming at present time, making an application to a large global dataset not (easily) feasible - in the period range of this study (at least for short periods).

In this PhD study, we have bet that, although not perfect, applying the finite-frequency formalism of Dahlen *et al.* (2000) could help to readily build a more accurate picture of the Earth's mantle. In this chapter, we then briefly present[2] the finite-frequency approach developed by the Princeton's seismological group (e.g. Dahlen *et al.* 2000).

1.2 The ray theoretical interpretation of travel times

In the ray theoretical interpretation of travel times, one considers that seismic waves have an infinite-frequency. Therefore, the travel time of a body wave is assumed to depend on the 3-D perturbation in wave speed along its infinitesimally narrow geometrical ray path (cf. figure 1.1). The travel time of such an infinite-frequency wave can then be written as

$$T = \int_P \frac{ds}{c(\mathbf{r})} \tag{1.3}$$

, where P indicates the "true" ray path of the wave in the "true" 3-D Earth, and $c(\mathbf{r})$ is the seismic velocity at point location $\mathbf{r}$.

Because the ray path also depends on $c(\mathbf{r})$, the relationship between the travel time, T, and the model velocity, c, is non-linear. However, this relationship can be linearized, using Fermat's Principle[3].

[2]This short presentation - of the finite-frequency formalism of Dahlen *et al.* (2000) - relies heavily on chapter 7 of Nolet's book 'A Breviary of Seismic Tomography' (2008).

[3]Fermat's Principle stipulates that the ray path taken by a seismic ray between two points is the path of least time (in fact, this is the path of stationary time). From Wikipedia.

Firstly, one assumes that the travel time for the background model is T_0, defined as

$$T_0 = \int_{P_0} \frac{ds}{c_0(\mathbf{r})}. \tag{1.4}$$

In this study, we will always consider the reference Earth model to be spherically symmetric (i.e. 1-D), such as we have $c_0(\mathbf{r}) = c_0(r)$.

Secondly, according to Fermat's Principle, one may compute the integral in (1.3) not over the "true" ray path (i.e.P) but over the ray path computed for the reference model (i.e. P_0), so that

$$T \simeq \int_{P_0} \frac{ds}{c(\mathbf{r})}. \tag{1.5}$$

Finally, using (1.4) and (1.5), we get a linearized relationship[4] between the travel time residual, δT, and the model velocity anomaly, $\delta c(\mathbf{r})$, as

$$\delta T = T - T_0 \simeq \int_{P_0} (\frac{1}{c} - \frac{1}{c_0}) ds \simeq \int_{P_0} (-\frac{\delta c}{c^2}) ds. \tag{1.6}$$

1.3 Limitations of ray theory

The frequency content of teleseismic body waves is far from the infinite frequency necessary to make ray theory valid - at least in the case of small scale heterogeneities (cf. section 1.1). During their journey through the Earth's interior, "finite-frequency" waves have their waveforms affected by wave diffraction effects. Travel time (and amplitude) measurements, made by cross-correlation of broadband seismic waveforms, are then sensitive to these "finite-frequency effects". In this section, we present two frequency-dependent diffraction phenomena, that may strongly influence such travel times: wavefront-healing and scattering. Clearly, such frequency-dependency cannot be handled by ray theory.

1.3.1 Wavefont-healing

Wavefront-healing is a ubiquitous diffraction phenomenon (see Nolet & Dahlen 2000; Hung *et al.* 2001; Nolet *et al.* 2005). It occurs whenever the scale of any geometrical irregularities in a wavefront are comparable to the wavelength of the wave (Gudmundsson, 1996), and affects cross-correlation travel time measurements (Hung *et al.* 2001).

[4]The last linearization in (1.6) can be avoided if we use the slowness, c^{-1}, instead of the velocity, c, as the model parameter.

Figures 1.3 and 1.4 illustrate the wavefront-healing phenomenon, with a 2-D finite-difference simulation, for a wavefront that has passed through a low and high velocity anomaly, respectively. Screen-shots of the wavefront are displayed after 10, 35, 60 and 85 time steps; the direction of propagation is upward.

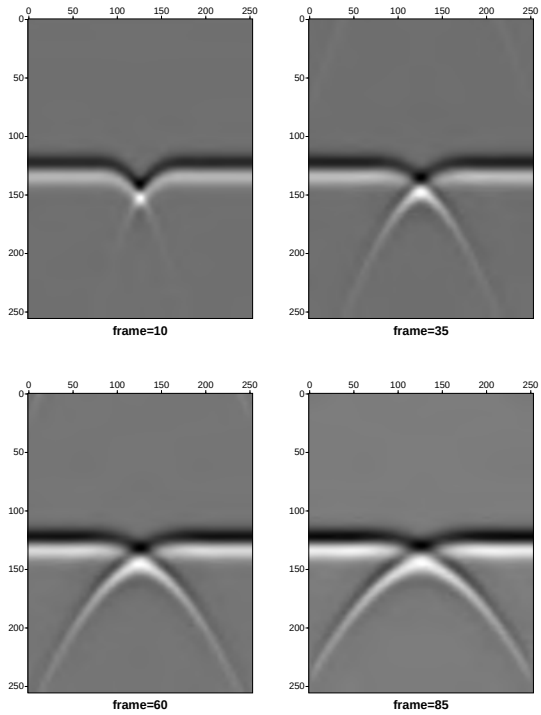


Figure 1.3: Simulation of the "healing" of a wavefront that has passed through a low velocity anomaly. From Nolet (2008).

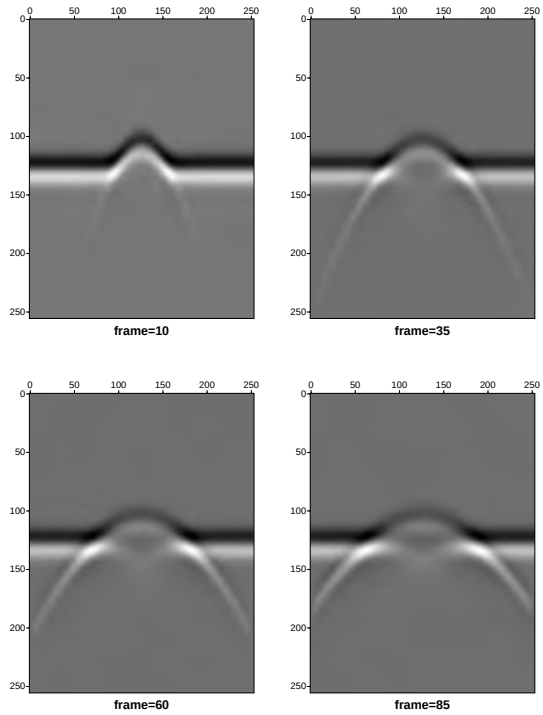


Figure 1.4: Simulation of the "healing" of a wavefront that has passed through a high velocity anomaly. From Nolet (2008).

Figure 1.3 shows that the low velocity anomaly delays the wavefront, which creates a gap that may be filled in (i.e. healed) by energy radiating from the sides, using Huygens Principle[5].

Figure 1.4 shows that high velocity anomaly advances the wavefront. The advanced wavefront quickly loses amplitude, because of the increased geometrical spreading. However, the magnitude of the negative time residual does not seem to decrease. On a noise-free seismogram, one could even pick the advanced seismic phase close to the predicted ray-theoretical arrival time (Nolet 2008).

[5]The Huygens Principle (also named Huygens-Fresnel Principle) stipulates that each point of an advancing wavefront is the source of a new train of waves. The advancing wave, as a whole, may be regarded as the sum of all the secondary waves arising from points in the medium already traversed. From Wikipedia.

Thus, wavefront-healing is different (i.e. non-linear) for positive and negative anomalies - a point first made by Wielandt (1987) and known as the 'Wielandt effect'. It is worth noting that the linear finite-frequency formalism of Dahlen *et. al* (2000) cannot handle the non-linearity part of the wavefront-healing phenomenon.

Wavefront-healing depends upon the wave's frequency and the anomaly size, which can be seen in figure 1.5. That is, figure 1.5 shows how an initial delay time diminishes after a wave, with wavelength λ, has crossed a low velocity anomaly, of width $2L$.

First, we see that wavefronts of longer period waves heal more quickly[6] as a function of distance from the perturbation. That is, for a fixed value of $L = 1$, we see in figure 1.5 that the healing is much more complete for long period wave ($L/\lambda = 0.5 \Rightarrow \lambda = cT = 2$, where c is the phase velocity and T the period) than for short period wave ($L/\lambda = 2.0 \Rightarrow \lambda = cT = 0.5$).

On the other hand, for a fixed value of $\lambda = 1$, we see in figure 1.5 that the healing is much more important for a small size anomaly ($L/\lambda = 0.5 \Rightarrow L = 0.5$) than for a large one ($L/\lambda = 2.0 \Rightarrow L = 2.0$).

Wavefront-healing is then a serious shortcoming of ray theory, if we wish to image small scale anomalies, with the halfwidth (L) of the same order as the wavelength (λ).

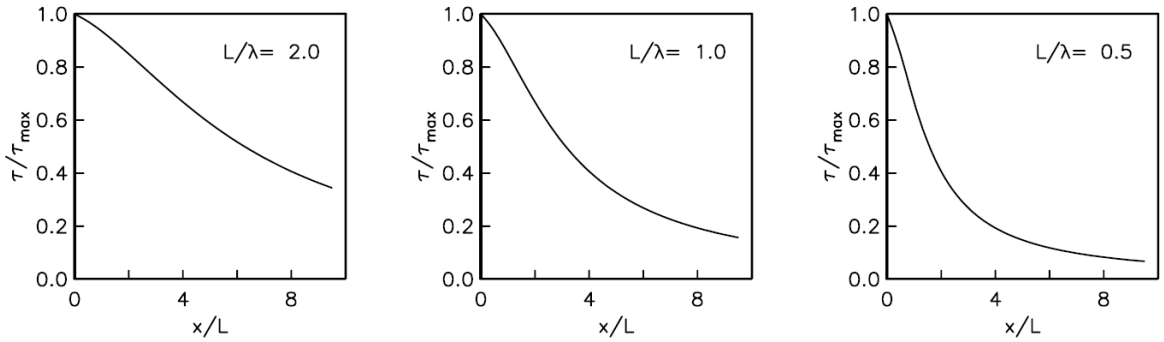


Figure 1.5: Evolution of body-wave delay times τ along a ray traversing the center of a spherical low velocity anomaly, as a function of the distance x traveled past the anomaly, scaled by the halfwidth L of the anomaly. The three curves are for different values of the halfwidth-to-wavelength ratio L/λ. The delay time, τ, is normalized by the initial delay time, τ_{max}, acquired by the seismic wave just after its passage through the anomaly. From Nolet & Dahlen (2000).

[6]It is worth noting that if a seismic wave passes through a low velocity anomaly, the longer the wave period is, the more important the healing will be, and therefore the less the wave will be apparently delayed at the receiver. The corresponding time residuals, δt, measured by cross-correlation at different filtering periods, T, will then lead to a *decreasing* dispersion curve $\delta t(T)$. This point will be the corner stone of our discussion on the analysis of wavefront-healing effect in our data (cf. section 2.3.1 of chapter 2).

1.3.2 Scattering

Scattering is another ubiquitous diffraction phenomenon, which is frequency-dependent[7]. It occurs if the target heterogeneities are smaller than the Fresnel zones of seismic waves. Here, following Nolet *et al.* (2005), we briefly explain the basic ideas of scattering effects on travel times measured by cross-correlation of broadband seismic waveforms.

We consider here a seismogram $s(t)$ as a succession of pulse-like arrivals $u_i(t)$. In the framework of first-order Born theory (i.e. single scattering), we add the contribution $\delta u_i(t)$ of waves scattered from the wavefront around ray i (see figure 1.6). We assume that the scattered part of the wave energy is small, such that one may ignore that the scattered waves themselves induce scattered energy (i.e. multiple-scattering) - this is the "Born approximation" (cf. figures 1.2 and 1.6).

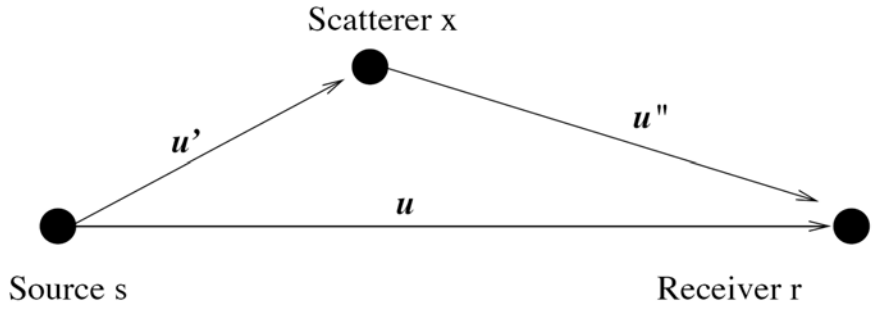


Figure 1.6: This cartoon illustrates how Born theory works in a homogeneous medium. A wave u' is scattered off a point scatterer and generates a small perturbation u'' that adds to the direct wave u. The detour time of the scattered wave is $\Delta T(\boldsymbol{x}) = T_{sx} + T_{xr} - T_{sr}$, with T_{sx}, T_{xr} and T_{sr} the travel times between source/scatterer, scatterer/receiver and source/receiver, respectively. From Nolet et al. (2005).

If we consider a S-wave striking a seismic heterogeneity, since the S-wave itself travels the path of minimum time, the scattered wave cannot arrive earlier than the direct wave (cf. figure 1.6). However, the scattered wave does not always have a delaying influence on the time residual, measured by cross-correlation of the observed waveform - the perturbed wave $u_i(t) + \delta u_i(t)$ - with the synthetic waveform - the unperturbed wave $u_i(t)$.

In the following, we denote by δt_{max} the time residual corresponding to the time-lag for which the cross-correlation, $C(t)$, reaches its maximum. Figure 1.7 shows that the added scattered wave, δu_i, deforms the wave shape of the direct wave, u_i, and may have a delaying or an advancing effect on the location of the cross-correlation maximum, i.e.

[7]The discussion on the frequency-dependency of scattering is postponed to section 2.3.2 of chapter 2, and is illustrated in figures D.1 and D.2 of appendix D.

on the time residual δt_{max}, depending on the sign of the scattered wave. The sign of the scattered wave is determined by the sign of the velocity anomaly that causes the scattered wave. High and low velocity scatterers generate scattered waves with negative and positive polarities, respectively. Figure 1.7.a shows that adding a scattered wave, $\delta u(t)$, with a negative polarity to the unperturbed wave, $u(t)$, leads to a time advance of the cross-correlation maximum, i.e. $\delta t_{max} < 0$. On the other hand, figure 1.7.b shows that a positive scattered wave leads to a delayed cross-correlation maximum, i.e. $\delta t_{max} > 0$.

Figure 1.7 also shows that the *onset* of the perturbed waveform is not affected by the scattered wave. The onset time remains the same as for the unperturbed waveform (i.e. zero). This is in agreement with the common assumption that onsets of teleseismic waves are dominated by very high-frequencies. Ray theory is then (more) justified to interpret onset travel times.

Finally, scatterers *off* the ray may not influence the *onset* of the waveform, but they do advance or delay the full waveform. Time residuals, measured by cross-correlation that uses all or part of the waveform, are sensitive to these waveform distortions. Therefore, they should not be inverted with ray theory.

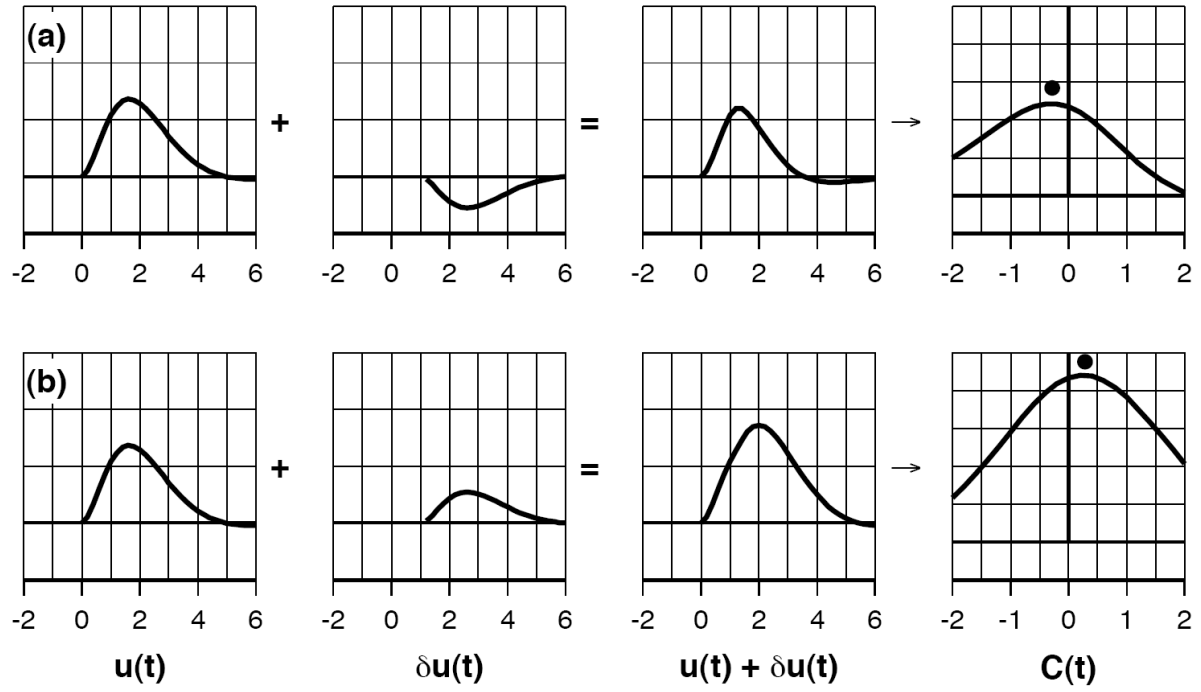


*Figure 1.7: The location of the maximum of the cross-correlation of a perturbed wave $u(t) + \delta u(t)$ with the unperturbed wave $u(t)$ is either advanced (**a**) or delayed (**b**), depending on the sign of the scattered wave. From Nolet et. al (2005).*

1.4 A finite-frequency interpretation of travel times

Wave diffraction effects (e.g. wavefront-healing, scattering) make travel time anomalies dependent on Earth structure in a 3-D region around the geometrical ray path, which cannot be handled by ray theory. In an effort to image small scale anomalies, that are not (easily) visible with ray theory, we should use a finite-frequency interpretation of travel times. We have then chosen to apply the "finite-frequency" formalism of Dahlen *et al.* (2000), which can model some of the first-order and linear effects of wave diffraction.

In this finite-frequency approach, geometrical ray paths are replaced by volumetric sensitivity (Fréchet) kernels - named "banana-doughnut kernels" by Marquering *et al.* (1999), because of their particular shape. These sensitivity kernels are designed for time residuals measured by cross-correlation, between observed and synthetic waveforms. Here, following Nolet (2008), we present a brief overview of the derivation of these banana-doughnut kernels.

1.4.1 Cross-correlating seismic waveforms

In the formalism of Dahlen *et al.* (2000), the 3-D Earth structure is represented by a smoothly varying reference model, for which ray theory is assumed to be valid. An heterogeneity is then added to the smooth medium, which scatters a (small) part of the wave energy (cf. figure 1.6). Here, following Nolet (2008), we aim to determine: How does the scattered wave influence the travel time of the direct wave?

We assume that the observed signal at the receiver, $d(t)$, consists of the direct, $u(t)$, and scattered, $\delta u(t)$, wave arrivals. As the synthetic[8] signal, $s(t)$, is calculated with ray theory, in a smooth reference model, it does not account for wave diffraction. Therefore, the observed and synthetic waveforms are, respectively:

$$\begin{cases} d(t) = u(t) + \delta u(t) \\ \\ s(t) = u(t). \end{cases} \tag{1.7}$$

The cross-correlation function between observed, $d(t)$, and synthetic, $s(t)$, waveforms is defined as

$$\gamma_{d,s}(\tau) = \int_{-\infty}^{+\infty} d(t)s(t-\tau)dt. \tag{1.8}$$

[8]In this thesis, synthetic seismograms will be calculated with ray theory, using the WKBJ method of Chapman (1978), in the radial (1-D) reference model IASP91 (Kennett & Engdahl 1991).

The autocorrelation of the unperturbed wave, $u(t)$, is given by

$$\gamma_{s,s}(\tau) = \int_{-\infty}^{+\infty} u(t)u(t-\tau)dt. \tag{1.9}$$

As defined by Dahlen *et al.* (2000), the time residual, $\delta t = t^{obs} - t^{syn}$, maximizes the cross-correlation function between the perturbed, $u(t) + \delta u(t)$, and the unperturbed, $u(t)$, waves:

$$\gamma_{d,s}(\tau) = \int_{-\infty}^{+\infty} [u(t) + \delta u(t)]u(t-\tau)dt \tag{1.10}$$

which leads to:

$$\gamma_{d,s}(\tau) = \gamma_{s,s}(\tau) + \delta\gamma(\tau) \tag{1.11}$$

with:

$$\delta\gamma(\tau) = \int_{-\infty}^{+\infty} \delta u(t)u(t-\tau)dt. \tag{1.12}$$

For the unperturbed wave, the cross-correlation reaches its maximum at zero lag-time:

$$\dot{\gamma}_{s,s}(0) = 0, \tag{1.13}$$

and for the perturbed wave the maximum is reached for δt:

$$\dot{\gamma}_{d,s}(\delta t) = \dot{\gamma}_{s,s}(\delta t) + \delta\dot{\gamma}(\delta t) = 0, \tag{1.14}$$

where the dot denotes the time differentiation. Developing $\dot{\gamma}$ to first order, we obtain (e.g. Marquering *et al.* 1999):

$$\dot{\gamma}_{s,s}(\delta t) + \delta\dot{\gamma}(\delta t) = \dot{\gamma}_{s,s}(0) + \ddot{\gamma}_{s,s}(0)\delta t + \delta\dot{\gamma}(0) + O(\delta^2) = 0, \tag{1.15}$$

which leads to:

$$\delta t = -\frac{\delta\dot{\gamma}(0)}{\ddot{\gamma}_{s,s}(0)} = \frac{\int_{-\infty}^{+\infty} \dot{u}(t)\delta u(t)dt}{\int_{-\infty}^{+\infty} \ddot{u}(t)u(t)dt}. \tag{1.16}$$

It is more convenient to express equation (1.16) in the frequency domain. Using the Parseval's theorem[9], we obtain:

$$\delta t = \frac{\int_{-\infty}^{+\infty} [\dot{u}(\omega)]^* \delta u(\omega)d\omega}{\int_{-\infty}^{+\infty} [\ddot{u}(\omega)]^* u(\omega)d\omega}. \tag{1.17}$$

[9]The Parseval's Theorem is: $\int_{-\infty}^{+\infty} f(t)g(t)dt = \int_{-\infty}^{+\infty} f^*(\omega)g(\omega)d\omega$, where an asterisk denotes the complex conjugate.

Now, we aim to express δt in function of u and δu. From the Fourier Analysis[10] theory, we have[11] $\dot{u}(\omega) = -i\omega u(\omega)$, which leads to

$$\delta t = \frac{\int_{-\infty}^{+\infty}[-i\omega u(\omega)]^* \delta u(\omega) d\omega}{\int_{-\infty}^{+\infty}[(-i\omega)^2 u(\omega)]^* u(\omega) d\omega} = \frac{\int_{-\infty}^{+\infty} i\omega u(\omega)^* \delta u(\omega) d\omega}{\int_{-\infty}^{+\infty} -\omega^2 u(\omega)^* u(\omega) d\omega} = \frac{N}{D}. \tag{1.18}$$

Using the spectral property of real signals: $u(-\omega) = u(\omega)^*$, we have:

$$N = \int_0^{+\infty} i\omega\{u(\omega)^* \delta u(\omega) - [u(\omega)^* \delta u(\omega)]^*\} d\omega = 2\mathrm{Re}\int_0^{+\infty} i\omega u(\omega)^* \delta u(\omega) d\omega, \tag{1.19}$$

and:

$$D = \int_0^{+\infty} -\omega^2\{u(\omega)^* u(\omega) + [u(\omega)^* u(\omega)]^*\} d\omega = 2\int_0^{+\infty} -\omega^2 u(\omega)^* u(\omega) d\omega. \tag{1.20}$$

Finally, with $|u(\omega)|^2 = u(\omega)^* u(\omega)$, the time residual δt may be expressed as:

$$\delta t = -\frac{\mathrm{Re}\int_0^{+\infty} i\omega u(\omega)^* \delta u(\omega) d\omega}{\int_0^{+\infty} \omega^2 |u(\omega)|^2 d\omega}. \tag{1.21}$$

1.4.2 Banana-doughnut sensitivity kernels

We have previously established, from equation (1.21), how the travel time of an unperturbed wave is influenced by a scattered wave - in the formalism of Dahlen *et al.* (2000). That is, the scattered wave effect is to induce a time residual, δt, which is function of both the unperturbed, u, and perturbed, δu, waves. From this result, we now aim to find an expression of the travel time sensitivity (Fréchet) kernel, that relates the measured time residual to the velocity anomalies in the Earth's interior. Here, we give a brief overview on the derivation of such (banana-doughnut) sensitivity kernels for S-waves, in the formalism of Dahlen *et al.* (2000), following Nolet *et al.* (2005) and Nolet (2008).

Firstly, we will only consider the case of purely forward scattering, i.e. scattering angle θ=0, for simplicity reason. However, this particular case is of interest, as we aim to measure time residuals over limited cross-correlation time windows (cf. section 1.4.4). That is, large scattering angle θ correspond to waves having been scattered from anomalies too

[10]Here, we use the following convention for the time-frequency Fourier Transform (FT): $u(\omega) = \int_{-\infty}^{+\infty} u(t) e^{i\omega t} dt$, where $\omega = 2\pi f$ is the angular frequency, with f the frequency in Hz. The inverse Fourier Transform is then: $u(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} u(\omega) e^{-i\omega t} d\omega$.

[11]We have $\dot{u}(\omega) = \mathrm{FT}(du(t)/dt) = \int_{-\infty}^{+\infty} du(t)/dt e^{i\omega t} dt = -i\omega u(\omega)$, using integration by parts and assuming that $u(t) \to 0$ when $|t| \to \infty$

much distant from the geometrical ray path. Their too large detour time causes them to arrive outside of the cross-correlation time window. Hence, they cannot affect the measured time residuals. Secondly, mode conversions upon scattering will also not be considered, as in this thesis we focus on *SH*-waves, and an incoming *SH*-wave that is scattered can only give rise to an outgoing *SH*-wave. Thirdly, scattering from density perturbations will be neglected. That is, Hung *et al.* (2000) show that travel time anomalies produced by realistic variations in density (ρ) are small, and perturbations of V_P and V_S (P- and S-wave velocity, respectively) are much more effective than perturbations of ρ in influencing the measured travel time. For the more general case of Fréchet kernels (i.e. scattering under arbitrary angle θ, and for arbitrary rays and anomalies in ρ, V_P and V_S), including the effects of caustics, we refer the reader to Dahlen *et al.* (2000) or Nolet (2008).

At this stage, we realize that we need to calculate the zero order wave field $u(\omega)$, i.e. the synthetic waveform. Different strategies have so far been explored by seismologists. For instance, Zhao *et al.* (2000) used summation of discrete normal modes, and Marquering *et al.* (1999) used summation of surface waves in the frequency domain. Even if they may provide an accurate estimation of $u(\omega)$, these methods are still too much time consuming, which makes them not well suited for global tomography, in our frequency range of analysis (10–51 s). In the formalism of Dahlen *et al.* (2000), the zero order wave field is calculated using ray theory - e.g. with the WKBJ method (Chapman 1978). Though this gives rise to some limitations of the kernels (cf. section 1.5.4), it is much faster, and makes possible - with the computer facilities available for this PhD - the computation of hundreds of thousands of kernels necessary in this study.

The spectrum of a far-field[12] S wave recorded at the receiver $\mathbf{r}$ from a moment tensor source with time behaviour $m(t)$ in a *homogeneous* medium with velocity β_0 and density ρ_0 is defined as (e.g. Aki & Richards 1980; Nolet 2008)

$$u^S(\mathbf{r}_r;\omega) = \frac{F^S \dot{m}(\omega)}{4\pi\rho_0\beta_0^3 r_{rs}} e^{-i\omega r_{rs}/\beta_0}, \tag{1.22}$$

where the geometrical spreading is given by the distance r_{rs} from the source s to the receiver r, and where F^S denotes an amplitude factor that includes the radiation pattern. The derivative of $m(t)$ is $\dot{m}(t)$ and $\dot{m}(\omega)$ is the Fourier transform of $\dot{m}(t)$. In a *heterogeneous* medium, the 'ray approximation' involves modeling the amplitude by a local factor $(\beta_r\rho_r)^{-1/2}R_{rs}^{-1}$, where R_{rs} replaces r_{rs} as the geometrical spreading factor, and where $(\beta_r\rho_r)^{-1/2}$ models the effect of changes in impedance, such that the energy flux is

[12]Banana-doughnut kernels only contain the far-field expression of the seismic wavefield. This poses problem for scatterers located near the source or beneath the receiver (cf. section 1.5.2).

conserved. The travel time from source to receiver is also generalized from r_{rs}/β_0 to T_{rs}. The effect of gradual changes in the medium is then obtained by replacing

$$\rho_0\beta_0^3 r_{rs} \rightarrow \beta_s^2 R_{rs}\sqrt{\rho_s\rho_r\beta_s\beta_r}. \tag{1.23}$$

This leads to the following generalized expression (e.g. Aki & Richards 1980; Nolet 2008)

$$u^S(\mathbf{r}_r;\omega) = \frac{F^S\dot{m}(\omega)}{4\pi\beta_s^2 R_{rs}\sqrt{\rho_s\rho_r\beta_s\beta_r}}e^{-i\omega T_{rs}}. \tag{1.24}$$

If we replace the receiver r by a scatterer x, the equation (1.24) can also be used to find the wavefield $u^S(\mathbf{r}_x;\omega)$ that impacts on a point scatterer x, that is:

$$u^S(\mathbf{r}_x;\omega) = \frac{F^S\dot{m}(\omega)}{4\pi\beta_s^2 R_{xs}\sqrt{\rho_s\rho_x\beta_s\beta_x}}e^{-i\omega T_{xs}}. \tag{1.25}$$

The amplitude of a scattered wave from a small volume dV with perturbations in density $\delta\rho$ and in Lamé parameter $\delta\mu$, for $S \rightarrow S$ scattering from an incoming wave of amplitude 1, in the approximation of 'forward scattering', is (e.g. Nolet *et al.* 2005)

$$\delta u^{S\rightarrow S}(\omega) = -\frac{\omega^2}{2\pi\beta^2}\frac{1}{r}\frac{\delta\beta}{\beta}dV. \tag{1.26}$$

where r is the propagation distance between the scatterer and the receiver. Replacing the source s with the scatterer x, we may use the equation (1.24) to generalize the expression for the scattered wave $\delta u^{S\rightarrow S}(\omega)$ - equation (1.26) - to *heterogeneous* media by adding the appropriate geometrical decay and a phase delay. This leads to the new expression

$$\delta u^{S\rightarrow S}(\mathbf{r}_x;\omega) = -\frac{\omega^2\rho_x dV}{2\pi R_{rx}\beta_x\sqrt{\rho_x\rho_r\beta_x\beta_r}}(\delta\beta/\beta)_x e^{-i\omega T_{rx}}. \tag{1.27}$$

Multiplying the scattered wave (1.27) for an incoming wave, with unit amplitude, with the wavefield that impacts on a point scatterer (1.25) results in

$$\delta u^{S\rightarrow S}(\mathbf{r}_x;\omega) \leftarrow \delta u^{S\rightarrow S}(\mathbf{r}_x;\omega)u^S(\mathbf{r}_x;\omega), \tag{1.28}$$

which leads to:

$$\delta u^{S\rightarrow S}(\mathbf{r}_x;\omega) = -\frac{F^S\dot{m}(\omega)\omega^2}{8\pi^2\beta_s^{5/2}\rho_s^{1/2}\beta_r^{1/2}\rho_r^{1/2}}\frac{(\delta\beta/\beta)_x dV}{\beta_x^2 R_{rx}R_{xs}}e^{-i\omega(T_{rx}+T_{xs})}. \tag{1.29}$$

There is no need to compute the geometrical spreading R_{rx} for each possible scattering location x. If we compute the geometrical spreading from the receiver to every point in the model (R_{rx}), the amplitude reciprocity principle gives us the spreading factor R_{xr}:

$$\beta_r R_{xr} = \beta_x R_{rx}. \tag{1.30}$$

Applying the equation (1.30) results in the following expression:

$$\delta u^{S\to S}(\mathbf{r}_x;\omega) = -\frac{F^S \dot{m}(\omega)\omega^2}{8\pi^2 \beta_s^{5/2}\rho_s^{1/2}\beta_r^{3/2}\rho_r^{1/2}} \frac{(\delta\beta/\beta)_x dV}{\beta_x R_{xr} R_{xs}} e^{-i\omega(T_{rx}+T_{xs})}. \tag{1.31}$$

We find the cross-correlation time residual induced by such a point scatterer if we insert (1.25) and (1.31) into the expression (1.21). On the one hand, we have

$$u(\omega)^* \delta u(\omega) = -\frac{(F^S)^2|\dot{m}(\omega)^2|\omega^2(\delta\beta/\beta)_x dV}{2^5\pi^3\beta_s^5 R_{rs}R_{xr}R_{xs}\rho_s\rho_r\beta_r^2\beta_x} e^{-i\omega\Delta T(\mathbf{r}_x)} \tag{1.32}$$

where $\Delta T(\mathbf{r}_x) = T_{rx} + T_{xs} - T_{rs}$ is the detour time of the scattered wave. On the other hand, we have

$$|u(\omega)|^2 = \frac{(F^S)^2|\dot{m}(\omega)^2|}{2^4\pi^2\beta_s^5 R_{rs}^2\rho_s\rho_r\beta_r}. \tag{1.33}$$

This leads to:

$$\delta t = -\frac{\int_0^{+\infty}\omega^3|\dot{m}(\omega)^2|\mathrm{Re}\{ie^{-i\omega\Delta T(\mathbf{r}_x)}\}d\omega}{\int_0^{+\infty}\omega^2|\dot{m}(\omega)^2|d\omega} \frac{(\delta\beta/\beta)_x dV R_{rs}}{2\pi\beta_x\beta_r R_{xr}R_{xs}}. \tag{1.34}$$

And using the fact that $\mathrm{Re}\{ie^{-i\omega\Delta T(\mathbf{r}_x)}\} = \sin[\omega\Delta T(\mathbf{r}_x)]$, this results in:

$$\delta t = -\frac{(\delta\beta/\beta)_x dV}{2\pi\beta_x} \frac{R_{rs}}{\beta_r R_{xr}R_{xs}} \frac{\int_0^{+\infty}\omega^3|\dot{m}(\omega)^2|\sin[\omega\Delta T(\mathbf{r}_x)]d\omega}{\int_0^{+\infty}\omega^2|\dot{m}(\omega)^2|d\omega}. \tag{1.35}$$

For a more general heterogeneity, we integrate over all point scatterers:

$$\delta t = \int K_\beta(\mathbf{r}_x)\frac{\delta\beta}{\beta}dV \tag{1.36}$$

where $K_\beta(\mathbf{r}_x)$ is the travel time sensitivity (Fréchet) kernel:

$$K_\beta(\mathbf{r}_x) = -\frac{1}{2\pi\beta_x} \frac{R_{rs}}{\beta_r R_{xr}R_{xs}} \frac{\int_0^{+\infty}\omega^3|\dot{m}(\omega)^2|\sin[\omega\Delta T(\mathbf{r}_x)]d\omega}{\int_0^{+\infty}\omega^2|\dot{m}(\omega)^2|d\omega}. \tag{1.37}$$

1.4.3 Doughnut hole

Marquering *et al.* (1999) first noticed that one characteristic of sensitivity kernels, such as equation (1.37), is that the location of the ray itself appears to be a region of zero sensitivity for the travel time of the seismic wave. Thus, the time residual, measured by cross-correlation, has zero sensitivity for scatterers located *on* the geometrical ray path.

The null sensitivity on the ray is a result that may be counterintuitive, because it seems to contradict ray theory. That is, ray theory defines the geometrical ray path as the only region where the travel time is sensitive to the Earth's structure. However, numerous numerical tests (e.g. Hung *et al.* 2001) have proved that ray theory was more deficient than the finite-frequency formalism of Dahlen *et al.* (2000). In the following, we give an explanation of this paradox - following e.g. Nolet (2008).

On the ray itself the detour time is null ($\Delta T = 0$), which implies that the numerator term, in equation (1.37), $\sin(\omega \Delta T) = 0$, and so the kernel is zero: heterogeneities *on* the ray do not influence the time residual, δt. This null sensitivity on the ray has been named the "doughnut hole" by Marquering *et al.* (1999) - for a visual example of this "hole", see figure 3.24 (in section 3.3.7 of chapter 3). In fact, the only way to remove the zero sensitivity for $\Delta T \to 0$ is to let $\omega \to \infty$ (as assumed in ray theory).

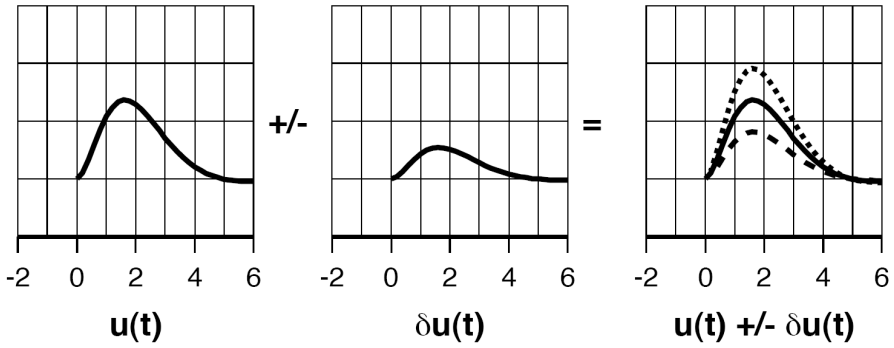


Figure 1.8: If the scattered wave $\delta u(t)$ has no delay because the heterogeneity is located on the ray path (i.e. the detour time is zero), only the amplitude of the direct wave $u(t)$ is affected by the scattered wave, not its phase. Dashed lines represent the effect of adding to $u(t)$ either a positive or negative $\delta u(t)$. From Nolet et. al (2005).

The zero sensitivity is more intuitive if we realize that a cross-correlation measurement involves the wave energy over a time window of finite length. The energy in such a window is moved forward or backward, depending on the sign of the scattered wave, δu, even though δu arrives after the direct wave, u (cf. section 1.3.2 and figure 1.7). But, if the

scatterer is located *on* the ray, there is no delay ($\Delta T = 0$). Therefore, δu can only perturb the amplitude of the wave u, not its phase, as depicted in figure 1.8.

Hung *et al.* (2001) showed that the presence of this "doughnut hole" enables a 3-D travel time sensitivity kernel to account for wavefront-healing effect (cf. figure 1.5). That is, as the propagation distance increases, the anomaly finds itself increasingly able to "hide" within the growing doughnut hole.

1.4.4 Effective width of kernels

Here, we briefly discuss on the effective width of the banana-doughnut kernels (Dahlen *et al.* 2000), following Nolet (2008). We aim to show that the effective 3-D volume of each kernel is directly related to both the length of the cross-correlation time window, and the passband filter, which are used for measuring the time residuals.

First, the numerator of equation (1.37) consists of the term $\sin(\omega\Delta T)$, modulated by the power spectrum $|\dot{m}(\omega)|^2$, and a factor ω^3. One may expect that the kernel has a maximum near $\omega_0\Delta T = \pi/2$, or for $\Delta T = T_0/4$, if T_0 is the dominant period of the signal. This high sensitivity suggests that we shall extend the length of the cross-correlation time window beyond a quarter of the dominant period, for taking into account the scattered energy that arises from this part of the kernel (cf. section 3.3.6 of chapter 3).

There are also secondary maxima in the kernel, corresponding to higher-order Fresnel zones. Their importance diminishes as we widen the bandwidth of the signal, with the passband filter used for the cross-correlation measurement (e.g. Tian *et al.* 2007b). In this study, we have used passband Gaussian filters, which mostly attenuate higher order Fresnel zones (cf. figure 3.23, in section 3.3.6 of chapter 3).

For scatterers near the surface, one should also include incoming rays that hit the scatterer directly, as well as those that first visit a boundary, as the surface or the CMB (cf. section 3.3.4 of chapter 3). Both may have a detour time small enough to allow the scattered wave, δu, to arrive in the cross-correlation time window. The same care should be taken for compound rays, such as for *SS* (cf. section 3.3.3 of chapter 3).

Finally, note that the Fréchet kernel should be set to zero for detour times larger than allowed by the windowing used in the cross-correlation.

1.4.5 Caustics

The presence of caustics slightly change the expression of the sensitivity kernel obtained in equation (1.37). For instance, surface reflected waves such as *SS* (but not *sS*)

acquire a $\pi/2$ phase shift when the reflection at the surface gives rise to an arrival with minimax[13] character. Because *SS* wave undergoes a $\pi/2$ phase shift (Hilbert transform), its waveform will be different from a scattered wave that has escaped the passage through a caustic. In this case, a scatterer *on* the ray will have a maximum effect on the time residual, δt, and the hole in the doughnut will disappear (cf. figure 3.17, in section 3.3.1 of chapter 3).

To acomodate for caustics, a slightly different expression of the kernel has to be used (Dahlen *et al.* 2000):

$$K_\beta(\mathbf{r}_x) = -\frac{1}{2\pi\beta_x}\frac{R_{rs}}{\beta_r R_{xr} R_{xs}}\frac{\int_0^{+\infty}\omega^3|\dot{m}(\omega)^2|\sin[\omega\Delta T(\mathbf{r}_x)-\Delta\Phi(\mathbf{r}_x)]d\omega}{\int_0^{+\infty}\omega^2|\dot{m}(\omega)^2|d\omega} \tag{1.38}$$

, where $\Delta\Phi(\mathbf{r}_x)$ is the phase shift due to passage through caustics. For *SS* waves, $\Delta\Phi(\mathbf{r}_x) = -\pi/2$ in between the two caustics located at $\Delta/3$ and $2\Delta/3$, with Δ the total epicentral distance, and $\Delta\Phi(\mathbf{r}_x) = 0$ elsewhere (e.g. Hung *et al.* 2001).

1.5 Limitations of banana-doughnut kernels

Until now, we have briefly mentioned some of the shortcomings of the "finite-frequency" formalism of Dahlen *et al.* (2000). Here, we aim to give a more complete overview of the major limitations of the corresponding (banana-doughnut) sensitivity kernels.

1.5.1 A first-order and linear finite-frequency theory

The "finite-frequency" formalism of Dahlen *et al.* (2000) is a first-order perturbation theory, which is fully linear.

The first-order aspect is related to the Born approximation. Only single-scattering is taken into account, which represents the first-order effect of wave diffraction. The underlying assumption is that the scattered part of wave energy is small, such that one may ignore the fact that the scattered waves themselves induce scattered energy. Multiple-scattering, which is considered as a second order diffraction effect, is then neglected.

The linearity aspect is intrinsically related to this single-scattering theory, which predicts that the magnitude of a time residual is proportional to the amplitude of the anomaly - cf. equation (1.36). In particular, high- and low-velocity anomalies of equal

[13]The SS phase is a "minimax" phase, because shifting the reflection point towards the source or receiver will actually result in *earlier* arrival times, whereas deflections out of the plane of propagation give the more usual positive delay of the arrival (see Choy & Richards 1975).

magnitude are predicted to have the same absolute time residual. However, some first-order diffraction effects are not linear, such as wavefront-healing. That is, the "healing" is different for high- and low-velocity anomalies (cf. section 1.3.1; figures 1.3 and 1.4). Hence, banana-doughnut kernels cannot model the non-linear part of wave diffraction effects.

1.5.2 Near-field scatterers

Banana-doughnut kernels have been derived only using the far-field expressions for the seismic wavefield. This poses problem for scatterers located near the source or directly beneath the receiver. The fact that the kernels do not contain near- and intermediate-field terms may be an important issue, especially at low frequency (e.g. Nolet *et al.* 2005). Therefore, these kernels are less accurate in regions close to source and receiver.

Morover, banana-doughnut kernels become singular at sources, receivers and caustics (e.g. Tian *et al.* 2007b). However, such a singularity may be handled by assuming a small region of constant velocity perturbation around it, and applying ray theory inside (cf. section 3.3.5 of chapter 3).

Favier *et. al* (2004) have developed finite-frequency kernels for travel time perturbations in the near-field. We refrained from using these more sophisticated kernels, because they are too much time consuming for a large scale tomographic study, in our period range of analysis (10–51 s).

1.5.3 Paraxial approximation

In the "finite-frequency" formalism of Dahlen *et al.* (2000), the geometrical spreading, R, and the detour time, ΔT, are needed to compute banana-doughnut kernels - cf. equation (1.37). An efficient method, to estimate R and ΔT, is dynamic ray tracing (e.g. Tian *et al.* 2007a), which requires that the paraxial approximation is valid. The paraxial approximation assumes that scatterers which influence the measured time residual are located close to the geometrical ray path, within the first few Fresnel zones.

Tian *et. al* (2007b) have extensively studied the accuracy of the paraxial approximation for teleseismic waves - in particular for S and SS. According to their study, the paraxial approximation is valid for most of the seismic waves we aim to use in this thesis (S, ScS and SS), at periods from 10 to 51 s. However, they also mention some cases where the paraxial approximation is not valid, which can lead to some errors, as briefly explained in the following.

Firstly, Tian *et. al* (2007b) point out that large errors may occur if the kernel sen-

sitivity extends over a too large volume, because the paraxial approximation breaks down at large distance from the geometrical ray. This is especially true for minimax phases as *SS* near their reflection point, at periods greater than 20 s. That is, *SS* sensitivity kernels become hyperbolic between their two caustics, instead of elliptical. This considerably extends the sensitivity over several thousands of km, which clearly makes the paraxial approximation not valid.

Secondly, Tian *et. al* (2007b) show that the paraxial approximation is not valid for *SS* waves near antipodal epicentral distances (i.e. >140°). That is, for a station at the antipode, waves departing the source in *all* azimuths arrive at this station. This makes the sensitivity of such kernel to extend over all azimuths in the mantle, which leads to large errors. They conclude that, if data at epicentral distance >140° are crucial in the dataset, efficient paraxial methods cannot be used, and alternative methods are needed to compute the sensitivity near the antipode. For instance, Calvet & Chevrot (2005) set up a table of travel times and geometrical spreading factors, for different source depths and epicentral distances, which may provide correct delocalized kernels near the antipode. We have not used this alternative method in this thesis, because it was not workable during the time of this PhD. Thus, we have refrained from using *SS* travel time measurements with a too large epicentral distance (>140°) in the inversion.

1.5.4 Ray theory based kernels

As previously mentioned (section 1.4.2), in the formalism of Dahlen *et al.* (2000), the zero order wave field (i.e. the synthetic waveform u) is calculated with ray theory - e.g. with the WKBJ method. Therefore, one cannot compute banana-doughnut kernels for evanescent waves (i.e. head waves) and diffracted waves (e.g. *Sdiff*).

The "finite-frequency" approach of Dahlen *et al.* (2000) does not allow for the analysis of *Sdiff*, which is a real issue in global tomography, from a data sampling point of view. That is, without *Sdiff* one cannot efficiently image the base of the mantle, especially in Southern Pacific and Africa. We have then encountered this problem in this thesis (see figure 4.6, in section 4.3.3 of chapter 4).

As an alternative, Nissen-Meyer et al. (2007) propose to use a 2-D spectral element method to compute the perturbed waveform, δu, for a series of source depths. This could allow to efficiently compute sensitivity kernels of core diffracted waves (e.g. *Sdiff*), and increase the resolution in the lowermost mantle.

1.6 Multiple-frequency tomography

In "finite-frequency" body-wave tomography, travel time anomalies are *frequency-dependent*. We can exploit this dependence by performing inversions with data from different frequency bands simultaneously, which is called: "multiple-frequency" tomography.

In multiple-frequency travel time tomography, the general form of the linear inverse problem is

$$\delta t_i(T) = \int_{V_i(T)} K_i(\mathbf{r};T) m(\mathbf{r}) d^3\mathbf{r} \tag{1.39}$$

where $\delta t_i(T)$ is a *frequency-dependent* delay-time measured by cross-correlation between the observed and synthetic waveforms, of the target seismic phase i, both filtered at several periods T. The volume integral $V_i(T)$ is, in practice, limited to the region where the (Fréchet) sensitivity kernel $K_i(\mathbf{r};T)$ has a significant amplitude. The model parameter $m(\mathbf{r})$ represents a velocity perturbation ($\delta c/c$).

At each period, the waveform is influenced by a different weighted average of the structure, through the corresponding 3-D sensitivity kernel (cf. figure 1.9). Measuring the travel time of a seismic phase at several periods should then increase the amount of independent informations in the inverse problem - provided that the kernel sampling of the Earth really differs at different periods.

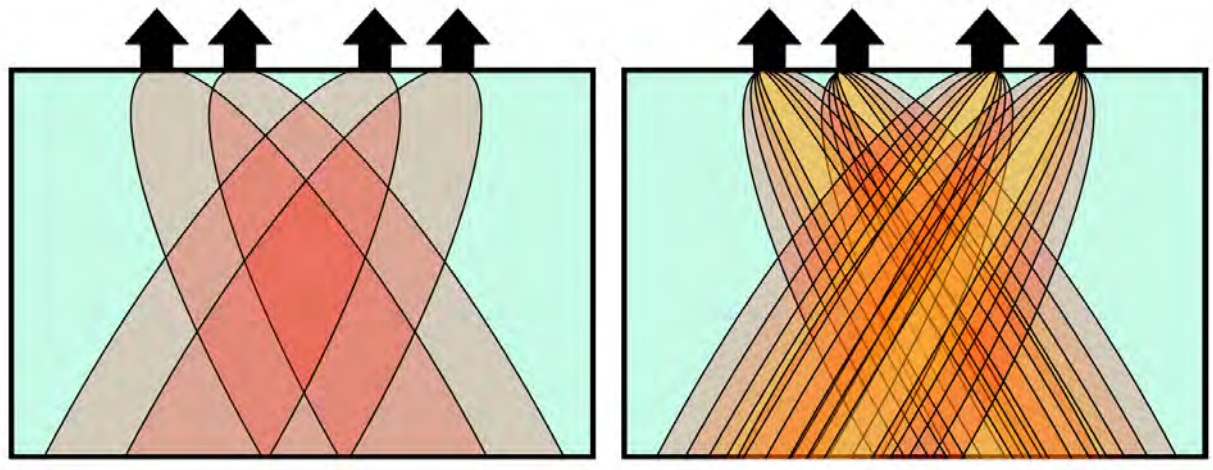


Figure 1.9: Illustration of multiple-frequency tomography: multiplying the constraints on the model, by using multi-band sensitivity kernels. Left: single-band sensitivity kernels. Right: multi-band sensitivity kernels (multiple-frequency tomography). Courtesy from G. Nolet.

1.7 Conclusion

Theoretically, interpreting frequency-dependent body-wave travel times should better constrain tomographic models (e.g. Dahlen 2004), and help us to better understand Earth's mantle dynamics.

Recently, Montelli *et al.* (2004a; 2004b; 2006b) published *P*- and *S*- wave global "finite-frequency" tomographic models, claiming to suggest the existence of deep mantle plumes. They mainly attributed this to an improvement in the resolving power of their "finite-frequency" approach - using the formalism of Dahlen *et al.* (2000) - which is still debated in the community (e.g. de Hoop & van der Hilst 2005a; Dahlen & Nolet 2005; de Hoop & van der Hilst 2005b; Montelli *et al.* 2006a; van der Hilst & de Hoop 2006).

Montelli *et al.* (2004a, 2004b, 2006b) only used travel times measured by matching the "first swing" of a long period (∼20 s) observed waveform with a synthetic (Bolton & Masters 2001). As noticed by Montelli *et al.* (2004b), such a measurement scheme presents a possible bias in dominant frequency, by emphasizing the early part of the waveform rather than the full period. Moreover, the global *S*-wave "finite-frequency" tomographic model obtained by Montelli *et al.* (2006b) is based on (*S*, *ScS*-*S* and *SS*-*S*) travel times measured in a "single" frequency band (∼20s). Hence, it does not benefit from the increased spatial resolution afforded by sensitivity kernels for a range of frequencies.

It is our view that, in order to obtain a "higher-resolution" *S*-wave tomographic model of the mantle - compared to Montelli *et al.* (2006b) - it is necessary to use travel times measured in different frequency bands (i.e. multiple-frequency tomography), and in a way which is fully consistent with the kernels. In this thesis, we have chosen to compute sensitivity kernels with the "finite-frequency" formalism of Dahlen *et al.* (2000). Though approximate, this formalism may handle the first-order and linear part of wave diffraction effects, while keeping manageable the kernels computational cost.

To our knowledge, there is no global dataset of *S*-wave travel times, measured by cross-correlation between observed and synthetic "full" waveforms - as needed in the formalism of Dahlen *et al.* (2000) - over a broad frequency range. We have then focussed on building such a frequency-dependent dataset. Travel times of single (*S*, *ScS*, *SS*) or groups of (*S*+*sS*, *ScS*+*sScS*, *SS*+*sSS*) seismic phases have been measured, within the 10–51 s period range. In chapter 2, we first present the dataset we have obtained, which consists of ∼400,000 frequency-dependent *S*-wave travel times. Then, in order to justify, from a data point of view, the use of a "finite-frequency" approach in body-wave tomography, we tackle the question: "Is there an explicit *structural* dispersion, related to mantle heterogeneities, in our data?".

CHAPTER 2

Frequency-dependent *SH*-wave travel times: observations

"The joy of being a seismologist comes to you when you find something new about the Earh's interior from the observation of seismic waves obtained on the surface, and you realize that you did it without penetrating the Earth or touching or examining it directly."
Keiiti Aki

Foreword

This chapter, which has been published in *Geophysical Journal International*, presents our data collection of S, ScS and SS travel times, measured at several periods between 10 and 51 s. Using this new database, we aim to show that the residual dispersion observed in our data is, at least partly, related to seismic heterogeneity and attenuation in the Earth's interior.

Our results explicitly show that data processing (seismograms recording, measurement method, etc) is accurate enough for retrieving a "structural" dispersion in global S-wave travel times. Therefore, from a data point of view, this favors the use of a "finite-frequency" approach - for global S-wave tomography - that may, in contrast to ray theory, take into acount some wave diffraction effects (e.g. wavefront-healing).

Geophys. J. Int. (2010) doi: 10.1111/j.1365-246X.2010.04667.x

Frequency-dependent effects on global *S*-wave travel times: wavefront-healing, scattering and attenuation

Christophe Zaroli,[1] Eric Debayle[1,2] and Malcolm Sambridge[3]

[1] *Institut de Physique du Globe de Strasbourg (UMR 7516 CNRS, Université de Strasbourg/EOST), 5 rue René Descartes, 67084 Strasbourg Cedex, France. E-mail: christophe.zaroli@eost.u-strasbg.fr*

[2] *Laboratoire de Science de la Terre, Centre National de la Recherche Scientifique, Université Lyon 1 et Ecole Normale Supérieure de Lyon, France*

[3] *Research School of Earth Sciences, Australian National University, Canberra ACT* 0200*, Australia*

Summary

We present a globally distributed dataset of ∼400,000 frequency-dependent *SH*-wave travel times. An automated technique is used to measure teleseismic *S*, *ScS* and *SS* travel times at several periods ranging from 10 to 51 s. The targeted seismic phases are first extracted from the observed and synthetic seismograms using an automated time window algorithm. Travel times are then measured at several periods, by cross-correlation between the selected observed and synthetic filtered waveforms. Frequency-dependent effects due to crustal reverberations beneath each receiver are handled by incorporating crustal phases into WKBJ synthetic waveforms.

After correction for physical dispersion due to intrinsic anelastic processes, we observe a residual travel time dispersion on the order of 1–2 s in the period range of analysis. This dispersion occurs differently for *S*, *ScS* and *SS*, which is presumably related to their differing paths through the Earth. We find that: (1) Wavefront-healing phenomenon is observed for *S* and to a lesser extent *SS* waves having passed through very low velocity anomalies. (2) A preferred sampling of high velocity scatterers located at the CMB may explain our observation that *ScS* waves travel faster at low-frequency than at high-frequency. (3) A frequency-dependent attenuation $q(\omega) \propto q_0 \times \omega^{-\alpha}$, with $\alpha \simeq 0.2$, is compatible with the globally averaged dispersion observed for *S* waves.

2.1 Introduction

Seismic tomography is a standard tool for constraining the structure of the Earth's interior. The resolution of global body wave seismic tomographic models has significantly improved in the last 25 years because of the growth in the number of seismic stations, increase in computational power and development of new analysis tools which extract more information from seismograms. Until recently, ray theory (RT) formed the backbone of all body wave tomographic studies. The long-wavelength structure of the Earth is similar across recent RT tomographic models, even though there is some disagreement on the amplitudes of even the most prominent structures (Romanowicz 2003). Global body wave tomography, based on RT, has revealed a variety of subducting slabs. Some remain stagnant around the 660 km discontinuity, whereas others penetrate into the lower mantle (e.g. Grand *et al.* 1997; Albaréde & van der Hilst, 1999; Fukao *et al.* 2001), which is inconsistent with the hypothesis of a two-layered convection in the Earth. The detection of hypothesized thin thermal plumes (Morgan, 1971) in the mantle has remained elusive in these RT based tomographic images (Romanowicz 2003), but would be of considerable value in understanding the Earth's mantle dynamics.

To be valid, ray theory requires that wavelengths are short and Fresnel zones narrow. Short-period (∼1s) P-wave travel times have so far been extensively used for global RT tomography of the Earth, but provide poor sampling of the upper mantle. Surface wave data, that are generally analysed at very long periods (∼40–300 s), may be combined with S-wave travel times measured at long periods (∼10–51 s). This provides a way to image the entire mantle, because surface waves are more sensitive to the upper mantle, and S-waves to the lower mantle. We have therefore chosen to focus on S-waves in this study. Long-period S-wave data have so far mostly been used in tomographic imaging of the very long-wavelength heterogeneity (>1000 km horizontally) in the Earth, for which a ray theoretical approach is acceptable. However, 1000 km represents a third of the Earth's mantle thickness, and RT breaks down when used for imaging smaller heterogeneities that are of considerable interest in geophysics, such as: parts of slabs sinking in the mantle, or hot rising plumes. These objects are likely to have dimensions that are rather limited in size (∼200 km horizontally). They are very difficult to constrain with RT since wave scattering and wavefront-healing effects are ignored. The effect of wave diffraction phenomena is to make travel time anomalies dependent on Earth structure in a 3-D region around the geometrical ray path, rather than only on the infinitesimally narrow ray path itself. Since this is not taken into account in RT, it seems progress toward obtaining higher-resolution images of small heterogeneities in the mantle requires a movement away from RT.

In an effort to improve upon the infinite-frequency approximations of RT, that are only applicable to the time of the wave *onset*, finite-frequency (FF) approaches have recently emerged in seismic tomography (e.g. Dahlen *et al.* 2000; Tromp *et al.* 2005). For instance, the FF theory developed by Dahlen *et al.* (2000) takes the effects of wave diffraction into account (single scattering only), which makes the imaging of smaller objects or anomalies possible. The RT ray paths are replaced by volumetric sensitivity (Fréchet) kernels. Delay-times (or time residuals) observed in different frequency bands contain information on the size of the heterogeneity. For instance, the healing of a wavefront depends on the ratio between wavelength and size of heterogeneity. In FF tomography, travel time (and amplitude) anomalies are therefore frequency-dependent. In principle one can exploit this dependence, by performing inversions with data from different frequency bands simultaneously. This may lead to an increase in resolution of the tomographic imaging. In FF tomography, the general form of the linear inverse problem is:

$$dt_i(T) = \int_{V_i(T)} K_i(\mathbf{r};T) \cdot m(\mathbf{r}) d^3\mathbf{r} \tag{2.1}$$

where $dt_i(T)$ is a *frequency-dependent* delay-time measured between the observed and synthetic waveforms of the target seismic phase i, both filtered around the period T. The volume integral $V_i(T)$ is theoretically over the entire Earth, but in practice limited to the region where the Fréchet kernel $K_i(\mathbf{r};T)$ has a significant amplitude. The model parameter $m(\mathbf{r})$ represents a velocity perturbation ($\delta c/c$). By measuring the travel time of a seismic phase at several periods, there is a potential for increasing the amount of independent informations in the inverse problem, as at each period the waveform is influenced by a different weighted average of the structure, through the corresponding 3D sensitivity kernel. Recently, Montelli *et al.* (2004a,b; 2006b) published P- and S- wave FF global tomography models, claiming to confirm the existence of deep mantle plumes below a large number of postulated hotspots. They mainly attributed this to an improvement in the resolving power of their FF approach (Dahlen *et al.* 2000).

The potential benefit of using an FF theory for tomography, as well as its significance in mapping mantle plumes, is somewhat controversial. A number of recent studies (e.g. Sieminski *et al.* 2004; de Hoop & van der Hilst 2005a; Dahlen & Nolet 2005; de Hoop & van der Hilst 2005b; Julian 2005; Trampert & Spetzler 2006; Montelli *et al.* 2006a; van der Hilst & de Hoop 2006; Boschi *et al.* 2006) suggest that the effect of such an FF theory could be smaller than that of practical considerations, such as: the level of damping, the weighting of different data sets and the choice of data fit. Another important

factor, which could limit the benefits of the new theory, is the number of FF data used, which has until now remained relatively small compared to the large number of travel times analysed with RT. For instance, FF tomographic models of Montelli *et al.* (2004a; 2004b) are constrained by about 90,000 long-period (∼20 s) *P*, *P*-*PP* and *pP*-*P* travel times, compared to about 1,500,000 short-period (∼1 s) *P* and *pP* travel times, commonly analyzed using RT (van der Hilst *et al.* 2005). Moreover, Montelli *et al.* (2004a, 2004b, 2006b) use travel times measured by matching the 'first swing' of a long period (∼20 s) observed waveform with a synthetic (Bolton & Masters 2001). As noticed by Montelli *et al.* (2004b), such a measurement scheme presents a possible bias in dominant frequency caused by the correlation operator emphasizing the early part of the waveform rather than the full period. Only analysing the early part of the waveform, which is closely related to the wave's onset, may prevent tomographers from taking full advantage of such FF approach. For instance, Dahlen *et al.* (2000) show that if scatterers located off the ray path do not affect the onset of the wave, they can still advance or delay the full waveform. The global *S*-wave FF tomographic model obtained by Montelli *et al.* (2006b) is also based on (*S*, *ScS*-S and *SS*-S) travel times measured in a single frequency band (∼20 s), and hence does not benefit from the increased spatial resolution afforded by sensitivity kernels for a range of frequencies. Recently, Sigloch & Nolet (2006) presented an approach for measuring FF body wave amplitudes and travel times of teleseismic *P* waves, between periods of 2 s to 24 s. They model the first 25 s of a seismogram after the direct *P*-wave arrival, including the depth phases *pP* and *sP*. The best source parameters (source time function, moment tensor, depth) are determined for each earthquake, with a cluster analysis which needs many stations having recorded the same event. This approach is however better suited for local or regional tomographic studies, rather than for global ones.

To take full advantage of using an FF approach, it is necessary to use travel times measured in a way which is fully consistent with the kernels. It is our view that for significant progress to be made, a new global dataset of multiple-frequency body wave travel times is needed. One measured by cross-correlation over a broad frequency range. In this study, we focus on *S*-wave travel times, because they may be readily combined with surface wave data, to obtain a high-resolution tomographic image of the entire mantle. To our knowledge, there is no global database of *S*-wave travel times measured at different frequencies. We aim to measure travel times of single (*S*, *ScS*, *SS*) or groups of (*S*+*sS*, *ScS*+*sScS*, *SS*+*sSS*) phases, within the 10–51 s period range. We use 30 years of broadband seismograms recorded at the Global Seismological Networks (GSN) and distributed by the IRIS and GEOSCOPE data centers.

In section 2.2, we present how we obtained a global dataset of ∼400,000 frequency-

dependent S-wave travel times. An automated scheme for measuring long period S-wave travel times in different frequency-bands has been developed. Automation was necessary to process the type of massive dataset needed for global seismic tomography applications. Travel time measurements have been corrected for elliptical, topographic, crustal, and attenuation effects (Tian *et al.* 2007a). Frequency-dependent effects due to crustal reverberations beneath each receiver have been handled by incorporating crustal phases into WKBJ (Chapman, 1978) synthetic waveforms. A good control of the frequency content of the waveforms, associated with a given travel time, enables us to associate each measurement with a kernel carrying the same frequency information. The resulting multiple-frequency travel times are fully compatible to be inverted with volumetric sensitivity (Fréchet) kernels, irrespective of whether these kernels are computed with adjoint, mode-coupling or paraxial methods (e.g. Tromp *et al.* 2005); travel times measured at a single period can also be inverted using ray theory.

In section 2.3, we focus on frequency-dependent effects occuring on global S-wave travel times in the mantle. If a residual structural travel time dispersion is indeed observable, we would have a new constraint on the nature of seismic heterogeneity and attenuation in the Earth's interior. We are then especially interested in pointing out, in our global dataset, frequency-dependent effects associated to wavefront-healing, scattering and attenuation.

2.2 A global dataset of frequency-dependent S-wave travel times

In this section, we describe our method for building a global dataset of frequency-dependent body-wave travel times. Our automated scheme consists of two main stages. The first involves an automated selection of time windows around a set of target phases, which are present on both the observed and synthetic seismograms. The second stage involves measurements of multiple-frequency travel times by cross-correlating the observed and synthetic waveforms, filtered within the 10–51 s period range, which are contained in the previously selected time windows.

Readers mainly interested in our observations of frequency-dependent effects in our global dataset, such as wavefront-healing, scattering and attenuation, may skip to section 2.3.

2.2.1 Time windows selection and seismic phases isolation

Our time windows selection algorithm follows several of the ideas developed by Maggi *et al.* (2009). These authors present an approach for automated window selection designed for adjoint tomography studies. This class of studies involves 3D numerical simulations of the seismic wavefield and 3D sensitivity (adjoint) kernels (e.g. Komatitsch & Tromp 1995; Komatitsch *et al.* 2002; Tromp *et al.* 2005). The advantage of this approach is that the adjoint kernel is obtained from a numerical calculation, with no need to identify specific seismic phases. That is, the kernel can be computed for any part of the seismogram and takes care of the relevant sensitivities. For this reason, Maggi *et al.* (2009) define time windows covering as much as possible of a given seismogram, while avoiding portions of the waveforms that are dominated by noise. They select time windows on the synthetic waveform only, without identifying specific seismic phases. Each time window on the synthetic seismogram is then associated with the same time window on the observed seismogram, assuming that they contain the same patterns of interference between seismic phases. This assumption will be valid for accurate synthetic seismograms, calculated by 3D propagation through a good 3D Earth model. However, it may not be fulfilled with more approximate spherical-Earth synthetics, computed in a 1D Earth model like IASP91 (Kennett & Engdahl 1991), as used in this study. Since strong 3D heterogeneities, present at the top and bottom of the mantle, can produce large delay-times between observed and 1D synthetic waveforms, we chose to focus our time windows selection on both the observed *and* 1D synthetic waveforms, in order to isolate well identified seismic phases. In appendix a, we describe our time windows selection and seismic phase isolation methodology, which largely makes use of the ideas of Maggi *et al.* (2009), tuned for our particular application (cf. table 2.3).

Finally, before entering into the measurement process (section 2.2.2), the selected observed and synthetic waveforms are tapered and extrapolated outside their isolation time windows with an amplitude set to zero. This is possible because body waves are finite duration pulses. A similar approach was followed by Pollitz (2007), who also used cross-correlation measurements based on narrow-window tapers.

2.2.2 Frequency-dependent travel time measurements

Our time windows selection scheme has allowed us to isolate, in an automated way, a pair of observed and synthetic waveforms, associated with each target seismic phase. We now aim to measure multiple-frequency travel times by cross-correlating the observed and

synthetic waveforms, filtered at different periods, which are contained in the previously selected time windows. We chose to build a global dataset of multiple-frequency travel times within the 10–51 s period range. It is our experience that S-waves are generally prominent compared to seismic noise in this period range. However, this is not always the case for the entire 10–51 s period range. In the following section, we determine the appropriate frequency range of analysis for each target phase to be measured.

2.2.2.a Frequency range of analysis

Seismic body waves, associated with long paths through the Earth, have their high frequency content more severely attenuated than those associated with shorter paths. For instance, S and ScS phases are generally associated with shorter ray paths and higher frequency content than SS. The long period nature of SS is also related to its longer journey in the shallow mantle, compared to S and ScS, which is strongly attenuating for high-frequencies. Moreover, seismic noise has a peak in amplitude at short periods ($\sim 6s$), mainly caused by the oceanic swell, that may significantly pollute the high frequency content of broadband seismograms recorded at oceanic stations.

In our time windows selection and seismic phases isolation scheme (see section 2.2.1 and appendix a), broadband seismograms recorded at the GSN are first band-pass filtered between 7 and 85 s with a non-causal Butterworth filter, whose short and long period corners are denoted by T_1 (7 s) and T_2 (85 s), respectively. In this study, we cover the period range between 7 and 85 s with five overlapping Gaussian filters, whose centre periods T are 10, 15, 22.5, 34 and 51 s (cf. table 2.1). Our aim is to determine the largest frequency range, associated to each targeted waveform, for which travel times can be measured.

We first determine the minimum short period corner of the Butterworth filter, denoted by T_1', for which the target phase can be isolated on the observed seismogram with the approach described in section 2.2.1 and appendix a. T_1' is chosen among three trial values : 7, 11 and 16 s. In the following, we will only consider those of the five Gaussian filters whose centre periods T are greater than T_1'.

A second selection is then performed by computing, for each selected Gaussian filter, the signal-to-noise ratio $SNR(T)$ between the absolute amplitude maxima of the isolated target observed waveform and of the seismic noise. Seismic noise is evaluated from a 100 s time window (see figure 2.10.a), taken on the observed seismogram, before the first arrival time among the S, ScS and SS phases. Among the selected Gaussian filters, we only keep those for which a ratio $SNRT(T)$ greater than 3 is found. In the following section 2.2.2.b, we describe how we measure time residuals at several periods. For each couple of observed

and synthetic waveforms, measurements are made at the periods for which Gaussian filters have been selected.

Table 2.1: Passband Gaussian filters for which frequency-dependent travel times were measured.

band index	**1**	**2**	**3**	**4**	**5**
Centre period (s), T	10.0	15.0	22.5	34.0	51.0
Centre frequency (mHz), ν	100.0	66.7	44.5	29.4	19.6
1σ–corner freq. (mHz), $\nu \pm \sigma$	81.0–119.0	54.0–79.4	36.0–53.0	23.8–35.0	15.9–23.3
2σ–corner freq. (mHz), $\nu \pm 2\sigma$	62.0–138.0	41.4–92.0	27.6–61.4	18.2–40.6	12.2–27.0

Where ν is the central frequency, $T = 1/\nu$ is the central period, and σ the standard deviation. These filters are Gaussian in the frequency domain.

2.2.2.b Measuring time residuals

We now aim to measure multiple-frequency time residuals for each optimal pair of observed and synthetic waveforms, within the frequency range of analysis corresponding to each target phase (cf. sections 2.2.1 and 2.2.2.a). These multiple-frequency time residuals are designed in order to be fully compatible with an FF approach for tomography, such as the one developed by Dahlen *et al.* (2000). That is, the FF time residual τ_m, associated with the period T, is defined as the time maximizing the cross-correlation function, $\gamma_{d,s}(\tau)$, between the observed, $d(t)$, and synthetic, $s(t)$, waveforms, both filtered around the period T. The cross-correlation function is defined as:

$$\gamma_{d,s}(\tau) = \int_{-\infty}^{+\infty} d(t) \cdot s(t-\tau) dt. \tag{2.2}$$

Picking the maximum of the cross-correlation is usually accurate. However, in some cases the highest absolute value of the cross-correlation corresponds to a cycle skip and leads to large error. By eye, such cycle-skips can often be recognized because the rest of the signal, away from the maximum, has a mismatch. We have noticed that cycle-skipping problems are more likely to occur when the two filtered signals, $d(t)$ and $s(t)$, are strongly dominated by one particular period.

In order to minimize biases due to cycle-skipping problems, our multiple-frequency time residuals are measured by determining the maximum position, by curve-fitting, of the new function: $F_3(\tau)$. This function $F_3(\tau)$ is designed to provide an accurate estimation of τ_m, while minimizing cycle-skipping. We show, in appendix b, that the function $F_3(\tau)$ and the cross-correlation function $\gamma_{d,s}(\tau)$ are maximized for the same time residual τ_m.

Therefore, Fréchet kernels designed to be used with time residuals measured by maximizing $\gamma_{d,s}(\tau)$ (e.g. Dahlen *et al.* 2000) can also be used with residuals obtained by maximizing $F_3(\tau)$. In the following, we define the function $F_3(\tau)$ and show how it assists in minimizing cycle-skips.

Following Ritsema & van Heijst (2002), we define two quantities, function of the time shift τ, that describe the similarity between the filtered observed, $d(t)$, and time-shifted synthetic, $s(t-\tau)$, waveforms. These two quantities are noted $F_1(\tau)$ and $F_2(\tau)$. They are defined such as:

$$F_i(\tau) = \begin{cases} F_i(\tau) & \text{if } \ 0 \leq F_i(\tau) \leq 1, \quad i = 1, 2 \\ \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

The first quantity $F_1(\tau)$ is the least-squares misfit between $d(t)$ and $s(t-\tau)$:

$$F_1(\tau) = 1 - \frac{\int_{-\infty}^{+\infty}[d(t) - s(t-\tau)]^2 dt}{\int_{-\infty}^{+\infty} d^2(t) dt}. \tag{2.4}$$

Note that $F_1(\tau)$ is close to 1 when τ minimizes the misfit between $d(t)$ and $s(t-\tau)$. The second quantity $F_2(\tau)$ measures the wave shape similarity between $d(t)$ and $s(t-\tau)$:

$$F_2(\tau) = \frac{min[A_1(\tau), A_2(\tau)]}{max[A_1(\tau), A_2(\tau)]} \tag{2.5}$$

where $A_1(\tau)$ and $A_2(\tau)$ minimize the functions f and g, respectively. The functions f and g are defined as:

$$\begin{cases} f(x) = \int_{-\infty}^{+\infty}[(d(t) - x \cdot s(t-\tau)]^2 dt \\ \\ g(x) = \int_{-\infty}^{+\infty}[x^{-1} \cdot d(t) - s(t-\tau)]^2 dt. \end{cases} \tag{2.6}$$

Which leads to:

$$\begin{cases} A_1(\tau) = \gamma_{d,s}(\tau)/\gamma_{s,s}(0) \\ \\ A_2(\tau) = \gamma_{d,d}(0)/\gamma_{d,s}(\tau) \end{cases} \tag{2.7}$$

where $\gamma_{d,d}(0)$ and $\gamma_{s,s}(0)$ are the autocorrelation values, at zero lag-time, of the observed and synthetic waveforms, respectively. Note that $F_2(\tau)$ is close to 1 when τ maximizes the wave shape similarity between $d(t)$ and $s(t-\tau)$. Finally, the function $F_3(\tau)$ is defined as:

$$F_3(\tau) = \frac{F_1(\tau) + F_2(\tau)}{2}. \tag{2.8}$$

The function $F_3(\tau)$ includes a more sophisticated information on the misfit, through $F_1(\tau)$, and on the wave shape similarity, through $F_2(\tau)$, than the cross-correlation function, $\gamma_{d,s}(\tau)$. When cycle-skips occur, it is statistically easier to find, in an automated way, the appropriate residual time by using the function $F_3(\tau)$, because its global maximum is enhanced and its secondary maxima (corresponding to cycle-skips) are minimized, compared to the ones of the cross-correlation function. An example of comparison between the two functions $F_3(\tau)$ and $\gamma_{d,s}(\tau)$ is shown in figure 2.11. By experimentation, we only retained the time residuals corresponding to a function $F_3(\tau)$ with a unique maximum greater than 80%, and with no secondary maximum greater than 70%. The use of the function $F_3(\tau)$, rather than a simple cross-correlation, has proved to be very useful for the automation of our measurement process (cf. figures 2.1 and 2.11). The function $F_3(\tau)$ mimics very well the seismologist's, often visual, decision in choosing the appropriate time residual.

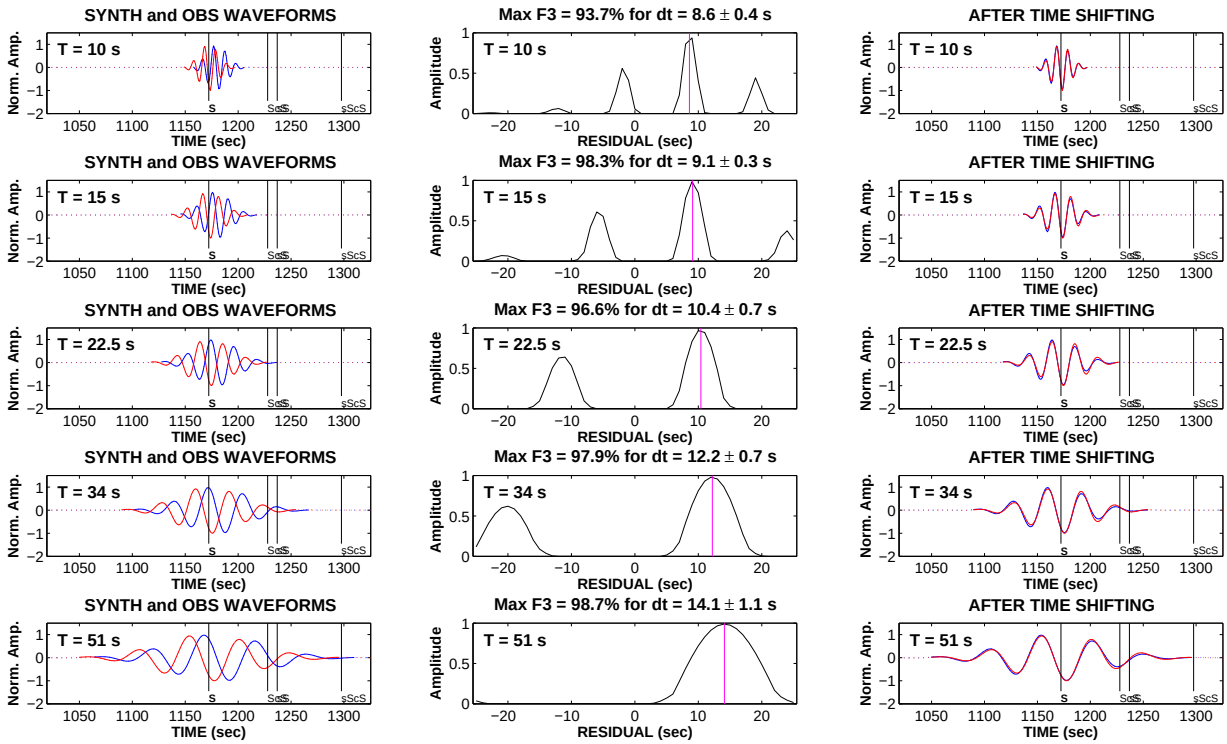


Figure 2.1: Measuring multiple-frequency time residuals. Left column: observed (blue) and synthetic (red) isolated and tapered S phase waveforms (corresponding to figure 2.10.a). Middle column: for each filtering period (T), $F_3(\tau)$ is maximized for a delay-time $\tau_m(T)$. Right column: observed and synthetic waveforms after time shifting of $\tau_m(T)$. Note that each raw corresponds to a different filtering period T.

2.2.2.c Measurement errors

Errors on our time residual estimates can result from waveform distortion, owing to the effects of both the noise and the approximations made in the synthetics computation. We aim here to approximate the standard deviation (σ) related to each measured time residual (τ_m). Following Chevrot (2002), we first compute the correlation coefficient, $\gamma_{d,s}(\tau_m)$, between the observed and time-shifted synthetic waveforms. This coefficient is then compared with the autocorrelation function, $\gamma_{s,s}(\tau)$, of the synthetic waveform. Finally, we approximate the error (σ) by the time lag at which the correlation coefficient, $\gamma_{d,s}(\tau_m)$, is observed in the autocorrelation function, $\gamma_{s,s}(\tau)$. That is:

$$\sigma = \{\tau \mid \gamma_{d,s}(\tau_m) = \gamma_{s,s}(\tau)\}\,. \tag{2.9}$$

Hence, observed waveforms exhibiting a strong correlation with the synthetic waveform will be attributed low errors. On the other hand, signals strongly contaminated by noise will produce larger travel time residual errors.

2.2.2.d Travel time corrections for global seismic tomography

In this study we aim to build a global dataset of multiple-frequency time residuals, suitable for imaging the Earth's mantle structure using inversion schemes based on equation (2.1). Time residuals are determined by cross-correlating observed and synthetic waveforms (cf. section 2.2.2.b). The synthetic waveforms used in this study are computed in a spherical Earth with the WKBJ method for the IASP91 velocity model extended with the Q model from PREM (Dziewonski & Anderson 1981).

We need to apply corrections to the predicted travel times, computed in the radial IASP91 reference velocity model, to account for known deviations from spherical symmetry in the Earth. We use the software by Tian *et al.* (2007a) to compute the ellipticity, dt_{ell}, crustal, dt_{cru}, and topographic, dt_{top}, travel time corrections, for each seismic phase (*S*, *sS*, *ScS*, *sScS*, *SS*, *sSS*) present in the WKBJ synthetic. The travel time after correction, for each seismic phase, is:

$$T_{corr} = T_{BG} + dt_{ell} + dt_{cru} + dt_{top} \tag{2.10}$$

where T_{BG} is the predicted travel time for the spherically reference (background) model (IASP91 in this study). We use the 3-D global crustal model CRUST2.0 (2°x2°), by Gabi Laske, which is an updated version of an earlier model CRUST5.1 (5°x5°), by Mooney *et al.* (1998).

As they propagate through the Earth, seismic waves experience attenuation and

dispersion resulting from microscopic dissipative processes, operating at a variety of relaxation times. Intrinsic attenuation causes dispersion of seismic velocities, decreasing the velocities of longer period waves, compared to shorter period ones. Properly correcting for the dispersion effect is crucial as we aim to use our multiple-frequency delay-times, determined in different frequency bands, to constrain velocities in the Earth. Frequency dependence of attenuation q can be represented by a power law:

$$q \propto q_0 \cdot \omega^{-\alpha}. \tag{2.11}$$

Seismic studies routinely assume that, within the seismic band, α cannot be resolved and thus implicitly rely on the frequency-independent attenuation model, i.e. $\alpha = 0$, of Kanamori & Anderson (1977). Usually, the difference in wave-speeds due to an attenuation value q at two frequencies $\omega_{1,2}$ is calculated using the expression:

$$\frac{V(\omega_2)}{V(\omega_1)} = 1 + \frac{q}{\pi} \cdot \ln(\omega_2/\omega_1) \tag{2.12}$$

which is only valid when $\alpha = 0$ (Kanamori & Anderson 1977). However, non-zero values of α (see section 2.3.3) require the use of a different expression (Anderson & Minster 1979):

$$\frac{V(\omega_2)}{V(\omega_1)} = 1 + \frac{q(\omega_1)}{2} \cdot \cot(\alpha\pi/2) \cdot [1 - (\omega_2/\omega_1)^{\alpha}]. \tag{2.13}$$

The values of α and $q(\omega_1)$ may significantly affect the magnitude of the dispersion correction. If one relies on a frequency-independent attenuation model, i.e. $\alpha = 0$, one should correct the multiple-frequency time residuals, measured by cross-correlation, by adding the physical dispersion correction, $dt^{\alpha=0}_{disp}$, to equation (2.10), with:

$$dt^{\alpha=0}_{disp}(T) = -\frac{t^*}{\pi} \cdot \ln(\frac{T_0}{T}). \tag{2.14}$$

T_0 is the reference period of the velocity model (T_0=1 s for IASP91), and T is the centre period of the Gaussian filter used to analyse the target phase. The parameter t^* is determined by kinematic ray tracing (Tian *et al.* 2007a):

$$t^* = \int_0^L \frac{dl}{c \cdot Q}. \tag{2.15}$$

The integration is along the ray path and Q is the quality factor ($Q = 1/q$) from the PREM model. On the other hand, if one relies on a frequency-dependent attenuation model, i.e.

$\alpha \neq 0$, one should correct the multiple-frequency time-residuals by adding a different physical dispersion correction, $dt_{disp}^{\alpha \neq 0}$, to equation (2.10), with:

$$dt_{disp}^{\alpha \neq 0}(T) = -\frac{t^*}{2} \cdot \cot(\alpha\pi/2) \cdot [1 - (T/T_0)^{\alpha}]. \tag{2.16}$$

2.2.2.e Frequency-dependent crustal effects

Removing the crustal signature from teleseismic travel times is very important to reduce the trade-off between crustal and mantle velocity heterogeneities in seismic tomography.

Yang & Shen (2006) discussed frequency-dependent effects due to continental crustal reverberations on teleseismic P wave travel times, if strong reverberations arrive early enough to influence the cross-correlation. They observed a difference of travel time up to 0.6 s between P waves filtered at short (i.e. 0.5–2 s) and long (i.e. 10–33 s) periods. Ritsema *et al.* (2009) also show that at relatively low frequencies, when the wavelengths of P- and SH-waves are similar to the thicknesses of the crust, crustal travel times are frequency dependent, either for continental or oceanic crustal structures. They use a global crustal model to calculate maps of frequency-dependent crustal effects for distinct SH-waves (e.g. S, SS). However, such crustal corrections should only be applied to 'single' seismic phases, free of interference with other seismic phases. When two phases interfere, such as S and sS in the case of a shallow earthquake, we cannot add linearly the frequency-dependent crustal travel time corrections of the two seismic phases. Indeed, the interference pattern between two phases has a very complex and non-linear frequency dependence.

In modern global tomographic studies aimed at improving image resolution, we wish to include the large amount of seismograms corresponding to shallow earthquakes. As a consequence, if one aims to correct for frequency-dependent crustal effects, we cannot ignore the non-linear problem of interference pattern between direct and depth phases. One way to account for this non-linearity is to incorporate crustal reverberations into the synthetic waveform used to cross-correlate, but this requires prior knowledge of crustal structure, not just its thickness (e.g. Nolet 2008).

In this study, we directly incorporate into the WKBJ synthetics crustal phases reverberated on the receiver side, using the global crustal model CRUST2.0. We incorporate S, sS, ScS, $sScS$, SS and sSS crustal reverberations, when they are present in our time windows with a significant amplitude. The WKBJ approach does not allow us to simultaneously model crustal phases reverberated in different crustal models. We therefore ignore crustal phases associated to surface reflection points at the source side (e.g. sS), or at the

bounce point in the case of *SS* waves.

Devilee *et al.* (2003) observed that interference between direct (e.g. *S*) and depth (e.g. *sS*) phases can also give apparent travel time dispersion. Our modelling of the dispersion associated to crustal effects, and to direct and depth phases interference, may not be perfect, but we believe that our measured multiple-frequency travel times are accurate enough to assess frequency-dependent effects occuring on global *S*-wave travel times in the Earth's mantle. The validity of this assumption can be checked by using deep earthquakes, for which crustal phases reverberated at the source side arrive too late to influence the cross-correlation, so that there is no interference between direct and depth phases. Therefore, *S* and *ScS* waves are only affected by crustal phases reverberated at the receiver side, which are incorporated in our WKBJ synthetics. We checked that all the results of this study about wavefront healing (section 2.3.1), scattering (section 2.3.2) and attenuation (section 2.3.3), remain the same if only deep earthquakes are used. Even if *SS* waves are affected by a surface reflection at the bounce point, whose associated crustal phases are not modeled here, this leads to the same result.

2.2.3 Data selection

A total of 28,810 earthquakes, with a body wave magnitude $m_b \geq 5.5$, were preselected from the Harvard centroid moment tensor (CMT) catalog, between 1976/01/01 and 2008/03/31. We obtained broadband seismograms (LH channel), associated with the selected events, from the IRIS and GEOSCOPE data centers, at almost 270 stations of the International Federation of Digital Seismograph Networks (FDSN). Only earthquakes with a magnitude such as $5.5 \leq m_b \leq 6.5$, and with a source half duration time $h_{dur} < 6$ s, were used in this study. These criteria reject waveforms strongly complicated by the earthquake rupture process (e.g. Ritsema & van Heijst 2002). Moreover, Devilee *et al.* (2003) show that an asymmetric source time function may cause significant dispersion at periods shorter than the source duration time, but that this dispersion is small at greater periods. As we aim to measure multiple-frequency delay-times within the long period range 10–51 s, our measurements are not expected to be biased by this kind of dispersion. Therefore, we assume the source time function to be Gaussian, and use the expression given by Komatitsch & Tromp (2002). This assumption is appropriate for global teleseismic seismograms, but for local or regional studies, one should instead try to determine the exact source time function (Sigloch & Nolet 2006).

2.2.4 Dataset robustness

Our multiple-frequency dataset includes single-phase travel-times (*S*, *ScS* and *SS*), completed with travel-time measurements for which the target phase interferes with its depth phase (*S*+*sS*, *ScS*+*sScS* and *SS*+*sSS*). This kind of interference is often associated with shallow earthquakes, while single-phase travel-times generally correspond to deep events. We have specifically rejected measurements associated with waveforms that could be contaminated by other kind of interference. This is important for the tomographic inversion, as we aim to associate our multiple-frequency travel-times with the appropriate sensitivity kernels.

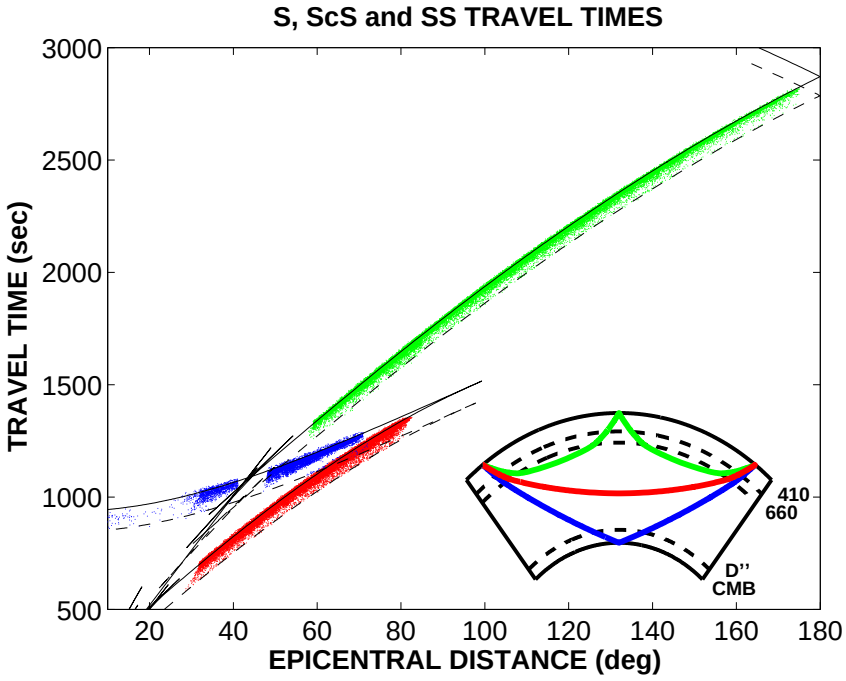


Figure 2.2: Travel time versus epicentral distance plot showing our distribution of travel times for S (red), ScS (blue) and SS (green) seismic phases, superimposed to their corresponding theoretical travel-time curves. Travel times are given for the shortest filtering period T for which each target phase has been measured (cf. section 2.2.2.b). Theoretical travel time curves are shown with solid lines for 0 km source depth and dashed lines for 410 km source depth. We also show an example of S (red), ScS (blue) and SS (green) ray paths into the Earth.

Figure 2.2 summarizes our travel time observations for each target phase, superimposed on the theoretical travel time curves, as a function of epicentral distance. Although we measure travel times of *SS* phases up to distances reaching $\sim 170°$, those *SS* measure-

ments near antipodal epicentral distances ($\geq 140°$) should not be used with a kernel based upon the paraxial approximation (Tian *et al.* 2007b). Nevertheless, these near antipodal *SS* measurements could be used with more sophisticated kernels (Calvet & Chevrot, 2005).

As in previous studies (e.g. Bolton & Masters 2001), the most subtle problem that we face is the accidental measurement of a depth phase (e.g. *sS*) when the direct phase (e.g. *S*) is poorly excited. Engdahl *et al.* (1998) show how the problem can be reduced using statistical methods. Bolton & Masters (2001) measure the arrival polarity to identify depth phase problems. Their measurements are based on the cross-correlation between the first swing of the observed and synthetic direct phases. Our analysis relies on the entire waveform(s) of the target phase(s). We impose (see appendix a, section a.6) a high correlation coefficient ($CC_{max} \geq 80\%$) between the observed and synthetic waveforms. Therefore, if a pattern of interference between two phases is present on the synthetic waveform, a similar pattern must also be met on the observed data. If this pattern is not found, the travel-times are not measured. Uncertainties on Harvard centroid moment tensor (CMT) solutions are likely to affect the relative amplitudes of the direct and depth phases, especially when one of the take-off azimuth is near a nodal plane. In this case, the observed and synthetic two-phase waveforms are expected to differ, i.e. CC_{max} is low, and the data are rejected. The cross-correlation criteria, i.e. $CC_{max} \geq 80\%$, associated with a waveform search in a specific time window (see appendix a, section a.5), enable us to reject a large number of data for which the CMT source mechanism is not reliable. This is especially true for near nodal measurements for which inacuracies in the CMT solution often imply significant differences between the observed and synthetic waveforms. This allows us to discard most accidental measurements of a depth phase when the direct phase is poorly excited.

Although our database has been built for the transverse component (*SH*), our approach can also be easily extended to *P*- or *SV*-components, provided that: (1) new crustal phases, as for instance from *P* to *S* conversions in the crust, are modeled and added in the synthetics; (2) new depth phases, such as *pS*, are added in the synthetics; (3) new seismic phases interference patterns, such as *S* with *SKS*, are taken into account.

2.2.5 Global patterns in the data

Figure 2.3 displays the ray coverage achieved with our database, for different depth ranges covering the entire lower mantle. The current coverage of seismic stations allows us to achieve a good sampling of the Northern Hemisphere for all but the rays with the shallowest lower mantle turning depths. Coverage in the Southern Hemisphere still remains

a problem, but many areas appear to be sampled well enough to reveal consistent patterns.

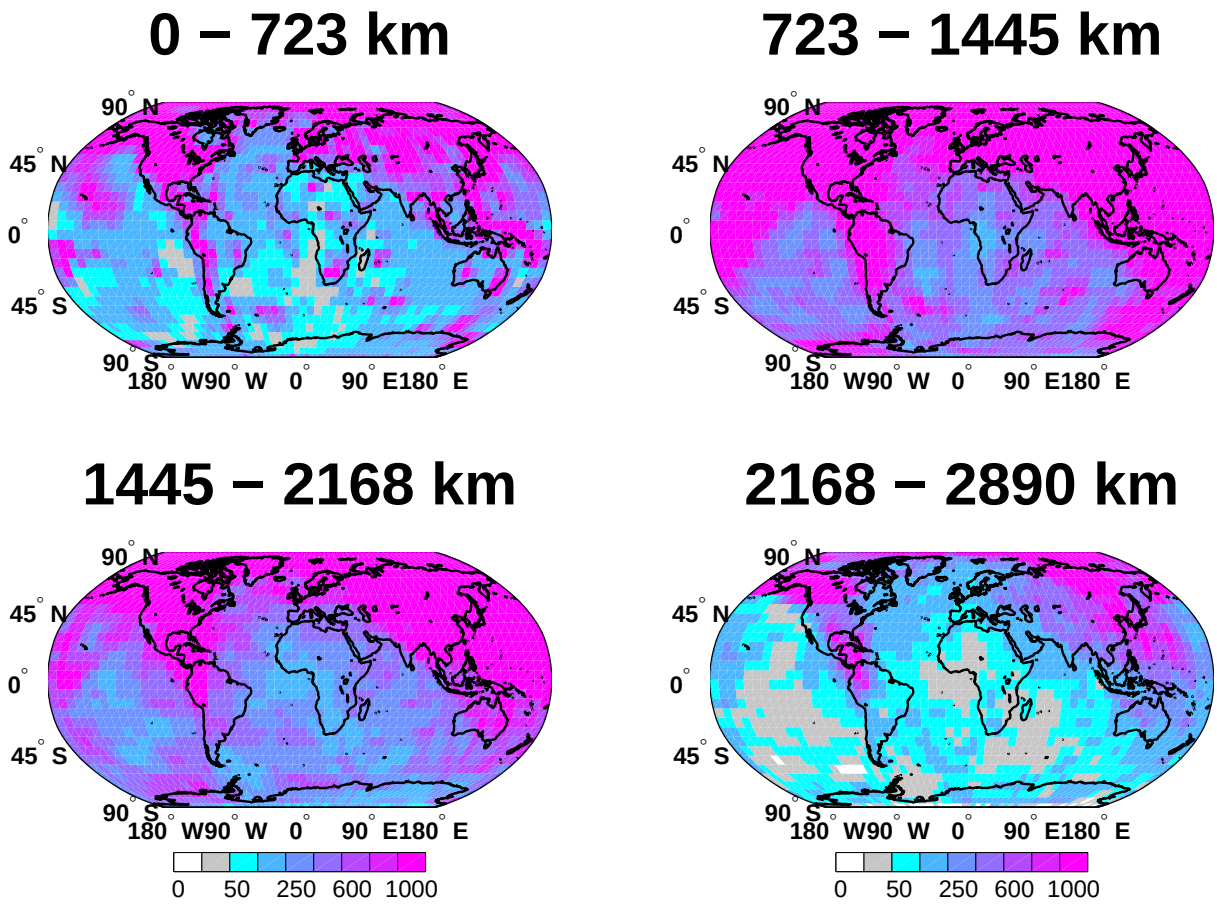


Figure 2.3: Ray density maps per 6°×6° cells, shown at four different depth slices. The colorscale represents the number of rays in each cell, normalized by the cell size. Cells with less than 10 rays are shown in white.

Figure 2.4 shows the geographic distributions of the S and ScS residuals, plotted at the surface projection of the ray turning points. Residuals are averaged in 6°x6° cells and shown over four ray turning depth ranges. As in previous studies (e.g. Bolton & Masters 2001; Houser *et al.* 2008), we observe large-scale patterns in both sign and amplitude. These patterns are clearly associated with the long wavelength structure seen in global tomography. Our observations suggest fast regions beneath Asia, Arctic, North and South America in the depth range between 650 and 1700 km. For example, strong fast residuals observed at turning points between 650 and 1700 km, below the northern part of South America, correspond to the subduction of the Nazca plate (van der Hilst *et al.* 1997). These

time residuals are consistent with the high velocity ring around the Pacific seen in most S-wave tomographic models (e.g. Dziewonski 1984; Masters *et al.* 1996, 2000). The deep turning rays, deeper than 1700 km are delayed by the slow areas seen in global tomography (e.g. Ritsema *et al.* 1999) at the base of the mantle over much of the central Pacific Ocean and beneath South Africa. The agreement between our observations and global tomography suggests that mantle structure in the region of the ray turning point is responsible for most of the observed patterns.

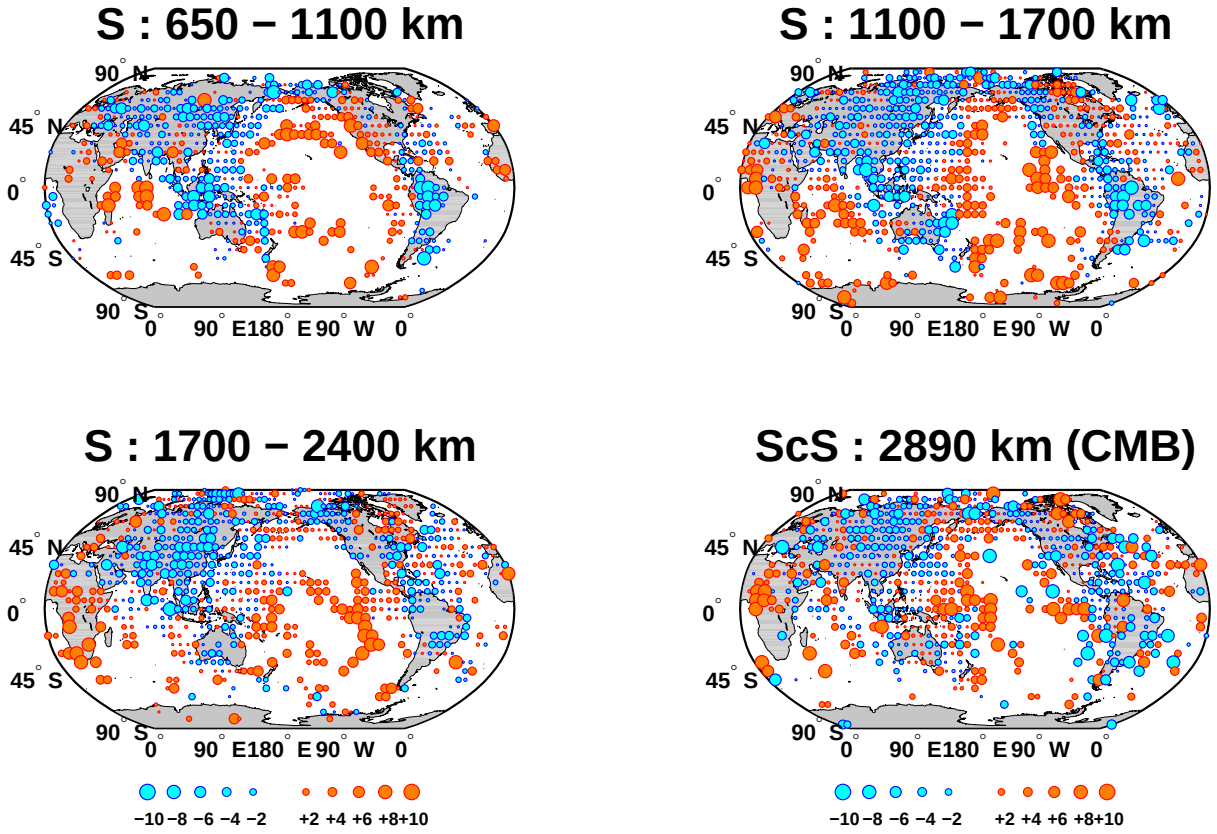


Figure 2.4: Data patterns: S and ScS (highest-frequency) time residuals (in s) are plotted at their turning points in 6°×6° cells (with 3 turning points at least in each cell), and shown at different depth slices. Time residuals correspond to the shortest filtering period T for which each target phase has been measured. The mean (μ) from each depth slice has been removed: μ_S(650-1100 km)=2 s, μ_S(1100-1700 km)=1.7 s, μ_S(1700-2400 km)=0.7 s and μ_{ScS}(2890 km)=-3.9 s. Since most cells contain residuals from many azimuths, we can infer that a majority of the signal shown here must be accumulated near the ray turning point.

2.3 Frequency-dependent effects on global *S*-wave travel times

In this section, we focus on frequency-dependent effects occuring on global S-wave travel times in the mantle. If a residual travel time dispersion is indeed structural and observable, we would have a new constraint on the nature of seismic heterogeneity and attenuation in the Earth's interior.

Table 2.2 summarizes the mean and standard deviation of our S, ScS and SS multiple-frequency time residual measurements, in our period range of analysis. After correction for physical dispersion (cf. section 2.3.3 and figure 2.9) due to intrinsic anelastic processes, under the hypothesis of a frequency-independent attenuation (i.e. $\alpha = 0$), we observe a clear frequency-dependency in our measurements. For instance, when the period increases, the mean delay-time decreases for ScS phases but increases for S and SS (cf. table 2.2). At first glance, the frequency-dependency observed in our global measurements is not directly related to specific seismic heterogeneity or attenuation in the mantle.

Table 2.2: Summarized global multiple-frequency time residuals

Period (s)	10	15	22.5	34	51
S waves					
N	19,008	36,708	49,089	46,000	38,238
μ (s)	0.6	0.9	1.4	1.7	2.6
σ (s)	$\pm$5.5	$\pm$5.6	$\pm$5.8	$\pm$5.8	$\pm$6.4
ScS waves					
N	4,939	10,094	13,069	11,480	8,801
μ (s)	-2.8	-3.8	-4.7	-6.1	-8.3
σ (s)	$\pm$8.1	$\pm$8.4	$\pm$8.3	$\pm$8.1	$\pm$8.5
SS waves					
N	4,189	21,777	49,963	50,041	40,882
μ (s)	0.2	0.7	1.2	1.2	1.5
σ (s)	$\pm$8.5	$\pm$8.1	$\pm$7.7	$\pm$7.5	$\pm$7.9

Where N is the number of measurements, μ is the mean, and σ is the standard deviation of the best fitting Gaussian function of each histogram of our S, ScS and SS datasets. Both single phase (e.g. S) and two-phase (e.g. S+sS) time residuals are considered.

So far, global tomographers have only relied on the inversion process - i.e. solving equation (2.1) - to unravel all the complex frequency-dependency information contained in their global multiple-frequency travel time measurements (e.g. Montelli *et al.* 2004a). In the following, we aim to give evidence that a residual structural dispersion is contained

in our data. We first point out, in section 2.3.1, that wavefront-healing produced by very low velocity anomalies is clearly observed in our S wave dataset and may contribute to the observed SS dispersion. We also report on our observation that ScS waves of our global dataset travel faster at low-frequency than at high-frequency. We suggest, in section 2.3.2, that a preferred sampling of high velocity scatterers located at the CMB, may explain the peculiar ScS dispersion pattern. Finally, we argue, in section 2.3.3, that the globally averaged dispersion observed for S and SS travel times is compatible with a frequency-dependent attenuation model for the average mantle.

2.3.1 Evidence for wavefront-healing from local to global scale

An important effect caused by the wave's frequency being finite is wavefront-healing (Nolet & Dahlen 2000; Hung *et al.* 2001; Nolet *et al.* 2005). Wavefront-healing is a ubiquitous diffraction phenomenon, which depends upon the wave's frequency and the anomaly size. It occurs whenever the scale of any geometrical irregularities in a wavefront are comparable to the wavelength of the wave (Gudmundsson, 1996), and affects cross-correlation travel-time measurements (Hung *et al.* 2001). That is, a low velocity anomaly creates a delayed wavefront with an unperturbed zone that may be filled in (i.e. healed) by energy radiating from the sides, using Huygens' Principle (Nolet *et al.* 2005). Wavefronts of longer waves heal more quickly as a function of distance from the perturbation (e.g. Nolet 2008). Therefore, if a seismic wave passes through a low velocity anomaly, the longer the wave period is, the more important the healing will be, and therefore the less the wave will be apparently delayed at the receiver. The corresponding time residuals, dt, measured by cross-correlation at different filtering periods, T, will then lead to a decreasing dispersion curve $dt(T)$.

2.3.1.a Wavefront-healing at local scale

Here we focus on travel time dispersion of S waves recorded at the LKWY broadband seismic station (see figure 2.5), which belongs to the US network. This station has the particularity of being located above the Yellowstone hotspot, whose seismic signature is a very low velocity anomaly (see figure 2.5.c). When the wave's period increases, such as its wavelength grows to a length comparable to the dimension of the anomaly, wavefront-healing becomes significant even at short distance from the anomaly. A seismic wave traveling through the Yellowstone low-speed anomaly is then expected to be significantly affected by wavefront-healing when recorded at the LKWY receiver. The corresponding dispersion curve, $dt(T)$, is therefore expected to decrease. For comparison, we also analyse travel time

dispersion of S waves recorded at five other seismic stations located in the close vicinity of the LKWY station (see figures 2.5.a and 2.5.b). All these S waves are associated with earthquakes located in similar regions along the AA' profile (see figure 2.5.a), so that we can attribute the observed travel time differences to the receiver side.

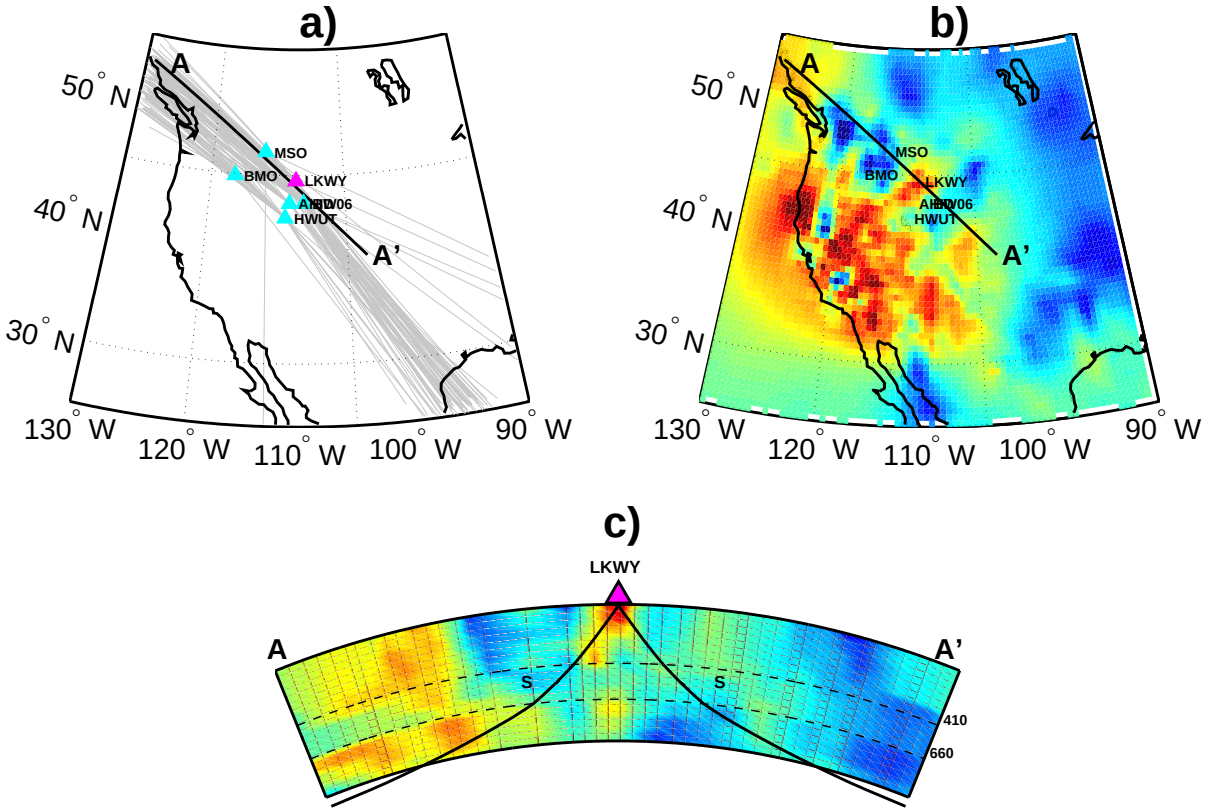


Figure 2.5: ***a)*** *Map showing all the ray paths used in figure 2.6 and six seismic stations located in the vicinity of the Yellowstone hostpot.* ***b)*** *P-wave velocity anomaly model at 200 km depth around the Yellowstone hotspot.* ***c)*** *Cross-section showing the Yellowstone hotspot (low velocity anomaly) beneath the seismic station LKWY. The tomographic model (MITP–USA–2007NOV) is from Burdick et al. (2008). The colorscale shows in red/blue the low/high velocity anomalies, respectively.*

Figure 2.6 shows the associated dispersion curves, measured within the 10–51 s period range. We plot $dt(T) - dt(T = 10s)$ so that increasing/decreasing dispersion curves are above/below zero of the y-axis. $dt(T = 10s)$ provides an information on the average velocity anomaly encountered by S waves between the source and the receiver. We therefore plot dispersion curves with $dt(T = 10s) < 4s$ in cool colors (blue, cyan and green) and dispersion curves with $dt(T = 10s) \geq 4s$ in warm colors (orange and red), where blue/red are for the lowest/highest values of $dt(T = 10s)$. Figure 2.6.a shows that 85% of the 59

dispersion curves recorded at LKWY are *decreasing* and associated with $dt(T = 10s) \geq 4s$ (warm colors). Figures 2.6.b–f also show that at the other stations, ∼83% of the dispersion curves are *increasing* and mainly associated with $dt(T = 10s) < 4s$ (cool colors). The large positive time residuals observed at station LKWY are likely to be due to the low-speed anomaly observed below Yellowstone (figure 2.5.c). Our observations suggest that the particular dispersion pattern recorded at LKWY is due to wavefront-healing and related to the crossing of the Yellowstone low-speed anomaly.

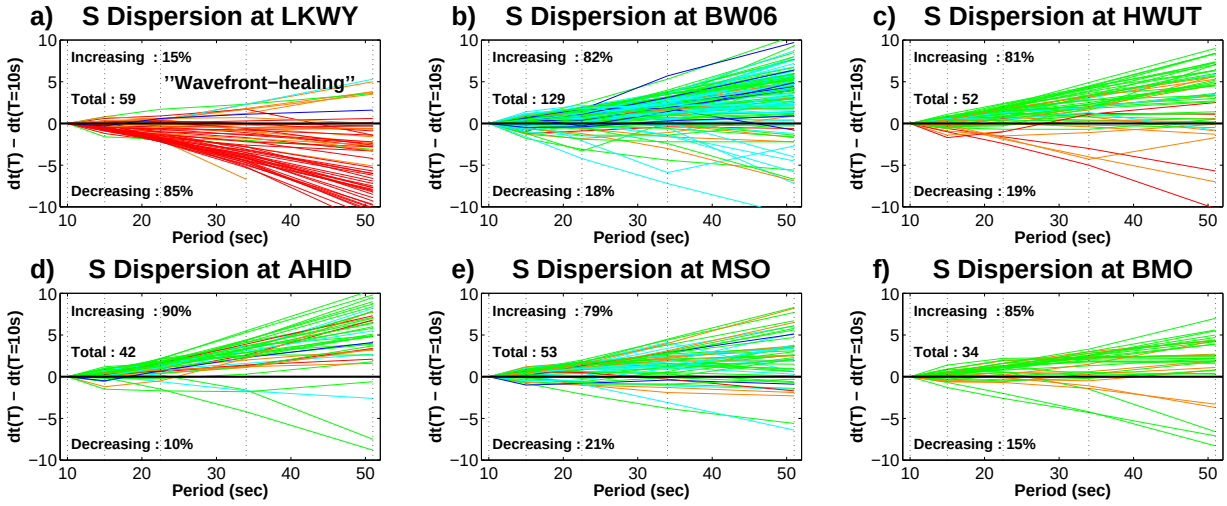


Figure 2.6: Dispersion curves $dt(T) - dt(T = 10s)$ of S waves recorded at six stations located in the vicinity of the Yellowstone hotspot. We plot dispersion curves with $dt(T = 10s) < 4s$ in cool colors (blue, cyan and green) and dispersion curves with $dt(T = 10s) \geq 4s$ in warm colors (orange and red), where blue/red are for the lowest/highest values of $dt(T = 10s)$. We see that: ***(a)*** *∼85% of the dispersion curves recorded at station LKWY, which is located above the Yellowstone hotspot, are decreasing and displayed in warm colors;* ***(b–f)*** *at the other stations, only ∼83% of the dispersion curves are increasing and mainly displayed in cool colors. This observation suggests that the particular dispersion pattern recorded at LKWY is due to wavefront-healing and related to the crossing of the Yellowstone low-speed anomaly.*

2.3.1.b Wavefront-healing at global scale

The case of Yellowstone hotspot (cf. section 2.3.1.a) suggests that, at least at local scale, our frequency-dependent S-wave travel times contain structural dispersion. In this section, we show that wavefront-healing effect is also present at global scale.

We first consider $\sim 32,000$ S dispersion curves for which S wave travel times have been successfully measured at 15, 22.5 and 34 s periods. Measurements at 10 s period were

omitted because the number of measurements was not important enough (cf. table 2.2), mainly because of the oceanic noise and mantle attenuation. Those at 51 s periods were also not used because of the often too large associated errors. With these two restrictions, we were able to extract a large subset of high quality S data (see figure 2.7.c). Figure 2.7.a shows the percentage of decreasing S dispersion curves, as a function of the time residual at 15 s period, $dt(T = 15s)$. That is, among all the S dispersion curves $dt(T)$ sharing the same value of $dt(T = 15s)$, we plot the relative number of them that are decreasing. For S waves having encountered velocity anomalies producing $-10s \leq dt(T = 15s) < 5s$, the percentage of decreasing dispersion curves is almost constant and equal to ∼25%. However, the percentage of decreasing dispersion curves linearly increases by a factor of 2.5 between $dt(T = 15s)$=5 s, where it is equal to ∼25%, and $dt(T = 15s)$=12 s, where it is equal to ∼65%. This observation suggests that, at global scale, S waves traveling across very low velocity anomalies experience wavefront-healing, a frequency-dependent effect which produces decreasing dispersion curves (cf. beginning of section 2.3.1).

We then consider ∼17,500 *SS* dispersion curves for which travel times have been successfully measured at 15, 22.5 and 34 s periods (figure 2.7.d). On figure 2.7.b, the percentage of decreasing dispersion curves associated with $dt(T = 15s) \leq -2s$, corresponding to *SS* waves having encountered high velocity anomalies, is almost constant and equal to ∼45%. Then, it increases linearly by a factor of 1.5 from ∼45%, at $dt(T = 15s)$=-2 s, to ∼65%, at $dt(T = 15s)$=13 s. This behaviour is more difficult to interpret than in the case of S waves. The fact that a smaller increase in the percentage of decreasing dispersion curves is observed over a broader interval of $dt(T = 15s)$ values, not always indicating very low velocity anomalies, is at first glance more difficult to associate with wavefront healing. However, it is important to keep in mind that *SS* waves have a surface reflection at their bounce points, whose associated frequency-dependent crustal effects are not modeled in our WKBJ synthetics (see section 2.2.2.e). The associated *SS* travel time sensitivity kernel is also more complex than for S waves, especially as *SS* waves encounter two caustics along their paths (e.g. Hung *et al.* 2000). Their longer journey into the lithosphere also make them more likely to be affected by strong scattering effects (scattering effects will be discussed in section 2.3.2). One part of the signal seen on figure 2.7.b may be due to wavefront healing effect. It is however likely that other effects compete and contribute to the *SS* dispersion. The behaviour of *SS* waves would reflect their more complex sensitivity to the 3D structure.

Finally, we consider ∼7,500 *ScS* dispersion curves for which travel times have been successfully measured at 15, 22.5 and 34 s periods. We find that ∼85% of these dispersion curves are decreasing. This tendency is also observed from the time residual of our entire

ScS dataset averaged at each period (table 2.2). However, the large majority of decreasing dispersion curves cannot be due to wavefront healing, as this would require a preferential sampling of low velocity anomalies. We will see in the next section that, although our *ScS* dataset provides a non-uniform sampling of the mantle, there are clear evidences for a preferential sampling of high velocity anomalies near the CMB.

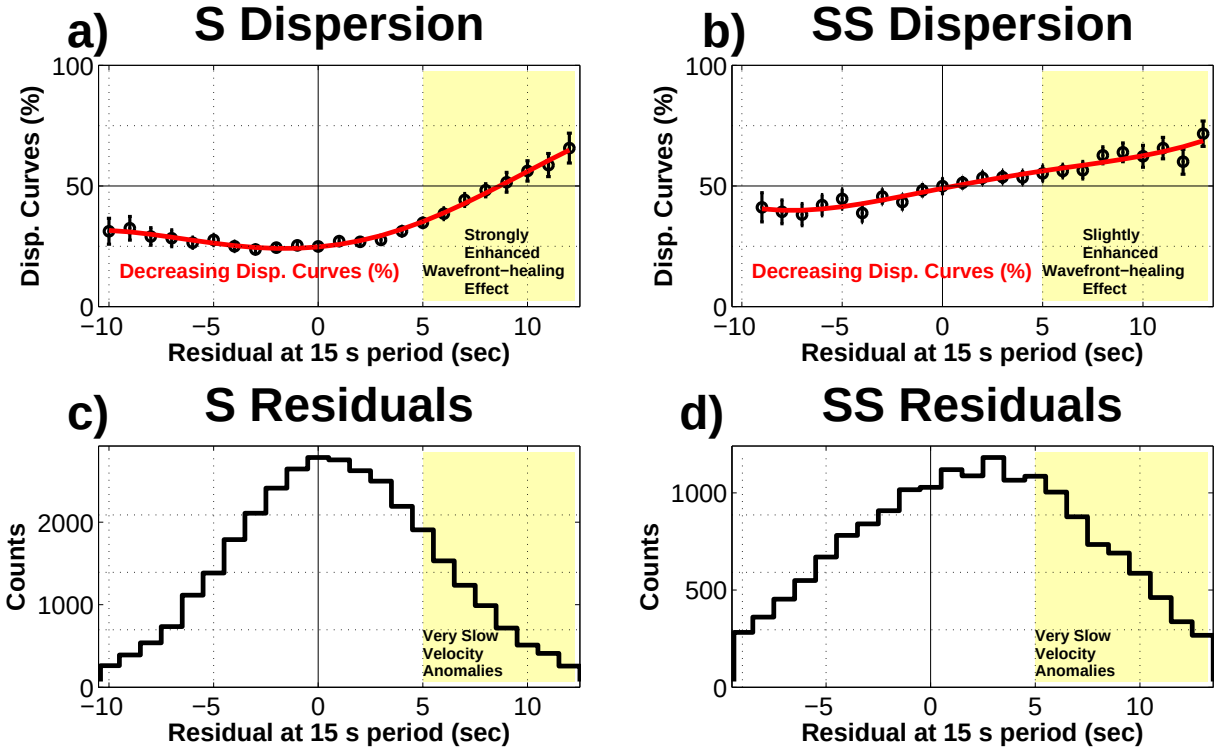


Figure 2.7: We consider ∼32,000 S and ∼17,500 SS dispersion curves for which time residuals have been successfully measured at 15, 22.5 and 34 s periods. ***a)*** *and* ***b)****: A drastic (smooth) increase in the percentage of decreasing dispersion curves is observed for S (SS) waves having traveled across very low velocity anomalies, associated to highly positive time residuals at 15 s period. This observation suggests that wavefront-healing effect is present at global scale. 2σ−error bars are determined by bootstrap technique.* ***c)*** *and* ***d)****: Histograms of S and SS time residuals at 15 s period, showing the very low velocity anomalies producing enhanced wavefront-healing effect.*

2.3.2 Scattering on *ScS* waves at CMB

Our *ScS* dataset shows a peculiar behaviour with a large majority of decreasing dispersion curves associated with negative time residuals. In addition to wavefront healing, we can reject intrinsic attenuation as a possible cause of this peculiar pattern. We show

in sections 2.2.4 and 3.3 that although intrinsic attenuation causes dispersion of seismic velocities, its effect is to produce increasing dispersion curves, by decreasing the velocity of long period waves compared to shorter period ones. In the following, we explore the possibility of explaining the dispersion pattern of our *ScS* data by scattering effect, related to high velocity scatterers located at the CMB.

We consider here a seismogram $s(t)$ as a succession of pulse-like arrivals $u_i(t)$, each with an amplitude A_i and a travel time τ_i, plus some noise $n(t)$. In the framework of Born theory, we add the contribution δu_i of waves scattered from the wavefront around ray i. If we consider a S-wave striking a seismic heterogeneity, since the S-wave itself travels the path of minimum time, the scattered signal cannot arrive earlier than the direct wave. However, this does not mean that it always has a delaying influence on the *measured* travel-time (Nolet 2008). The addition of δu_i to u_i deforms the wave shape and therefore may have a delaying or an advancing effect, depending on the sign of the scattered wave. The sign of the scattered wave is determined by the sign of the velocity anomaly that causes the scattered wave. High and low velocity scatterers generate scattered waves with negative and positive polarities, respectively (Nolet *et al.* 2005). The effect of adding δu_i is to re-distribute the energy within the cross-correlation window. Under the paraxial approximation, the sensitivity kernel of travel-time with respect to velocity perturbation (Dahlen *et al.* 2000) may be written as:

$$K_T^c(\mathbf{r}_x) = -\frac{1}{2\pi c(\mathbf{r}_x)} \cdot \frac{R_{rs}}{c_r R_{xr} R_{xs}} \cdot \xi \tag{2.17}$$

with:

$$\xi = \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \cdot \sin[\omega \Delta T(\mathbf{r}_x) - \Delta\Phi(\mathbf{r}_x)] d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2 d\omega}. \tag{2.18}$$

$\Delta\Phi$ is the phase shift due to passage through caustics or super critical reflection, R_{rs}, R_{xr} and R_{xs} are the geometrical spreading factors, and ΔT is the detour time of the scattered wave. Unless the wave is supercritically reflected with an angle-dependent phase shift, $\Delta\Phi$ takes three possible values: 0, $-\pi/2$, and $-\pi$ (Dahlen *et al.* 2000; Hung *et al.* 2000). If we only consider S and *ScS* phases, we have $\Delta\Phi = 0$. The numerator of equation (2.18) then consists of the term $\sin(\omega\Delta T)$ modulated by the power spectrum $|\dot{m}(\omega)|^2$ and a factor ω^3. One may expect that the kernel has a maximum near $\omega_0 \Delta T = \pi/2$, or for $\Delta T = T_0/4$, if T_0 is the dominant period of the signal (Nolet 2008). If there is no phase shift, one may assume (e.g. Nolet *et al.* 2005) that $\delta u_i(t)$ preserves the shape of $u_i(t)$ (they will only differ by their amplitudes). Let the polarity of the scattered wave be negative, as for a high velocity anomaly. The measurement process consists of cross-correlating the observed and

synthetic waveforms, for instance filtered around the period $T_0 = 10s$. The time residual dt corresponds to the maximum of the cross-correlation of the perturbed wave $u_i(t) + \delta u_i(t)$ (i.e. the observed waveform) with the unperturbed wave $u_i(t)$ (i.e. the synthetic waveform). The observed waveform is expected to be dominated by arrivals of scattered waves with detour times close to $\Delta T_i(T_0 = 10s) = T_0/4 = 2.5s$, corresponding to the maximum sensitivity of the associated kernel. The contribution of these scattered waves is to decrease the amplitude of the observed waveform, around the time $t \simeq \tau_i + \Delta T_i(T_0)$, compared to the synthetic waveform, such as this will have an advancing effect on the time residual dt. We have checked (cf. appendix A) that this advancing effect may be expected to increase with the period T_0. For instance, at $T_0 = 34s$, the observed waveform should be dominated by scattered waves with detour times close to $\Delta T_i(T_0 = 34s) = 8.5s$. This will then decrease the amplitude of a latter part of the observed waveform, which means a greater advancing effect on dt. Therefore, in regions where high velocity scatterers dominate, we expect an apparent dispersion with $dt(T_0 = 10s) > dt(T_0 = 34s)$, corresponding to a decreasing dispersion curve $dt(T)$. In regions where low velocity scatterers dominate, we have checked (cf. appendix A) that we may expect an increasing dispersion curve. In such low velocity regions, we also expect that wavefront-healing (cf. section 2.3.1) and scattering effects are competing.

A significant difference between *ScS* waves and the remaining part of our dataset is that *ScS* waves cross the D" discontinuity, which is located ∼300 km above the CMB. This D" discontinuity is associated with a sharp increase in *S*-wave velocity and marks the top of a very heterogeneous zone at the bottom of the mantle. This region is not sampled by our deepest *S* and *SS* waves, which bottom near 2400 km depth. Deeper *S* waves interfere with the *ScS* waveforms and have been rejected by our selection process (cf. appendix a). Using *Sdiff* waves would help to better understand frequency-dependent effects on global *S*-waves in the D" layer (e.g. To & Romanowicz 2009). However, *Sdiff* are not used in this study, as they cannot be properly synthesized with WKBJ synthetics.

We consider ∼3,300 earthquake-station couples in the epicentral distance range 55–70 degrees, with both *S* and *ScS* dispersion curves successfully measured at 15, 22.5 and 34 s periods. At these distances, *S* and *ScS* waves have very similar travel time sensitivity kernels except near the bottom of the mantle (see figures 2.8.b and 2.8.c), so that we can attribute their travel time differences to velocity anomalies located above the CMB. Figures 2.8.a and 2.8.d show that the high velocity ring around the Pacific and in eastern Asia at the CMB is preferentially sampled by our restricted *ScS* dataset. The fast anomalies at the CMB are thought to be a collection of slab material (van der Hilst *et al.* 1997), although this interpretation is still difficult to prove or disprove. Houser *et al.* (2008) also find fast

anomalies at the CMB surrounding the entire Pacific Plate, and attribute them to the cold remnants of past subduction. Very few of our *ScS* waves cross the low velocity anomalies present at the base of the mantle over much of the central Pacific Ocean and beneath South Africa (e.g. Ritsema *et al.* 1999). For this restricted dataset, we find that $\sim 85\%$ of the dispersion curves are decreasing for *ScS* waves, compared to $\sim 25\%$ for *S* waves (see figures 2.8.e–f). This suggests that scattering effect, related to a preferential sampling of high velocity scatterers located at the base of the mantle, is a plausible explanation for the peculiar dispersion observed for *ScS* waves.

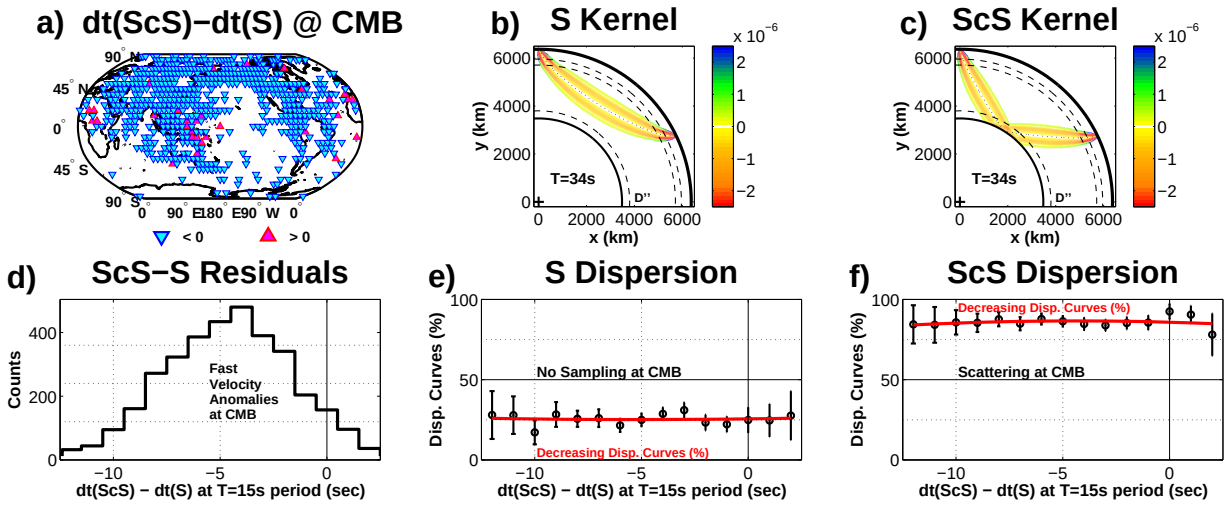


Figure 2.8: We selected a set of ∼3,300 epicenter-station couples in the distance range 55–70 degrees, with both S and ScS dispersion curves successfully measured at 15, 22.5 and 34 s periods. ***a)*** *Difference of time residuals at 15 s period between ScS and S waves, i.e.* $dt_{ScS}(T = 15s) - dt_S(T = 15s)$*, averaged in* $6° \times 6°$ *cells and geographically plotted at their corresponding CMB locations.* ***b)*** *Travel time sensitivity Fréchet kernel (in sec/km*^{3}*) for S wave, computed using the software by Tian et al. (2007b).* ***c)*** *Fréchet kernel for ScS wave.* ***d)*** *Histogram of ScS-S residuals at 15 s period.* ***e–f)*** *Our results show that* $\sim 85\%$ *of the dispersion curves are decreasing for ScS waves, compared to* $\sim 25\%$ *for S waves. This argues in favor of strong scattering effect occuring on ScS waves, owing to preferential sampling of high velocity scatterers at CMB.* 2σ*–error bars are determined by bootstrap technique.*

2.3.3 Frequency-dependent attenuation

The mantle acts as an absorption band for seismic waves (e.g. Anderson 1976) and attenuation q depends on the frequency of oscillation. Within the absorption band, attenuation is relatively high and its frequency-dependent effects are expected to be weak

for long period body waves (e.g. Sipkin & Jordan 1979), i.e. within the 10–51 s period range of analysis of this study. The frequency dependence of the attenuation q can be described by a power law $q \propto \omega^{-\alpha}$ (equation 11), with a model-dependent α, usually thought to be smaller than 0.5 (Anderson & Minster 1979). Constraining the frequency dependence of intrinsic seismic attenuation in the Earth's mantle is crucial in order to properly correct for velocity dispersion due to attenuation. Global tomographic models usually rely on a frequency-independent attenuation model (Kanamori & Anderson 1977), corresponding to the case $\alpha = 0$. A non-zero α implies that seismic waves of different frequencies are differently attenuated, and accordingly modifies the velocity dispersion relation (see section 2.2.2.d).

Despite observational and experimental advances, no clear consensus concerning the value of α for the Earth's mantle has emerged over the past 25 years. Nevertheless, theoretical predictions of $\alpha > 0$ have been systematically confirmed in various laboratory studies. A recent review by Romanowicz & Mitchell (2007) identifies a number of studies that collectively constrain α to the 0.1–0.4 range. Using normal mode and surface wave attenuation measurements, Lekic *et al.* (2009) find that $\alpha = 0.3$ should better approximate the α representative of the average mantle, at periods between 1 and 200 s. Their preferred model of frequency dependence of attenuation is also consistent with other studies that have relied upon body waves and have focused on higher frequencies. Looking at S/P ratios at periods lower than 25 s, several studies (Ulug & Berckhemer 1984; Cheng & Kennett 2002) have argued for α values in the 0.1–0.6 range. Shito *et al.* (2004) used continuous P-wave spectra to constrain α between 0.2 and 0.4 at periods shorter than 12 s. Flanagan & Wiens (1998) found an α value of 0.1–0.3 was needed to reconcile attenuation measurements on sS/S and pP/P phase pairs in the Lau basin.

In this study, we have measured globally distributed multiple-frequency time residuals of thousands of S waves, within the 10–51 s period range (see table 2.2). These measurements have been corrected from physical dispersion relying on a frequency-independent attenuation model (Kanamori & Anderson 1977). Sampling of the Earth's (lower) mantle corresponding to our S dataset is mostly global (cf. figure 2.3). Table 2.2 shows the globally averaged time residual of S waves at each period T between 10 and 51 s, denoted by $\mu_S(T)$ in the following. We observe that, when the period T increases, the globally averaged time residual $\mu_S(T)$ slightly increases (cf. the blue curve on figure 2.9). At first glance, it is very difficult to explain with scattering effect only that $\mu_S(T)$ is positive and increases within our period range. Indeed, this would require a preferred sampling of low velocity scatterers (see section 2.3.2) in the mantle, above 2400 km depth, for which there is no evidence at global scale. Wavefront-healing cannot explain such a positive and increasing averaged

dispersion curve $\mu_S(T)$ (see section 2.3.1).

By only considering attenuation effect, long period seismic waves should arrive later than short period ones (cf. section 2.2.2.d and figure 2.9). An underestimation of this effect in our attenuation correction may therefore account for a major part of the observed increasing behaviour of $\mu_S(T)$. Here, we propose to explain the averaged S residual dispersion, remaining after the common correction of physical dispersion with $\alpha = 0$, by taking into account the possible frequency-dependency of attenuation with a non-zero α. We find that a frequency-dependent attenuation with $\alpha \simeq 0.2$ better accounts for our frequency-dependent S travel times, as it predicts a globally averaged time residual $\mu_S(T)$ very close to zero at each period (cf. the red curve on figure 2.9). This value of $\alpha \simeq 0.2$ is close to the value of 0.3 found by Lekic *et al.* (2009) for the average mantle, at periods lower than 200 s (and longer than 1 s). An α value of 0.2 is also compatible with other studies (e.g. Romanowicz & Mitchell 2007).

We need however to consider that there is a trade-off between Q (i.e. t^*) and α, as shown by equation (2.16). That is, when considering a single S wave propagating in the mantle, we might also explain its residual dispersion by varying both Q and α. In this study, we use the radial PREM Q model, because 3D variations of Q are not well constrained in the Earth's mantle. We believe that considering a radial (1D) Q model, to interpret the observed globally averaged S residual dispersion (cf. figure 2.9), is reasonable because our thousands of S waves average the 3D variations of Q sufficiently well in the average mantle. Our results suggest that applying a frequency-dependent attenuation correction, with $\alpha \simeq 0.2$, is a plausible explanation for the averaged residual dispersion of S waves observed in the entire 10–51 s period range.

Table 2.2 also suggests a slight increase of the globally averaged time residual $\mu_{SS}(T)$ for SS waves in the 10–51 s period range. In this case, we find that a frequency-dependent attenuation, with $\alpha \simeq 0.1$, better accounts for our frequency-dependent SS travel times, as it predicts a globally averaged time residual $\mu_{SS}(T)$ very close to zero at each period. Compared with S waves, SS waves experience a longer journey into the lithosphere and upper mantle. It is therefore possible that the different α values obtained with S and SS waves reflect their different sampling of the Earth's mantle.

We observe a decrease of the averaged time residual $\mu_{ScS}(T)$ for ScS waves in the 10–51 s period range (see table 2.2). In this case, a frequency-dependent attenuation, with $\alpha > 0$, would reinforce the decreasing trend of the ScS residual dispersion. Our ScS dispersion pattern can therefore not be explained by a a frequency-dependent attenuation, with $\alpha > 0$. This favors scattering, instead of attenuation, to explain the particular ScS dispersion pattern (section 2.3.2).

Our observation that frequency-dependent effects of Q might explain the averaged residual dispersion of our global S dataset is compatible with the idea that other diffraction phenomena (e.g. wavefront-healing and scattering) can be predominant on individual data. As far as physical dispersion remains weak compared to the observed residual dispersion, the error that we make in the evaluation of this physical dispersion correction is unlikely to change the residual dispersion patterns we have observed and related to structural effects (cf. sections 2.3.1 and 2.3.2). We have checked that wavefront-healing is similarly observed in our S and SS datasets with a new attenuation correction corresponding to non-zero α. This conclusion supports other previous studies which suggest that incorporating anelastic dispersion cannot completely account for the observed S-wave discrepancy (e.g. Liu *et al.* 1976; Baig & Dahlen, 2004).

2.4 Conclusion

We have built a global database of ∼400,000 S, ScS and SS travel times measured at five different periods (10, 15, 22.5, 34 and 51 s). An automated scheme for measuring long period body wave travel times in different frequency bands has been presented. The scheme comprises of two main parts. The first involves an automated selection of time windows around the target phases present on both the observed and synthetic seismograms. The second stage involves measurements of multiple-frequency travel times by cross-correlating the selected observed and synthetic filtered waveforms. Frequency-dependent effects due to crustal reverberations beneath each receiver are handled by incorporating crustal phases into WKBJ synthetic waveforms. The obtained multiple-frequency S-wave travel times are well suited for global multiple-frequency tomographic imaging of the Earth's mantle.

After correction for physical dispersion due to intrinsic anelastic processes, we observe a residual dispersion on the order of 1–2 s in the period range of analysis. This dispersion occurs differently for S, ScS and SS, which is presumably related to their differing paths through the Earth. Our results show that: (1) Wavefront-healing phenomenon, produced by very low velocity anomalies, is observed in our S and, to a lesser extent, SS travel times. (2) A preferred sampling of high velocity scatterers located at the CMB may explain our observation that ScS waves travel faster at low-frequency than at high-frequency. (3) The globally averaged dispersion observed for S and SS travel times favor a frequency-dependent attenuation model $q(\omega) \propto q_0 \cdot \omega^{-\alpha}$, with an α value of $\simeq 0.2$ for S waves and $\simeq 0.1$ for SS waves.

Our results therefore suggest that the residual dispersion observed in our data is,

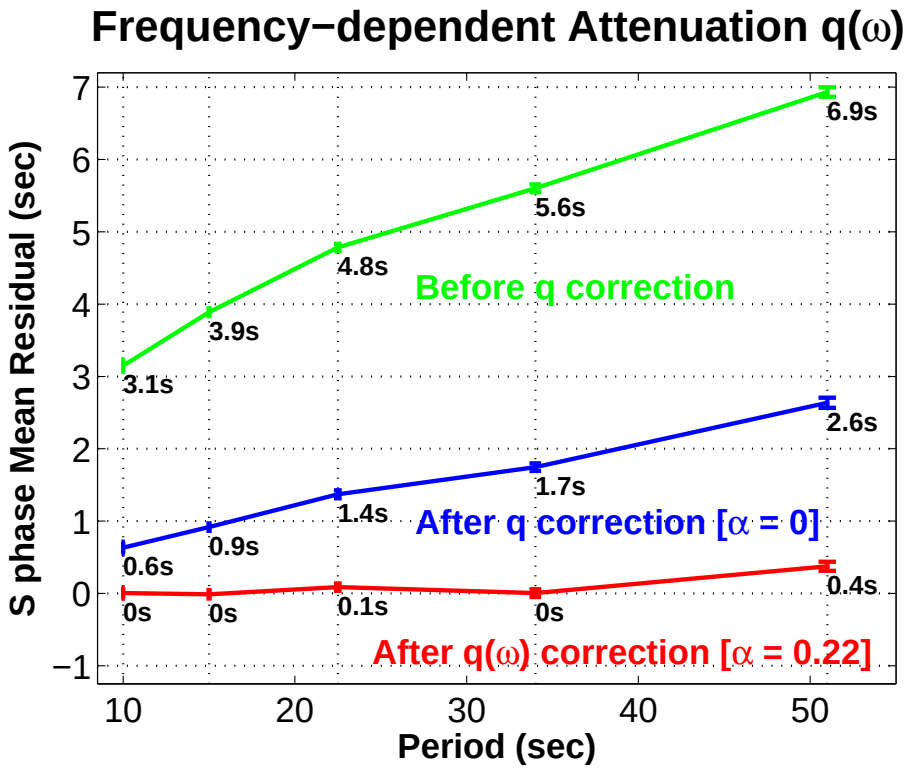


Figure 2.9: The green curve represents the globally averaged S wave time residual, $\mu_S(T)$, at each period T between 10 and 51 s, with no attenuation correction applied to S travel times. The blue curve represents $\mu_S(T)$ corrected with a frequency-independent attenuation model, corresponding to $\alpha = 0$. The red curve represents $\mu_S(T)$ corrected with a 'frequency-dependent' attenuation model, $q(\omega) \propto q_0 \cdot \omega^{-\alpha}$, corresponding to a non-zero value of α. Our observations show that $\alpha \sim 0.2$ better accounts for our S observations, as it predicts $\mu_S(T) \sim 0$ in the full 10– 51 s period range. 2σ–error bars are determined by bootstrap technique.

at least partly, related to seismic heterogeneity and attenuation in the Earth's interior. With this, we feel that tomographic reconstruction schemes, that explicitly take account of this frequency-dependency, should help to build a more accurate picture of the Earth's mantle. Our expectations are that, with the newly processed observations, one may be able to shed light on some key small scale features present in the mantle, and in doing so, better constrain mantle dynamics.

Acknowledgments

This work was supported by the young researcher ANR TOMOGLOB no ANR-06-JCJC-0060. We thank the Iris and Geoscope data centres for providing seismological data. Discussions with M. Cara, J-J. Leveque, A. Maggi, L. Rivera and B. Tauzin have been stimulating at various stages of this research. The authors thank J. Ritsema, J. Trampert and an anonymous reviewer for helpful reviews that improved the original paper.

Appendix a: Time windows selection

We describe here our time windows selection and seismic phase isolation methodology, which largely makes use of the ideas of Maggi *et al.* (2009) . The main differences are the following: (1) in step 1 we work on the rotated SH-component; (2) in steps 2, 3, 4 and 5, we perform all the operations on the observed and synthetic seismograms (i.e. not on the synthetic only); (3) in step 6, we test all possible combinations of time windows before deciding on an optimal pair of observed and synthetic waveforms, corresponding to the target seismic phase. For the present study, our codes have been tuned to measure travel times of SH-waves, which have the advantage of being free of P energy. There is no difficulty in applying the same approach to SV- and P-waves, although it is likely that interference between S and P energy would result in fewer windows surviving the selection criteria.

a.1 Step 1: pre-selection

The purpose of this step is to pre-process input seismograms and reject noisy records. Three components seismograms are first rotated along the SH-component. The observed seismograms are then band-pass filtered with a non-causal Butterworth filter, whose short and long period corners are denoted by T_1 and T_2, respectively. Following Maggi *et al.* (2009), we define the time-normalized power in the signal and noise portions of the data by:

$$P_{signal} = \frac{1}{t_E - t_A} \cdot \int_{t_A}^{t_E} d^2(t)dt \tag{2.19}$$

and

$$P_{noise} = \frac{1}{t_A - t_0} \cdot \int_{t_0}^{t_A} d^2(t)dt \tag{2.20}$$

where $d(t)$ is the observed seismogram, t_0 is the start time, t_A is set to be slightly before the time of the first arrival and t_E is the end of the main signal. We compute the power signal-

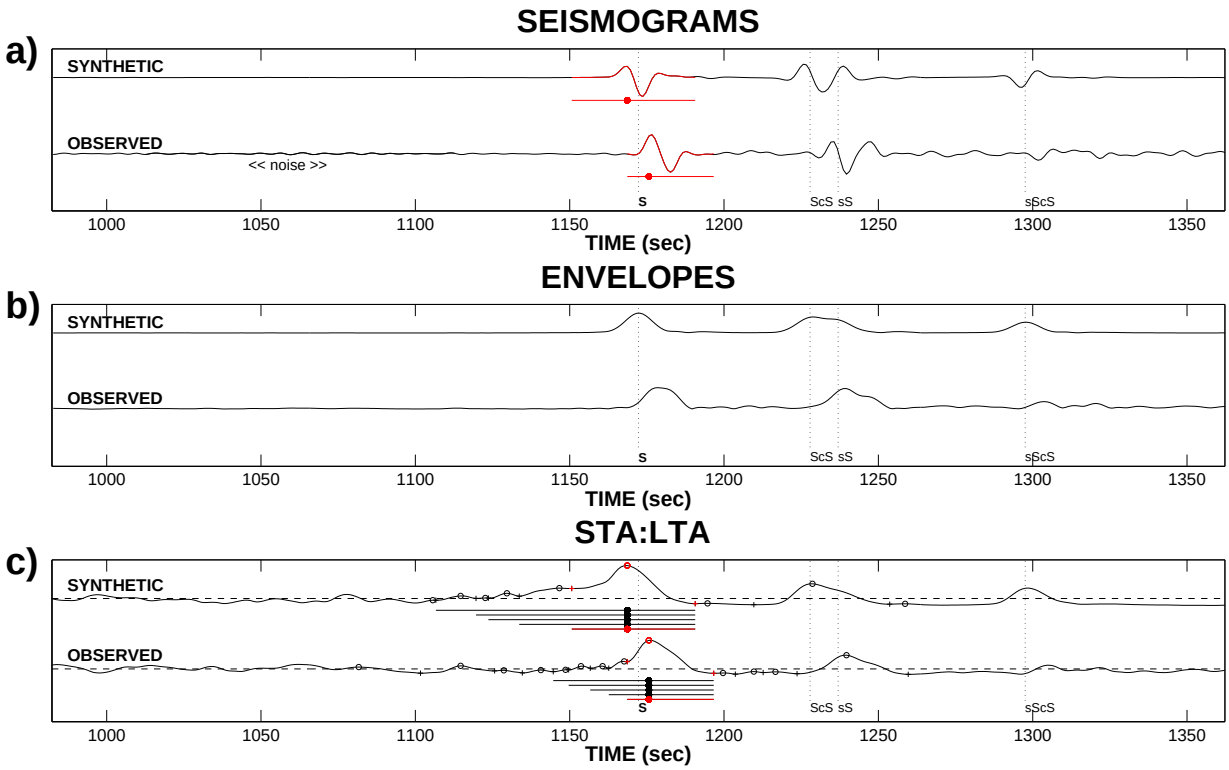


Figure 2.10: Time windows selection process, illustrated for the observed and synthetic S waveforms isolation. ***a)*** *Observed and synthetic seismograms for a deep earthquake located at Kyushu, Japan (date: 2005/11/21, centroid depth: 155 km, body wave magnitude: 5.9) and recorded at CAN station (Geoscope network). Seismograms are discretized with a time step δt of 1 sec, and filtered with a Butterworth filter (corner frequencies: T_1=7 s and T_2=85 s). The optimal observed and synthetic time windows are shown in red.* ***b)*** *Observed and synthetic envelopes.* ***c)*** *Corresponding STA:LTA waveforms. In black dashed line overlaid on the STA:LTA waveforms is the water level (WL). Local maxima and minima of STA:LTA waveforms are denoted by circles and crosses, respectively. All the observed and synthetic time windows represented in black (and red for the optimal pair), correspond to the candidate windows having survived several selection criteria (i.e. steps 4 and 5 of sections A4 and A5). The selection of the optimal pair (represented in red) among these candidate time windows is explained in section A6. Amplitudes of observed and synthetic S waveforms are normalized to 1.*

to-noise ratio $SNR_P = P_{signal}/P_{noise}$ and the amplitude signal-to-noise ratio $SNR_A = A_{signal}/A_{noise}$,where A_{signal} and A_{noise} are the maximum values of $|d(t)|$ in the signal and noise time-spans, respectively. We reject records for which $SNR_P < r_P$ or $SNR_A < r_A$ (see table 2.3).

a.2 Step 2: short-term / long-term average ratios

Seismic phase arrivals are detected using standard short-term / long-term (STA:LTA) average ratios. In contrast to Maggi *et al.* (2009), this STA:LTA analysis is applied to both the observed and synthetic seismograms. We first compute the envelopes of the observed and synthetic seismograms. The envelope $e(t)$ of a seismogram $s(t)$, whose Hilbert transform is noted $H[s(t)]$, is given by:

$$e(t) = |s(t) + i \cdot H[s(t)]|. \tag{2.21}$$

Assuming that both observed and synthetic waveforms are discretized with time-step δt, we compute the short-term $S(i)$ and long-term $L(i)$ averages for each time sample i of the envelope, using:

$$S(i) = C_S \cdot S(i-1) + e(i) \tag{2.22}$$

$$L(i) = C_L \cdot L(i-1) + e(i) \tag{2.23}$$

and evaluate their ratios:

$$E(i) = S(i)/L(i). \tag{2.24}$$

The constants C_S and C_L determine the decay of the relative weighting of earlier parts of the signal in the calculation of the current average. Following Bai & Kennett (2001), we use $C_S = 10^{-\delta t/T_1}$ and $C_L = 10^{-\delta t/12 \cdot T_1}$. Figure 2.10.a shows an example of observed and synthetic waveforms. The corresponding envelopes $e(t)$ and STA:LTA waveforms $E(i)$ are shown in figure 2.10.b and 2.10.c, respectively.

a.3 Step 3: time windows isolation

At this stage, the intention is to list all possible time windows present on the observed and synthetic STA:LTA waveforms $E(i)$. As underlined by Maggi *et al.* (2009), the agreement between local maxima in $E(i)$ and the position of seismic phases on the observed and synthetic seismograms, as well as the correspondence between local minima and the transitions between successive phases, suggest that time windows should start and end at local minima surrounding a maximum in $E(i)$ (see figure 2.10). We first select all maxima in $E(i)$ lying above a given water level WL (see table 2.3) on the observed and synthetic waveforms. The water level is identical for the observed and synthetic waveforms $E(i)$. Each maximum is then taken as a 'seed' maximum about which all possible candidate time windows can be created around it. The time windows start and end at local minima of the STA:LTA waveforms $E(i)$. We consider all local minima before the seed maximum as

a potential start time for the window, and all local minima after the seed maximum as a potential end time. Therefore, each candidate time window is defined by three times: its start time, its end time and the time of its seed maximum.

a.4 Step 4: shape based time windows rejection

At this stage, we are left with a list of possible time windows surrounding the target phases present on the observed and synthetic seismograms. We first reject windows based on the shape of the STA:LTA waveforms $E(i)$. The aim of this shape-based window rejection is to extract observed and synthetic time windows with well-developed single phases (e.g. S) or groups of phases (e.g. S+sS). We use the same criteria as in Maggi *et al.* (2009), except that we apply them on both the synthetics and observed STA:LTA waveforms. First, we reject all time windows that contain internal local minima in $E(i)$ whose amplitude is less than $C_0 \cdot WL$ (see table 2.3). This choice forces a partition of unequivocally distinct seismic phases into separate time windows. Second, we reject short windows whose length is smaller than $C_1 \cdot T_1$ (see table 2.3). This criteria allows us to reject windows which are too short to contain useful information. Third, we reject time windows whose seed maximum rises by less than $C_2 \cdot WL$ above either of its adjacent minima (see table 2.3). Finally, we reject time windows containing at least one strong phase arrival that is well separated in time from the seed maximum time. This allows us to distinguish inseparable phase groups from distinct seismic phases that arrive close in time.

a.5 Step 5: SNR and time interval based windows rejection

At this stage, we are left with several pairs of observed and synthetic time windows containing well-developed single (or groups of) seismic phase(s). We wish to extract the optimal pair of observed and synthetic time windows for each target phase. This task is not trivial to implement in an automated way. One of the main difficulties is that, in most cases, the observed and synthetic time windows corresponding to the same target phase have different start and end times. This is especially true when the time residual, between the observed and 1D synthetic seismic phases, becomes large. In addition, when the target phase interferes with other phases, our automated scheme should ideally ensure that the observed and synthetic waveforms, present in the retained time windows, do carry the same pattern of interference.

First, we compute for each (observed and synthetic) candidate time window a signal-to-noise ratio (SNR): $SNR_w = A_{window}/A_{noise}$, where A_{window} and A_{noise} are the maximum absolute amplitude values of the seismic signal contained in the candidate time window

and in the noise time-span, respectively. We reject each (observed and synthetic) candidate time window if $SNR_w < r_0$ (see table 2.3). Second, on the synthetic seismogram, we retain time windows around the predicted arrival time (t_p) of the target phase. Third, on the observed seismogram, we retain time windows whose seed maxima are contained in the time interval $w_{obs} = [t_p - 25s - T, t_p + 25s + T]$, where T is the dominant period of the target phase. This choice is based on the fact that, for global S-wave tomography, delay-times have been observed to vary in the interval $[-25s, +25s]$ (e.g. Bolton & Masters, 2001). Therefore, for a target phase with a dominant period $T \sim 10s$, we span the time interval $w_{obs} = [t_p - 35s, t_p + 35s]$, which must only contain the target single phase (e.g. S), or groups of phases (e.g. $S{+}sS$), for avoiding unwanted phases interference (e.g. ScS with SS). The remaining (observed and synthetic) candidate time windows, at the end of step 5, are shown in figure 2.10.c.

Table 2.3: Overview of standard and fine tuning parameters, as defined by Maggi et al. (2009), with the values used in this study.

Standard tuning parameters:	
$r_P = 1$	power signal-to-noise ratio threshold
$r_A = 2.5$	amplitude signal-to-noise ratio threshold
$r_0 = 2.5$	signal-to-noise ratio single windows
$WL = 0.1$	water level
Fine tuning parameters:	
$C_0 = 0.7$	for rejection of internal minima
$C_1 = 3$	for rejection of short windows
$C_2 = 0.1$	for rejection of un-prominent windows
$C_{3a} = 1$, $C_{3b} = 2$	for rejection of multiple distinct arrivals
$C_{4a} = 3$, $C_{4b} = 12$	for curtailing of windows

Where the tuning parameters C_{3a}, C_{3b}, C_{4a} and C_{4b} are used in this study as described by Maggi *et al.* (2009).

a.6 Step 6: selection of the optimal pair of time windows

At this stage, we may still be left with several candidate time windows, around a given target phase (see figure 2.10.c.). To select the optimal pair among them, we first test all combinations of cross-correlation between all the remaining pairs of observed and synthetic waveforms. The aim of this cross-correlation step is to help with the association of a synthetic time window with its best equivalent on the observed seismogram. For each pair of observed and synthetic waveforms, we obtain a cross-correlation maximum (CC_{max}) and a corresponding delay-time (dt_{max}). We keep those pairs of candidate time windows

whose CC_{max} is greater than 80%. Although this choice ensures a strong similarity between observed and synthetic waveforms, it does not always guarantee that they include the same portion of signal. We use the delay-time dt_{max} for discriminating wrong pairs of candidate time windows among those with $CC_{max} > 80\%$. We then compute the ratio:

$$P = \frac{CC_{max}}{\max(\varepsilon, |dt_{max}|)} \tag{2.25}$$

and select as our optimal pair of observed and synthetic time windows the one with the highest value of parameter P. We use ε=0.1 s for avoiding to divide by zero, and because our delay times are determined with a precision down to $\pm 0.1s$. If several observed waveforms present a high degree of similarity with several synthetic waveforms, this choice is a compromise that favors small delay-times, because they are closer to the reference model.

Appendix b: Time residual

We aim to prove that the function $F_3(\tau)$ and the cross-correlation function $\gamma_{d,s}(\tau)$ are maximized for the same time residual. We call τ_m^{CC} and $\tau_m^{F_3}$ the time residuals maximizing $\gamma_{d,s}(\tau)$ and $F_3(\tau)$, respectively. The recorded signal at the receiver consists of a direct wave arrival, $u(t)$, and a scattered wave arrival, $\delta u(t)$. Therefore, the observed and synthetic waveforms are, respectively:

$$\begin{cases} d(t) = u(t) + \delta u(t) \\ \\ s(t) = u(t). \end{cases} \tag{2.26}$$

The autocorrelation of the unperturbed wave u is given by:

$$\gamma_{s,s}(\tau) = \int_{-\infty}^{+\infty} u(t) \cdot u(t-\tau) dt. \tag{2.27}$$

The time residual τ_m^{CC} is defined as maximizing the following cross-correlation function, between the observed signal $(u + \delta u)$ and the unperturbed wave (u):

$$\gamma_{d,s}(\tau) = \int_{-\infty}^{+\infty} [u(t) + \delta u(t)] \cdot u(t-\tau) dt \tag{2.28}$$

which leads to:

$$\gamma_{d,s}(\tau) = \gamma_{s,s}(\tau) + \delta\gamma(\tau) \tag{2.29}$$

with:

$$\delta\gamma(\tau) = \int_{-\infty}^{+\infty} \delta u(t) \cdot u(t-\tau)dt. \tag{2.30}$$

For the unperturbed wave, the cross-correlation reaches its maximum at zero lag-time, so:

$$\dot{\gamma}_{s,s}(0) = 0, \tag{2.31}$$

and for the perturbed wave the maximum is reached for τ_m^{CC}, so:

$$\dot{\gamma}_{d,s}(\tau_m^{CC}) = \dot{\gamma}_{s,s}(\tau_m^{CC}) + \delta\dot{\gamma}(\tau_m^{CC}) = 0, \tag{2.32}$$

where the dot denotes the time differentiation. Developing $\dot{\gamma}$ to first order, we find (e.g. Marquering *et al.* 1999):

$$\dot{\gamma}_{d,s}(\tau_m^{CC}) = \dot{\gamma}_{s,s}(0) + \ddot{\gamma}_{s,s}(0)\tau_m^{CC} + \delta\dot{\gamma}(0) + O(\delta^2) = 0 \tag{2.33}$$

which leads to:

$$\tau_m^{CC} = -\frac{\delta\dot{\gamma}(0)}{\ddot{\gamma}_{s,s}(0)}. \tag{2.34}$$

We have previously defined the function $F_3(\tau)$, which is an average of the two quantities $F_1(\tau)$ and $F_2(\tau)$ - cf. section 2.2.2.b and equation (2.8). We can write the quantity $F_1(\tau)$ - cf. equation (2.4) - as:

$$F_1(\tau) = \frac{2\gamma_{d,s}(\tau) - \gamma_{s,s}(0)}{\gamma_{d,d}(0)}, \tag{2.35}$$

and the quantity $F_2(\tau)$ - cf. equation (2.5) - as:

$$F_2(\tau) = \frac{\gamma_{d,s}^2(\tau)}{\gamma_{d,d}(0) \cdot \gamma_{s,s}(0)} \quad \text{, if } A_1(\tau) < A_2(\tau). \tag{2.36}$$

The maximum of the functions $F_1(\tau)$, $F_2(\tau)$ and $F_3(\tau)$ are reached for $\tau_m^{F_1}$, $\tau_m^{F_2}$ and $\tau_m^{F_3}$, respectively, such as:

$$\dot{F}_1(\tau_m^{F_1}) = 0, \ \dot{F}_2(\tau_m^{F_2}) = 0, \ \dot{F}_3(\tau_m^{F_3}) = 0. \tag{2.37}$$

Note that if $A_1(\tau) > A_2(\tau)$ (see section 2.2.2.b), we should analyse the maximum of $1/F_2(\tau)$. This maximum will be reached for the same time residual $\tau_m^{F_2}$, as:

$$\frac{d}{d\tau}\frac{1}{F_2(\tau)} = -\frac{1}{F_2(\tau)^2} \cdot \dot{F}_2(\tau) = 0 \ \Rightarrow \ \tau = \tau_m^{F_2}. \tag{2.38}$$

We then have:

$$\begin{cases} \dot{F}_1(\tau_m^{F_1}) = \frac{2\dot{\gamma}_{d,s}(\tau_m^{F_1})}{\gamma_{d,d}(0)} = 0 \\ \\ \dot{F}_2(\tau_m^{F_2}) = \frac{2\gamma_{d,s}(\tau_m^{F_2})\dot{\gamma}_{d,s}(\tau_m^{F_2})}{\gamma_{d,d}(0)\gamma_{s,s}(0)} = 0, \end{cases} \tag{2.39}$$

which leads to:

$$\begin{cases} \dot{F}_1(\tau_m^{F_1}) = 0 \;\Rightarrow\; \dot{\gamma}_{d,s}(\tau_m^{F_1}) = 0 \\ \{\dot{F}_2(\tau_m^{F_2}) = 0 \text{ and } \gamma_{d,s}(\tau_m^{F_2}) > 0\} \Rightarrow\; \dot{\gamma}_{d,s}(\tau_m^{F_2}) = 0. \end{cases} \tag{2.40}$$

As previously, by developing $\dot{\gamma}$ to first order, we find:

$$\tau_m^{F_1} = \tau_m^{F_2} = \tau_m^{F_3} = -\frac{\delta\dot{\gamma}(0)}{\ddot{\gamma}_{s,s}(0)}. \tag{2.41}$$

Finally, the equality $\tau_m^{CC} = \tau_m^{F_3}$ is verified.

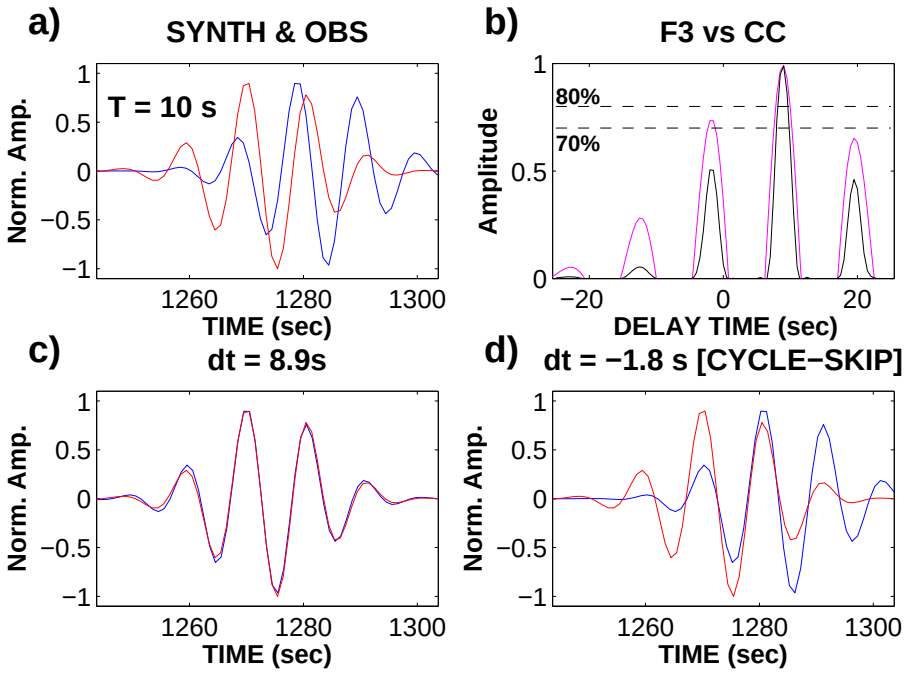


Figure 2.11: $F_3(\tau)$ versus $\gamma_{d,s}(\tau)$ for a S wave recorded at station ATD (Geoscope network), event 2004/10/15. ***a)*** *Observed and synthetic waveforms, filtered at 10 s period.* ***b)*** *Secondary maxima of $F_3(\tau)$ (in black) are minimized compared to the ones of $\gamma_{d,s}(\tau)$ (in magenta).* ***c)*** *Waveforms after appropriate time-shifting.* ***d)*** *Waveforms after wrong time-shifting, corresponding to a cycle-skip.*

Afterword

The presented study marks a new point, from a data point of view, in the debate (e.g.Van der Hilst & De Hoop 2005) on the relevance of using a "finite-frequency" approach in global S-wave tomography. For instance, our results show that wavefront-healing phenomenon is clearly observed for S waves having passed through very low velocity anomalies. As ray theory cannot take into account this observed *structural* dispersion, we shall exploit this new observable in a global multiple-frequency tomography.

We expect that this new information - the frequency-dependency of body-waves - may help to better constrain the 3-D elastic structure of the Earth's mantle. We have then chosen to simultaneously invert all our dataset, using the "finite-frequency" formalism of Dahlen *et. al* (2000). In the following chapter 3, we explain how we have posed the corresponding inverse problem. In chapter 4, we will present the inversion results, and assess, from a model point of view, the actual benefits from using multi-band data, rather than single-band data as Montelli *et al.* (2006b).

CHAPTER 3

Multiple-frequency tomography: posing the inverse problem

"Many people, if you describe a train of events to them, will tell you what the result would be. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backwards."

Sherlock Holmes

3.1 Introduction

What can we learn from the Earth's interior from seismological observations at its surface? That is, starting with our frequency-dependent S-wave travel time measurements (cf. chapter 2), can we work backwards to characterize[1] the 3-D elastic structure of the Earth's mantle, through which the waves passed?

Inverse problems can be posed by assuming that we understand the physics of a process which, for a set of model parameters ($\mathbf{m}$), gives rise to a set of observed data ($\mathbf{d}$). The data are then considered as the result of a - linear or non-linear - function (G) acting on the model parameters:

$$\mathbf{d} = G(\mathbf{m}). \tag{3.1}$$

Provided that we understand the process (G), it is easy to predict the data ($\mathbf{d}$) that would result from a given model ($\mathbf{m}$). However, the corresponding inverse problem, that is finding what model ($\mathbf{m}$) gave rise to a specific set of observed data ($\mathbf{d}$), is much more difficult. We need to assume that some physical model[2] describes the process, and then use the data to estimate a set of model parameters - of the Earth's mantle - that are consistent with the data. Inverse problems may be solved using mathematical techniques to find $\mathbf{m}$ directly from $\mathbf{d}$ (cf. chapter 4).

In multiple-frequency travel time tomography, the general form of the inverse problem is (cf. chapter 1):

$$\delta t_i(T) = \int_{V_i(T)} K_i(\mathbf{r};T) m(\mathbf{r}) d^3\mathbf{r} \tag{3.2}$$

where $\delta t_i(T)$ is the delay-time, measured at several periods T, of the target seismic phase i. The volume integral $V_i(T)$ is, in practice, limited to the region where the (Fréchet) kernel $K_i(\mathbf{r};T)$ has a significant amplitude. The model parameter $m(\mathbf{r})$ represents a velocity perturbation ($\delta c/c$). This "linear" inverse problem may be formulated as

$$\mathbf{d} = G \cdot \mathbf{m}. \tag{3.3}$$

In previous chapter 2, we have built the data vector $\mathbf{d}$, which comprises of frequency-

[1]Note that body-wave travel times depend on *both* the earthquakes (i.e. the sources) that generated the seismic waves and the medium through which the waves passed. In this thesis, we have assumed that sources are "perfectly" known (from the Harvard Centroid Moment Tensor catalog), which is far from true! We justify this assumption, and discuss on earthquake mislocation, in section 4.2.6 of chapter 4.

[2]In this thesis, we have chosen the "finite-frequency" formalism of Dahlen *et al.* (2000) to relate our frequency-dependent data to the 3-D structure of the Earth's mantle. That is, this formalism takes into account some wave diffraction effects (e.g. wavefront-healing and scattering), which cause body-wave travel times to be frequency-dependent (cf. chapter 1).

dependent *S*, *ScS* and *SS* travel times, measured in the 10–51 s period range. This chapter 3 is then devoted to build the G matrix (section 3.4), which represents the projection of the "banana-doughnut" sensitivity kernels (Dahlen *et al.* 2000) onto the model parameterization. The model parameterization used in this tomographic study is explained in section 3.2, and the kernel computation is presented in section 3.3. The resulting G matrix will be inverted in chapter 4 to get tomographic images of the Earth's mantle, i.e. the model parameters vector $\mathbf{m}$.

3.2 Parameterizing the Earth's mantle

In this section, we present how we have parameterized the model - i.e. the whole Earth's mantle. Our strategy heavily relies on the work of Nolet & Montelli (2005).

3.2.1 Irregular parameterization

The type of parameterization, used in a tomographic experiment, clearly limits the size of lateral variations in the resulting model. In most tomographic studies, the (lateral) spatial variability in the data's "resolving power" is largely ignored. That is, authors have often opted to build Earth models in terms of uniform basis functions (e.g. cubic cells or spherical harmonics).

More recently, tomographic inversions have been performed with parameterizations which are themselves laterally varying, and "tuned" to the (assumed) resolving power of the data (cf. Sambridge & Rawlinson 2005, and references therein). Such irregular parameterizations attempt to maximize the extraction of structural information from data.

However, the use of an irregular parameterization in global tomography is complicated by several factors. For instance, sophisticated computational algorithms are required for building, storing and searching through an irregular 2-D or 3-D mesh. In global tomography, massive datasets need to be inverted, so that using an irregular parameterization implies a serious extra time for building the G matrix - equation (3.3).

Is it worth using an irregular parameterization? First, one may suspect that uniform parameterizations do not extract all the available information in the data, because the minimum scale length (e.g. block size or number of harmonics) is often chosen as a compromise between data constraints and computational convenience. Moreover, even if we chose a uniform block model with length scales equivalent to the smallest size of structures that can be resolved with the data, then the increased number of unknown model parameters ($\mathbf{m}$) would make the tomographic system computationally prohibitive.

A clever way, of capturing the range of scale lengths contained in the data, is then to vary the density of the mesh spatially (e.g. Sambridge & Rawlinson 2005). We have therefore chosen to build an irregular parameterization of the Earth's mantle for this tomographic study. We explain how we have proceeded in the following.

3.2.2 Spherical Triangular Prism parameterization

In our multiple-frequency inverse problem, we assume that the data d_i $(i = 1, \cdots, N)$ are linearly related to the Earth model $m(\mathbf{r})$, as $d_i = \int_{V_i} F_i(\mathbf{r}) m(\mathbf{r}) d^3\mathbf{r}$, where $F_i(\mathbf{r})$ is a (Fréchet) sensitivity kernel (Dahlen *et al.* 2000). Here, we assume that the model is described by a finite number of parameters m_j $(j = 1, \cdots, M)$, and the continuous Earth model is given by:

$$m(\mathbf{r}) = \sum_{j=1}^{M} m_j h_j(\mathbf{r}). \tag{3.4}$$

The basis functions h_j represent the model parameterization. Many choices of the basis functions h_j are possible (cf. section 3.2.4).

In this study, we aim to build a global parameterization of the Earth's mantle which is adapted to the estimated "local resolution" of the data. The local resolution is the smallest structure's length one may expect to resolve with the data. Its estimation is based on the ray density (cf. section 3.2.3.g).

In the following, we present the global Spherical Triangular Prism (STP) irregular parameterization of the mantle we have built for our global dataset. This STP parameterization consists of:

1. *Radially*
   $\Rightarrow$ 18 constant-depth spherical layers[3] (cf. table 3.1 and figure 3.1), ranging from the surface to the CMB, and filled in with spherical triangular prisms (figure 3.2)

2. *Laterally*
   $\Rightarrow$ In each layer, the nodes (i.e. the vertices of the top spherical triangle of each prism, as shown in figure 3.2) attempt to fit a "resolution function" based upon the ray density. The nodes are then spatially varying.

[3]In the context of the TOMOGLOB project (cf. general introduction), we aim to jointly invert multiple-frequency S-wave travel times (chapter 2) with surface wave data (Debayle *et al.* 2005). Thus, our parameterization has been designed to take into account that surface wave data require large velocity gradients in the uppermost mantle. That is, we have chosen to put thinner layers (thickness of 100 km) in the upper mantle, than in the lower mantle (thickness of $\sim$200 km). The resulting (uppermost mantle) parameterization is also consistent if only S-wave travel times are used in the inversion.

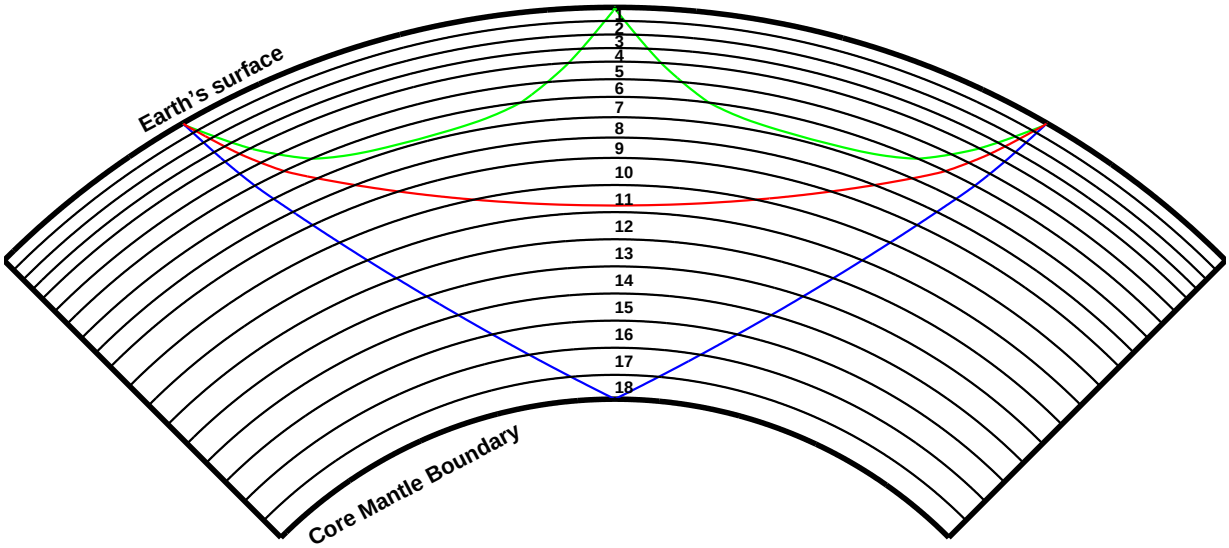


Figure 3.1: 18 constant-depth spherical layers used in this study for radially parameterizing the Earth's mantle. S, ScS and SS ray paths are shown in red, blue and green solid line, respectively. See table 3.1 for detailed characteristics of each layer.

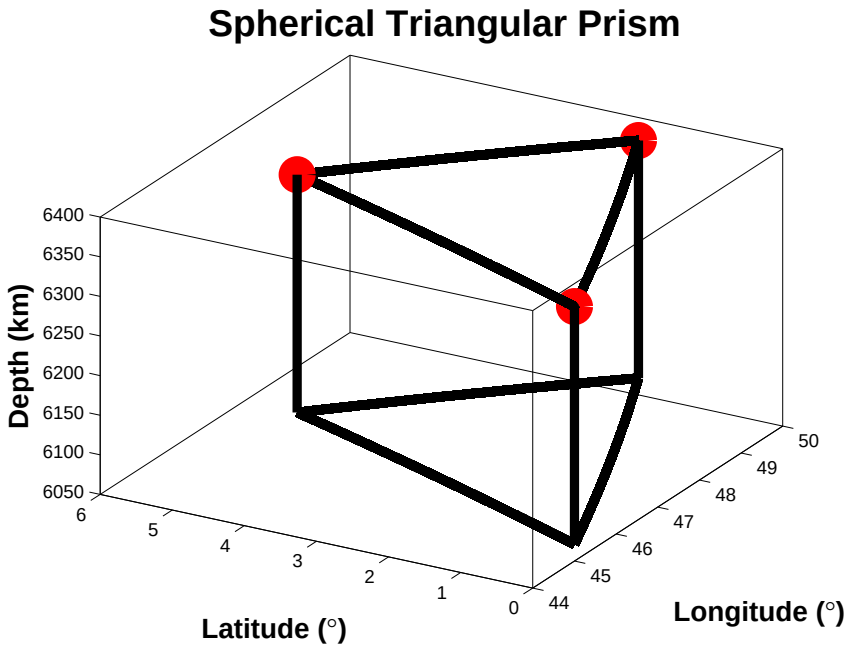


Figure 3.2: Representation of a spherical triangular prism. Red dots, which are the three vertices of the top triangle of this prism, represent three nodes of the model parameterization (i.e. model parameters).

Table 3.1: Characteristics of the starting configuration grid and resolution function used for the 18 layers in this study

Layer	**Depth (km)**	Δz **(km)**	**Iter**	L_f **(km)**	M_z	ℓ **bounds (km)**	ε
1	0–100	100	4–6	233–932	1,577	230–**930**	1
2	100–200	100	4–6	229–917	2,209	225–915	1
3	200–300	100	4–6	225–903	2,425	220–900	1
4	300–400	100	4–6	222–888	2,762	218–900	1
5	400–530	130	4–6	218–874	2,312	215–870	1
6	530–660	130	4–6	213–855	2,734	210–850	1
7	660–810	150	4–6	209–835	3,741	205–830	0.8
8	810–960	150	4–6	203–814	4,282	200–810	0.8
9	960–1110	150	4–6	198–792	3,910	195–790	0.8
10	1110–1310	200	4–6	192–770	2,772	189–765	0.8
11	1310–1510	200	4–6	185–740	2,809	**182**–735	0.8
12	1510–1710	200	4–6	178–711	2,177	190–700	0.8
13	1710–1910	200	4–5	341–682	1,072	338–680	0.8
14	1910–2110	200	4–5	326–652	948	323–650	0.8
15	2110–2310	200	4–5	312–623	704	310–620	0.8
16	2310–2510	200	4–5	297–594	696	295–590	0.8
17	2510–2710	200	3–5	282–1,127	456	280–**930**	0.8
18	2710–2889	179	3–5	268–1,069	539	265–**930**	0.8

- **Layer**: index of each layer.
- **Depth**: depth bounds of each layer.
- Δz: thickness of each layer.
- **Iter**: minimum and maximum number of triangle subdivisions (i.e. number of iterations from an initial icosahedron).
- L_f: minimum and maximum triangle edge lengths of the nodes agency in each layer.
- M_z: number of nodes present in each layer.
- ℓ **bounds**: minimum and maximum resolution length bounds for each layer.
- ε: values used in the refining algorithm of equation 3.13.

3.2.3 Towards an optimal nodes agency

3.2.3.a The optimization problem

Our model parameterization consists of 18 layers (cf. figure 3.1). Each layer is filled in with spherical triangular prisms, whose spatial distribution is (laterally) irregular. The (three) vertices of the spherical triangle at the top of each prism will be denoted as "nodes" (cf. figure 3.2). The vector of the model parameters, $\mathbf{m}$, is defined by the total set of nodes. In the following, we use the strategy of Nolet & Montelli (2005) in order to find the optimal set of nodes, for each layer, that will maximize the extraction of structural information from our data.

Consider that we know the local "resolution length", $\ell(\mathbf{r})$, of our data, at each spatial location, $\mathbf{r}$, in the mantle. We postpone the discussion on the data resolution estimate to section 3.2.3.g. Our aim is then to space nodes near $\mathbf{r}$ at a distance close to the resolving length $\ell(\mathbf{r})$.

Each node has a set of natural neighbours[4] (e.g. Sambridge *et al.* 1995), for which we aim to apply the previous distance criteria. Let z stands for the layer index, that is $z = 1, \cdots, 18$. Let N_j be the set of natural neighbors of node j. We aim to minimize the following penalty function (Nolet & Montelli 2005), for each layer z:

$$E_z = \sum_{j=1}^{M_z} \sum_{k \in N_j} \frac{(L_{jk} - \ell_{jk})^2}{\ell_{jk}^2} \tag{3.5}$$

where L_{jk} is the actual distance between nodes j and k, ℓ_{jk} is the average resolving length between nodes j and k, and M_z is the total number of nodes in layer z. The spatial positions of the nodes (i.e. model parameters) will then be adjusted to minimize E_z - equation (3.5).

3.2.3.b Delaunay mesh

Consider a set of nodes spatially located on a spherical shell (i.e. one layer of our parameterization). By using the QHULL software (Barber *et al.* 1996) - distributed by the Geometry Center of Minneapolis - we can construct its associated convex hull. This will result[5] in a set of spherical triangles spanning the spherical shell and forming a triangulation. The resulting mesh is called a Delaunay mesh. The QHULL program gives us a list of all the triangle numbers of such a Delaunay mesh, with also the node number of the three

[4]Natural neighbors are the nodes connected in a Delaunay triangulation.

[5]Here, we assume that nodes are covering "both hemispheres". That is, if nodes were only located over a finite region, then there is an awkward set of convex hull facets lying within the Earth.

vertices of each triangle. We first compute once, and store in computer's memory, all the triangle neighbors for each triangle of the Delaunay mesh. Then, it is straightforward to have access to all natural neighbors (e.g. Sambridge *et al.* 1995) of each node j.

3.2.3.c The need for a starting configuration of nodes

Nolet & Montelli (2005) discuss the non-linearity of the optimization problem described in section 3.2.3. Therefore, rather than using methods of non-linear optimization (e.g. simulated annealing), we construct a starting grid that is close enough to the global minimum (Nolet & Montelli 2005). That is, the final (optimal) mesh will then be obtained with a simple conjugate gradient algorithm (cf. section 3.2.3.h).

Many methods[6] could be used for generating a starting configuration of nodes. Following Nolet & Montelli (2005), we use the method of the tessellation of a sphere (e.g. Wang & Dahlen 1995; Sambridge & Faletic 2003). Triangles are subdivided until their side's length is "close" to the desired node spacing, imposed by the local resolution length.

The number of nodes, M_z, for each layer z is then determined by the starting grid. For each individual layer z, the number of nodes available to minimize E_z, that is M_z, determines how close we can get to the ideal solution $E_z = 0$. The starting configuration of nodes is then a crucial issue.

3.2.3.d Tessellation of a sphere from subdivided icosahedron

We aim at constructing a starting configuration of nodes, for each layer z, by iteratively refining an initial icosahedron, driven by the assumed resolution length of the data.

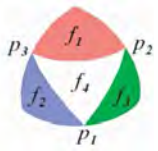


Figure 3.3: Four subdivision triangles coming from the same parent triangle. From Wu et al. (2005).

An icosahedron is a regular polyhedron with 20 identical equilateral triangular faces, 30 edges and 12 vertices (cf. figure 3.4.a). A local subdivision of single faces of the icosahedron can be implemented easily, since all faces are triangles at all stages of iteration. One subdivision step of a single triangle consists of the introduction of 3 new nodes at the middle points of all triangle edges (cf. figures 3.4 and 3.3). Together with the three nodes of the original (i.e. the "parent") triangle, these new nodes define 4 smaller child triangles, which together cover the parent triangle exactly.

[6]For instance, Montelli *et al.* (2004ab) generate a specified number of random node locations within a set of subregions in the mantle, and reject those that are too close.

With this subdivision scheme, it is possible to describe complex shapes on a spherical shell by iterated refining. It is then possible to construct a starting configuration grid that has small meshes in "good resolution" regions.

Table 3.2 shows that the number of grid nodes is multiplied by about 4 with each subdivision step. From the subdivision scheme, one could further expect that the number of nodes is 3 times the number of triangles. However, since each node is used by 5 or 6 triangles, there are only about 3/8 nodes per triangle. This property is one of the reasons that lead us to choose the nodes themselves, rather than the triangles, as the model parameters (cf. section 3.2.6).

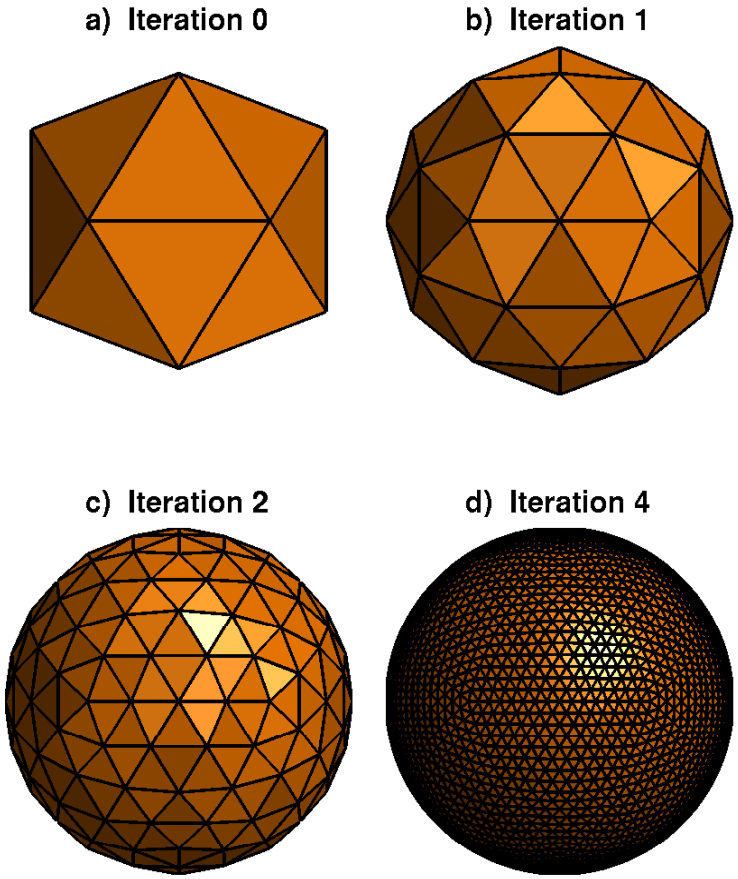


*Figure 3.4: Icosahedron grid in different stages of face subdivision. For these plots, all triangles are subdivided equally to show the global uniformity of the resulting grid. (**a**) Initial icosahedron, (**b**) first iteration: new corners were introduced at the middle of each edge of the icosahedron, (**c**) second iteration and (**d**) fourth iteration.*

Table 3.2: Number of nodes and triangles in a globally refined icosahedron grid

Iteration	Nodes	Triangles	Nodes/Triangles
0	12	20	0.6
1	42	100	0.42
2	162	420	0.3857
3	642	1,700	0.3776
4	2,562	6,820	0.3757
5	10,242	27,300	0.3752
6	40,962	109,220	0.3750

3.2.3.e Refining triangle criteria

Here, we present the criteria we have applied for refining spherical triangles (cf. figure 3.3), based on the assumed local resolution length of the data.

Let f_n be one triangle of the icosahedron grid at the n-*th* iteration. We need to design a criteria, for refining or not the triangle f_n, based on the 'averaged' resolution function value $\bar{\ell}_{f_n}$ within the triangle f_n. That is, the resolution length value can change within the same triangle.

In order to estimate an "averaged" value $\bar{\ell}_{f_n}$ for the triangle f_n, we do the following:

1. first, we calculate the coordinates of the 4 barycenters G_i, with $i = 1, \cdots, 4$, of the four child triangles of the parent triangle f_n.

2. Then, we calculate the resolution values $\ell(G_i)$ at the 4 barycenters G_i locations.

3. Finally, we compute: $\bar{\ell}_{f_n} = \frac{1}{4} \cdot \sum_{i=1}^{4} \ell(G_i)$.

Let L_{f_n} be the edge length of the (equilateral) triangle f_n. The edge length of its three child triangle is: $L_{f_{n+1}} = \frac{L_{f_n}}{2}$. The criteria we use for the triangle refining process is:

$$\begin{cases} \text{if } \; L_{f_{n+1}} \geq \varepsilon \cdot \bar{\ell}_{f_n} \\ \Rightarrow \text{subdivide triangle} \\ \text{otherwise} \\ \Rightarrow \text{triangle is ok} \end{cases} \tag{3.6}$$

where ε is a number that is adapted to the resolution length bounds for each layer (cf. table 3.1), as explained in section 3.2.3.f.

It is worth noting that, because we only have to subdivide 2-D spherical triangles, the parameterization refining is well behaved. That is, we only introduce 3 more unknowns (i.e. nodes) at each refining step, and always have to deal with spherical triangles. The equivalent in 3-D tetrahedra, used by other authors (e.g. Sambridge & Faletic 2003), creates many more unknowns at each step. This may be a serious disadvantage with highly irregular 3-D meshes.

Our approach avoids the complexities of the 3-D approach. This lead us to favor the "layered style" at irregular parametrization, that is: 2-D laterally (spherical triangles) and 1-D radially (i.e. constant-depth layers).

3.2.3.f Resolution length bounds

We need to define a "resolution length" function everywhere in the mantle. As we use global S-wave travel times, measured in the 10–51 s period range, we may have an *a priori* on the upper and lower bounds of this resolution function (e.g. Nolet 2008; Montelli *et al.* 2006b; Vasco *et. al* 2003).

Firstly, we expect that in highly sampled regions in the mantle, we can resolve laterally varying structure of ∼200 km size. Secondly, in very low sampled regions, we cannot resolve laterally varying structure smaller than ∼1,000 km.

Table 3.1 shows, for each layer z, the minimum and maximum number of refining iterations from an initial icosahedron, with the associated edge length L_f, used for building the starting grids. The refining process starts from an initial icosahedron which is based on a sphere with radius R_z. Therefore, if one applies the same number of refining iterations to two different layers (i.e. with not the same radius R_z), we will get a different triangle edge length, L_f. This, in conjunction with the upper and lower bounds of the resolution length, lead us to choose an ε value, in equation (3.12), that is not constant for the 18 layers (cf. table 3.1).

It is our experience that the resolution bounds, and the ε values given at table 3.1, enable us to put small (large) enough mesh in highly (poorly) sampled regions - as far as our dataset is concerned.

3.2.3.g Estimating the resolution function

We aim to estimate the resolution function, $\ell(\mathbf{r})$, which gives the (assumed) resolution length of the data, at each spacial point $\mathbf{r}$ in the mantle.

For a linear problem of the form $\mathbf{d} = G \cdot \mathbf{m}$, with a generalized inverse G^{-}, the solution estimate is given by $m_e = G^{-} \cdot d = G^{-} \cdot G \cdot m = R \cdot m$, where R is the resolution

matrix (Wiggins 1972). If R can be explicitly computed, the width of the region along the jth row of R, where non-diagonal terms are significantly different from zero, defines the resolution length near node j.

Unfortunately, global tomographic systems are usually so large, that the resolution matrix, R, cannot explicitly be computed. In this case, the best compromise is to "estimate" the resolution length (cf. Nolet & Montelli 2005, and references therein). Vasco *et. al* (2003) show that the ray density provides a first-order estimate of local resolution, for an extensive set of body waves traversing the mantle and core of the Earth. That is, except in the region of the Pacific, where significant trade-offs occur along the dominant ray direction. In this thesis, we have therefore chosen to approximate[7] the local resolution length, of our data, to the local ray density.

We need to calculate the ray density function, of our global dataset, in each layer z. One aims to do it without distortion problems at the Northern and Southern poles, that usually happen if a regular grid (e.g. 6° by 6°) is used. In this thesis, we have chosen to use the mesh provided by a 4 times iterated icosahedron. This leads to a uniform mesh of (mostly) equilateral spherical triangles (cf. figure 3.4.d). Each layer z is meshed by regular spherical triangular prisms, whose volume is

$$V_{prism} \simeq \frac{\sqrt{3}}{4} {L_f}^2 \Delta z, \tag{3.7}$$

where L_f is the triangle edge and Δz is the layer thickness. It is then straightforward to calculate the ray density in each prism, over the entire mantle. One simply counts the number of geometrical rays passing through each prism, and then normalize this number by the prism volume, V_{prism} - equation (3.13).

At this stage, we know the ray density function, and the upper and lower bounds of

[7]Using the "kernel density", rather than the ray density, would be more consistent for the purpose of multiple-frequency tomography. That is, one aims to obtain a model parameterization driven by (frequency-dependent) data, which are related to 3-D sensitivity "finite-frequency" kernels (Dahlen *et al.* 2000). However, when we performed this work, our automated program, for projecting a large amount of 3-D kernels onto a Delaunay mesh, was not already coded. This prevented us from calculating the kernel density. Even if using 3-D kernels, instead of geometrical rays, will lead to a (slightly) different mesh, we believe that the resulting parameterization will be unlikely to modify our tomographic results (cf. chapter 4). Nevertheless, we plan to build a new parameterization, that will be driven by the 3-D kernels, because it is now feasible (and so there is no reason for not doing it). It is worth noting that the kernel density should take into account the absolute kernel values, instead of simply counting the number of kernels passing through each cell - as we do for the ray density (cf. section 3.2.3.g). For instance, consider the 3-D sensitivity kernel of one S wave, that passed through the Earth's mantle. If an heterogeneity, with spatial position $\mathbf{r}$, is located *on* the geometrical ray path, it will be "hidden" in the doughnut hole of the associated sensitivity kernel (cf. section 1.4.3, of chapter 1). Thus, the resolution length $\ell(\mathbf{r})$ should be set to zero, because the anomaly is not "sensed" by the S wave, and cannot be resolved.

the resolution length function, for each layer z. The next step is to assign to each spatial point $\mathbf{r}$ of the mantle, with a given ray density, the corresponding resolution length $\ell(\mathbf{r})$. Thus, for each individual layer z, our criteria are:

1. The highest and lowest ray density regions are associated to the upper and lower resolution length bounds, respectively (cf. table 3.1).

2. Concerning regions with intermediate ray density value, the associated resolution length is computed by linear interpolation, between the resolution length bounds, in function of the ray density value.

This procedure allows us to define a resolution function, at every point, $\mathbf{r}$, in the mantle. We can then use $\ell(\mathbf{r})$ in our refining mesh algorithm (section 3.2.3.e), and get a starting configuration grid for each layer z. Table 3.1 shows the resulting number of nodes, M_z, of each starting grid. We obtain a total number of 38,125 nodes, spanning the 18 layers.

3.2.3.h Optimization by conjugate gradient

One aims to optimize the agency of the 38,125 starting nodes, by minimizing the total energy, E_z, of the system, which is based on the previously defined resolution length function, $\ell(\mathbf{r})$ (cf. section 3.2.3.c) We give, in appendix B, the outline of our optimization scheme.

Following Nolet & Montelli (2005), the minimization of E_z - equation (3.5) - is performed using a conjugate gradient algorithm. For each layer z, all the nodes lie on a sphere of radius R_z. We can then use their spherical coordinates θ (colatitude) and ϕ (longitude) to find the gradient of E_z. As we deal with large radius R_z, we may use the great-circle distance rather than the line-of-sight distance. We then obtain, for each layer z, the following equations to implement in the conjugate gradient algorithm (Nolet & Montelli 2005):

$$E_z = \sum_{j=1}^{M_z} \sum_{k \in N_j} \frac{(R_z \Delta_{jk} - \ell_{jk})^2}{\ell_{jk}^2}, \tag{3.8}$$

$$\frac{\partial E_z}{\partial \theta_l} = 4 \sum_{k \in N_l} \frac{R_z (R_z \Delta_{lk} - \ell_{lk})}{\ell_{jk}^2} \frac{\partial \Delta_{lk}}{\partial \theta_l}, \tag{3.9}$$

$$\frac{\partial E_z}{\partial \phi_l} = 4 \sum_{k \in N_l} \frac{R_z (R_z \Delta_{lk} - \ell_{lk})}{\ell_{jk}^2} \frac{\partial \Delta_{lk}}{\partial \phi_l}, \tag{3.10}$$

where Δ_{lk} is the geodesic distance between nodes l and k, which satisfies:

$$\cos(\Delta_{lk}) = \cos(\phi_l - \phi_k)\sin(\theta_l)\sin(\theta_k) + \cos(\theta_l)\cos(\theta_k). \tag{3.11}$$

Note that every node pair - i.e. nodes j and k in equation (3.6) - occurs twice in the sum over pairs, which leads to the '4' term in the partial derivatives of E_z (cf. Nolet 2008). We then find the derivatives of Δ_{lk} by differentiating equation (3.9), with respect to longitude ϕ and colatitude θ (Nolet & Montelli 2005):

$$\frac{\partial \Delta_{lk}}{\partial \theta_l} = \frac{\sin(\theta_l)\cos(\theta_k) - \cos(\phi_l - \phi_k)\cos(\theta_l)\sin(\theta_k)}{\sin(\Delta_{lk})}, \tag{3.12}$$

$$\frac{\partial \Delta_{lk}}{\partial \phi_l} = \frac{\sin(\phi_l - \phi_k)\sin(\theta_l)\sin(\theta_k)}{\sin(\Delta_{lk})}, \tag{3.13}$$

Figure 3.5 shows the convergence of the energy minimization of the nodes agency of the sixth layer (cf. table 3.1). As in the case of the sixth layer (cf. figure 3.5), less than twenty iterations were needed for the minimization to converge for the starting grids of the 18 layers. Moreover, it is also our experience that only seven iterations were enough for the optimization to almost fully converge (cf. figure 3.5). This fast convergence is related to the use of a good starting configuration - which is already "close" to the optimal grid.

3.2.3.i Quality of the optimized grid

Here, one aims to compare the nodes agency of both the starting and optimized grids, by inspecting the distribution of the parameter ξ, defined as (Nolet & Montelli 2005):

$$\xi = \frac{L_{jk}}{\ell_{jk}} \tag{3.14}$$

, where nodes j and k are natural neighbors, L_{jk} is the actual distance between j and k, and ℓ_{jk} the average resolving length between j and k. The parameter ξ should be close to 1 for a good nodes agency.

Figure 3.6 shows histograms of ξ, before and after the optimization, in the case of the sixth layer:

1. ξ averages 0.97 for the starting grid, with a distribution quite sparse around $\xi = 1$.

2. ξ averages 1.001 for the optimized grid, with a distribution more symmetric around $\xi = 1$, and with smaller variance.

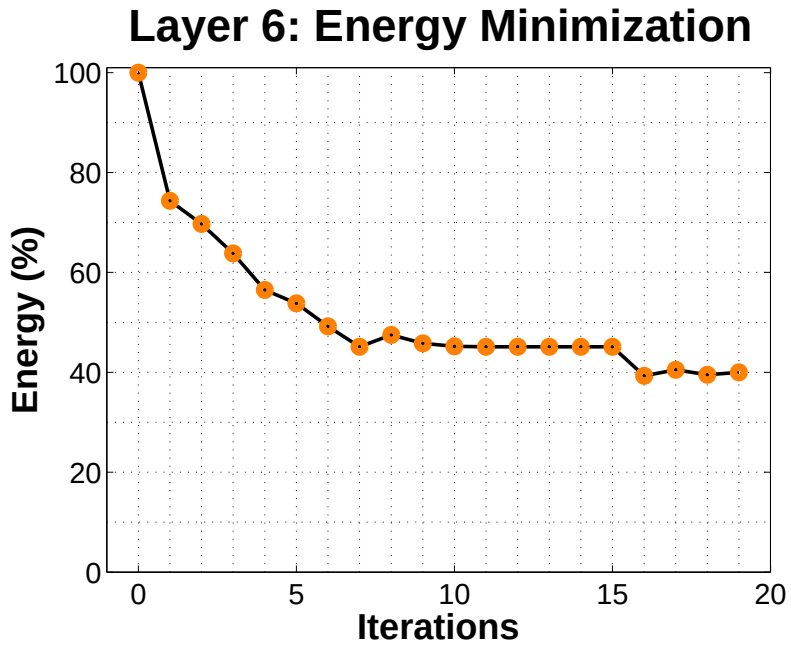


Figure 3.5: Convergence of the energy minimization of the starting nodes agency, in the case of the sixth layer (530–660 km depth), by using a conjugate gradient algorithm. The starting grid corresponds to the iteration 0.

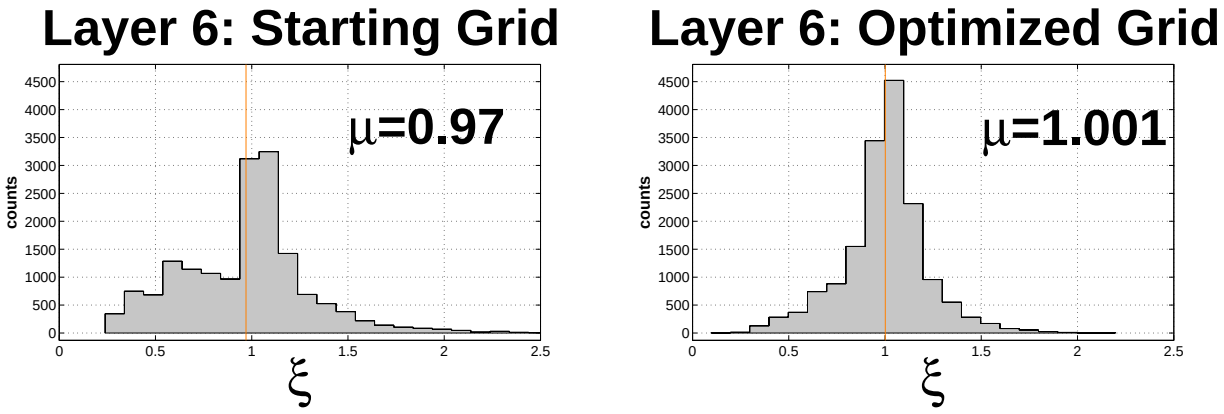


Figure 3.6: Comparison between starting and optimized grids, in the case of the sixth layer (530–660 km depth). The parameter ξ is defined in section 3.2.3.i - equation (3.14). The mean of each histogram of ξ is denoted by μ.

Finally, figure 3.7 shows the actual starting and optimized (model) parameterizations, in the case of the sixth layer (530–660 km depth).

Figure 3.7: Starting (upper frame) and optimized (lower frame) grids, displayed in black, of the sixth layer (530–660 km depth). The resolution length function, $\ell(\boldsymbol{r})$, is superimposed, with a linear colorscale ranging from 210 km (in cyan) to 850 km (in magenta).

3.2.3.j Application to our global dataset

Figure 3.8 shows the optimized parameterizations (i.e. nodes), corresponding to our global dataset, obtained for six layers (cf. table 3.1). The current coverage of seismic stations generally allows us to have closely spaced nodes in the Northern Hemisphere, for all but the rays with the shallowest lower mantle turning depths. Nodes spacing is more coarse in the Southern Hemisphere, as stations coverage is still generally poorer. At depth greater than ~1,700 km, most of the ray coverage comes from *ScS* waves, with a preferential sampling around the Pacific ring.

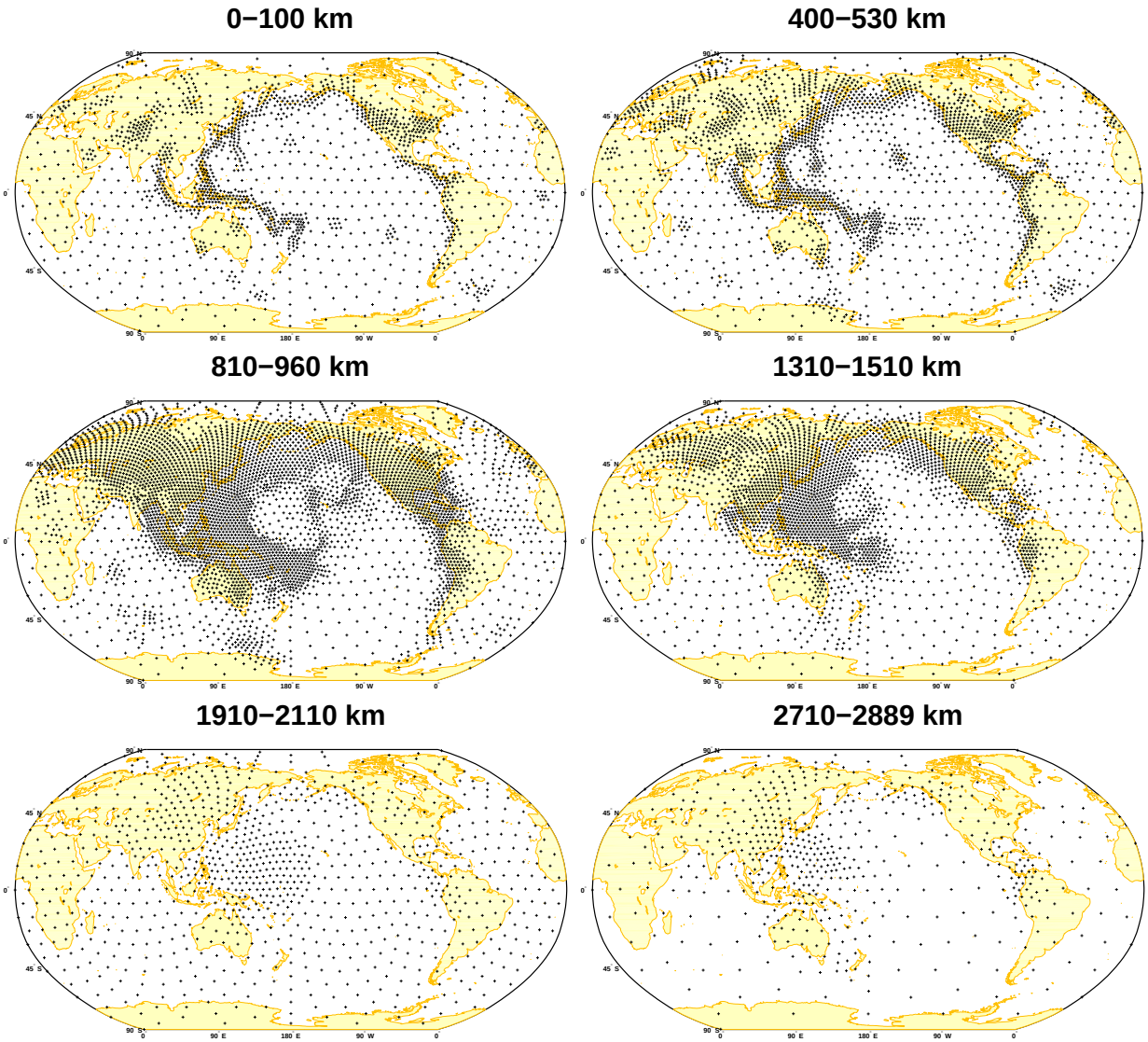


Figure 3.8: Optimized nodes agency, obtained for six layers (cf. table 3.1).

3.2.4 Interpolation

The interpolation consists of relating the model value $m(\mathbf{r})$, at an arbitrary location $\mathbf{r}$ inside the Earth, in function of the model parameters m_j. Thus, with the use of some basis functions h_j, we have: $m(\mathbf{r}) = \sum_{j=1}^{M} m_j h_j(\mathbf{r})$ - equation (3.4). Here, we only consider basis functions with "local support", which means that they are only non-zero within a limited sub-region of the model. Because we use a parameterization made of "blocks", we have two main choices for the basis functions:

1. The model parameters, m_j, represent constant velocity perturbation within each block j. The basis functions are then defined as:

$$\begin{cases} \text{if } \mathbf{r} \in \text{block } j \\ \Rightarrow h_j(\mathbf{r}) = 1 \\ \text{otherwise} \\ \Rightarrow h_j(\mathbf{r}) = 0. \end{cases} \tag{3.15}$$

2. The basis functions are "linear" interpolation functions.

In this thesis, we have chosen to use linear interpolation functions h_j, as explained in the following.

We have parameterized the Earth's mantle with 18 spherical layers, filled in with spherical triangular prisms, corresponding to a total number of 38,125 nodes. Each prism is defined by three nodes, corresponding to three model parameters m_j (cf. figure 3.2).

Let us consider the (spherical triangular) prism $P_f(\mathbf{r})$ enclosing a point $\mathbf{r}$, with $f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$ the spherical triangle located at the top of this prism. We note $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ the three vertices of the triangle f. Tomographic equations - equation (3.2) - are integrals, and insensitive to small detail in the basis functions that describe the model. Therefore, a "linear" interpolation between the three vertices $\mathbf{p}_0$, $\mathbf{p}_1$ and $\mathbf{p}_2$ is appropriate (e.g. Nolet & Montelli 2005). This can be accomplished using the concept of "barycentric coordinates" (cf. appendix C). Consider $\mathbf{r}'$ as the radial projection of $\mathbf{r}$ onto the spherical triangle $f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$. Thus, the three barycentric coordinates, b_i $(i = 1, 2, 3)$, are defined as

$$\mathbf{r}' = b_1\mathbf{p}_1 + b_2\mathbf{p}_2 + b_3\mathbf{p}_3. \tag{3.16}$$

To make them unique, they must be normalized as

$$b_1 + b_2 + b_3 = 1. \tag{3.17}$$

Equation (3.16) shows that the barycentric coordinates themselves are the appropriate weights to use in a linear interpolation scheme. If $\mathbf{r}'$ is at vertex k, all b_i equal zero except $b_k = 1$. Moreover, the b_i on a triangle's edge between a pair of vertices are only dependent on those two vertices. This guarantees the continuity of the interpolated field between two adjacent triangles that share one edge.

By summing over the contributions of the 3 vertices of the triangle $f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$, associated to the prism $P_f(\mathbf{r})$ enclosing $\mathbf{r}$, the seismic model value $m(\mathbf{r})$ can be linearly interpolated:

$$m(\mathbf{r}) = m(\mathbf{r}') = \sum_{i=1}^{3} b_i m(\mathbf{p}_i). \tag{3.18}$$

Note[8] that $m(\mathbf{r}) = m(\mathbf{r}')$ as our parameterization is made of constant-depth layers. The correspondence with equation (3.2) is now obvious. We only need to map back each triangle vertex $m(\mathbf{p}_i)$ to its original parameter index m_j.

Locating a point in a highly irregular mesh is clearly more complicated than in a regular mesh. Here, in order to find the prism that encloses $\mathbf{r}$ we have developed an algorithm, named the "*Where Am I*" algorithm, which is presented in appendix C.

Our locating algorithm is largely based on the "*spherical point location*" algorithm of Wu *et. al* (2005), but is tuned for our own application. Note that the algorithm of *Wu et al.* (2005) largely makes use of the previous ideas of the "*walking triangle*" algorithm, developed by Sambridge *et. al* (1995).

3.3 Calculating the Fréchet kernels

In this section, following the formalism of Dahlen *et al.* (2000), we derive analytical expressions for single phase (*S*, *ScS*, *SS*) and two-phase interference (*S*+*sS*, *ScS*+*sScS*, *SS*+*sSS*) "finite-frequency" sensitivity kernels, in a spherical (1-D) reference model (e.g. IASP91).

Analytical expressions allow us to compute the Fréchet kernels extremely fast (generally less than 1 s for each kernel). This is very valuable as we aim to calculate hundreds of thousands of kernels on a very fine "kernel grid" (cf section 3.4.2). Such calculations are feasible because one assumes that the source time function is Gaussian, so that we can represent it with an analytical expression. This assumption is appropriate for calculating

[8]For instance, this would not be the case if 3-D tetrahedra were used for parameterizing the model.

Fréchet kernels corresponding to teleseismic waves - at the global scale.

Concerning local or regional studies, one should instead try to determine the exact source time function (e.g. Sigloch & Nolet 2006). However, with more realistic source time functions, we cannot obtain analytical expressions for the Fréchet kernels, and so they have to be numerically calculated (e.g. Tian *et al.* 2007a). The computation cost of such a numerical estimation is much higher. In addition, we have not attempted to determine exact source time functions in building our global dataset (cf. chapter 2). Therefore, we have chosen to use analytical kernels for this global multiple-frequency tomography. Nevertheless, it is our view that this is unlikely to make a significant difference in the tomographic results of this thesis.

3.3.1 Single phase kernels

Under the paraxial approximation, the travel time sensitivity kernel of a "single" body wave, with respect to velocity perturbation ($\delta c/c$), is (*Dahlen et al.* (2000):

$$K(\mathbf{r}_x) = -\frac{1}{2\pi c(\mathbf{r}_x)} \frac{R_{rs}}{c_r R_{xr} R_{xs}} \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin[\omega \Delta T(\mathbf{r}_x) - \Delta\Phi(\mathbf{r}_x)] d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2 d\omega}, \tag{3.19}$$

where $\mathbf{r}_x$ is the spatial position of scatterer x, $\Delta\Phi$ is the phase shift due to passage through caustics or super critical reflection, R_{rs}, R_{xr}, R_{xs} are the geometrical spreading factors, $|\dot{m}(\omega)|^2$ is the source power spectrum, ΔT is the detour time of the scattered wave, c_r and $c(\mathbf{r}_x)$ are the velocities at receiver and scatterer position, respectively.

Unless the wave is supercritically reflected, with an angle-dependent phase shift, $\Delta\Phi$ takes three possible values: 0, $-\pi/2$, and $-\pi$ (e.g. Dahlen *et al.* 2000; Hung *et al.* 2000). $\Delta\Phi$ is always equal to 0 for S and ScS waves. Concerning SS waves, $\Delta\Phi(\mathbf{r}_x) = -\pi/2$ in between the two caustics located at $\Delta/3$ and $2\Delta/3$, with Δ the total epicentral distance, and $\Delta\Phi(\mathbf{r}_x) = 0$ elsewhere.

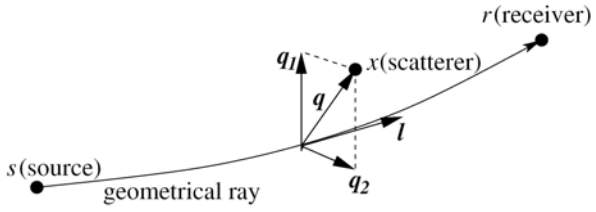


Figure 3.9: Ray centered coordinates of scatterer $\boldsymbol{r}_x = (l; q_1, q_2)^t$, where l is the length of the geometrical ray measured from the source. Vectors $\boldsymbol{l}$, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are mutually perpendicular, with $\boldsymbol{l}$ and $\boldsymbol{q}_1$ in the source-receiver plane. From Tian et.al (2007b).

In the following, we explain how to calculate each term of the single-phase kernel expression - equation (3.19) - for *S*, *ScS* and *SS* seismic waves. Note that we will often use the ray centered coordinates of scatterer x, as depicted in figure 3.9, that is: $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2)^t$.

3.3.1.a Calculating the coordinates of scatterer x

Consider the basis $F' = (O, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, in which the Fréchet kernels will be calculated. The choice of F' is explained in section 3.4.3. Here, one aims to calculate the coordinates of any scatterer $\mathbf{x}$ in such a basis F'.

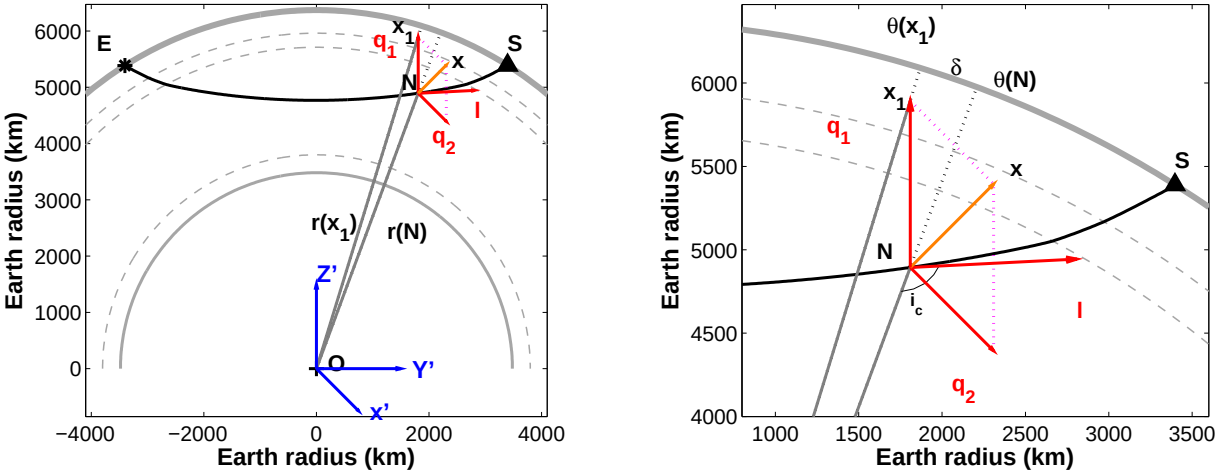


Figure 3.10: Coordinates of scatterer $\boldsymbol{x}$ in the basis $F' = (O, \boldsymbol{X}', \boldsymbol{Y}', \boldsymbol{Z}')$. O is the Earth centre, E is the epicenter, S is the station, i_c is the ray incidence angle. See text for other notations.

Figure 3.10 shows the geometry associated to a scatterer $\mathbf{x}$ in the basis F'. Let Δ be the total epicentral distance, between the epicenter and the station. In the basis F', we define the angular distance, θ, such as: θ is equal to $-\Delta/2$ at the epicenter, zero at the mid-point (between epicenter and station), and $+\Delta/2$ at the station. Consider $\mathbf{x}_1$ as the projection of $\mathbf{x}$ onto the geometrical ray plane (i.e. the plane defined by this page). Let $\theta(\mathbf{x}_1)$ and $\theta(N)$ be the angular distance of the points $\mathbf{x}_1$ and N, respectively. Let $r(\mathbf{x}_1)$ and $r(N)$ be the radial distances $|\mathbf{Ox}_1|$ and $|\mathbf{ON}|$, respectively. The coordinates $(X'_x; Y'_x; Z'_x)_{F'}$ of the scatterer $\mathbf{x}$, in the basis F', are then given by:

$$\begin{cases} X'_x = q_2(\mathbf{x}) \\ Y'_x = r(\mathbf{x}_1)\sin[\theta(\mathbf{x}_1)] \\ Z'_x = r(\mathbf{x}_1)\cos[\theta(\mathbf{x}_1)] \end{cases} \tag{3.20}$$

It is now clear, from equation (3.20), that we have to calculate $r(\mathbf{x}_1)$ and $\theta(\mathbf{x}_1)$. We explain how we have proceeded in the following.

First, we need to determine the radius $|\mathbf{r}_x|$ of the scatterer $\mathbf{x}$. We have:

$$\mathbf{r}_x = \mathbf{Ox} = \mathbf{ON} + \mathbf{q}_1 + \mathbf{q}_2 \tag{3.21}$$

which leads to

$$|\mathbf{r}_x|^2 = |\mathbf{Ox}|^2 = (\mathbf{ON} + \mathbf{q}_1 + \mathbf{q}_2)(\mathbf{ON} + \mathbf{q}_1 + \mathbf{q}_2) \tag{3.22}$$

and

$$|\mathbf{r}_x|^2 = |\mathbf{ON}|^2 + 2\mathbf{ON}\cdot\mathbf{q}_1 + 2\mathbf{ON}\cdot\mathbf{q}_2 + 2\mathbf{q}_1\cdot\mathbf{q}_2 + |\mathbf{q}_1|^2 + |\mathbf{q}_2|^2. \tag{3.23}$$

Let i_c the incidence angle of the geometrical ray (cf. figure 3.10). Using the fact that $\mathbf{q}_1$ is perpendicular to $\mathbf{q}_2$, and $\mathbf{ON}$ is perpendicular to $\mathbf{q}_2$ (their scalar product is null), we obtain:

$$\begin{cases} |\mathbf{ON}|^2 = |\mathbf{r(N)}|^2 \\ \mathbf{ON}\cdot\mathbf{q}_1 = |\mathbf{r(N)}||\mathbf{q}_1|\cos(i_c - \pi/2). \end{cases} \tag{3.24}$$

Therefore, we have:

$$|\mathbf{r}_x| = \sqrt{|\mathbf{r(N)}|^2 + 2|\mathbf{r(N)}||\mathbf{q}_1|\cos(i_c - \pi/2) + |\mathbf{q}_1|^2 + |\mathbf{q}_2|^2}. \tag{3.25}$$

It is now straightforward to obtain the distance $r(\mathbf{x}_1) = \mathbf{Ox}_1$, if we realize that it corresponds to the distance $|\mathbf{r}_x|$ with $\mathbf{q}_2 = \mathbf{0}$. This leads to:

$$r(\mathbf{x}_1) = \sqrt{|\mathbf{r(N)}|^2 + 2|\mathbf{r(N)}||\mathbf{q}_1|\cos(i_c - \pi/2) + |\mathbf{q}_1|^2}. \tag{3.26}$$

Second, we aim to calculate $\theta(\mathbf{x}_1)$. Let δ the angle difference between $\theta(\mathbf{x}_1)$ and $\theta(N)$. We realize that $\theta(\mathbf{x}_1)$ may be derived from $\theta(N)$ and δ, in function of the incidence angle i_c of the geometrical ray. That is:

$$\begin{cases} \text{if } \; i_c > \pi/2 \\ \Rightarrow \; \theta(\mathbf{x}_1) = \theta(N) + \text{sign}[q_1(\mathbf{x})]|\delta| \\ \text{otherwise, if } \; i_c < \pi/2 \\ \Rightarrow \; \theta(\mathbf{x}_1) = \theta(N) - \text{sign}[q_1(\mathbf{x})]|\delta| \\ \text{otherwise, if } \; i_c = \pi/2 \\ \Rightarrow \; \theta(\mathbf{x}_1) = \theta(N). \end{cases} \tag{3.27}$$

Finally, one needs to calculate δ. Consider the triangle defined by the three points: O, N

and $\mathbf{x}_1$. The three edge lengths of this triangle are known: $r(\mathbf{x}_1)$ between O and x_1; $r(N)$ between O and N; $|q_1(\mathbf{x})|$ between N and x_1. The angle between **ON** and $\mathbf{Ox}_1$ is δ. We then apply the "Al-Kashi theorem" in this triangle:

$$|q_1(\mathbf{x})|^2 = r^2(N) + r^2(\mathbf{x}_1) - 2r(N)r(\mathbf{x}_1)\cos(\delta). \tag{3.28}$$

This leads to

$$\delta = \text{acos}(\frac{r^2(N) + r^2(\mathbf{x}_1) - |q_1(\mathbf{x})|^2}{2r(N)r(\mathbf{x}_1)}). \tag{3.29}$$

3.3.1.b Calculating the scatterer velocity $c(\mathbf{r}_x)$

We aim to calculate the scatterer velocity $c(\mathbf{r}_x) = c(|\mathbf{r}_x|)$ in a 1D (radial) reference seismic velocity model (e.g. IASP91). $|\mathbf{r}_x|$ is the radial distance of the scatterer $\mathbf{x}$ from the Earth centre, given by equation (3.25). Therefore, by simply interpolating in the radial reference model $c(\mathbf{r})$, we obtain the scatterer velocity $c(\mathbf{r}_x)$.

3.3.1.c Calculating the ratio $R_{rs}/(c_r R_{xr} R_{xs})$

Following Dahlen *et al.* (2000), this ratio can be written as:

$$\frac{R_{rs}}{c_r R_{xr} R_{xs}} = \sqrt{|\det(\mathbf{H})|}. \tag{3.30}$$

The matrix $\mathbf{H}$ is called the Hessian matrix, and is defined as:

$$\mathbf{H} = \begin{pmatrix} H_{11} & 0 \\ 0 & H_{22} \end{pmatrix} \tag{3.31}$$

with:

$$\begin{cases} H_{11} = (\partial^2 t/\partial q_1^2)_{\mathbf{q}=\mathbf{0}} \\ H_{22} = (\partial^2 t/\partial q_2^2)_{\mathbf{q}=\mathbf{0}} \end{cases} \tag{3.32}$$

where t stands for the travel time field. The Hessian matrix is defined for each point N located on the geometrical ray (i.e.with $\mathbf{q} = \mathbf{0}$). The ratio $R_{rs}/(c_r R_{xr} R_{xs})$ is the same for every scatterer x located in the plane $(N, \mathbf{q}_1, \mathbf{q}_2)$. We compute the Hessian values H_{11} and H_{22} using the software by Tian *et. al* (2007b).

3.3.1.d Calculating the detour time $\Delta T(\mathbf{r}_x)$

The detour time, $\Delta T(\mathbf{r}_x)$, represents the extra time needed for the ray to visit the scatterer, at spatial position $\mathbf{x}$. That is: $\Delta T(\mathbf{r}_x) = T_{sx} + T_{xr} - T_{sr}$, with T_{sx}, T_{xr} and T_{sr} the travel times between source/scatterer, scatterer/receiver and source/receiver, respectively (cf. figure 1.6, in chapter 1). Provided that we know the Hessian matrix H everywhere on the ray path, the detour time can be expressed as (Dahlen *et al.* 2000):

$$\Delta T(\mathbf{r}_x) = \frac{1}{2}\mathbf{q}^t \cdot \mathbf{H} \cdot \mathbf{q}, \tag{3.33}$$

which can be reformulated as:

$$\Delta T(\mathbf{r}_x) = \frac{1}{2}(H_{11}q_1^2 + H_{22}q_2^2). \tag{3.34}$$

3.3.1.e Calculating the source power spectrum $|\dot{m}(\omega)|^2$

Here, we assume a Gaussian source time function, in order to derive analytical expressions for the finite-frequency kernels. This assumption is appropriate for global tomography, in our period range of interest (i.e. 10–51 s). Following Hung *et. al* (2001), we use a source time function of the form:

$$m(t) = e^{-2\pi^2(t/T-1/2)^2}. \tag{3.35}$$

Its corresponding time derivative is $\dot{m}(t) = -4\pi^2 T^{-2}(t-T/2)e^{-2\pi^2(t/T-1/2)^2}$, with a visually upswing at $t \simeq 0$ and a characteristic period equal to T (i.e. T is the visual period of the seismic wave). The power spectrum of the source time function is then:

$$|\dot{m}(\omega)|^2 = \frac{\omega^2 T^2}{2\pi} e^{-\omega^2 T^2/4\pi^2}. \tag{3.36}$$

3.3.1.f Calculating $N(\Delta\Phi)$

Here, one aims to calculate $N(\Delta\Phi) = \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin[\omega\Delta T - \Delta\Phi] d\omega$, in the presence ($\Delta\Phi = -\pi/2$) or not ($\Delta\Phi = 0$) of caustics.

1. $\underline{\Delta\Phi = 0}$
   We have:
   $$N(0) = T^2/2\pi \int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \sin[\omega\Delta T] d\omega. \tag{3.37}$$

In appendix A.1, we show how to calculate the integral $\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \sin[\omega\Delta T]d\omega$. This leads to:

$$N(0) = \sqrt{\pi}e^{-(\Delta T\pi/T)^2} 16\pi^{10}\Delta T/T^9(\Delta T^4 - 5\Delta T^2T^2/\pi^2 + \frac{15}{4}\frac{T^4}{\pi^4}). \tag{3.38}$$

2. $\underline{\Delta\Phi = -\pi/2}$
As $\sin[\omega\Delta T - (-\pi/2)] = \cos[\omega\Delta T]$, we now have to calculate:

$$N(-\frac{\pi}{2}) = T^2/2\pi \int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \cos[\omega\Delta T]d\omega. \tag{3.39}$$

In appendix A.2, we show how to calculate the integral $\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \cos[\omega\Delta T]d\omega$. This leads to:

$$\begin{array}{l} N(-\frac{\pi}{2}) = T^2/(2\pi)\{2^5(\frac{\pi}{T})^{10}[\Delta T^4 - \frac{9}{2}\Delta T^2T^2/\pi^2 + 2T^4/\pi^4] \\ -\sqrt{\pi}2^5(\frac{\pi}{T})^{11}\Delta T e^{-(\Delta T\pi/T)^2}[\Delta T^4 - 5\Delta T^2T^2/\pi^2 + \frac{15}{4}T^4/\pi^4]\text{erfi}(\Delta T\pi/T)\}. \end{array} \tag{3.40}$$

In appendix A.4, we also show how to efficiently calculate the imaginary error function $\text{erfi}(z) = 2/\sqrt{\pi}\int_0^z e^{+t^2}dt$.

3.3.1.g Calculating D

Here, one aims to calculate $D = \int_0^\infty \omega^2|\dot{m}(\omega)|^2 d\omega$. We have:

$$D = T^2/(2\pi) \int_0^\infty \omega^4 e^{-\omega^2(T/(2\pi))^2} d\omega \tag{3.41}$$

In appendix A.3, we show how to calculate the integral $\int_0^\infty \omega^4 e^{-\omega^2(T/(2\pi))^2} d\omega$. This leads to

$$D = \frac{3\sqrt{\pi}}{8}(\pi/T)^5 2^5 T^2/(2\pi). \tag{3.42}$$

Finally, 2-D views of single phase *S*, *ScS* and *SS* "finite-frequency" travel time sensitivity kernels, analytically calculated at 10 and 34 s periods, are shown at figures 3.11, 3.12 and 3.13, respectively.

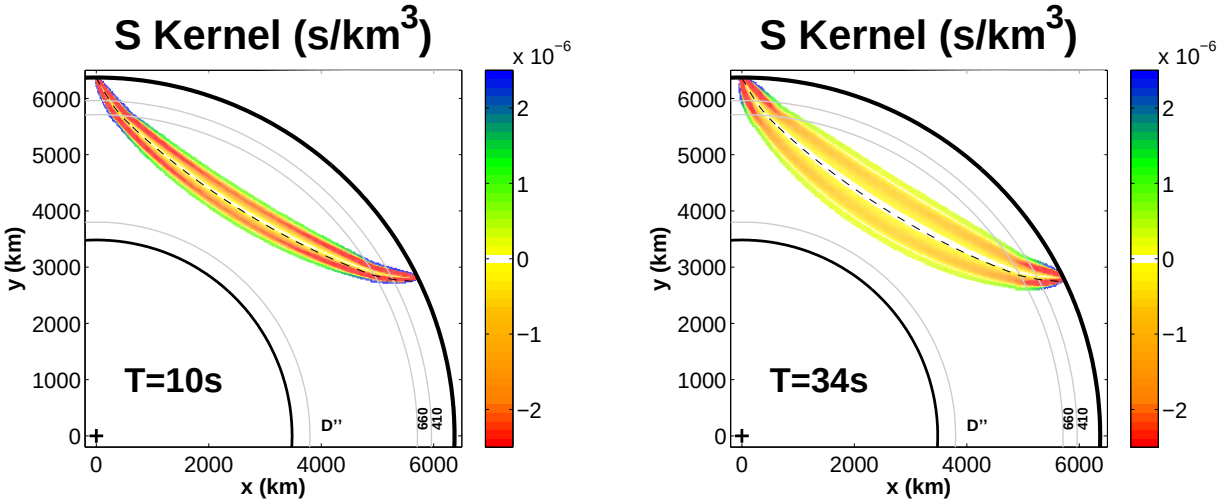


Figure 3.11: S phase Fréchet kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 0 km; epicentral distance: 64.5°. Black dashed line: geometrical ray path.

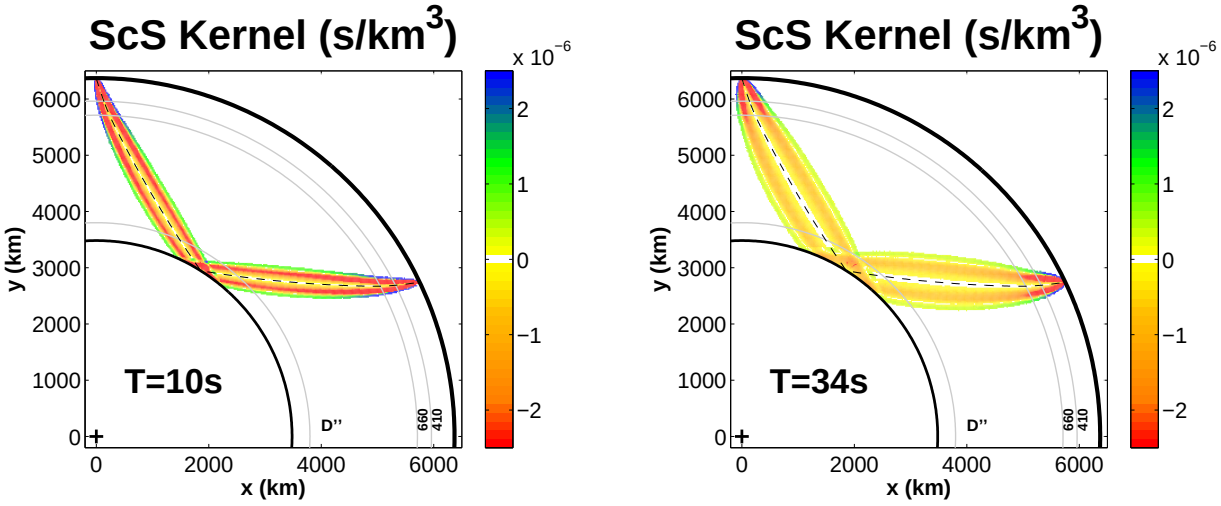


Figure 3.12: ScS phase Fréchet kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 0 km; epicentral distance: 64.5°. Black dashed line: geometrical ray path.

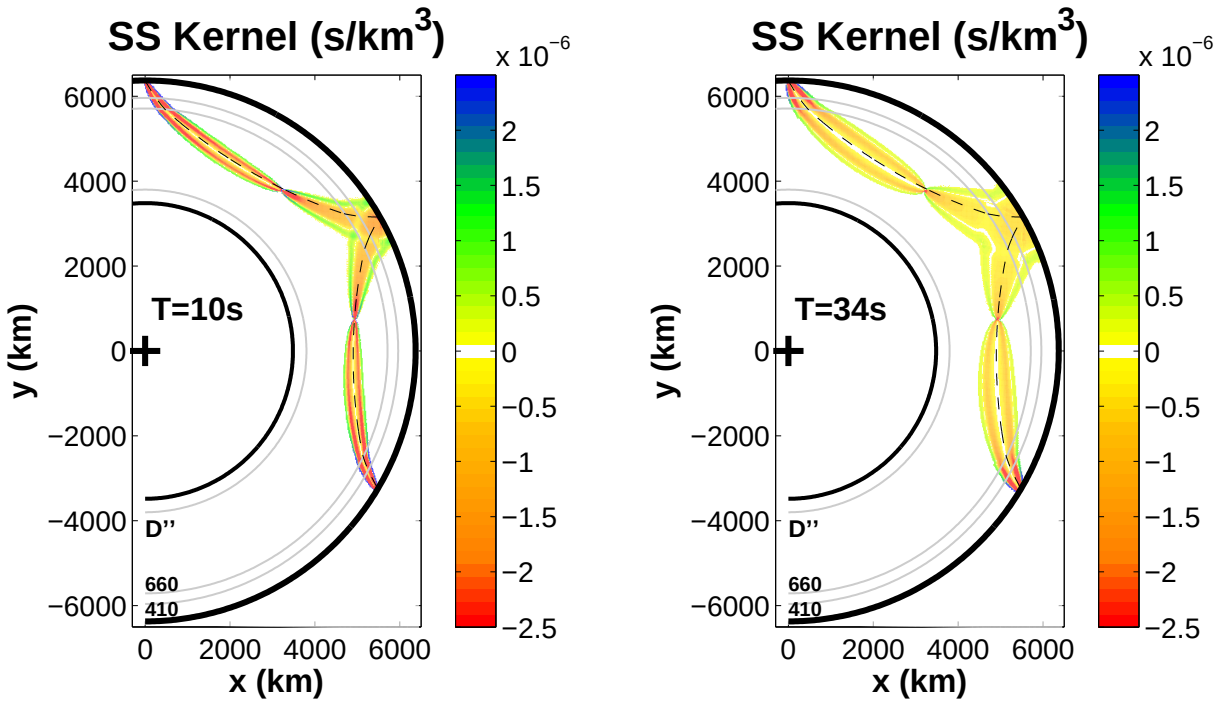


Figure 3.13: SS phase Fréchet kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 0 km; epicentral distance: 120.6°. Black dashed line: geometrical ray path.

3.3.2 Two-phase interference kernels

There are numerous instances in which there may be more than one seismic phase arriving within the cross-correlation time window. For example, in the case of a source at shallow depth, the seismic waveforms of the "direct" *S*, *ScS* and *SS* arrivals are contaminated by a superposition of later-arriving phases, such as *sS*, *sScS* and *sSS*, respectively. Since a majority of shallow earthquakes (source depth ∼15 km) have been used to build our global dataset, we need to use "two-phase interference kernels" to invert for measurements of interfering waveforms, such as *S*+*sS*, *ScS*+*sScS* and *SS*+*sSS*. Dahlen *et. al* (2000) give theoretical expressions of "finite-frequency" sensitivity kernels for such interfering or overlapping phases. In this section, we derive analytical expressions for two-phase interference kernels, in the presence (*SS*+*sSS*) or not (*S*+*sS* and *ScS*+*sScS*) of caustics.

Here, we only consider the interference of direct (e.g. *S*) and depth (e.g. *sS*) phases. Thus, the corresponding source depth is quite shallow, and we can consider that the direct and depth phases have very similar ray paths. Therefore, one is justified in making the

following assumption (Hung *et al.* 2000):

$$\begin{cases} A_{direct} \sim A_{depth} \\ R_{direct} \sim R_{depth} \end{cases} \tag{3.43}$$

where A is the amplitude factor and R is the geometrical spreading factor. The travel time sensitivity kernel, associated to the interference of direct and depth phases, may then be expressed as (e.g. Dahlen *et al.* 2000; Hung *et al.* 2000):

$$K_{1+2}(\mathbf{r}_x) = -\frac{1}{2\pi c(\mathbf{r}_x)} \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 [\cdots]_{1+2} d\omega}{D_{1+2}} \tag{3.44}$$

where '1' and '2' stand for the direct and depth phases, respectively, and with

$$\begin{cases} [\cdots]_{1+2} = [\cdots]_{11} + [\cdots]_{12} + [\cdots]_{21} + [\cdots]_{22} \\ [\cdots]_{11} = \sqrt{|\det(\mathbf{H}_1)|} \sin[\omega \Delta T_1 - \Delta\Phi] \\ [\cdots]_{12} = \sqrt{|\det(\mathbf{H}_1)|} \sin[\omega(\Delta T_1 + t_1 - t_2) - \Delta\Phi] \\ [\cdots]_{21} = \sqrt{|\det(\mathbf{H}_2)|} \sin[\omega(\Delta T_2 + t_2 - t_1) - \Delta\Phi] \\ [\cdots]_{22} = \sqrt{|\det(\mathbf{H}_2)|} \sin[\omega \Delta T_2 - \Delta\Phi] \\ D_{1+2} = \int_0^\infty 2\{1 + \cos[\omega(t_1 - t_2)]\} \omega^2 |\dot{m}(\omega)|^2 d\omega, \end{cases} \tag{3.45}$$

where t_i, ΔT_i, and $\mathbf{H}_i$ are the predicted arrival times, detour times, and Hessian matrices of the direct (i=1) and depth (i=2) phases, respectively; $\Delta\Phi$ represents the phase shift due to passage through caustics.

3.3.2.a Calculating $N_{1+2}(\Delta\Phi)$

Here, one aims to calculate $N_{1+2}(\Delta\Phi) = \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 [\cdots]_{1+2} d\omega$, in the presence ($\Delta\Phi = -\pi/2$) or not ($\Delta\Phi = 0$) of caustics - cf. equation (3.45).

1. $\underline{\Delta\Phi = 0}$
   We have:

   $$N_{1+2}(0) = \sqrt{|\det(\mathbf{H}_1)|}\{I(\Delta T_1) + I(\Delta T_1 + t_1 - t_2)\} + \sqrt{|\det(\mathbf{H}_2)|}\{I(\Delta T_2) + I(\Delta T_2 + t_2 - t_1)\} \tag{3.46}$$

   with $I(\alpha) = \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin(\alpha\omega) d\omega$. We have already calculated the integral I - equation (3.38). An analytical expression of $N_{1+2}(0)$ can therefore be derived.

2. $\underline{\Delta\Phi = -\pi/2}$
   As $\sin(\alpha\omega + \pi/2) = cos(\alpha\omega)$, we have:

$$N_{1+2}(-\frac{\pi}{2}) = \sqrt{|\det(\mathbf{H}_1)|}\{J(\Delta T_1)+J(\Delta T_1+t_1-t_2)\}+\sqrt{|\det(\mathbf{H}_2)|}\{J(\Delta T_2)+J(\Delta T_2+t_2-t_1)\} \tag{3.47}$$

   with $J(\alpha) = \int_0^\infty \omega^3|\dot{m}(\omega)|^2\cos(\alpha\omega)d\omega$. We have already calculated the integral J - equation (3.40). An analytical expression of $N_{1+2}(-\pi/2)$ can therefore be derived.

3.3.2.b Calculating D_{1+2}

Here, one aims to calculate $D_{1+2} = \int_0^\infty 2\{1+\cos[\omega(t_1-t_2)]\}\omega^2|\dot{m}(\omega)|^2 d\omega$. We can write

$$\begin{cases} D_{1+2} = 2\{d+\gamma\} \\ d = \int_0^\infty \omega^2|\dot{m}(\omega)|^2 d\omega \\ \gamma = \int_0^\infty \omega^2|\dot{m}(\omega)|^2\cos[\omega(t_1-t_2)]d\omega. \end{cases} \tag{3.48}$$

We have already calculated the integral d - equation (3.42). We still need to calculate the integral γ, which can be written as

$$\begin{cases} \gamma(a) = \int_0^\infty \omega^2|\dot{m}(\omega)|^2\cos(a\omega)d\omega \\ a = t_1 - t_2. \end{cases} \tag{3.49}$$

Replacing the source power spectrum, $|\dot{m}(\omega)|^2$, by equation (3.36), we obtain:

$$\begin{cases} \gamma(a) = \frac{T^2}{2\pi}\int_0^\infty \omega^4 e^{-(b\omega)^2}\cos(a\omega)d\omega \\ b = \frac{T}{2\pi} \end{cases} \tag{3.50}$$

where T is the dominant wave period (which is the same for both the direct and depth phases). In appendix A.4, we show how to calculate the integral $\int_0^\infty \omega^4 e^{-(b\omega)^2}\cos(a\omega)d\omega$. This leads to:

$$\gamma(a) = \frac{T^2}{2\pi 2^5(T/2\pi)^9}e^{-(a\pi/T)^2}\sqrt{\pi}(a^4 - 12(T/2\pi)^2a^2 + 12(T/2\pi)^4). \tag{3.51}$$

An analytical expression of D_{1+2} can therefore be derived:

$$D_{1+2} = 2\{3\sqrt{\pi}2^2(\pi/T)^5T^2/2\pi + \gamma(t_1-t_2)\}. \tag{3.52}$$

Finally, 2-D views of two-phase interference *S+sS*, *ScS+sScS* and *SS+sSS* "finite-frequency" travel time sensitivity kernels, analytically calculated at 10 and 34 s periods, are shown at figures 3.14, 3.15 and 3.16, respectively.

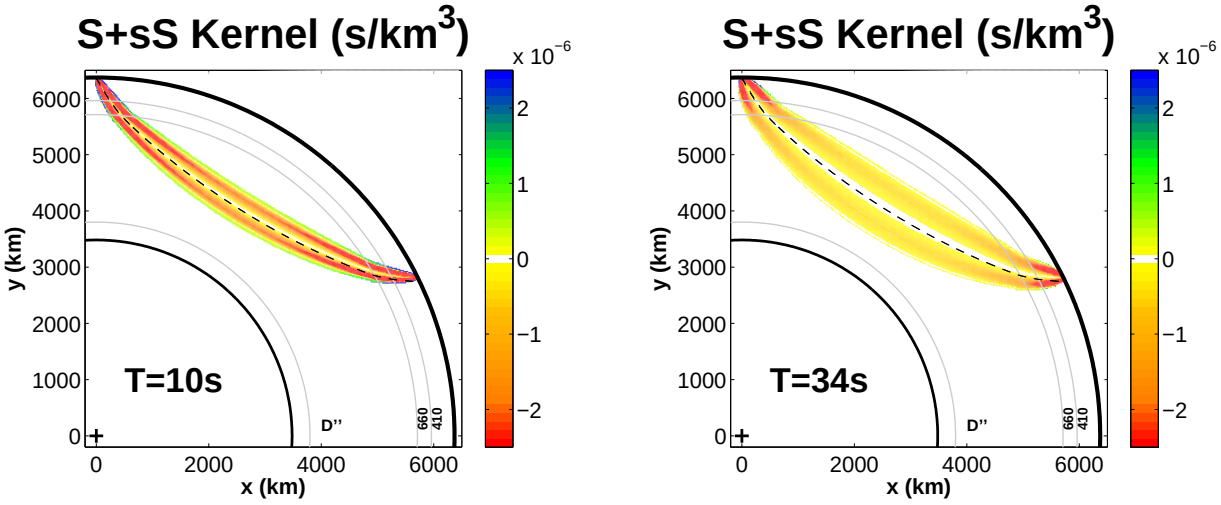


Figure 3.14: S+sS two-phase interference kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 15 km; epicentral distance: 64.5°. Black dashed line: geometrical ray path.

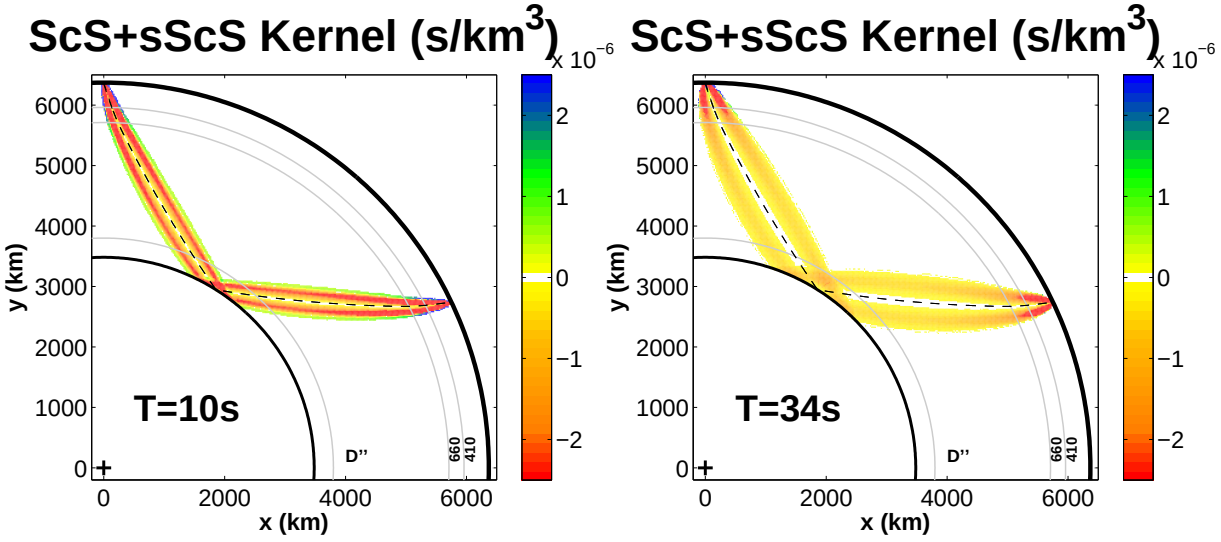


Figure 3.15: ScS+sScS two-phase interference kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 15 km; epicentral distance: 64.5°. Black dashed line: geometrical ray path.

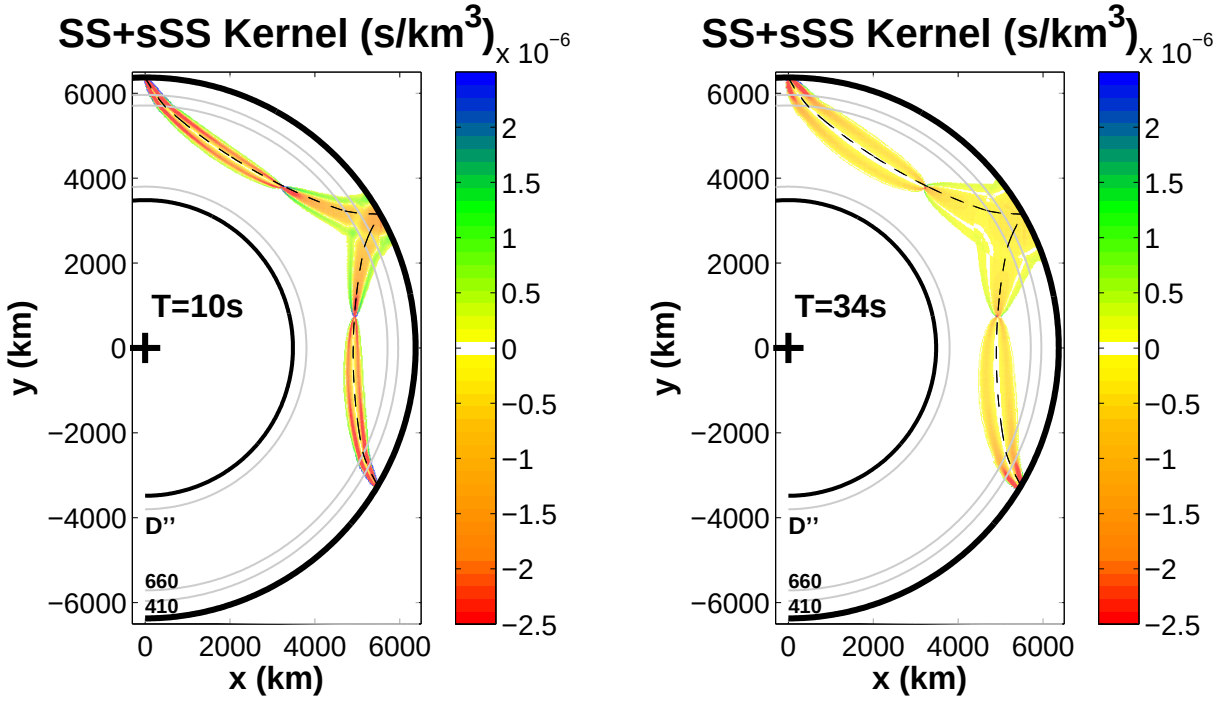


Figure 3.16: SS+sSS two-phase interference kernel at 10 and 34 s period. Source/receiver is on the left/right side of each frame. Source depth: 15 km; epicentral distance: 120.6°. Black dashed line: geometrical ray path.

3.3.3 Compound rays

In the paraxial approximation, we consider only forward scattering on nearby, like-type source-to-scatterer-to-receiver ray path. This leads to paraxial (Fréchet) kernels, such as defined in sections 3.3.1 and 3.3.2. However, in the case of a compound ray, such as *sS*, *ScS* or *SS*, one scatterer x can be associated to more than one perpendicular projection point ξ (on the geometrical ray path), as illustrated in figure 3.17. Therefore, the paraxial kernel of a compound ray should be calculated as a *sum* over all central ray points ξ (e.g. Hung *et. al* 2000).

Here, we briefly explain how we have implemented such a *sum* in our codes for computing paraxial kernels. Firstly, we define a kernel grid of closely spaced points $\mathbf{r}_p$ (cf. section 3.4.2). Secondly, we find the reflection point(s), at the surface or the CMB, of each compound ray. There may be one (e.g. for sS, ScS, SS) or several (e.g. for sScS, sSS) reflection points. These reflection points divide the ray path in several parts: 2 parts for sS, ScS and SS ray paths (as there is one reflection point); 3 parts for sScS and sSS ray paths

(as there are two reflection points). We then calculate the kernel contribution for each part of the ray path on the kernel grid. Finally, we sum all the contributions, calculated for each cell of the kernel grid.

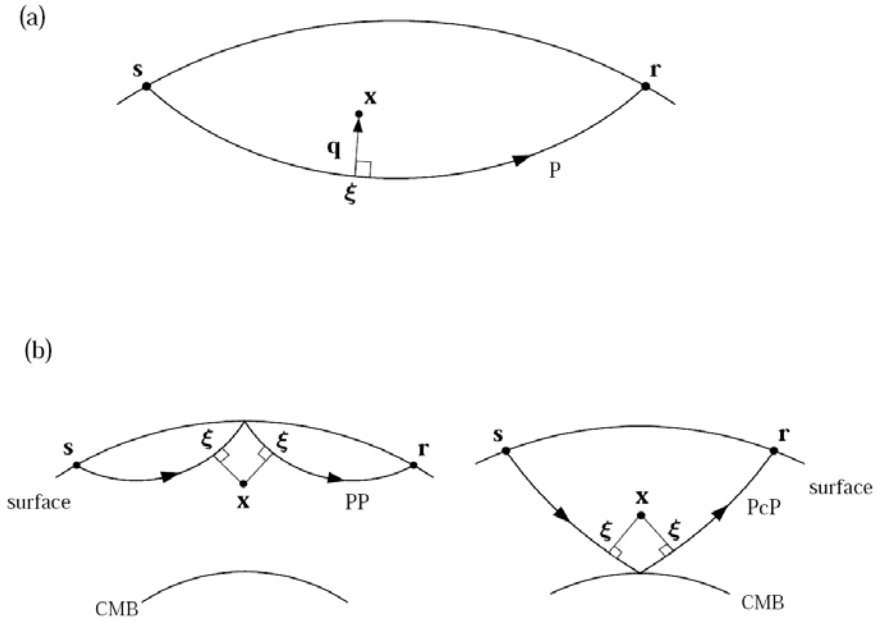


Figure 3.17: ***(a)*** *In the paraxial approximation, every scatterer* $\boldsymbol{x}$ *is perpendicularly projected onto the nearest point* ξ *on the central geometrical ray from the source* $\boldsymbol{s}$ *to the receiver* $\boldsymbol{r}$*. The difference vector is* $\boldsymbol{q} = \boldsymbol{x} - \xi$*.* ***(b)*** *In many circumstances, such as sS, ScS, sScS, SS and sSS waves in this study, a scatterer can be projected onto more than one paraxial point* ξ*. (Left) A scatterer in the vicinity of the surface reflection point of a SS (or PP as shown here) wave. (Right) A scatterer in the vicinity of the Core-Mantle Boundary (CMB) reflection point of a ScS (or PcP as shown here) wave. From Hung et. al (2000).*

3.3.4 Near discontinuous interfaces

For scatterers near the surface, one should take into account incoming rays that hit the scatterer directly, as well as those that first visit a boundary, as the surface or the CMB. Both may have a detour time small enough to allow the scattered wave to arrive in the cross-correlation time window. That is, for reflected phases, such as *sS*, *sScS*, *sSS*, *ScS* and *SS*, an heterogeneity near the reflection point will contribute more than one scattered wave. The situation is depicted in figure 3.18 for *SS*, where a scatterer near the Earth's surface produces two forward scattered waves (we neglect the backward scattering).

Therefore, we need to mirror scatterers near reflective boundaries. Following Tian

et. al (2007b), we explain the "mirroring" procedure, in the case of an *SS* wave (cf. figure 3.18). The same analysis works for core reflections (e.g. *ScS*).

Consider the first scattered wave (cf. figure 3.18a). The scatterer $\mathbf{x}$ sits on its incoming leg. The detour time ΔT_x is determined by the Hessian matrix H_a and by the scatterer's displacement $\mathbf{q}_a$, from the incoming leg l_i of the geometrical ray. Thus, we have $\Delta T_x = \frac{1}{2}\mathbf{q}_a^t H_a \mathbf{q}_a$.

Now consider the second scattered wave (cf. figure 3.18b). The scatterer $\mathbf{x}$ sits on its outgoing leg. The detour time ΔT_x is determined by H_b evaluated at the projection of $\mathbf{x}$ onto the outgoing leg l_o of the geometrical ray, and by $\mathbf{q}_b$, the scatterer's displacement from l_o. Though the projection of $\mathbf{x}$ is out of the Earth, we take the mirror image **x'** of $\mathbf{x}$ about the Earth's surface. Thus, we have $\Delta T_x = \Delta T_{x'}$, because $\mathbf{q}_b \equiv \mathbf{q}_b^{'}$. That is, we use the fact that $\mathbf{x}$ and its mirror image **x'** have the same detour time and geometrical spreading.

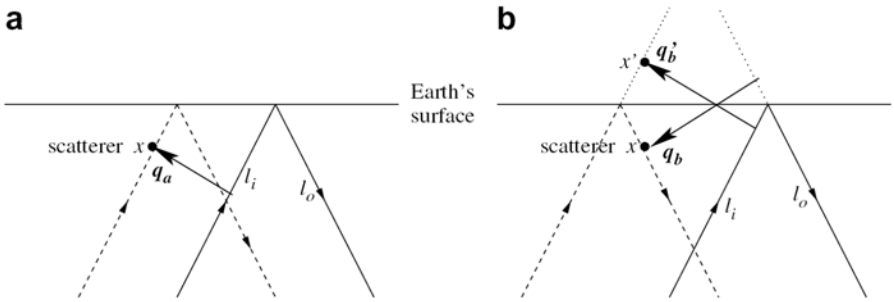


*Figure 3.18: The 'mirroring' procedure for scatterers near reflective boundaries, in the case of an SS wave. Thin solid arrows represent geometrical rays, with l_i denoting the incoming leg, and l_o the outgoing leg. Dashed arrows represent scattered rays. Dotted lines extend rays out of the reflective boundary. The vectors $\boldsymbol{q}_a$, $\boldsymbol{q}_b$ and $\boldsymbol{q}_b^{'}$ are off ray vectors in ray centered coordinates (cf. figure 3.9). Under the paraxial approximation, there are two scattered waves from the same scatterer x near the surface reflection point: (**a**) x on the incoming leg of the scattered wave; (**b**) x on the outgoing leg of the scattered wave. The scatterer x and its mirror image x' about the reflective boundary have the same detour time and geometrical spreading. From Tian et. al (2007b).*

3.3.5 Singularities at sources, receivers and caustics

At sources, receivers and caustics (e.g. of SS), the Hessian matrix H - equation (3.32) - becomes singular, and so do sensitivity kernels. To avoid this, following Tian *et al.* (2007b), we isolate a small sphere V_0 around a singular point and apply ray theory in

this sphere. The delay time due to passage through the sphere is then (cf. section 1.2 of chapter 1):

$$\delta t = \int_{l\in V_0} \frac{dl}{c+\delta c} - \int_{l\in V_0} \frac{dl}{c} \simeq \int_{l\in V_0} -c^{-1}\frac{\delta c}{c} dl \tag{3.53}$$

where c is the background velocity and the line integral is along the ray path in the sphere. The "ray-theoretical travel time kernel" is then $K = -c^{-1}$.

3.3.6 Paraxial kernel: computational aspects

In the formalism of Dahlen *et al.* (2000), "finite-frequency" kernels are calculated under the paraxial approximation, which assumes that scatterers influencing the data are located close to the ray, within the first few Fresnel zones. Here, one aims to determine the maximum detour time, ΔT_m, for which the paraxial kernel becomes negligible.

The detour time of a scattered wave is $\Delta T = \frac{1}{2}(H_{11}q_1^2 + H_{22}q_2^2)$ - cf. equation (3.34). The maximum detour time, ΔT_m, may be defined as (Tian *et al.* 2007b)

$$\Delta T_m = \min(\Delta T_{Xcorr}, \Delta T_{lim}), \tag{3.54}$$

where:

1. ΔT_{Xcorr} is the maximum length of the cross-correlation time window (used to measure time residuals).

2. ΔT_{lim} is the detour time for which the kernel becomes negligible.

Let ρ be the radial distance of the scatterer from the geometrical ray. We have $\rho = \sqrt{q_1^2 + q_2^2}$. The largest radial distance, ρ_m, of scatterers for which the paraxial kernel should be calculated is then defined as (Tian *et al.* 2007b):

$$\rho_m = \min(\sqrt{2\Delta T_m/|min(H_{11}, H_{22})|}, \rho_{lim}). \tag{3.55}$$

It is worth noting that ρ_m varies along the ray. The limiting radial distance, ρ_{lim}, avoids the extension of the kernel calculation beyond the limits where the paraxial approximation is valid. Following the numerical study of Tian *et al.* (2007b), on the accuracy of the paraxial approximation, we adopted ρ_{lim}=1000 km for calculating all the kernels.

Now, we need to determine the limiting detour time, ΔT_{lim}, for which the paraxial kernel becomes negligible. That is: "How many Fresnel zones should we take into account in the computation of paraxial kernels?".

Figure 3.19 shows the paraxial sensitivity kernel of an S wave, at T=34 s period, calculated for different values of the limiting detour time, ΔT_{lim}. We see that:

1. The kernel calculated for detour times lower than $\Delta T_{lim} = T/6$ only includes the first Fresnel zone (figure 3.19a).

2. The kernel calculated for detour times lower than $\Delta T_{lim} = T/4$ includes the first and second Fresnel zones (figure 3.19b).

3. The kernel calculated for detour times lower than $\Delta T_{lim} = T/2$ includes the first, second and third Fresnel zones (figure 3.19c).

It is clear that the kernel sensitivity becomes negligible beyond the second Fresnel zone (cf. figures 3.19b and 3.19c). For this reason, we use the criteria $\Delta T_{lim} = T/4$ for computing all the kernels, which is consistent with the paraxial approximation. This choice neglects third and higher order Fresnel zones.

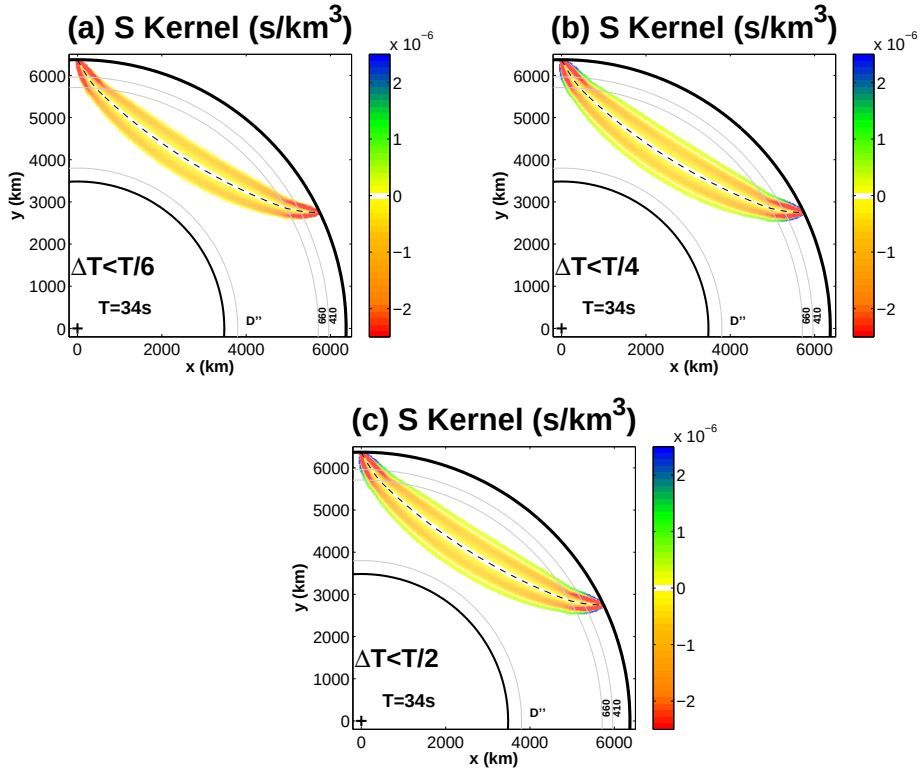


Figure 3.19: S phase kernel including: a)1st order Fresnel zone only, corresponding to $\Delta T \leq T/6$; b)1st and 2nd order Fresnel zones, corresponding to $\Delta T \leq T/4$; c)1st, 2nd and 3rd order Fresnel zones, corresponding to $\Delta T \leq T/2$.

3.3.7 3-D characteristics of S, ScS and SS kernels

Here, we present 3-D views of single phase *S*, *ScS*, and *SS* travel time sensitivity kernels, calculated with the "finite-frequency" formalism of Dahlen et al. (2000). We aim to illustrate several interesting features of these "banana-doughnut" kernels.

First, the typical "banana-dougnut" character of the kernels is apparent. For instance, we see that on the geometrical ray itself (i.e. the dashed line in figures 3.20, 3.21 and 3.22), the kernel sensitivity is null. The "doughnut hole" is visible along almost the entire ray path for *S* and *ScS* waves (cf. figures 3.20 and 3.22). Concerning the *SS* wave, the hole disappears in between the two caustics, located at $\Delta/3$ and $2\Delta/3$, with Δ the total epicentral distance (cf. figure 3.22).

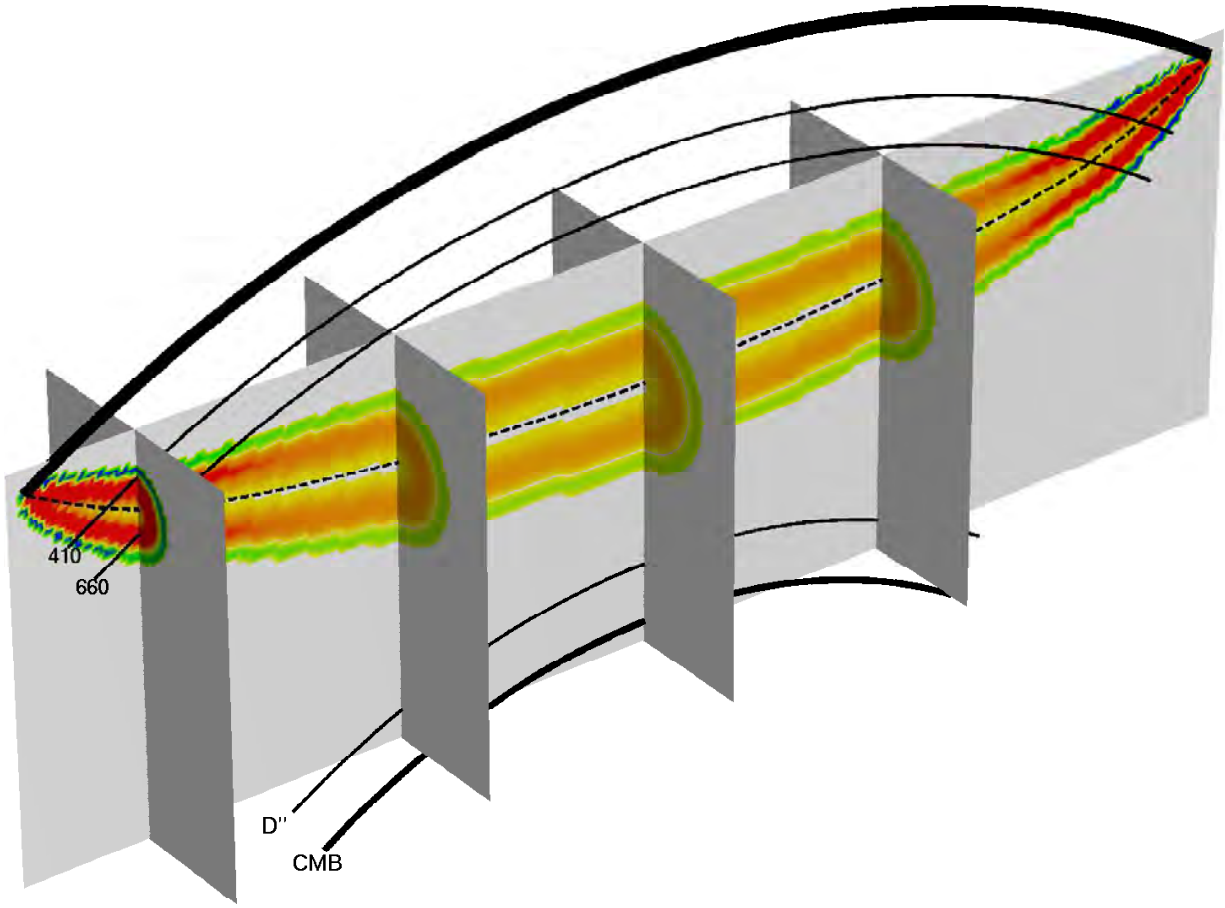


Figure 3.20: 3D view of S phase Fréchet kernel at 20 s period. Source/receiver is on the left/right side. Source depth: 0 km; epicentral distance: 64.5°. Same color scale as in figure 3.11 (i.e. $\pm 2.5 \cdot 10^{-6}\ s \cdot km^{-3}$).

For a typical S wave, we see that the finite-frequency effect is strongest at the center of the ray (cf. 3.20). That is, the doughnut is largest where the ray is deepest. Closer to the source or receiver, the doughnut is smaller but the sensitivity per km^3 is higher.

We also observe that the sensitivity is negative in the first Fresnel zone (e.g. figure 3.20). This is understandable[9]: an increase in the velocity perturbation, $\delta c/c$, decreases the time residual, δt. The first Fresnel zone is also surrounded by a second Fresnel zone with weaker but positive sensitivity (e.g. figure 3.20). The sensitivity becomes negligible beyond the second Fresnel zone.

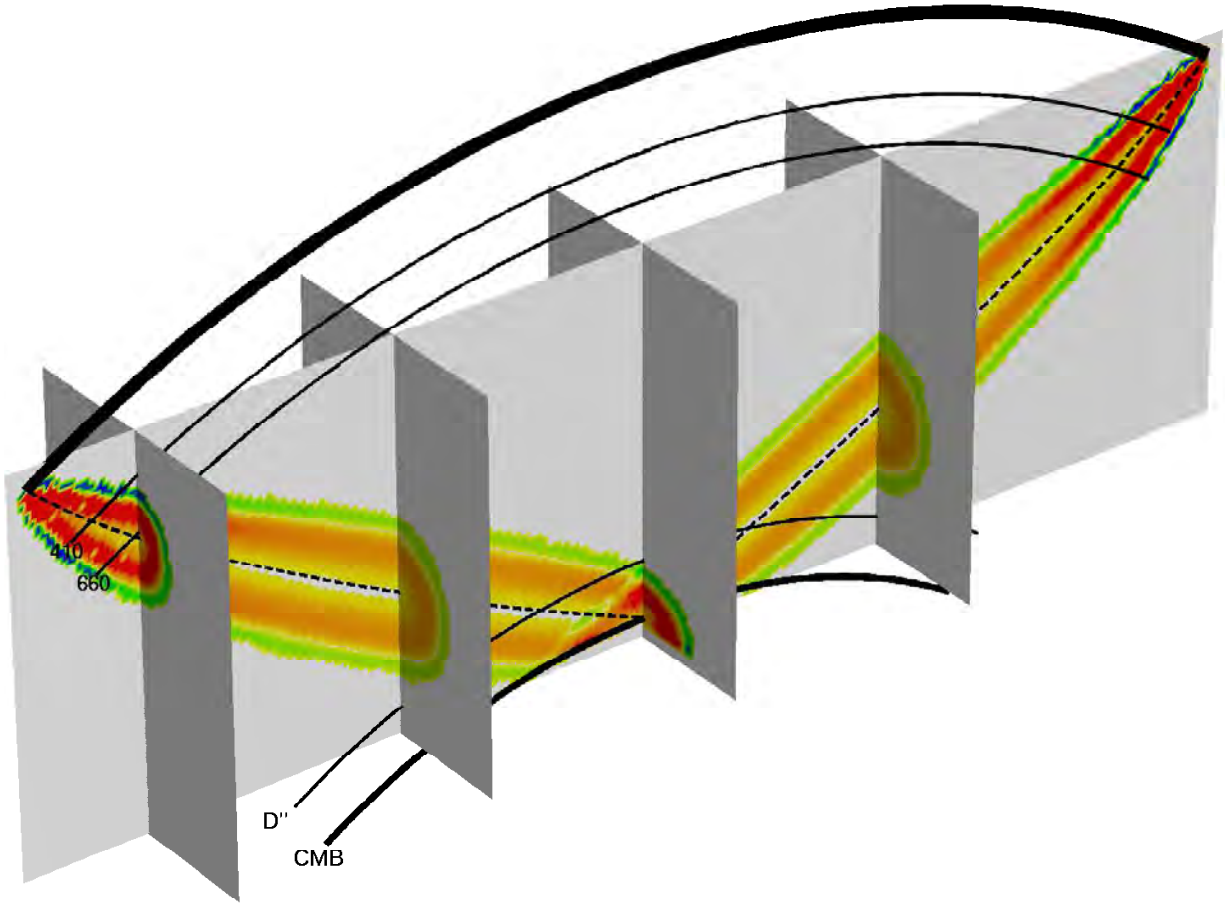


Figure 3.21: 3D view of ScS phase Fréchet kernel at 20 s period. Source/receiver is on the left/right side. Source depth: 0 km; epicentral distance: 64.5°. Same color scale as in figure 3.12 (i.e. $\pm 2.5 \cdot 10^{-6}\ s \cdot km^{-3}$).

[9]We have $\delta t \propto 1/\delta c \Rightarrow$ if δc increases, then δt decreases. Therefore, as $\delta t \equiv K \delta c$, K has to be negative, at least in the first Fresnel zone.

The particular pattern of the *SS* kernel, in between the two caustics, is better observed in 3-D. The *SS* wave sensitivity, from the source, or receiver, to the nearest caustic[10], has a similar elliptical pattern as the *S* and *ScS* wave sensitivity. However, the *SS* sensitivity between the two caustics has a more complex hyperbolic pattern, which is a consequence of the minimax nature of the *SS* wave (e.g. Choy & Richards 1975).

Finally, 3-D kernel views also show the effect of scatterers near reflective boundaries, such as the Earth's surface or the CMB, on the sensitivity pattern of *ScS* and *SS* waves (cf. figures 3.21 and 3.22). For instance, the *ScS* kernel has a more complicated pattern near the CMB, compared to elsewhere in the mantle (see figure 3.21), owing to "mirrored" scatterers near the CMB.

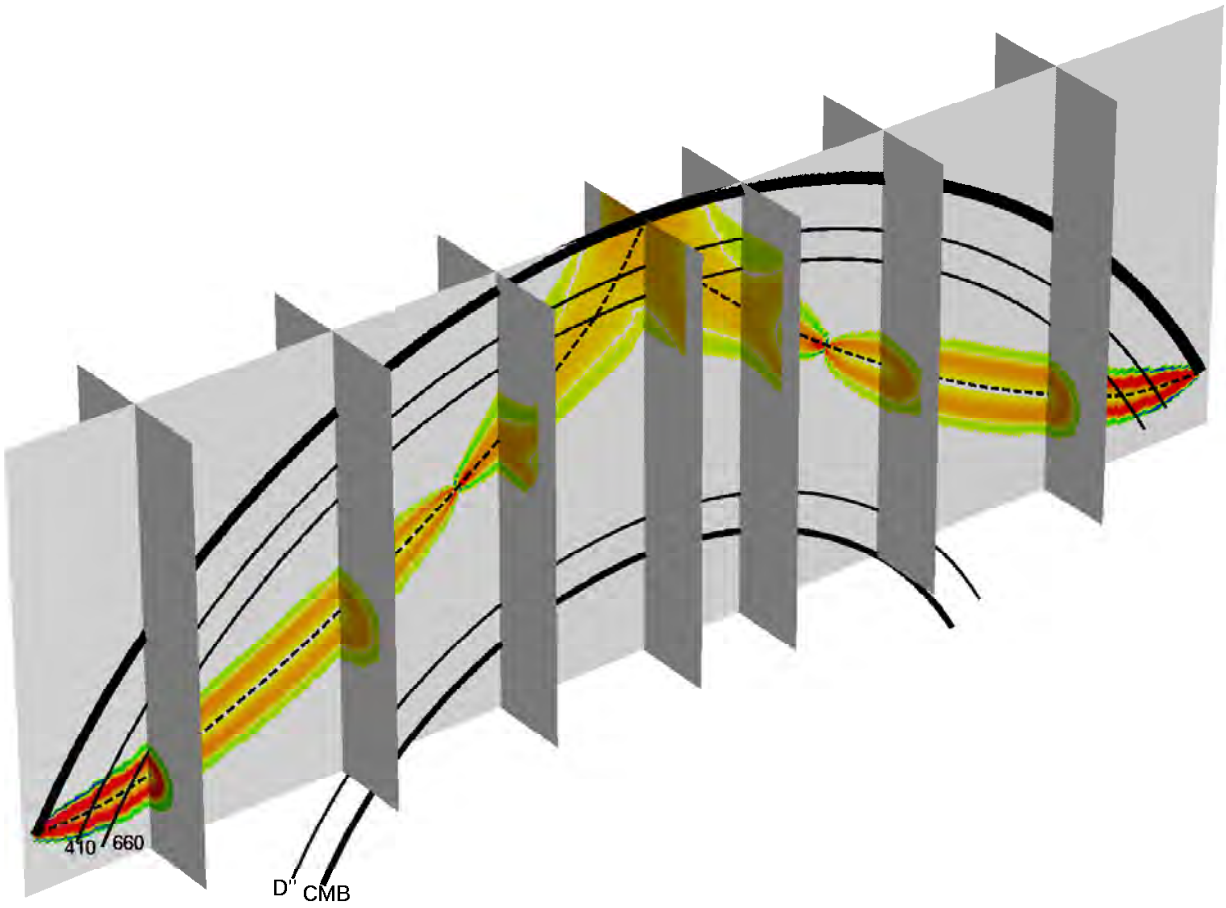


Figure 3.22: 3D view of SS phase Fréchet kernel at 20 s period. Source/receiver is on the left/right side. Source depth: 0 km; epicentral distance: 120.6°. Same color scale as in figure 3.13 (i.e. $\pm 2.5 \cdot 10^{-6}\ s \cdot km^{-3}$).

[10]In figure 3.22, the two caustics of the *SS* wave are at the epicentral distances $\phi = 40°$ and $\phi = 80°$.

3.4 Constructing the sensitivity matrix G_{ij}

In multiple-frequency travel time tomography, the general form of the inverse problem is $\delta t_i(T) = \int_{V_i(T)} K_i(\mathbf{r};T)m(\mathbf{r})d^3\mathbf{r}$ - equation (3.2). As this is a "linear" inverse problem, it may be formulated as $\mathbf{d} = G \cdot \mathbf{m}$ - equation (3.3).

In this section, we aim to build the G matrix, which represents the projection of the "banana-doughnut" sensitivity kernels (calculated in section 3.3) onto the model parameterization (developed in section 3.2).

3.4.1 Discretizing the inverse problem

In "finite-frequency" tomography, the general form of the linear inverse problem is

$$d_i = \int_V K_i(\mathbf{r})m(\mathbf{r})d^3\mathbf{r} \tag{3.56}$$

where the i-th datum is, in our case, a (frequency-dependent) time residual - cf. equation (3.2). This continuous system - equation (3.56) - can be discretized as (e.g. Tian *et al.* 2007b):

$$\begin{cases} m(\mathbf{r}) = \sum_j m_j b_j(\mathbf{r}) \\ d_i = \sum_j G_{ij} m_j \\ G_{ij} = \int_V K_i(\mathbf{r}) b_j(\mathbf{r}) d^3\mathbf{r}, \end{cases} \tag{3.57}$$

where the functions $b_j(\mathbf{r})$ represent the model parameterization. In this thesis, they are linear interpolation functions, between a grid of nodes (cf. section 3.2.4).

We compute the integral in equation (3.57) by a simple Riemann sum. That is, we split the volume V into volumes ΔV_p, much smaller than the spherical triangular prisms used for the model parameterization, around a regularly spaced grid of points $\mathbf{r}_p$, called the "kernel grid":

$$d_i = \sum_p K_i(\mathbf{r}_p)m(\mathbf{r}_p)\Delta V_p. \tag{3.58}$$

To relate $m(\mathbf{r}_p)$ to the parameter values m_j at model nodes, we find the spherical triangular prism t_p that encloses $\mathbf{r}_p$ (cf. appendix C), and interpolate between the three vertices (i.e.

nodes) of its top spherical triangle (cf. section 3.2.4):

$$m(\mathbf{r}_p) = \sum_{k=1}^{3} b_k^{t_p}(\mathbf{r}_p) m_k^{t_p}. \tag{3.59}$$

Thus, the equation (3.58) becomes:

$$d_i = \sum_p K_i(\mathbf{r}_p)[\sum_{k=1}^{3} b_k^{t_p}(\mathbf{r}_p) m_k^{t_p}]\Delta V_p = \sum_p \sum_{k=1}^{3} [K_i(\mathbf{r}_p) b_k^{t_p}(\mathbf{r}_p)\Delta V_p] m_k^{t_p}. \tag{3.60}$$

To construct the sensitivity matrix G_{ij}, in equation (3.57), each spherical triangle vertex, $m_k^{t_p}$, is mapped back to its original parameter index, m_j, so that each volume element ΔV_p contributes to three elements of row i of matrix G_{ij}:

$$G_{ij} \leftarrow G_{ij} + K_i(\mathbf{r}_p) b_k^{t_p}(\mathbf{r}_p)\Delta V_p \ , \quad k = 1, 2, 3. \tag{3.61}$$

Each element G_{ij} is initially set to zero and is "updated" whenever there is contribution from the integral volume element ΔV_p, according to equation (3.61). Therefore, each row i of the G matrix represents the projection of the individual kernel K_i onto the model parameterization.

At sources, receivers and caustics, sensitivity kernels become singular. We then use[11] ray theory in a small sphere V_0 around each singular point (cf. section 3.3.5). In this case, the equations (3.53) leads to:

$$G_{ij} \leftarrow G_{ij} - c^{-1}(\mathbf{r}_l) b_k^{t_l}(\mathbf{r}_l) dl \ , \quad k = 1, 2, 3 \tag{3.62}$$

where $\mathbf{r}_l$ represents the point on the ray in sphere V_0, which is enclosed in the spherical triangular prism t_l.

[11] We have made an approximation in the equation (3.62), which is used to handle the kernel singularities. This approximation concerns the evaluation of dl, which represents the length of the i-*th* geometrical ray path in the sphere V_0. First, kernels are computed on a very fine grid (cf. section 3.4.2), which consists of small size cubic cells (20 km × 20 km × 20 km). For simplicity, we have approximated the sphere V_0 by such a cubic cell. Therefore, dl should represent the length of the ray path in the cubic cell. In fact, dl has been also approximated to the edge length of the cubic cell, that is 20 km, for simplicity reasons. However, the resulting error is of second order for the G_{ij} value. It is then unlikely to (significantly) modify our tomographic results. We plan to implement, in our codes, an algorithm to compute dl more accurately.

3.4.2 Kernel grid

We need to calculate the kernel values, $K_i(\mathbf{r}_p)$, on a regularly spaced grid of points, $\mathbf{r}_p$, called the "kernel grid" - cf. equation (3.58). This grid must be much smaller than the model parameterization[12], which is called the "model grid".

Here, we address the problem of choosing an "optimal" kernel grid. That is, we should consider that:

1. The kernel grid points, $\mathbf{r}_p$, cannot be too largely spaced, because the kernel would be down-sampled, and important features of the wave sensitivity would be no more taken into account (e.g. the doughnut hole).

2. The kernel grid points, $\mathbf{r}_p$, cannot be too closely spaced, because the kernel would be over-sampled, which is not necessary, and may also become prohibitive in term of computing time.

It is our experience that a good choice for the kernel grid is to take a regular cartesian grid of points, $\mathbf{r}_p$, spaced of ∼20 km along the X-, Y-, and Z-directions, as shown in figure 3.23. The integral volume element is then: $\Delta V_p \sim 20$ km^3.

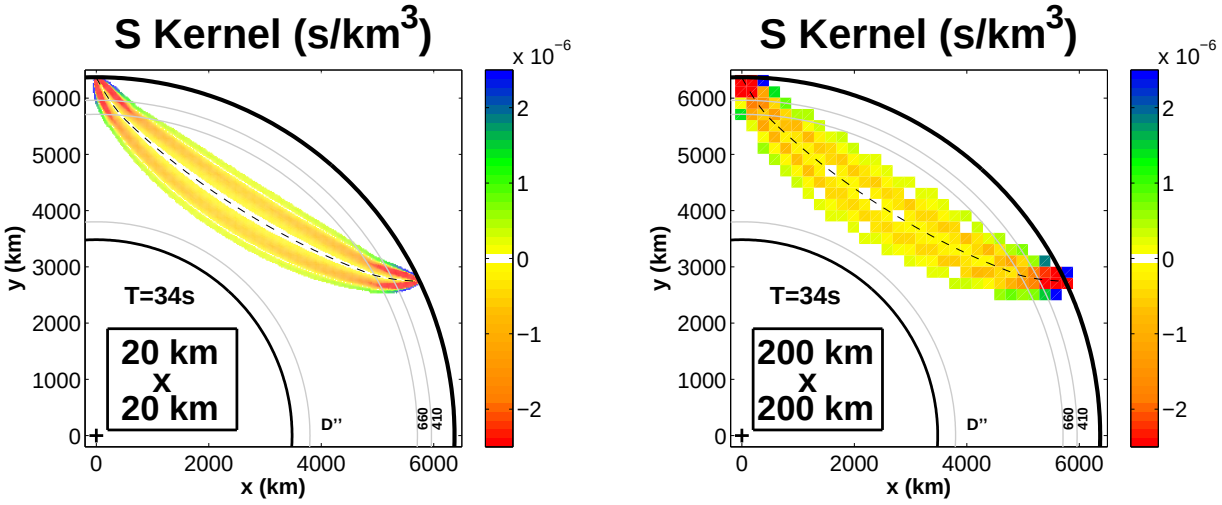


Figure 3.23: S phase Fréchet kernel (T=20 s period) calculated on different "kernel grid" of points, $\mathbf{r}_p$, regularly spaced of 20 km (on the left frame) and 200 km (on the right frame). Source/receiver is on the left/right side of each frame. Source depth: 0 km; epicentral distance: 64.5°.

[12]The size of the smallest/largest cell of the model grid is a spherical triangular prism with an edge length of ∼ 200/1000 km (laterally), and a thickness of ∼ 100/200 km (vertically), respectively

3.4.3 Kernel to model grid

The model grid is defined in the usual basis $F = (O, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, associated to the spherical Earth. $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are the unit vectors of F. We can define, for any spatial point $\mathbf{M}$ inside the Earth, its cartesian $(x, y, z)_F$ and spherical $(r, \theta, \phi)_F$ coordinates in F, where r is the radius, $\theta \in [0, +\pi]$ the co-latitude and $\phi \in [-\pi, +\pi]$ the longitude of $\mathbf{M}$. The formulas relating the spherical and cartesian coordinates are:

$$\begin{cases} x = r\sin(\theta)\cos(\phi) \\ y = r\sin(\theta)\sin(\phi) \\ z = r\cos(\theta) \end{cases} \tag{3.63}$$

and

$$\begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \theta = \arccos(z/r) \\ \phi = \text{atan2}(y, x). \end{cases} \tag{3.64}$$

At first glance, we may think that one should define the cartesian kernel grid, $\mathbf{r}_p$, in the same basis F. Projecting each kernel value, $K_i(\mathbf{r}_p)$, onto the model grid - cf. equation (3.61) - would then be straightforward, by simply using equation (3.64). Such a kernel grid should then contain the entire globe. The edge of the smallest cube containing the whole Earth is 6,371×2=12,742 km. Thus, this represents a kernel grid of ∼637×637×637=258,474,853 points $\mathbf{r}_p$ (if spacing is 20 km), with a large majority of zero kernel values, $K_i(\mathbf{r}_p)$ (corresponding to the unsampled part of the Earth by the kernel).

We have preffered to reduce the computer memory required to store the kernel grid, and to minimize the amount of zero kernel values, $K_i(\mathbf{r}_p)$. That is, we have chosen to define the kernel grid, $\mathbf{r}_p$, only in a rectangular region surrounding the 3-D kernel (cf. figure 3.24). This allows us to mainly store "non-zero" kernel values $K_i(\mathbf{r}_p)$ (corresponding to the sampled model by the kernel). The size of the new kernel grid is ∼50 times smaller than in the previous case. This "optimized" kernel grid, $\mathbf{r}_p$, is defined in the new basis $F' = (O, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, where $\mathbf{X}'$, $\mathbf{Y}'$ and $\mathbf{Z}'$ are unit vectors of F' (cf. figure 3.24).

We now aim to project each kernel value, defined in $F' = (O, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, onto the model grid, defined in $F = (O, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ (cf. figure 3.25). That is, if the coordinates of a point $\mathbf{r}_p$ in the kernel grid basis F' are $(x'_{r_p}, y'_{r_p}, z'_{r_p})^t_{F'}$, we need to find its corresponding coordinates $(x_{r_p}, y_{r_p}, z_{r_p})^t_F$ in the model grid F. We explain how to do it in appendix E.

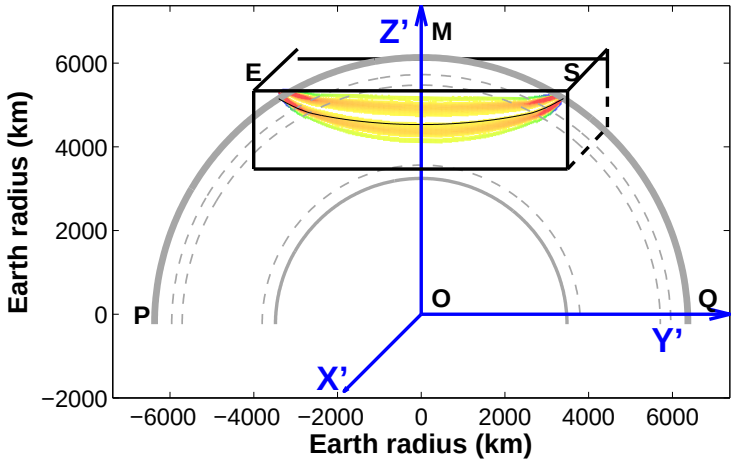


Figure 3.24: The kernel grid $\boldsymbol{r}_p$ is defined in the basis ($\boldsymbol{X}$',$\boldsymbol{Y}$',$\boldsymbol{Z}$'). This allows us to store in the computer memory the minimal number of points $\boldsymbol{r}_p$ that will be effectively calculated under the paraxial approximation, i.e. the rectangular region shown in black.

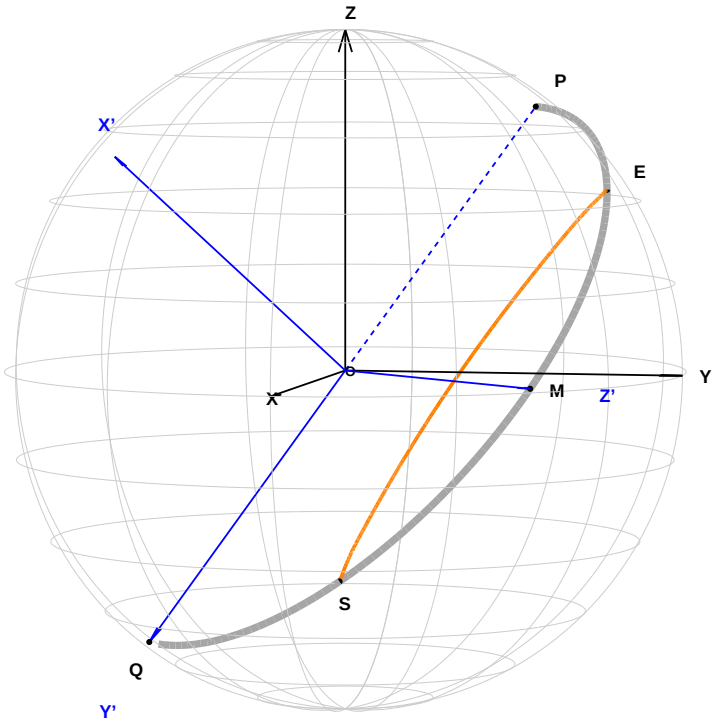


Figure 3.25: The model and kernel grids are defined in the basis $F = (O, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ and $F' = (O, \boldsymbol{X}', \boldsymbol{Y}', \boldsymbol{Z}')$, respectively. In orange is shown the geometrical ray path, between epicenter E and source S, corresponding to figure 3.24.

3.5 Conclusion

In this chapter, we have posed the inverse problem corresponding to multiple-frequency tomography - equation (3.2). We have used a "data driven" parameterization of the mantle, and "finite-frequency" sensitivity kernels, calculated with the formalism of Dahlen *et al.* (2000). We are now in the position to simultaneously invert our data, measured in different frequency bands, in a single inversion.

In chapter 4, we firstly present the 3-D *SH*-wave tomographic model of the Earth's mantle we have obtained, with respect to the radial (1-D) reference model IASP91 (Kennett & Engdahl 1991). Secondly, we aim to assess if, from a model point of view, multiple-frequency *S*-wave travel times do significantly provide additional constraints on the 3-D elastic structure of the mantle. That is, we try to quantify the benefits achieved, in our model, by inverting multi-band (measurements at 10, 15, 22.5 and 34 s periods) rather than single-band (measurements at 22.5 s period) data.

CHAPTER 4

Global multiple-frequency *SH*-wave tomography: inversion results

"Mantle:
the part of the interior of a terrestrial planet
and especially the earth that lies beneath
the crust and above the central core."
Webster dictionary

4.1 Introduction

In chapter 2, we have presented a globally distributed dataset of ∼400,000 frequency-dependent *S*, *ScS* and *SS* travel times, measured at several periods ranging from 10 to 51 s. We have shown that a residual "structural" dispersion is observed in our data. The use of this new observable, in seismic tomography, is expected to improve the imaging of seismic heterogeneities in the mantle.

In chapter 3, we have built the sensitivity matrix (G), which linearly relates our frequency-dependent data ($\mathbf{d}$) to the model parameters ($\mathbf{m}$), with the "finite-frequency" formalism of Dahlen *et al.* (2000). We are now able to invert our multiple-frequency travel times, and to obtain tomographic images of the Earth's mantle.

In this chapter 4, we first present details of the inversion (section 4.2). We then briefly discuss the general patterns of the resulting (3-D) *SH*-wave velocity perturbation tomographic model of the mantle, extending from the surface to the CMB (section 4.3). Finally, we present a comparison of single- versus multi-band tomographic models (section 4.4). That is, we aim to assess, from a model point of view, the actual benefits of inverting multi-band rather than single-band data.

4.2 Inversion

4.2.1 Squeezing the inverse problem

We face a very large inverse problem, as it consists of M=38,125 model parameters and N ∼400,000 data. The size of the G matrix is then $M \times N \sim 15,250,000,000$. Even if only $\sim 2\%$ of the elements of G are non-null (i.e.$\sim 305,000,000$), it is problematic to construct the G matrix itself, and then to invert it on a single CPU computer. The central issue is the storage of G.

In a first approach, we chose to reduce the size of the linear system by not taking into account measurements at 51 s period. Those very long period *S*-wave data are often associated with large errors (cf. section 4.2.4), because of small signal-to-noise ratio (cf. chapter 2). This reduces the number of data to N ∼300,000, which is sufficient to make the storage and inversion of G practical.

This "squeezing" of the inverse problem is consistent with our observation that short-period (i.e. 10 and 15 s period) data are, statistically, better fitted by our multi-band model, compared to long-period (i.e. 34 s period) data (cf. section 4.4.2 and table 4.2).

In the future, we aim to assess if these very long-period (51 s) data could signifi-

cantly improve our shear-wave tomographic model. Theoretically, this should be the case, because the associated sensitivity kernels sample larger volumes in the mantle, so that finite-frequency effects are expected to be more important. However, if the measurement errors of the very long-period data are too large, this may counter the benefit expected from improved sampling of the Earth's interior.

Finally, our reduced data vector consists of N=297,465 data, with time residuals measured at 10 s, 15 s, 22.5 s and 34 s periods.

4.2.2 Finding a model

Here, we aim to solve the linear inverse problem, that we have posed in chapter 3,

$$\mathbf{d} = G\mathbf{m} \tag{4.1}$$

for the model parameters vector $\mathbf{m}$, and with $\mathbf{d}$ the data vector. As G is non-square, this matrix does not have an inverse. If our system of equations (4.1) were well-conditioned (i.e. with a small condition number[1]), we could use the least-squares (estimated) solution:

$$\mathbf{m}_{est} = G^{-}\mathbf{d} = (G^t G)^{-1} G^t \mathbf{d} \tag{4.2}$$

which minimizes $|G\mathbf{m} - \mathbf{d}|^2$. Here, G^{-} is the generalized inverse of G.

In general for large tomographic systems, as in this thesis, the G matrix is not well-conditioned, mostly because of the uneven ray (or kernel) sampling of the Earth. G^tG is even worse, as the condition number is squared. In any case, making G^tG can itself be time consuming and remove the sparsity of G. Hence, the previous least-squares solution - equation (4.2) - is not a good choice for inverting our global dataset.

The process of throwing away small singular values of G (i.e. TSVD: Truncated Singular Value Decomposition) can improve its conditioning (e.g. Nolet 2008). This is an example of "regularization" of the inverse problem. However, TSVD is not commonly used in large problems, because the computational cost of the Singular Value Decomposition is too much as M and N increase, but would be desirable if possible.

From a Bayesian point of view, assuming that the *a priori* covariance matrices of the data, C_d, and of the model parameters, C_m, follow Gaussian probability functions, we may obtain the maximum-likelihood estimate of the model solution by minimizing the

[1]The condition number associated with the linear equation $\mathbf{d} = G\mathbf{m}$ is defined as the maximum ratio of the relative error in $\mathbf{m}$ divided by the relative error in $\mathbf{d}$. The condition number can be calculated as the absolute ratio between the largest and the smallest eigenvalues of the matrix G.

following function f (e.g. Tarantola & Nercessian, 1984; Tarantola, 1987):

$$f(\mathbf{m}) = (G\mathbf{m} - \mathbf{d})^t C_d^{-1}(G\mathbf{m} - \mathbf{d}) + \mathbf{m}^t C_m^{-1}\mathbf{m}. \tag{4.3}$$

Expanding out the brackets, taking the derivative of f with respect to $\mathbf{m}$, and setting to zero gives:

$$\mathbf{m}_{est} = (G^t C_d^{-1} G + C_m^{-1})^{-1} G^t C_d^{-1}\mathbf{d}, \tag{4.4}$$

which can also be written as the following system:

$$\begin{pmatrix} C_d^{-1/2} G \\ C_m^{-1/2} \end{pmatrix} \mathbf{m} = \begin{pmatrix} C_d^{-1/2}\mathbf{d} \\ \mathbf{0} \end{pmatrix}. \tag{4.5}$$

By again multiplying (4.5) by $C_d^{+1/2}$, we obtain

$$\begin{pmatrix} G \\ C_d^{+1/2} C_m^{-1/2} \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix}. \tag{4.6}$$

The model covariance, C_m, means how model parameters are correlated. High correlation should usually be achieved for nearby parameters. In the following, we use a very simple model covariance of the form

$$C_m = \sigma_m^2 I_M, \tag{4.7}$$

where σ_m is the standard deviation of all the model parameters (which is constant), and I_M the identity matrix of size $M \times M$. The parameter σ_m influences the solution, by damping the model norm, and allows us to "regularize" the inverse problem.

The equation (4.5) recalls us that we need to scale the system (4.1) with the data uncertainty (i.e. $C_d^{-1/2}\mathbf{d}$). If we had previously scaled the data to be univariant (by dividing each datum d_i by its uncertainty σ_i), the data covariance matrix (C_d) would be the identity.

In this thesis, we have not previously scaled the data to be univariant[2]. However, we have considered the same uncertainty for all our data (i.e. $\sigma_i = \sigma_d$, cf. section 4.2.5).

[2]The solver (i.e. LSQR) used for the inversion of the system (4.6) treats univariant data (i.e. with the same weights). In this tomographic study, we assume that our "raw" data have the same uncertainties (cf. sections 4.2.4 and 4.2.5), so that they have equal weights and can be used in the inversion of (4.6). It is our view that the tomographic results of this thesis can be interpreted (cf. section 4.2.6). However, in reality, data uncertainties differ, and "raw" data should be weighted prior to be used in the inversion of (4.6). In the future, we aim to improve our tomographic model by taking into account different data uncertainties. Therefore, we will need to weight the data "before" using the solver, which will be done.

Therefore, we use a data covariance, C_d, of the form

$$C_d = \sigma_d^2 I_N, \tag{4.8}$$

where I_N is the identity matrix of size $N \times N$.

Finally, in order to solve the "damped least-squares" problem (4.6), we use the LSQR method (Paige & Saunders 1982), with $\lambda = \sigma_d/\sigma_m$ the "damping parameter". LSQR is an iterative row action[3] method that converges to the minimum of $\| \begin{pmatrix} G \\ \lambda I_M \end{pmatrix} \mathbf{m} - \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix} \|_2$.

4.2.3 Measuring data misfit

In data analysis, as in Earth sciences, the most commonly used misfit criterion is that of least-squares, in which we minimize χ^2 as a function of the model $\mathbf{m}$:

$$\chi^2(\mathbf{m}) = \sum_{i=1}^{N} \frac{(d_i^{obs} - d_i^{pred})^2}{\sigma_i^2}, \tag{4.9}$$

with $d_i^{pred} = \sum_{j=1}^{M} G_{ij} m_j$, and σ_i the standard deviation of i-*th* datum d_i^{obs}.

χ^2 is a direct measure of the data misfit, in which we weight the misfits ($d_i^{obs} - d_i^{pred}$) inversely with their standard errors (σ_i). In the perfect case that every datum is satisfied with a misfit of one standard deviation, we find $\chi^2 = N$.

Clearly, $\chi^2 >> N$ is unwanted, because the misfit is higher than could be expected from the knowledge of data errors. $\chi^2 << N$ is also unwanted, as it means that the model is trying to fit the data errors rather than the general trend in the data (e.g. Nolet 2008).

Because we want $\chi^2 \simeq N$, it is convenient to work with the reduced χ^2, denoted by χ^2_{red}, which is defined as $\chi^2_{\text{red}} = \chi^2/N$, so that the optimal model solution is found for $\chi^2_{\text{red}} \simeq 1$. If the total uncertainty σ_i on each datum i is constant, and equal to σ_d, therefore one may write

$$\chi^2_{\text{red}} = \frac{1}{N\sigma_d^2} \sum_{i=1}^{N} (d_i^{obs} - d_i^{pred})^2. \tag{4.10}$$

[3]In large scale tomography, finding a solution of the general system (4.1), using techniques that require explicit memory storage of G, is often not possible. Row action methods like LSQR require only access to one row of G at a time and consequently G may reside on secondary storage. Row action methods are iterative methods. At each iterative step, an approximate solution is obtained, which is used as a starting point for the solution of the next iteration.

4.2.4 Estimating data errors

It is important to obtain accurate estimates of the uncertainties (or standard errors), σ_i, of our data. Bolton & Masters (2001) extensively discuss the assignment of quantitative errors in the case of a global S wave dataset. Following their analysis, we aim to identify the separate contributions to the overall data variance (σ^2_{total}) in our global S-wave dataset. This total variance may be modeled as:

$$\sigma^2_{total} = \sigma^2_{\text{3-D}} + \sigma^2_{source} + \sigma^2_{measure}, \tag{4.11}$$

where σ^2_{3D} is due to 3-D seismic heterogeneities (that we aim to image), σ^2_{source} is due to earthquake mislocation, and $\sigma^2_{measure}$ is attributed to measurement[4] errors. Bolton & Masters (2001) estimate $\sigma_{\text{3-D}}$=3.2 s for S waves.

In the tomographic inversion presented in this chapter, we consider that all the data have the same uncertainty (i.e. $\sigma_i = \sigma_d$), as previously mentioned in section 4.2.2. Here, we aim to justify this approximation.

Firstly, the measurement errors ($\sigma_{measure}$) of our global dataset (cf. section 2.2.2.c, of chapter 2) have an average value ($\bar{\sigma}_{measure}$) equal to 0.54 s, 0.67 s and 0.75 s, at short (10–15 s), intermediate (22.5 s) and long (34 s) periods, respectively. Thus, on average, $\sigma_{measure}$ only slightly increases with the period. Moreover, we have observed that histograms of $\sigma_{measure}$ have a distribution that is mostly concentrated nearby $\bar{\sigma}_{measure}$, so that there is only a few data for which $|\sigma_{measure} - \bar{\sigma}_{measure}| > 0.5$s.

Secondly, in this thesis, we have not been able to determine individual errors related to earthquake mislocation (σ_{source}), as explained in section 4.2.5. Therefore, we cannot have an accurate estimation of the total uncertainty (i.e. $\sigma_i = \sigma_{source} + \sigma_{measure}$) on each individual data. However, Bolton & Masters (2001) estimate σ_{source} for S waves to be 1.6–2.5 s, assuming a typical depth uncertainty of about 10 km, at epicentral distances of about 70°, and for mislocation vectors of length 10–20 km. Therefore, we know that, on average, σ_{source} is mostly twice larger than $\sigma_{measure}$.

Finally, this justifies to approximate $\sigma_i = \bar{\sigma}_{source} + \bar{\sigma}_{measure} = \sigma_d$. As $\bar{\sigma}_{measure} \simeq 0.7$ s for data measured between 10 and 34 s periods, we have chosen $\bar{\sigma}_{source} \simeq 2.3$ s - which is a bit smaller than the upper bound value of 2.5 s found by Bolton & Masters (2001) - so that $\sigma_d \simeq 3$ s.

[4]The measurement errors may be related to different factors, such as: the effect of both the noise and the approximations made in the synthetics computation, incorrect crustal modeling, unidentified station timing problems, etc.

4.2.5 Earthquake mislocation

The previous discussion (cf. section 4.2.4) suggests that quite a lot of the observed variance in our data could be due to event mislocation. Earthquake mislocation is a major source of error in every long-period body-wave global dataset. It is important to treat this properly, either in the inversion stage (e.g. Bolton & Masters 2001; Houser *et. al* 2008), or in the measurement stage (e.g. Sigloch & Nolet 2006). This is particularly true of teleseismic P wave data, where mislocation effects are similar in size to the signal from 3-D structure (a typical mislocation error is $\sim$0.9 s, and the signal from 3-D structure is $\sim$1 s), but is less important for S waves. In this study, we have not accounted for the earthquake mislocation effects, as explained in the following.

In several other tomographic studies, data are associated to the onset, or first swing, of the target phase (e.g. Masters *et al.* 2000; Montelli *et. al* 2006b; Houser *et. al* 2008). We can then estimate the uncertainty due to source mislocation by considering:

$$\delta t = \frac{\partial t}{\partial x}\delta x + \frac{\partial t}{\partial y}\delta y + \frac{\partial t}{\partial z}\delta z + \delta t_0, \tag{4.12}$$

where δx, δy and δz are errors in event location, δt_0 is the error in origin time, and δt is the resulting error in travel time.

However, using equation (4.12) may not be a good choice, if data are time residuals measured by cross-correlation of observed and synthetic "full" waveforms, as in this thesis. As an example, consider a shallow event, with a typical source depth of 15 km. Clearly, the direct (e.g. S) and depth (e.g. sS) phases will interfere, and the associated measured time residual will reflect this interference pattern. Therefore, if one adds a perturbation (δx, δy, δz; δt_0) to the source location and origin time, this will modify the original two-phase interference pattern, and hence the associated delay-time. The relationship between the source perturbation and the induced delay-time perturbation is, in this case, strongly *non-linear*. This is a real issue in "finite-frequency" tomography, because we aim to interpret such two-phase interference patterns. This problem disappears if only single-phase measurements are used, corresponding to deep events.

Equation (4.12) can be used if we recalculate the synthetics for each new trial source (δx, δy, δz; δt_0). It is then possible to estimate, at each period of analysis, the actual change on the observed time-residual. This task was computationnaly prohibitive during this thesis, even with parallel computation. Another possibility for overcoming the source mislocation problem, in "finite-frequency" tomography, is to follow the approach of Sigloch & Nolet (2006), who determine the best source parameters (source time function, moment

tensor, depth) for each earthquake, using a cluster analysis which needs many stations having recorded the same event. In this case, the mislocation problem is treated at the stage of data measurement. This approach seems easier to implement in local (regional) tomographic studies, and is an interesting direction to be explored in the future.

A limiting factor of our study is then that earthquake mislocation effects are not handled, which should be kept in mind when interpreting the inversion results. This problem will be the subject of future work, in order to refine our tomographic imaging of the mantle. By way of reassurance, we note that Masters *et al.* (2000) report that their tomographic results of S wave inversions vary little when source mislocation effects are included. So, the effect of mislocation may not be an important factor in global S-wave tomography, or at least less than for P-waves.

4.2.6 Outliers

Any raw data set has inconsistent measurements of one kind or another (e.g. Bolton & Masters 2001). After a first inversion, about 16,000 outliers, with data misfit deviations larger than $\pm 2\sigma$, were rejected for subsequent inversions (cf. figure 4.1). This corresponds to about 5% of the entire dataset. The damping parameter used for this first inversion was λ=1,000. Though this damping value is not optimal (cf. section 4.2.7), it corresponds to a physically plausible model, compared to other studies (e.g. Houser *et al.* 2008, Montelli *et al.* 2006b). With the rejection of outliers, we are left with N=280,673 data.

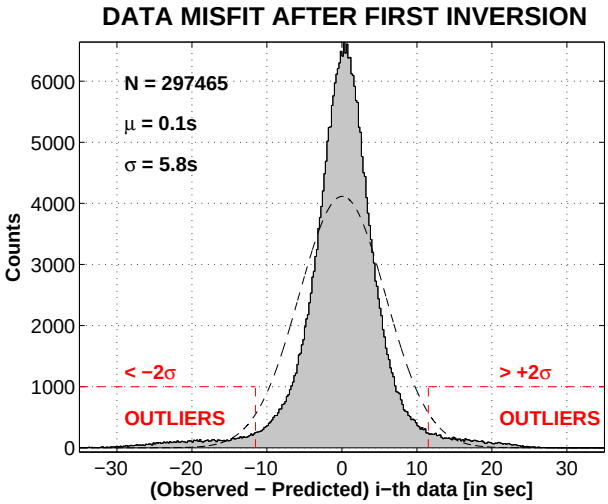


Figure 4.1: Outliers, with a misfit larger than $\pm 2\sigma$, have been rejected after a first inversion. σ is the standard deviation of the Gaussian function that better fits the histogram (in dashed black line). N is the initial number of data. μ is the mean of data misfits.

4.2.7 Model smoothness vs data fitting

We seek a model that minimizes $|G\mathbf{m}-\mathbf{d}|^2+\lambda^2|\mathbf{m}|^2$ (cf. section 4.2.2) The damping parameter, λ, controls the balance between minimizing the model norm, $|\mathbf{m}|$, which is often a criterion of smoothness of the reconstructed tomographic image, and minimizing the data misfit norm $|G\mathbf{m}-\mathbf{d}|$, which corresponds to fitting the data.

When λ is small, there is little weight put on the model norm, the data are fitted well and the reconstructed image is not smooth. When λ is large, the model norm dominates and the reconstructed image is smooth, at the expense of not fitting the data so well.

Perhaps the most convenient graphical tool for setting λ is the "L-curve" (or trade-off curve). That is, if estimates of standard errors are known, one may plot $|\mathbf{m}|$ in function of χ^2_{red} - cf. equation (4.10). We then get the characteristic L-shaped curve (figure 4.2), with a (often) distinct corner separating vertical and horizontal parts of the curve.

The rationale for using the L-curve is that regularization of the inverse problem, via the damping parameter λ, is a trade-off between the data misfit (i.e. χ^2_{red}) and the model norm. The L-curve can then be used to decide how close one wants to be to a model that fits the data at a level given by $\chi^2_{red}=1$. If the trade-off curve shows that one could significantly reduce the norm of the model, while paying only a small price in terms of an increase in χ^2_{red}, this is an indication that the standard errors in the data have been under-estimated (e.g. Nolet 2008), and means that we should accept a model with $\chi^2_{red}>1$. But, if a significant decrease in χ^2_{red} can be obtained at the cost of only a minor increase in model norm, this indicates an over-estimate of data errors, and means that we should accept a model with $\chi^2_{red}<1$. Depending on where on the L-curve we find that $\chi^2_{red}=1$, we know if we do, or do not, have a strong constraint on the norm of the model. In a well designed tomography experiment, $\chi^2_{red}\simeq 1$ near the bend in the L-curve.

Figure 4.2 shows the trade-off curve between χ^2_{red} and the model norm, corresponding to our global dataset of time residuals measured at 10, 15, 22.5 and 34 s periods. The model norm used here is the Euclidean norm, defined as:

$$\|m\|_2=\{\sum_{j=1}^{M} m_j^2\}^{1/2}. \tag{4.13}$$

We have chosen, on the L-curve displayed in figure 4.2, the model solution, $m_{\lambda=350}$, corresponding to a damping parameter $\lambda=350$. This model has $\chi^2_{\rm red}$=1.2189. Clearly, figure 4.2 shows that a further reduction in the model norm will lead to a much larger data misfit.

The reader might wonder why we did not choose the model with $\chi^2_{\rm red}$=1 (corre-

sponding to $\lambda \sim 150$). Figure 4.2 also shows that there is a significant increase in model norm, between the two points on the L-curve where χ^2_{red}=1.2189 and χ^2_{red}=1. This implies that, if we chose the model with χ^2_{red}=1, we would obtain a lot of extra details, compared to the model with χ^2_{red}=1.2189. However, those extra details only depend on our choice of the *a priori* errors contained in our data (cf. section 4.2.4).

Finally, we have preffered the much simpler tomographic images corresponding to the model $m_{\lambda=350}$, with χ^2_{red}=1.2189, which would be equivalent to χ^2_{red}=1 if we underestimated the standard errors in the data by only ∼10% (i.e. σ_d=3.3 s instead of 3 s). In the following, the 3-D shear-wave velocity perturbation (multi-band) model, $m_{\lambda=350}$, will be named *ZDS-S10*.

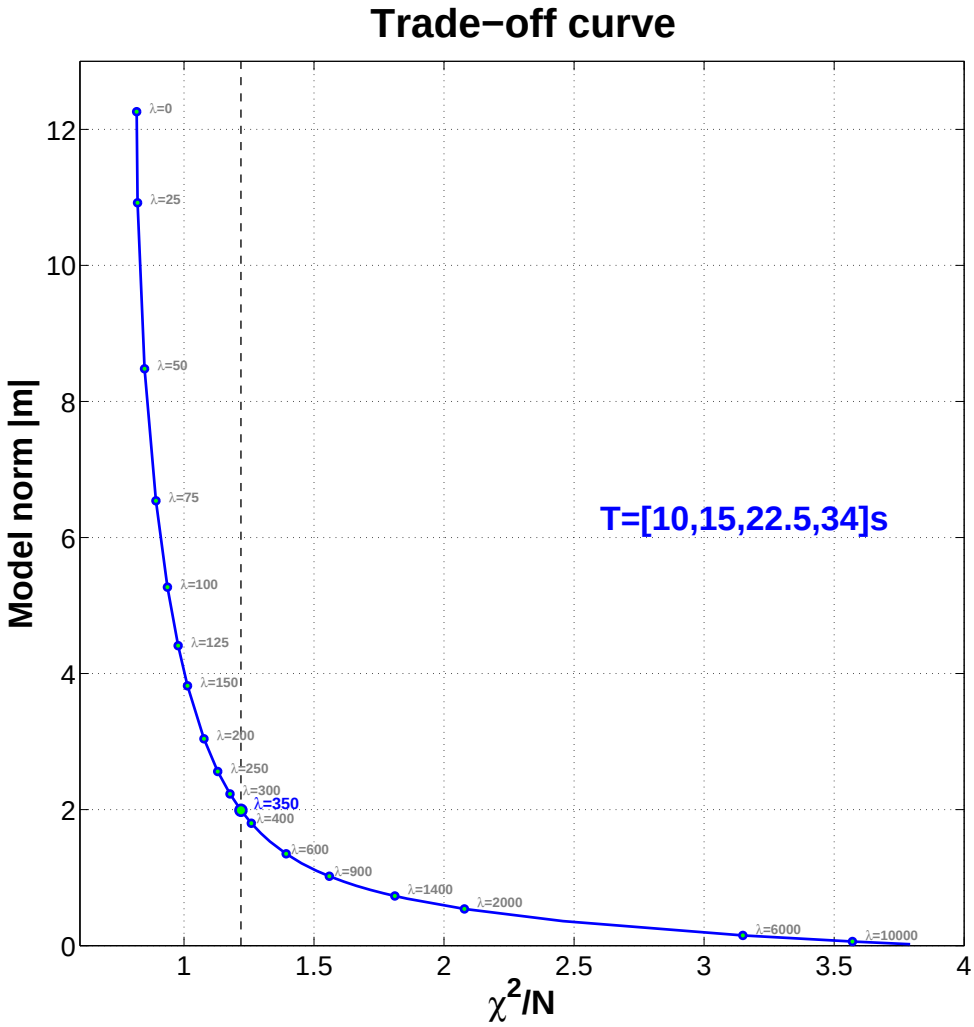


Figure 4.2: The L-curve (or trade-off curve) between χ^2/N and model norm, corresponding to our global dataset of measurements at 10, 15, 22.5 and 34 s periods.

4.3 Shear-wave velocity model *ZDS-S10*

Our 3-D *SH*-wave velocity perturbation (multi-band) tomographic model *ZDS-S10* is derived from the inversion of 280,673 *S*, *ScS* and *SS* frequency-dependent travel times, measured at 10 s, 15 s, 22.5 s and 34 s periods. In this section, we present some general characteristics of *ZDS-S10*, such as: global patterns, sensitivity tests, etc.

4.3.1 Global patterns

Figure 4.3 shows the root-mean-square (RMS) values of the velocity anomalies $\delta V_S/V_S$ of *ZDS-S10*, as a function of depth. The RMS value is calculated, for each layer z, as

$$\mathrm{RMS}(z) = \{\frac{1}{N}\sum_{k=1}^{N}(m_k^z)^2\}^{1/2}, \qquad (4.14)$$

where m_k^z are the model values interpolated on a $1^\circ \times 1^\circ$ regular grid. We see that, statistically, the major velocity variations occur, at global scale, in the lithosphere and the top of the upper mantle, as well as in the D" layer.

Figure 4.4 shows an overview of *ZDS-S10*, displayed at global scale for each layer of the parameterization.

In appendix F, we also show nice geophysical objects imaged in our model (e.g. subducting slabs, rising up low-velocity anomalies).

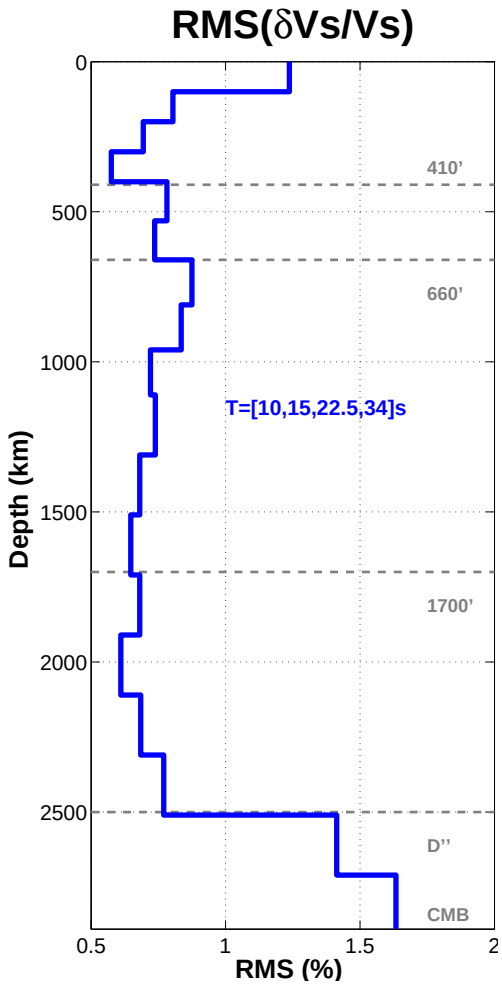
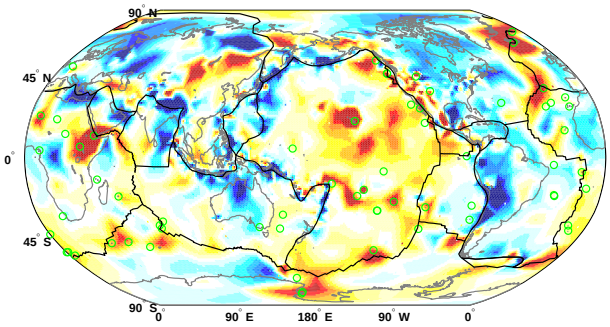
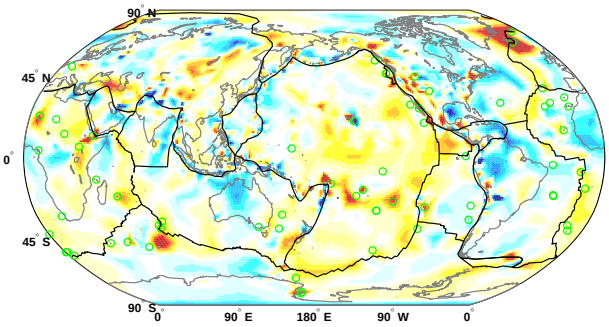


Figure 4.3: RMS velocity perturbation $\delta V_S/V_S$ versus depth for ZDS-S10.

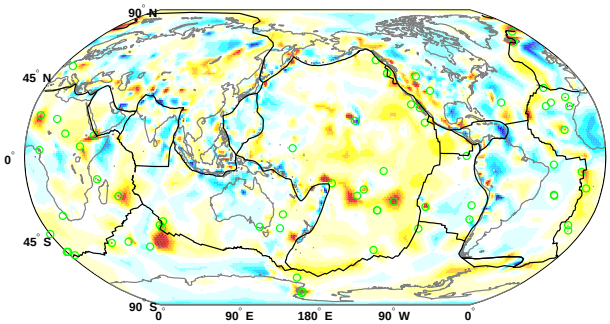
a) δVs/Vs (%), 0−100 km, T=[10−34]s



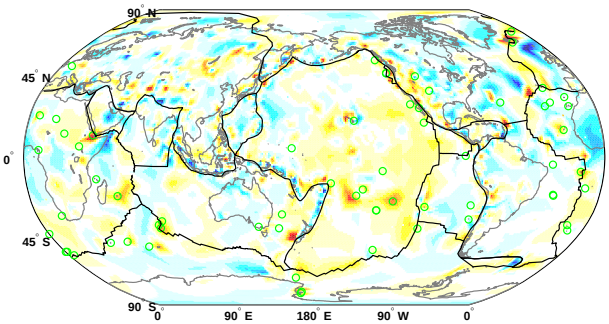
b) δVs/Vs (%), 100−200 km, T=[10−34]s



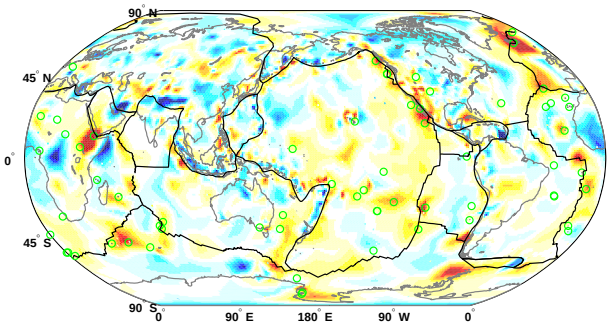
c) δVs/Vs (%), 200−300 km, T=[10−34]s



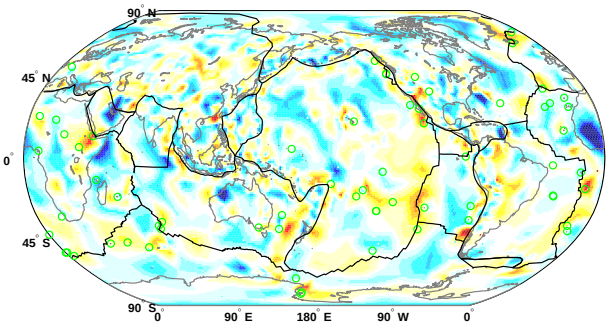
d) δVs/Vs (%), 300−400 km, T=[10−34]s



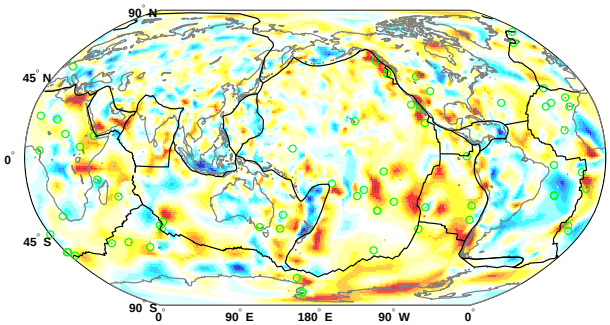
e) δVs/Vs (%), 400−530 km, T=[10−34]s



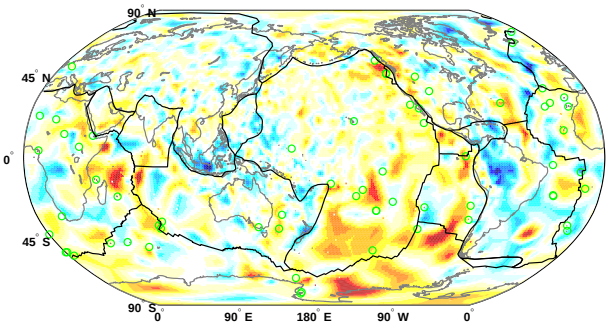
f) δVs/Vs (%), 530−660 km, T=[10−34]s

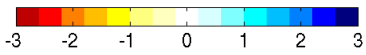


g) δVs/Vs (%), 660−810 km, T=[10−34]s

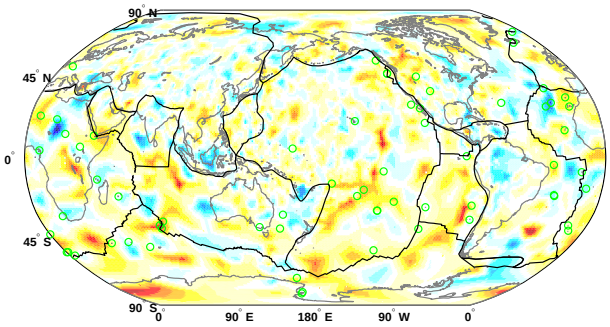


h) δVs/Vs (%), 810−960 km, T=[10−34]s

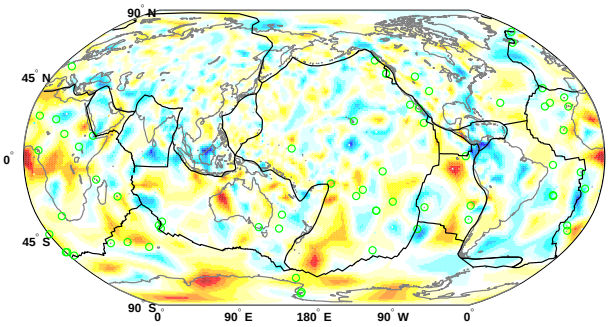




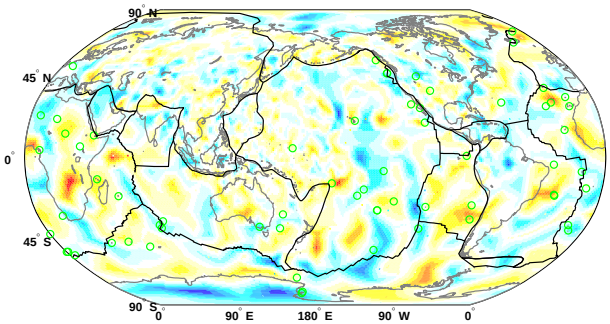
i) δVs/Vs (%), 960−1110 km, T=[10−34]s



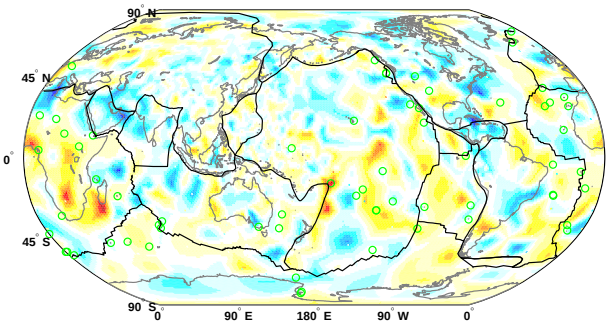
j) δVs/Vs (%), 1110−1310 km, T=[10−34]s



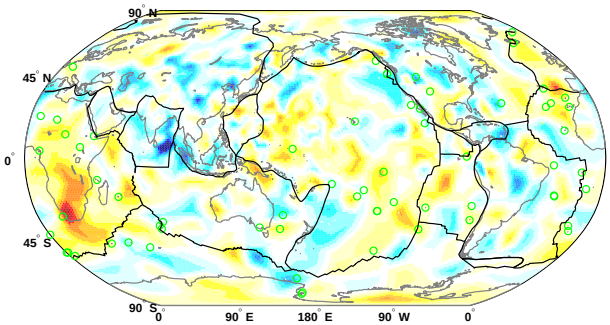
k) δVs/Vs (%), 1310−1510 km, T=[10−34]s



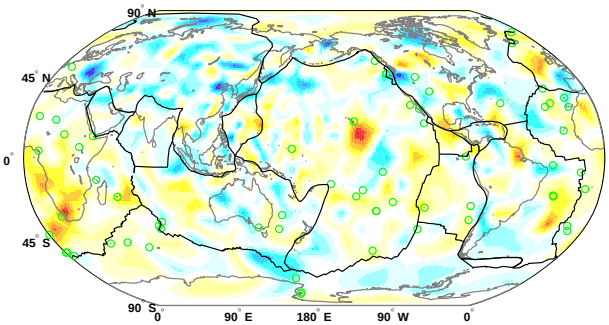
l) δVs/Vs (%), 1510−1710 km, T=[10−34]s



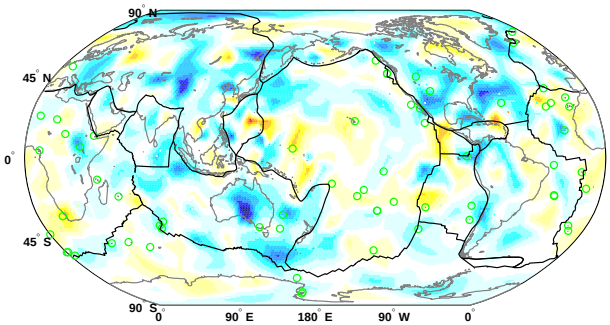
m) δVs/Vs (%), 1710−1910 km, T=[10−34]s



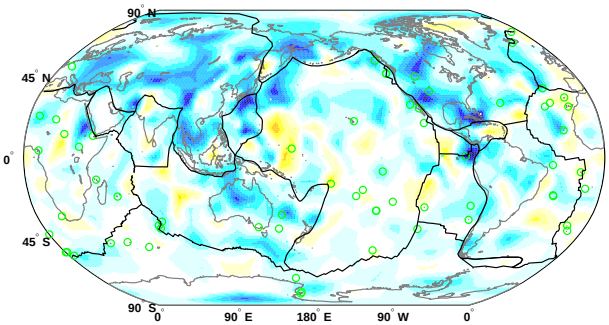
n) δVs/Vs (%), 1910−2110 km, T=[10−34]s

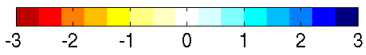


o) δVs/Vs (%), 2110−2310 km, T=[10−34]s



p) δVs/Vs (%), 2310−2510 km, T=[10−34]s





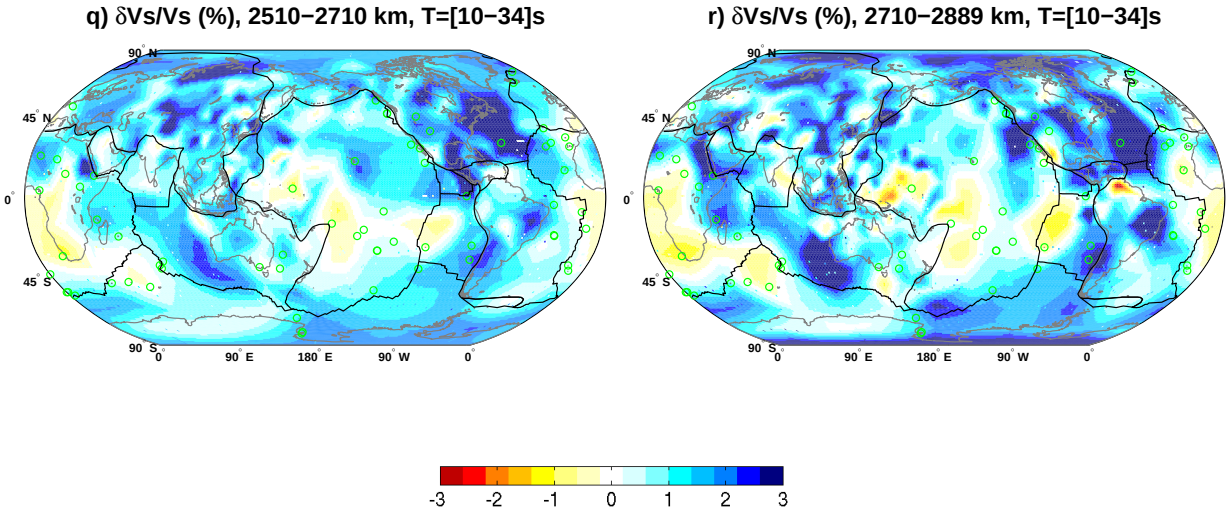


Figure 4.4: SH-wave velocity anomalies, in %, of our global multiple-frequency tomographic model ZDS-S10, shown for each layer of the parameterization. This model has been obtained from the inversion of 280,673 S, ScS and SS frequency-dependent time residuals, measured at 10, 15, 22.5 and 34 s periods. Warm colors: low-velocity anomalies (slow regions); cool colors: high-velocity anomalies (fast regions). Green circles: hotspots (Anderson 2005); gray solid line: continents; solid black line: tectonic plates.

4.2.3.a High-velocity anomalies

Figure 4.4 shows the circum-Pacific ring of fast material above the CMB (below ∼2500 km depth). A drastic increase in very high-velocity anomalies starts at about 2000 km depth, down to the CMB. These very fast regions are assumed to result from the accumulation of ancient subducted slabs (e.g. van der Hilst & Karason 1999). For instance, below 660 km depth and down to the CMB, the ancient Farallon slab is associated with high velocity anomalies beneath North America (e.g. Grand 1994) - see figure F.1, in appendix F. Beneath the Mediterranean and southern Eurasia, high-velocity anomalies accumulated at ∼1500 km depth, are believed to be remnants of the Tethys slab (e.g. van der Hilst & Karason 1999).

A variety of, more recent, subducting slabs are also present in our model. Some remain stagnant around the 660 km discontinuity, whereas others penetrate into the lower mantle. For instance, a high velocity region is clearly visible beneath the Tonga subduction region, from the surface and down to ∼1000 km depth - see also figure F.2, in appendix F. A short compilation of subducting slabs, as imaged from our multi-band global tomography, is shown in appendix F.

Our tomographic observations are in agreement with the widely accepted concept that slabs do pass through the transition zone into the lower mantle (e.g. Grand *et al.* 1997; AlbarÃ¨de & van der Hilst, 1999; Fukao *et al.* 2001; Nolet 2009), which is inconsistent with the hypothesis of a two-layered convection in the Earth.

4.2.3.b Low-velocity anomalies

Figure 4.4 clearly shows the slow regions beneath the Horn of Africa, in the upper mantle. These very low-velocity anomalies are in agreement with other tomographic studies (e.g. Debayle *et al.* 2001; Montelli *et al.* 2006b). In section 4.4.2, we plot a cross-section of our model through the complex shear-wave low-velocity anomaly present in the whole African mantle (figure 4.16). The connection of this low-velocity anomaly beneath Lake Victoria with the African superplume is very similar to other studies (e.g. Ritsema *et al.* 1999; Montelli *et. al* 2006b).

A closer look at figure 4.4 also shows that many significant low-velocity anomalies (displayed in red) are located near a known hotspot (green circles) - e.g. see figure 4.4.g. A majority of these hotspots are located in Africa and Pacific ocean (see also figure 4.5).

The presence of two large slow regions at the CMB, beneath Africa and the Pacific Ocean, is receiving increasing attention in geosciences community. These two regions are often referred to as "superplumes", and could be feeding up smaller low-velocity anomalies in the shallower mantle (e.g. Davaille 1999; Montelli *et. al* 2006b).

However, in this study, these two "superplumes" are less visible than in other studies (e.g. Ritsema *et al.* 1999; Montelli *et. al* 2006b; Houser *et al.* 2008). The reason is simply a lack of data sampling in these two regions, as explained in section 4.2.3.c.

4.2.3.c Non-uniform sampling at CMB

In every global dataset, the coverage of the two "superplumes" near the CMB is weaker beneath Africa and Pacific Ocean. That is, there is a small number of seismic stations located in oceans and Africa, and the distribution of earthquakes is also non-uniform at global scale.

In this study, only *ScS* waves provide a sampling at the CMB, because *S* and *SS* waves bottom shallower, near 2400 km depth. This poor CMB sampling is retrieved in the sensitivity tests (section 4.3.4, figure 4.12), and the "kernel column density" analysis (section 4.3.3, figure 4.6). We plan to use other seismic phases for improving our sampling in the deepest part of the mantle, such as *Sdiff*, *SKS*, *SSS*, etc. We will need to associate with each phase its appropriate sensitivity kernel, which may not be computable under

the paraxial approximation (cf. section 1.5.3, of chapter 1). Therefore, the use of more complicated kernels could be necessary (e.g. Calvet and Chevrot 2005).

Sampling of deep mantle could greatly be improved with new earthquake recorders massively deployed in oceans. Ocean Bottom Seismometers (OBS) are very useful for recording earthquakes on the ocean's floor. However, as they are quite expensive, they can only be used in limited regions.

Recently, Simons *et. al* (2006) showed that it is possible to record earthquakes in oceans, with an hydrophone mounted on a freely floating diver, submerged at 700 m depth. As the floating diver is drifted by ocean currents, this may provide a more uniform sampling in oceanic regions (Nolet 2008).

4.2.3.d Vertically continuous structures in lower mantle

To investigate the vertical continuity of the anomalies present in the lower mantle, we compute the vertical average of the S-velocity anomaly, $\delta V_S/V_S$, over the whole lower mantle, i.e. between 660 and 2889 km depth (see figure 4.5). Averaging emphasizes predominantly vertical features that are continuous with depth.

We clearly see, in figure 4.5, the high-velocity anomalies around Pacific Ocean. They are believed to be remnants of subducted slabs, such as the Farallon slab (e.g. Grand 1994) beneath America and the Thetys slab beneath Eurasia (e.g. van der Hilst & Karason 1999), which are interpreted as "cold" material sinking into the mantle.

We also see, in figure 4.5, low-velocity anomalies present beneath Southern Africa and Pacific Ocean. They are believed to correspond to deep plumes (e.g. Montelli *et. al* 2006b), which are associated with "hot" material rising up from the deep mantle. However, owing to a non-uniform sampling in the lowermost mantle, the contribution of the two "superplumes" is likely to be under-estimated in figure 4.5.

Figure 4.5 shows that the first-order convection, occuring in the lower mantle, is well retrieved in our model. In the future, we aim to explore smaller scale anomalies, in *ZDS-S10*, throughout the mantle (cf. appendix F), to better constrain the 3-D elastic structure of the mantle and its convection (e.g. Davaille 1999; Trampert *et al.* 2004; Davaille *et al.* 2005).

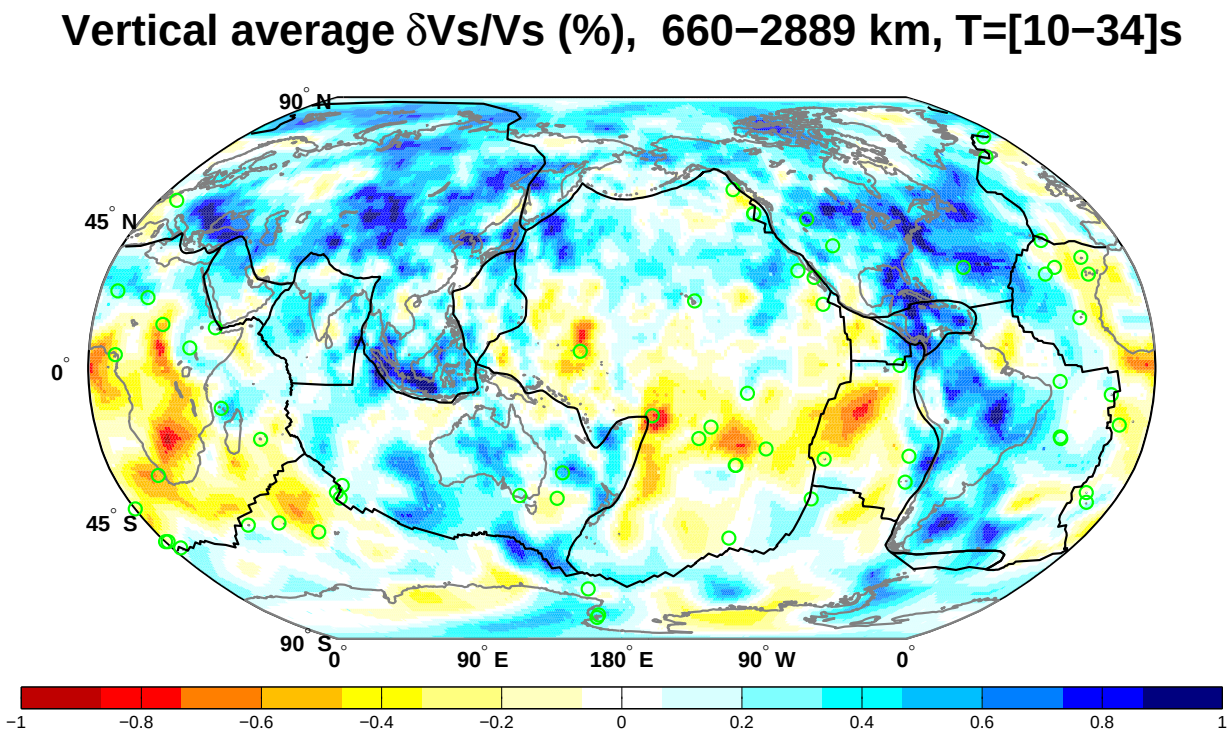


Figure 4.5: Vertical average over the whole lower mantle (i.e. between 660 km and 2889 km depth) of the shear-wave velocity perturbation (in %) of our multi-band model model ZDS-S10. The colorscale used here is only between ±1%, compared to the usual coloscale used for displaying tomographic images of our model ZDS-S10 (see figure 4.4 and appendix F), which is between ±3%. Our aim is here to enhance the amplitude of the averaged velocity anomalies, which have been decreased by the averaging process. It is thus easier to distinguish vertically continuous structures. Warm colors: low-velocity anomalies (slow regions); cool colors: high-velocity anomalies (fast regions). Green circles: hotspots (Anderson 2005); gray solid line: continents; dashed black line: tectonic plates.

4.3.2 Variance reduction

It is common, in tomographic studies, to associate the quality of the obtained model to an important variance reduction (VR), which is defined as:

$$\mathrm{VR}(\%) = \{1 - \frac{\sum_{i=1}^{N} (\delta t_i^{obs} - \delta t_i^{pred})^2}{\sum_{i=1}^{N} (\delta t_i^{obs})^2}\} \times 100, \tag{4.15}$$

where N is the number of data, δt_i^{obs} is the i-*th* observed time residual, and $\delta t_i^{pred} = \sum_{j=1}^{M} G_{ij} m_j$ is the i-*th* time residual, predicted by the model (*ZDS-S10*).

The variance reduction, which depends on the Euclidean norm of the misfit vector, relates how much a 3-D tomographic model satisfies the data when compared to a 1-D reference model (e.g. IASP91). This reduction is as much a function of the fit of the 1-D starting model as of the data fit itself (e.g. Nolet 2008). That is, the same 3-D model can have different variance reductions depending on the starting model. It is therefore quite useless as a statistical measure of quality for the tomographic model.

However, since other global tomographic models are often specified in terms of their variance reduction, we give our value here. The variance reduction of the multi-band model, *ZDS-S10*, is $\sim$ 70%. We believe that this value could slightly be improved, maybe up to $\sim$80%, by fully accounting for the effect of earthquake mislocation (cf. section 4.2.5).

4.3.3 Kernel column density

As a measure of how strongly a particular model parameter m_j in the Earth is sensed by the all the sensitivity kernels, corresponding to our global dataset, we compute the "kernel column density", defined as (Tian *et.al* 2009):

$$D_j = \frac{\sum_{i=1}^{N} |G_{ij}|}{\max_j \{\sum_{i=1}^{N} |G_{ij}|\}} \tag{4.16}$$

where G_{ij} is the sensitivity matrix of our inverse problem, and with $j = \{1, \cdots, M\}$.

Figure 4.6 shows $\log_{10}(D_j)$. High/low values correspond to highly/poorly sampled regions, shown in red/white, respectively. We see that the current coverage of seismic stations leads to large values of D_j in the Northern Hemisphere. As stations coverage in the Southern Hemisphere still remains a problem, we see lower values of D_j in these regions. The lowest values of D_j (displayed in white and light yellow) are observed in the deepest lower mantle (deeper than $\sim$2000 km), beneath Africa and Pacific Ocean. This again underlines the non-uniform kernel sampling of our dataset, in the lowermost 1000 km of the lower mantle.

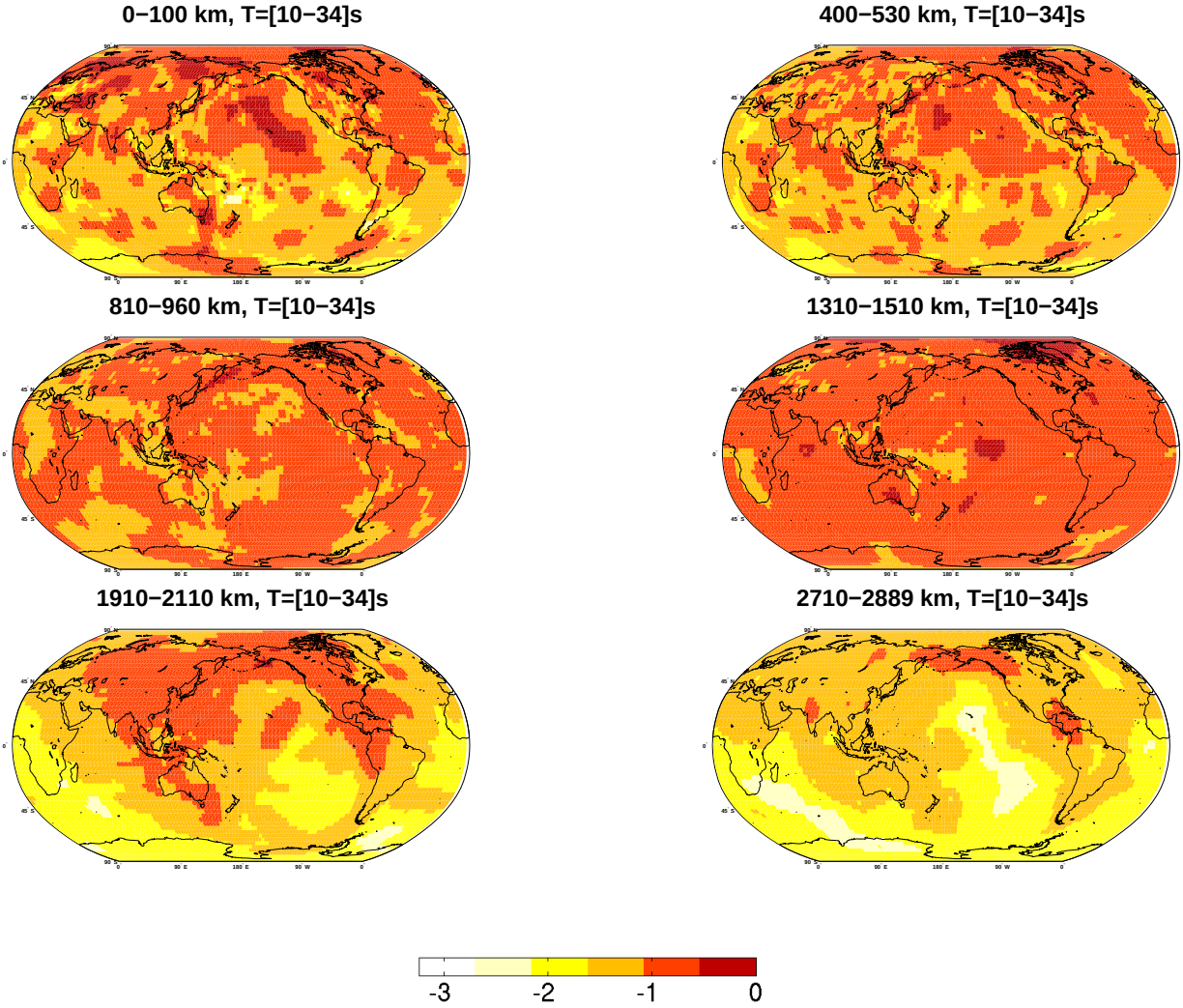


Figure 4.6: Plot of $\log_{10}(D_j)$ *at different depths, where* D_j *is the "kernel column density". In red: good kernel sampling; in white: poor kernel sampling.*

4.3.4 Sensitivity tests

Checkerboard sensitivity tests provide a visual evaluation of how well a particular input model is recovered in the inversion process, and an estimate of the influence of damping on the solution. If we deal with massive tomographic systems, as in this thesis, such tests are often the only feasible method for getting an idea of the "resolving power" of the designed tomographic experiment. This "resolving power" depends on the source-receiver geometry, the theory used to relate the data and model parameters, and the regularization used in the inversion.

A checkerboard test consists of calculating synthetic data, $\mathbf{d}_{syn}$, for a known input model, $\mathbf{m}_{syn}$, and then inverting the synthetic data to compare the output model, $\mathbf{m}_{est}$, with the input model. The synthetic data are computed as: $\mathbf{d}_{syn} = G\mathbf{m}_{syn}$. We then have:

$$\mathbf{m}_{est} = G^{-}\mathbf{d}_{syn} = G^{-}G\mathbf{m}_{syn} = R\mathbf{m}_{syn} \tag{4.17}$$

, where $R = G^{-}G$ is the resolution matrix. The checkerboard recovery often contains features that have spread out horizontally ("smearing"), or vertically ("leaking").

We aim here to get an estimate of the resolution of our multi-band model, *ZDS-S10*, in the transition zone (figures 4.7 and 4.8), mid-mantle (figures 4.9 and 4.10), and CMB (figures 4.11 and 4.12). We use a checkerboard test model, in a single layer of the mantle, to produce the synthetic data. The input layer has a width of about 250 km, 400 km and 380 km, for the transition zone, mid-mantle and CMB, respectively. The initial checkerboard is made of equally spaced blocks, with diameter of about 15° for both the transition zone and the mid-mantle, and about 30° for the CMB. The input anomaly amplitude decays horizontally with increasing radius, following a Gaussian curve.

The synthetic data inversion allows us to determine the ability of our 3-D kernels geometry to resolve the pattern and amplitude of the input checkerboard model, using the same damping as in the actual inversion (i.e. λ=350). Note that the "input" checkerboard pattern is distorted in some regions, such as in the Southern Hemisphere. This is related to the use of our irregular parameterization, which may be too coarse for the imposed checkerboard pattern in poorly sampled region.

In the transition zone, our data provide a good lateral resolution and amplitude recovery in the circum-Pacific area (see figures 4.7 and 4.8). This is not the case in southern oceans and beneath Africa. This test also shows that structures in the transition zone may be smeared at depth (leaking), owing to nearly vertical shallow rays.

In the mid-mantle, our data provide both good lateral resolution and good amplitude recovery mostly everywhere (see figures 4.9 and 4.10). The amount of vertical and horizontal smearing at depth is weak.

Near the CMB, our data provide a good lateral resolution and amplitude recovery around the Pacific ring, in the Northern Hemisphere. However, this is not the case in most of the Southern Hemisphere (see figures 4.11 and 4.12). That is, we are not able to recover the input model, beneath Africa and Pacific Ocean, where the two "superplumes" are believed to originate. Our tests also show some radial smearing in the D" layer.

Finally, we also observe smaller reconstructed amplitudes in the output models. We attribute this to the damping parameter (λ) used in the regularization of the inversion.

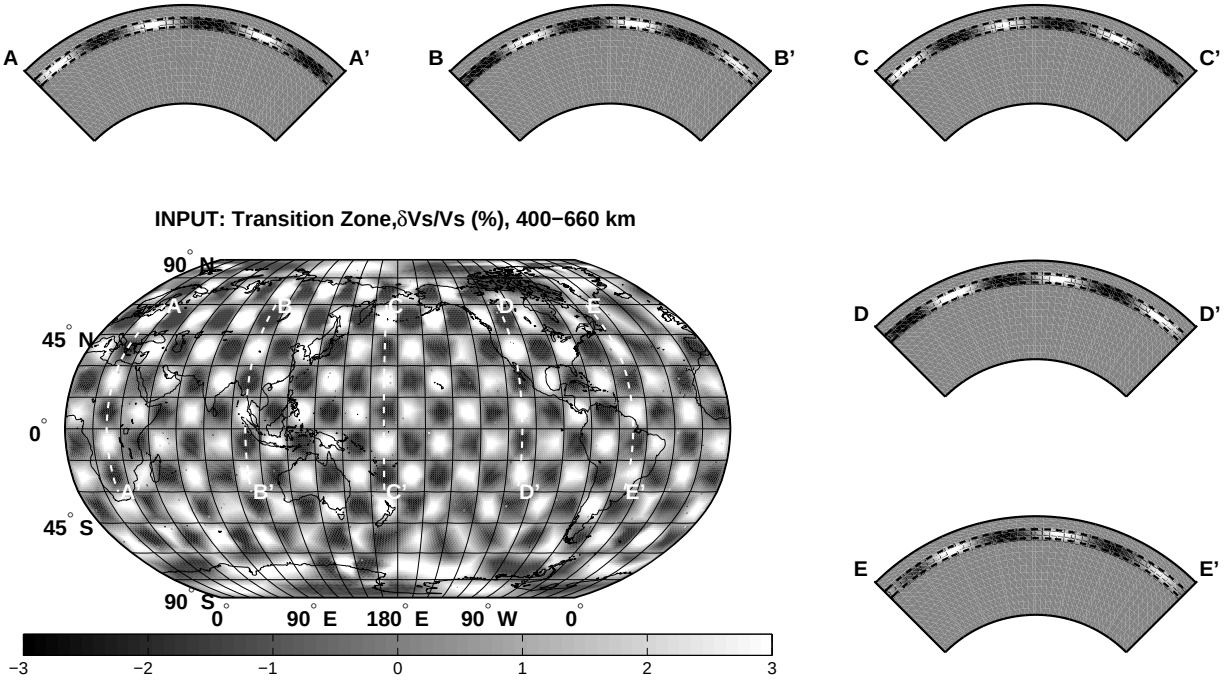


Figure 4.7: Checkerboard input, Transition zone, 400–660 km.

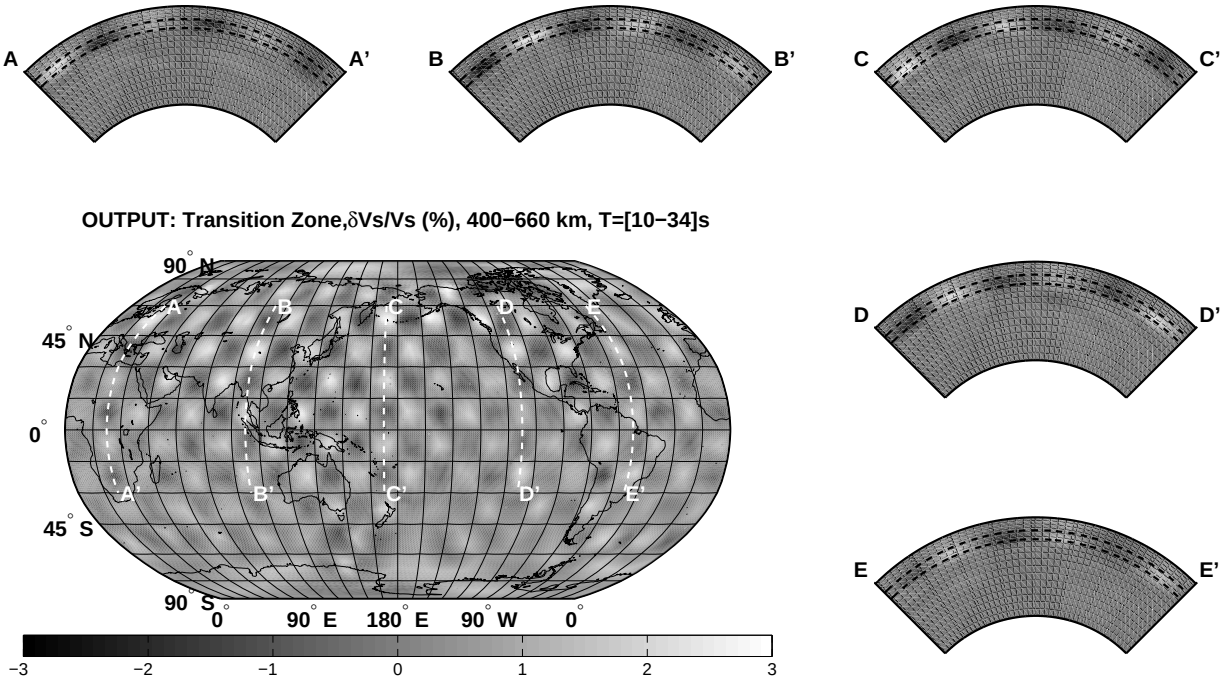


Figure 4.8: Checkerboard output, Transition zone, 400–660 km, T=[10–34] sec period.

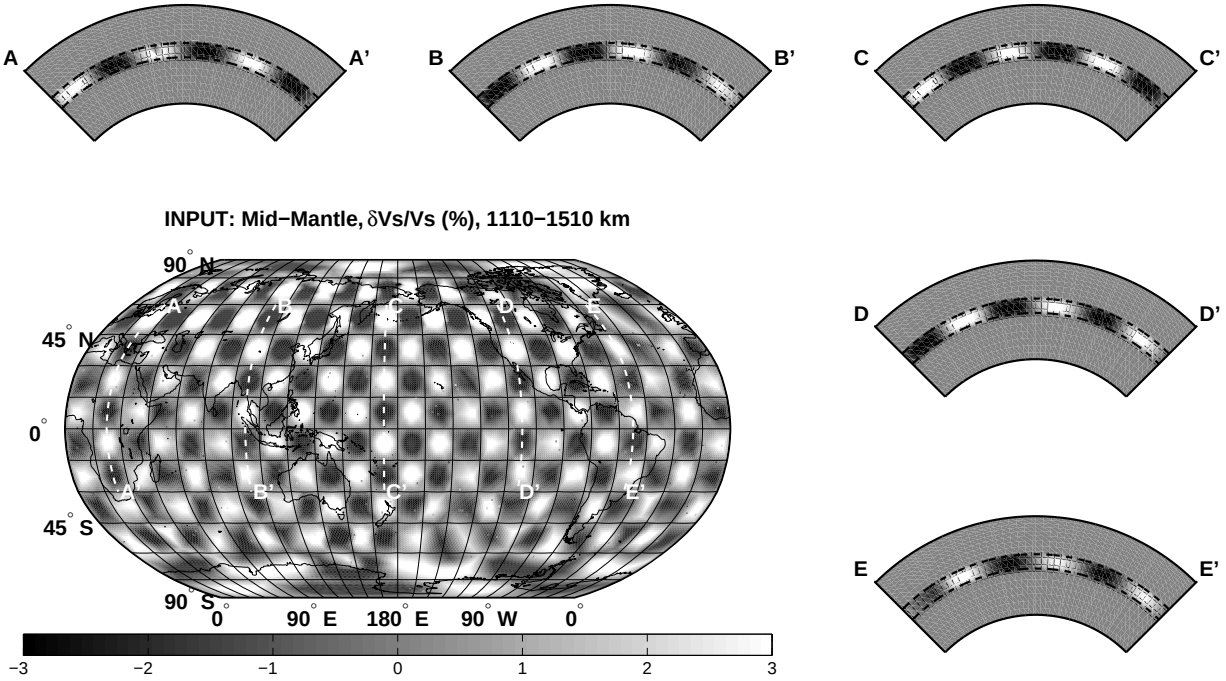


Figure 4.9: Checkerboard input, Mid-mantle, 1110–1510 km.

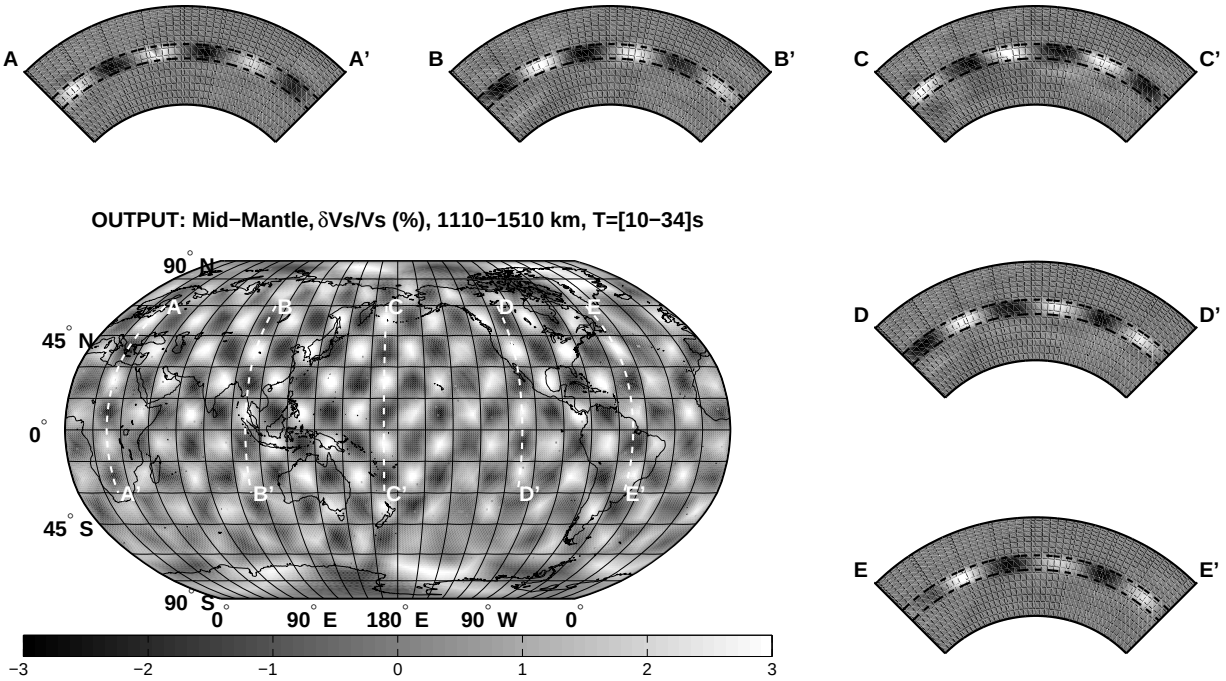


Figure 4.10: Checkerboard output, Mid-mantle, 1110–1510 km, T=[10–34] sec period.

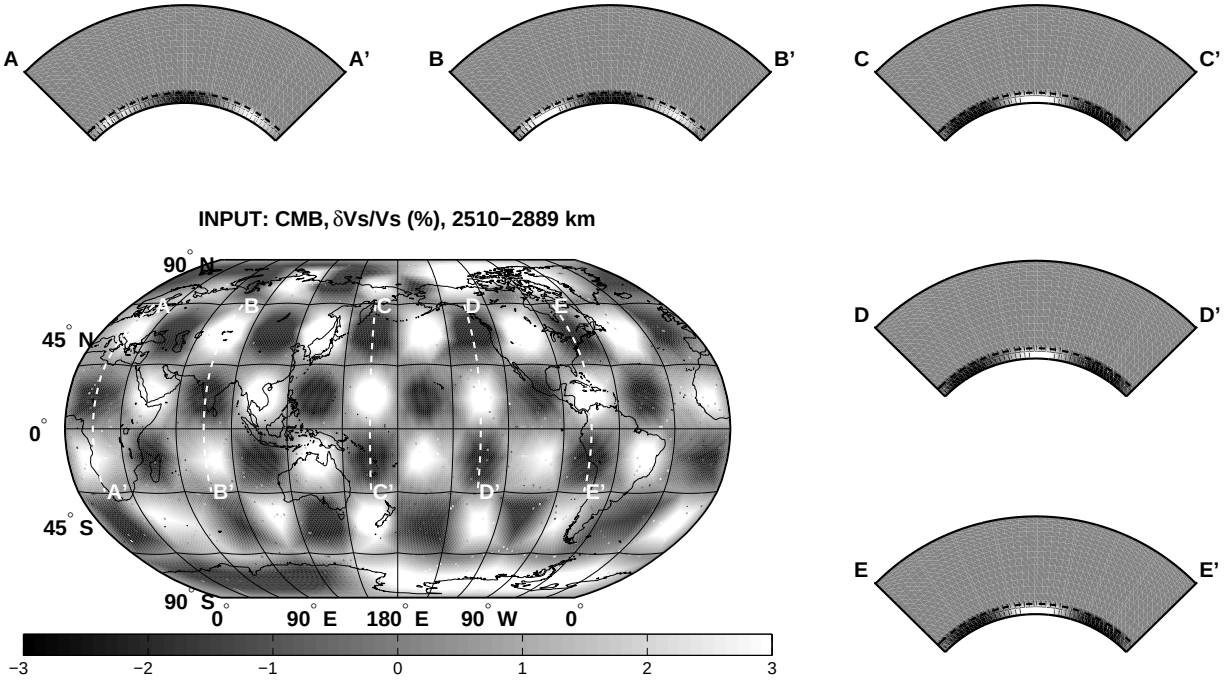


Figure 4.11: Checkerboard input, CMB, 2510–2889 km.

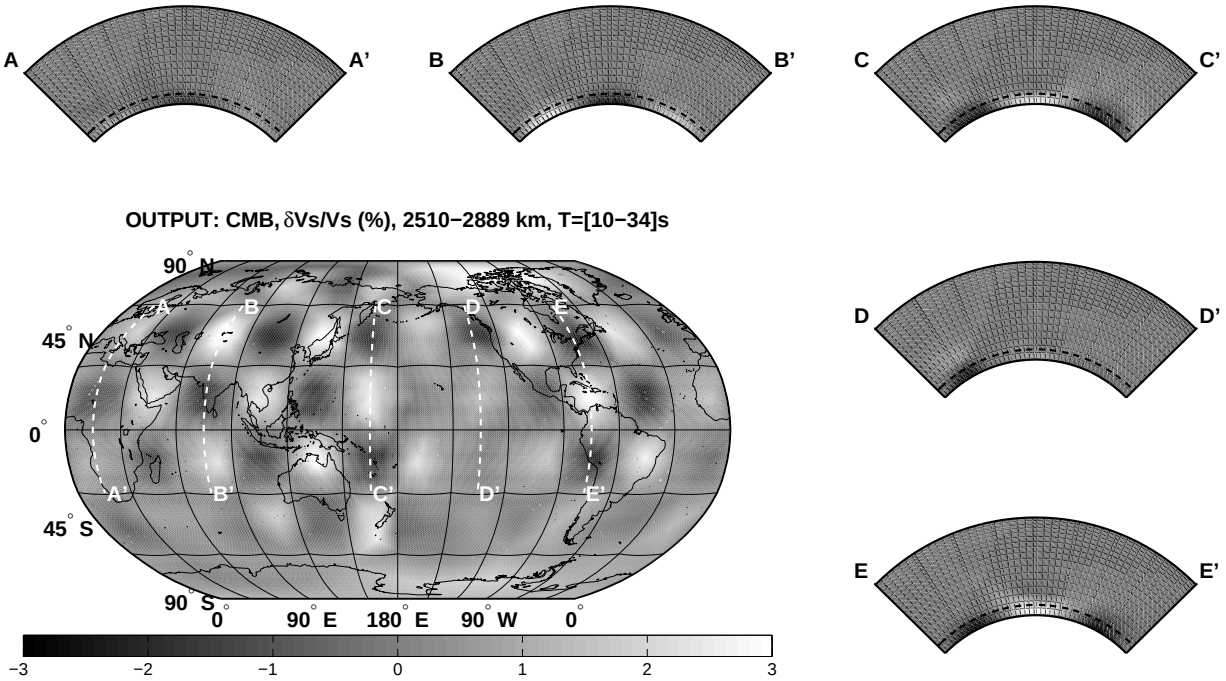


Figure 4.12: Checkerboard output, CMB, 2510–2889 km, T=[10–34] sec period.

4.4 Single-band vs multi-band tomography

In chapter 2, we have shown that a residual "structural" travel time dispersion is observed in our data. Here, we aim to test if, from a model point of view, this new observable can provide additional constraints on the 3-D elastic structure of the Earth's mantle.

We have previously presented, in section 4.3, our global multiple-frequency tomographic model, *ZDS-S10*. This model has been derived from the "finite-frequency" inversion (Dahlen *et al.* 2000) of 280,673 *S*, *ScS* and *SS* frequency-dependent time residuals, measured at periods 10, 15, 22.5 and 34 s. Montelli *et al.* (2006b) published a global *S*-wave tomographic model, hereafter referred to as *PRI-S05*, obtained from the "finite-frequency" inversion (Dahlen *et al.* 2000) of *S*, *ScS-S* and *SS-S* travel times measured in a single-frequency band (∼20s). The model *PRI-S05* hence does not benefit from the increased spatial resolution afforded by sensitivity kernels for a range of frequencies. Improving spatial constraints through the use of "multiple-frequency" tomography was the primary motivation for the present study. A second motivation factor was the use of frequency-dependent travel times measured in a fully consistent way with the kernels.

Theoretically, a multi-band inversion should allow to extract more independent information from the data than a single band inversion. This may translate into better resolved seismic structures. We are now in the position to verify this hypothesis, and then to answer the following question: "Does a multi-band tomographic model contain significantly more information, about the Earth's mantle seismic structure, than a single-band model?".

It would be misleading to directly compare the tomographic images of the single-band model *PRI-S05* with our multi-band model *ZDS-S10*. That is, the first-order differences would likely be related to: (1) non-similar ray sampling (source-receiver geometry) and measurement method, i.e. different datasets; (2) differences in the inversion (earthquake relocation, data weighting, regularization, etc). Therefore, we chose to compare our multi-band model, obtained with measurements at 10, 15, 22.5 and 34 s periods, with the corresponding single-band model, obtained from measurements at 22.5 s period only. In this way, all extraneous factors are common between the two inversions and a comparison is meaningful.

4.4.1 Finding a single-band model

Figure 4.13 shows the trade-off curves (or L-curves) for both the single-band (in red) and multi-band (in blue) models. We see that the two L-curves are very close to each other if the model norms are strongly damped, corresponding to values of $\chi^2_{red} \geq 1.6$. That is,

if the two models are very smooth, they must be statistically similar. On the other hand, figure 4.13 also shows that the less the models are damped, the more the two L-curves (i.e. the two models) differ. This situation corresponds to values of $\chi^2_{red} < 1.6$.

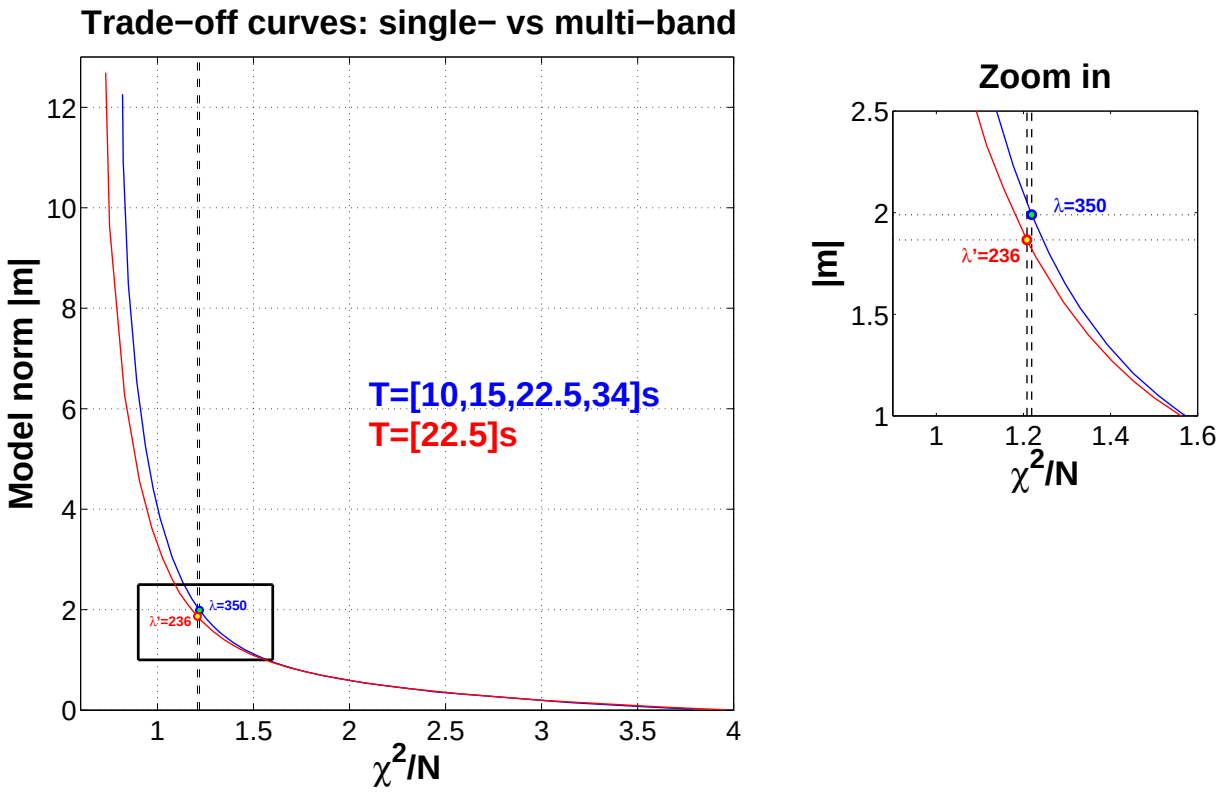


Figure 4.13: Trade-off curves for single- and multi-band tomography, obtained from ∼30 inversions, with λ ranging between 0 and 10,000.

Table 4.1: Data subsets for different periods.

	T	N
$\mathbf{d}_1$ (short-period data)	10 s & 15 s	87,827
$\mathbf{d}_2$ (intermediate-period data)	22.5 s	98,584
$\mathbf{d}_3$ (long-period data)	34 s	94,262

Where T is the period, and N the number of measurements.

The zoom in on the two L-curves (figure 4.13) shows that the multi-band model *ZDS-S10* (blue dot), obtained with a damping parameter $\lambda = 350$, should, on average, differ from its corresponding single-band model. The question is then: "How to choose the damping parameter for the single model, that could be usefully compared to the multi-

band model *ZDS-S10*?". Our choice is to select the single-band model that gives the same fit to the 22.5 s period data as the multi-band model.

We denote $\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3}$ the multi-band model (*ZDS-S10*) obtained from the inversion, with the damping parameter $\lambda = 350$, of the 10, 15, 22.5 and 34 s period data (i.e. $\mathbf{d}_1 + \mathbf{d}_2 + \mathbf{d}_3$, cf. table 4.1). We denote $\mathbf{m}^{\lambda'}_{\mathbf{d}_2}$ the single-band model obtained from the inversion, with the damping parameter λ', of the 22.5 s period data (i.e. $\mathbf{d}_2$, cf. table 4.1). χ^2_{red} is a measure of the misfit between the observed data $\mathbf{d}^{obs}$, and the predicted data $\mathbf{d}^{pred} = G\mathbf{m}$ for the model $\mathbf{m}$ (cf. section 4.2.3), so that we may write $\chi^2_{red}(\mathbf{m}; \mathbf{d}^{obs})$.

Therefore, we select the damping parameter λ', of the single-band model, such as

$$\chi^2_{red}(\mathbf{m}^{\lambda'}_{\mathbf{d}_2}; \mathbf{d}_2) = \chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3}; \mathbf{d}_2) = 1.2082. \tag{4.18}$$

We find that λ'=236 satisfies criterion (4.18).

Finally, we shall now compare the multi-band model $\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3}$ (figure 4.13: the blue dot on the blue curve), with the single-band model, $\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2}$ (figure 4.13: the red dot on the red curve).

4.4.2 Comparing single- and multi-band models

Figure 4.13 shows that the model norm is larger for the multi-band than the single-band model (cf. the two horizontal dotted lines on the "zoom in"). This feature is confirmed with the comparison of the RMS values of these two models, as a function of depth (cf. figure 4.14). Figure 4.14 shows that most differences are located in the shallowest (i.e. 0–100 km depth) and lowermost (below 2,500 km depth) mantle. If we assume[5] that the multi- and single-band models contain the same source-receiver geometry, the model differences should be related to different kernel sampling, new information contained in added data (i.e. $\mathbf{d}_1$ and $\mathbf{d}_3$), etc. In the following, we aim to better quantify the differences between the single- and multi-band models.

4.4.2.a Differences between multi- and single-band data fit

The single- and multi-band models both explain in the same way the intermediate-period data $\mathbf{d}_2$ - equation (4.18). However, the single-band model has no reason to fit the other data ($\mathbf{d}_1$ and $\mathbf{d}_3$) well, unless if all relevant information is already present in the

[5]We recently realized that there is actually ∼5% of seismic phases which have been measured at either 10, 15 or 34 s period, but not at 22.5 s period. We plan to update our codes in the future, in order to handle this discrepancy, although it is unlikely to modify the main results of our "single- vs multi-band tomography" comparison. That is, only strongly under-sampled regions of the mantle should be affected.

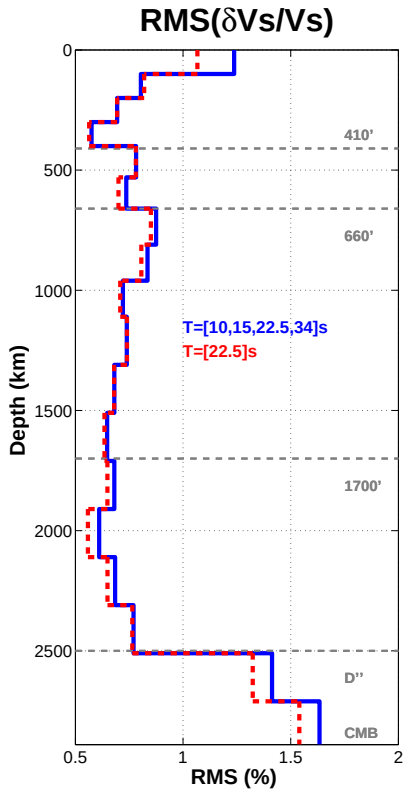


Figure 4.14: Root-mean-square velocity perturbation $\delta V_S/V_S$, versus depth, for both the multi- and single-band models.

22.5 s period data. We therefore compare values of the data misfit measure, $\chi^2_{red}(\mathbf{m};\mathbf{d}^{obs})$, obtained for the single- and multi-band models with the short- and long-period data subsets $(\mathbf{d}_1+\mathbf{d}_3)$. More precisely, three cases are considered:

1. Case a: $\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d}_1+\mathbf{d}_3) > \chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d}_1+\mathbf{d}_3)$
   $\Rightarrow$ The multi-band model better fits $\mathbf{d}_1+\mathbf{d}_3$ than the single-band model. This means that, statistically, the extra data do bear new information.

2. Case b: $\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d}_1+\mathbf{d}_3) \simeq \chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d}_1+\mathbf{d}_3)$
   $\Rightarrow$ Both models provide the same fit to $\mathbf{d}_1+\mathbf{d}_3$. Hence, statistically, no new information is extracted from the extra data.

3. Case c: $\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d}_1+\mathbf{d}_3) < \chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d}_1+\mathbf{d}_3)$
   $\Rightarrow$ The single-band model better fits $\mathbf{d}_1+\mathbf{d}_3$ than the multi-band model. This suggests that something is wrong in the tomographic experiment (data, inversion, etc).

Table 4.2: Comparison of single- and multi-band models.

$\mathbf{d}$	$\chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d})$	$\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d})$	$\delta\chi^2_{red}/\chi^2_{red}$	
$\mathbf{d}_1$ (10 s & 15 s)	1.1683	1.2415	6.3%	
$\mathbf{d}_2$ (22.5 s)	1.2082	1.2082	0%	eq. (4.18)
$\mathbf{d}_3$ (34 s)	1.3015	1.3614	4.6%	
$\mathbf{d}_1+\mathbf{d}_3$	1.2343	1.3033	5.6%	
$\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3$	1.2189	1.2695	4.1%	

Where $\delta\chi^2_{red}/\chi^2_{red} = \{\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d}) - \chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d})\}/\chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d})$.

Table 4.2 shows the $\chi^2_{red}(\mathbf{m};\mathbf{d}^{obs})$ values of both the single- and multi-band models, for different data subsets. We are in the situation of "case a", because $\chi^2_{red}(\mathbf{m}^{\lambda'=236}_{\mathbf{d}_2};\mathbf{d}_1+\mathbf{d}_3)$=1.2343 is lower than $\chi^2_{red}(\mathbf{m}^{\lambda=350}_{\mathbf{d}_1+\mathbf{d}_2+\mathbf{d}_3};\mathbf{d}_1+\mathbf{d}_3)$=1.3033. However, there is only a small difference, 5.6%, between these two values. As we have observed a quite small residual dispersion in our data, we might assume small differences between multi- and single band models. However, this is not necessary true, as it depends on the conditioning[6] of the G matrix.

From table 4.2, we also see that the short-period data ($\mathbf{d}_1$) are, statistically, better fitted by the multi-band model than the long-period data ($\mathbf{d}_3$). This is understandable, if we realize that short-period data are often associated with smaller uncertainties than long-period data (section 4.2.4), due to the frequency content of teleseismic S-waves (cf. section 2.2.2.a, of chapter 2). This means that short-period data are statistically more consistent, so that they are easier to map into the model, and then the model fits them better.

4.4.2.b Differences between multi- and single-band models

We shall now try to quantify the effect on the multi-band model of the new constraints provided by the extra data ($\mathbf{d}_1$ and $\mathbf{d}_3$). In figure 4.15, we plot (in black dots) the mean of $(\delta V_S/V_S)_{\text{multi}} - (\delta V_S/V_S)_{\text{single}}$, in function of $(\delta V_S/V_S)_{\text{single}}$, for six layers spanning the Earth's mantle. We consider the three following cases, before interpreting figure 4.15.

1. Case A: The points (black dots) are located on the y=0 axis.
   ⇒ This means that $(\delta V_S/V_S)_{\text{multi}}$ is statistically the same as $(\delta V_S/V_S)_{\text{single}}$. In this case, the effect on the multi-band model is on average negligible.

[6] If G is not well-conditioned, small data perturbation could lead to either small or large model perturbation.

2. Case B: The points are located in the upper-right and lower-left frames of each plot.
   ⇒ This means that on average we have $(\delta V_S/V_S)_{\text{multi}}/(\delta V_S/V_S)_{\text{single}} > 1$. In this case, the effect on the multi-band model is on average to enhance $\delta V_S/V_S$. That is, a high-velocity (low-velocity) anomaly in the single-band model will be higher (lower) in the multi-band model.

3. Case C: The points are located in the upper-left and lower-right frames.
   ⇒ This means that on average we have $(\delta V_S/V_S)_{\text{multi}}/(\delta V_S/V_S)_{\text{single}} < 1$. In this case, the effect on the multi-band model is, on average, either to decrease $\delta V_S/V_S$, or to change the sign of the anomaly.

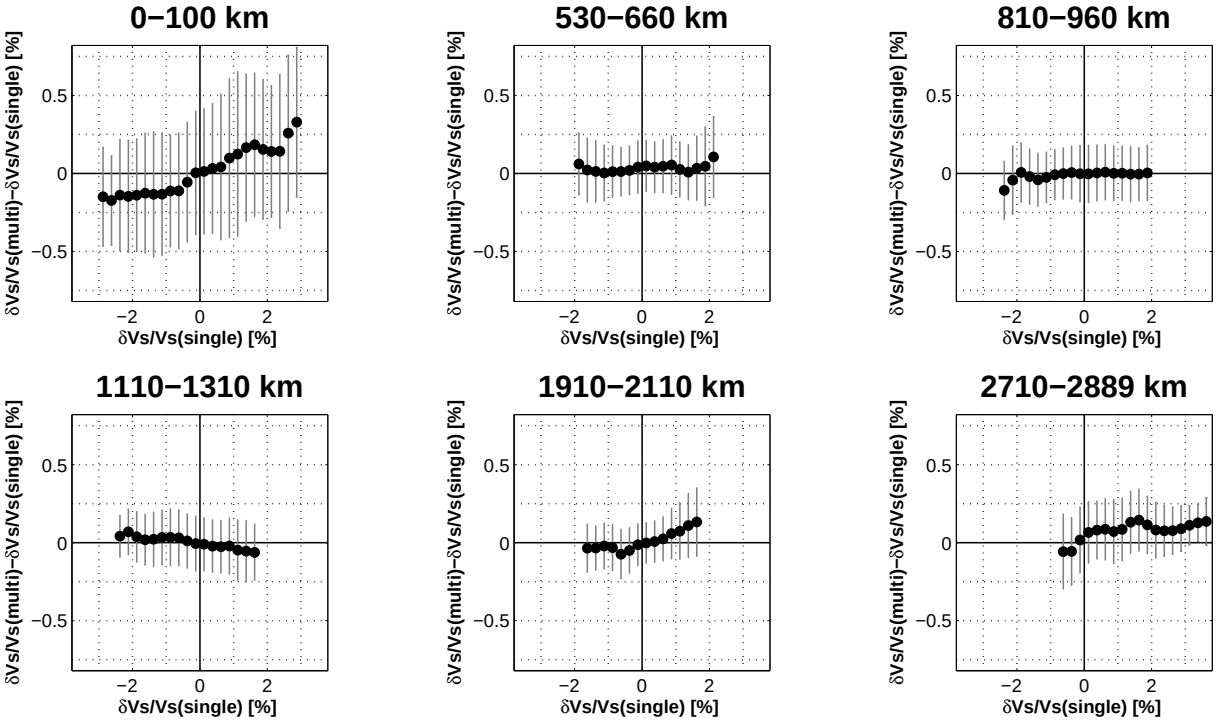


Figure 4.15: Difference of $\delta V_S/V_S$ between the multi- and single-band models, versus $\delta V_S/V_S$ for the single-band model. A larger $\delta V_S/V_S$ is observed for the multi-band model (positive slope) in a few layers (e.g. 0–100 km, 1910–2110 km, 2710–2889 km). In other layers velocity anomalies are on average similar (close to the $y = 0$ axis). However, important 1-σ bars suggest that the effect of extra data may be more significant for individual anomalies, than on average. Black dot: mean of $(\delta V_S/V_S)_{multi} - (\delta V_S/V_S)_{single}$, in function of $(\delta V_S/V_S)_{single}$. 1-$\sigma$ bars (in gray) are computed with bootstrap technique.

Figure 4.15 suggests that:

1. In few layers (e.g. 0–100 km, 1910–2110 km, and 2710–2889 km depth), the effect of extra data is, on average, to enhance $\delta V_S/V_S$ in the multi-band model (case B).

2. In most layers (e.g. 530–660 km, 810–960 km, 1110–1310 km depth, and other layers not shown here) the effect on the multi-band model is, on average, weak. That is, the points (black dots) are oscillating around the $y = 0$ axis (case A). Moreover, the case C is also observed (e.g. 1110–1310 km depth).

3. 1-σ bars show that, for individual anomalies, the effect of extra data on the multi-band model may significantly be larger than the average effect.

The strongest effect of extra data ($\mathbf{d}_1$ and $\mathbf{d}_3$) occurs for the shallowest layer (0–100 km depth) of the multi-band model, where the increase of $\delta V_S/V_S$ is up to $\sim \pm 0.2\%$ on average, and up to $\sim \pm 0.5\%$ locally (figure 4.15). Moreover, the strongest differences between multi- and single-band models, in the 0–100 km depth range, are observed for the largest velocity perturbations (figure 4.15).

4.4.2.c Multi- vs single-band tomography beneath Africa

It is, at first glance, quite difficult to assess the differences between our single- and multi-band models. Therefore, rather than showing our single-band model in the same format as in figure 4.4, we have chosen to present a comparison of both models in a limited region, that is beneath Africa.

Figure 4.16 shows a comparison of the imaging of the complex shear-wave low-velocity anomaly beneath Africa, using multi-band (top row) and single-band data (middle row). We see that, with the color scale used here ($\pm 3\%$), the single- and multi-band models are not that much different, except in the 0–100 km depth layer.

A closer look at the two cross sections shows that the low-velocity anomaly beneath Africa seems to be more connected, from the CMB to the surface, in the multi-band model. However, this difference is subtle, and may also depend on the used colorscale. Therefore, we also show, in figure 4.16 (bottom row), the ratio $(\delta V_S/V_S)_{\text{multi}}/(\delta V_S/V_S)_{\text{single}}$. If this ratio is greater than one, this means that the extra data used in the multi-band model have enhanced the velocity anomalies. For instance, this is observed (figure 4.16, cross-section of bottom row) in some regions, such as near the CMB where the low-velocity anomaly seems to be rooted.

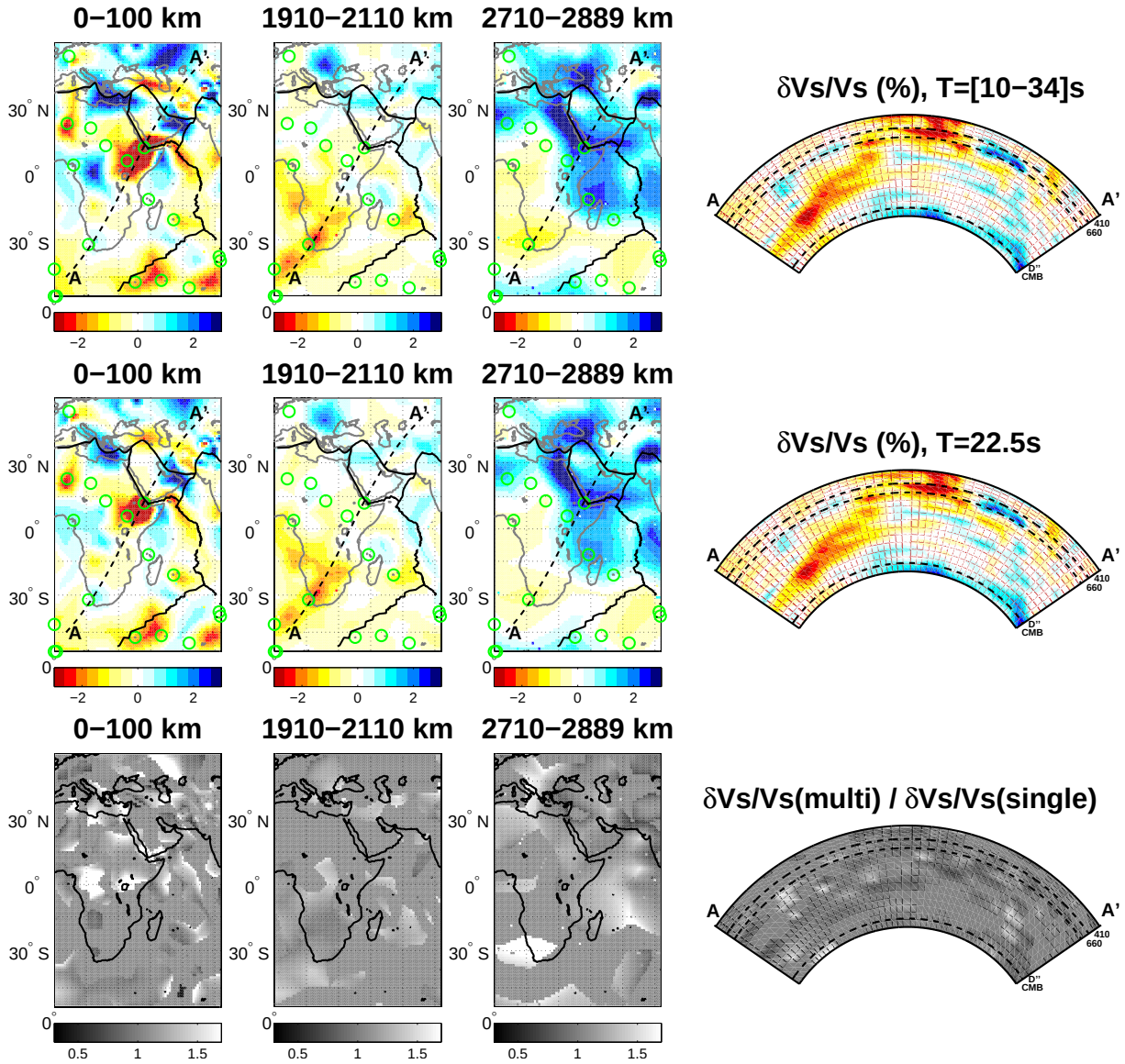


Figure 4.16: Enhancing velocity anomalies using multi-band data. Top row: multi-band model $(\delta V_S/V_S)_{multi}$. Middle row: single-band model $(\delta V_S/V_S)_{single}$. Bottom row: ratio $(\delta V_S/V_S)_{multi}/(\delta V_S/V_S)_{single}$, computed for anomalies greater than 0.4% (in absolute value). Ratio values greater than one (white) means an enhanced model perturbation in the multi-band model.

4.5 Conclusion

In this chapter, we have presented a global 3-D *SH*-wave velocity perturbation tomographic model, *ZDS-S10*, of the Earth's mantle. This multiple-frequency model has been obtained from a "finite-frequency" (Dahlen *et al.* 2000) inversion of ∼300,000 *S*, *ScS* and *SS* travel times measured at 10, 15, 22.5 and 34 s periods, using a data driven parameterization. Large scale anomalies are well retrieved in *ZDS-S10*, from the surface to the lowermost mantle. Though more investigations are needed, we believe that *ZDS-S10* could also provide complementary imaging of small scale features present in the mantle (cf. appendix F).

A comparison of single- versus multi-band tomographic models has been presented. Even if a structural dispersion is observed in our data, it has proven to be difficult to assess the differences between single- and multi-band models. Our results show that the multi- and single-band models are on average quite similar. However, the effect of simultaneously using multiple-frequency data in an inversion is not negligible. That is, it may significantly increase the contrast of individual anomalies (especially for those with significant amplitude). This could potentially help us to refine the tomographic imaging of small scale seismic heterogeneities in the mantle.

Finally, this tomographic study can certainly be improved in several aspects (e.g. earthquake mislocation uncertainties, data weighting in the inversion, etc). Such improvements should allow us to consider less damped models. One may then expect that the structural dispersion observed in multiple-frequency data would be better exploited, so that a multi-band model would contain significantly more information, on the 3-D elastic structure of the mantle, than a single-band model.

Conclusions

General conclusion

For better constraining the structure of the Earth's interior, new theoretical developments on seismic wave propagation have emerged in recent years, and received increasing attention in tomography. One of these new methods is the "multiple-frequency tomography", that we have applied in this study. Our aim was to obtain a "high-resolution" *SH*-wave tomographic model of the Earth's mantle, that could contribute to a better understanding of mantle dynamics.

Firstly, we have built a globally distributed dataset of ∼400,000 frequency-dependent *SH*-wave travel times. An automated technique is used to measure teleseismic *S*, *ScS* and *SS* travel times at several periods ranging from 10 to 51 s. The measurement scheme, which comprises two main parts, makes use of WKBJ synthetic seismograms (Chapman 1976), calculated with ray theory. The first stage is an automated selection of time windows around the target phases present on both the observed and synthetic seismograms. The second stage involves measurements of multiple-frequency travel times by cross-correlating the selected observed and synthetic filtered waveforms. Frequency-dependent effects due to crustal reverberations beneath each receiver are handled by incorporating crustal phases into synthetic waveforms.

After common correction for physical dispersion due to intrinsic anelastic processes (Kanamori & Anderson 1977), we observe a residual travel time dispersion on the order of 1–2 s in the period range of analysis. This dispersion occurs differently for *S*, *ScS* and *SS*, which is presumably related to their differing paths through the Earth. Our results show that:

1. Wavefront-healing phenomenon, produced by very low velocity anomalies, is observed in our *S* and, to a lesser extent, *SS* travel times.

2. A preferred sampling of high velocity scatterers located at the CMB may explain our observation that *ScS* waves travel faster at low-frequency than at high-frequency.

3. The globally averaged dispersion observed for *S* and *SS* travel times favor a frequency-dependent attenuation model $q(\omega) \propto q_0 \times \omega^{-\alpha}$, with an α value of $\simeq 0.2$ for *S* waves and $\simeq 0.1$ for *SS* waves.

Our results therefore suggest that the residual dispersion observed in our data is, at least partly, related to seismic heterogeneity and attenuation in the Earth's interior.

Secondly, in order to exploit this *structural* dispersion contained in our global dataset, we have used a "finite-frequency" approach to incorporate this new observable

in tomography, because ray theory cannot take into account this frequency-dependency. We have then built a multiple-frequency *SH*-wave tomographic model of the Earth's mantle, from the inversion of ∼300,000 *S*, *ScS* and *SS* travel times measured at 10, 15, 22.5, and 34 s periods, using the (first-order, linear) "finite-frequency" formalism of Dahlen *et al.* (2000) and a data driven parameterization. The resulting "multi-band" shear-wave velocity model, named *ZDS-S10*, seems to be consistent with other studies. For instance, the first-order convection occuring in the lower mantle is well retrieved in our model. Though more investigations are needed, we feel that *ZDS-S10* could provide complementary imaging of small-scale features present in the mantle (e.g. parts of slabs sinking in the mantle, hot rising plumes, etc), that are still debated in the community.

Thirdly, we have presented a comparison of single- versus multi-band tomographic models, in order to assess, from a model point of view, the actual benefits of using multi-band rather than single-band data. Our results show that the multi-band model is not very different, on average, from the single-band one. However, the effect of using multi-band data in a single inversion is not negligible, as it allows to increase the contrast of some individual velocity anomalies, especially for those with significant amplitude. Though more investigations are also needed, we believe that multiple-frequency tomography could help us to refine imaging of small-scale objects in the Earth's interior (e.g. plumes), which may play a key role in mantle dynamics.

To conclude, this thesis explicitly shows that data processing (seismograms recording, measurement method, etc) is accurate enough for retrieving a "structural" dispersion, at least partly, in global *SH*-wave travel times. This justifies, from a data point of view, the need for a "finite-frequency" approach that may, in contrast to ray theory, take into account diffraction effects undergone by teleseismic waves. Moreover, as this tomographic study can certainly be improved, we may expect to better exploit this observed structural dispersion, and thus maximize the benefits of inverting multi-band data in the deep Earth imaging.

Future directions

The success of a tomographic study depends on our ability to extract relevant information from seismograms and to map it into the model. Here we propose several directions for improving the 3-D *SH*-wave tomographic model of the mantle presented in this thesis.

1. *Data coverage*

   Data coverage clearly depends on the source-receiver geometry, which is far from uniformly distributed, and on the type of seismic waves used. In this study, we use

S, *ScS* and *SS* waves, which allow us to achieve a good sampling of the Northern Hemisphere, for all but the rays with the shallowest lower mantle turning depths. However, coverage in the Southern Hemisphere still remains a problem. In order to improve the coverage of the two "superplumes", near the CMB beneath Southern Africa and Pacific, we plan to use other seismic phases, such as *Sdiff*, *SKS*, etc. Nevertheless, if one aims to associate to each phase its appropriate sensitivity kernel, alternative kernels should be used (e.g. Calvet & Chevrot 2005).

As the spatial distribution of earthquakes is unlikely to vary a lot in the near future, we can only try to spatially improve the receivers distribution. Receivers mainly consist of many terrestrial seismometers, installed on continents and a few islands, and of a small amount of Ocean Bottom Seismometers (OBS), deployed on the ocean's floor. As the International Federation of Digital Seismograph Networks (FDSN) is slowly growing up, seismologists are currently investigating alternatives to conventional receivers. For instance, Simons *et. al* (2006) showed that it is possible to record earthquakes in oceans, with an hydrophone mounted on a freely floating diver and submerged at 700 m depth. As the floating diver is drifted by ocean currents, such Ocean Seismometers (OS) could provide more uniform sampling in oceanic regions.

2. *Earthquake mislocation and data weighting*

   A lot of the observed variance in our data could be due to event mislocation. It is then important to treat this properly, either in the inversion stage (e.g. Bolton & Masters 2001; Houser *et. al* 2008), or in the measurement stage (e.g. Sigloch & Nolet 2006). Several options to tackle this problem have been discussed in chapter 4 (section 4.2.5). A better treatment of earthquake mislocation should lead to a better assessment of the related error, σ_{source}, on the travel times.

   As the actual measurement error, $\sigma_{measure}$, have already been determined, we could take into account the proper uncertainty of each datum, that is $\sigma_{source} + \sigma_{measure}$. Thus, a better data weighting should be achieved in the inversion, and less damped models could be explored. One may then expect that the structural dispersion observed in multiple-frequency data would be better exploited, so that a multi-band model would contain significantly more information, on the 3-D elastic structure of the mantle, than a single-band model.

3. *Surface waves*

   Surface wave data, that are generally analysed at very long periods (∼40–300s), may be combined with *S*-wave travel times measured at long periods (∼10–51 s).

This provides a way to image the entire mantle, because surface waves are more sensitive to the upper mantle, and S-waves to the lower mantle. In the context of the TOMOGLOB project, we plan to jointly invert our global SH-wave dataset with Love waves. The joint inversion will be based on the "finite-frequency" formalism of Dahlen *et al.* (2000) for body waves, and on ray theory for surface waves. The resulting 3-D model, from the surface to the CMB, will essentially have a better resolution in the upper mantle, compared to *ZDS–S10*, because of the addition of surface waves. Our expectations are that this improved model will benefit to a broader community of geoscientists (geodynamics, geochemistry and mineral physics) by providing new seismological constraints on geodynamical models of the Earth's mantle.

4. *Beyond Born theory*

   It is our view that the previously mentioned future directions '2' and '3' should be tackled first, in order to "quickly" improve our model *ZDS-S10*. If significant progress are effectively obtained, we could then consider to use a more accurate "finite-frequency" approach.

   The "finite-frequency" formalism (Dahlen *et al.* 2000) we have used here is a first order and linear theory: multiple-scattering is neglected, and the non-linear part of diffraction effects (e.g. wavefront-healing) cannot be modeled. Moreover, though extremely fast, the use of ray theory to compute the kernels cannot handle headwaves or diffracted waves (e.g. *Sdiff*). As discussed in chapter 1 (section 1.5), those kernels are also computed under the paraxial approximation, which breaks down for long distance waves (e.g. *SS*) near the antipode, and do not take into account near-field effects.

   Alternatives methods, which may overcome some of the limitations raised in the Born theory (linearity and single-scattering), are available (e.g. Marquering *et al.* 1998; Zhao *et al.* 2000; Komatitsch *et. al* 2002; Calvet & Chevrot 2005; Tromp *et al.* 2005; Nissen-Meyer & Dahlen 2007). These methods enable us to compute more accurate Fréchet kernels. Unfortunately, they are often too much time consuming, making an application to large data set not (easily) feasible, in the period range of this study (10–51 s).

   We feel that, in the next decade, supercomputer facilities will make them more attractive in global tomography. Nevertheless, will they really make a difference? In the following, we give some thoughts on this pragmatic question.

5. *A philosophical strategy for future global tomography*

   We intuitively think that, for improving future global tomography, one should not go too fast at building new 3-D models with more sophisticated theory. That is, it would be clever to check first if we can, or cannot, retrieve a similar "structural" dispersion pattern in accurate 3-D synthetic seismograms[7], as observed in real data[8].

   Let us assume that the residual dispersion observed in real data does contain structural information, and that wave diffraction effects (e.g. scattering, wavefront-healing) can be modeled in 3-D synthetic seismograms. It would then be interesting to generate synthetic seismograms in our 3-D "multi-band" model *ZDS-S10* (with the same source-receiver geometry). That is, one could obtain a corresponding synthetic frequency-dependent dataset, using the same measurement scheme as for real data. Finally, one could easily verify if the structural dispersion, that we have observed in real data (cf. section 2.3, of chapter 2), is also found in synthetic data.

   If the same kind of dispersion is not observed in the synthetic data, this would suggest that different factors, in the inversion and/or the "finite-frequency" theory, still prevent us from "mapping" in *ZDS-S10* the structural information observed in real data. Theoretically, the "finite-frequency" theory of Dahlen *et al.* (2000) takes into account first-order, linear, wave diffraction effects (e.g. Hung *et al.* 2001; Dahlen 2004). Therefore, it is more likely that our unability in exploiting the structural dispersion of the data is mainly owing to the inversion itself (e.g. regularization, data weighting, under-determined problem, etc). In this case, using more accurate Fréchet kernels should not be a priority for future global tomography. One should rather focus on improving the inversion procedure and the data coverage.

   If a similar dispersion pattern (e.g. wavefront-healing effect) is observed in the synthetic data, this would suggest that both the inversion and the "finite-frequency" theory (Dahlen *et al.* 2000) may allow us to exploit the structural information contained in frequency-dependent S-wave travel times. This would mean that the 3-D heterogeneities, which produced the structural dispersion are, at least partly, well mapped in *ZDS-S10*. In this case, this would be more justified to use, in future global tomography, more accurate Fréchet kernels than "banana-doughnut" kernels (Dahlen *et al.* 2000), though they may be less digestible!

[7]For instance, spectral element methods enable us to accurately calculate synthetic seismograms in realistic 3-D Earth models, though it is still very much time consuming, in our period range of interest (e.g. Komatitsch *et al.* 2002; Tromp *et al.* 2005). Such synthetics are here denoted "3-D synthetic seismograms".

[8]Here, "real data" means our global dataset of frequency-dependent SH-wave time residuals, measured in the 10–51 s period range (cf. section 2.2, of chapter 2).

APPENDIX A

On the frequency-dependency of scattering

Here, one aims to show that, at first order, one may expect that high (low) velocity scatterers should lead to a decreasing (increasing) dispersion curve $\delta t_m(T)$, respectively.

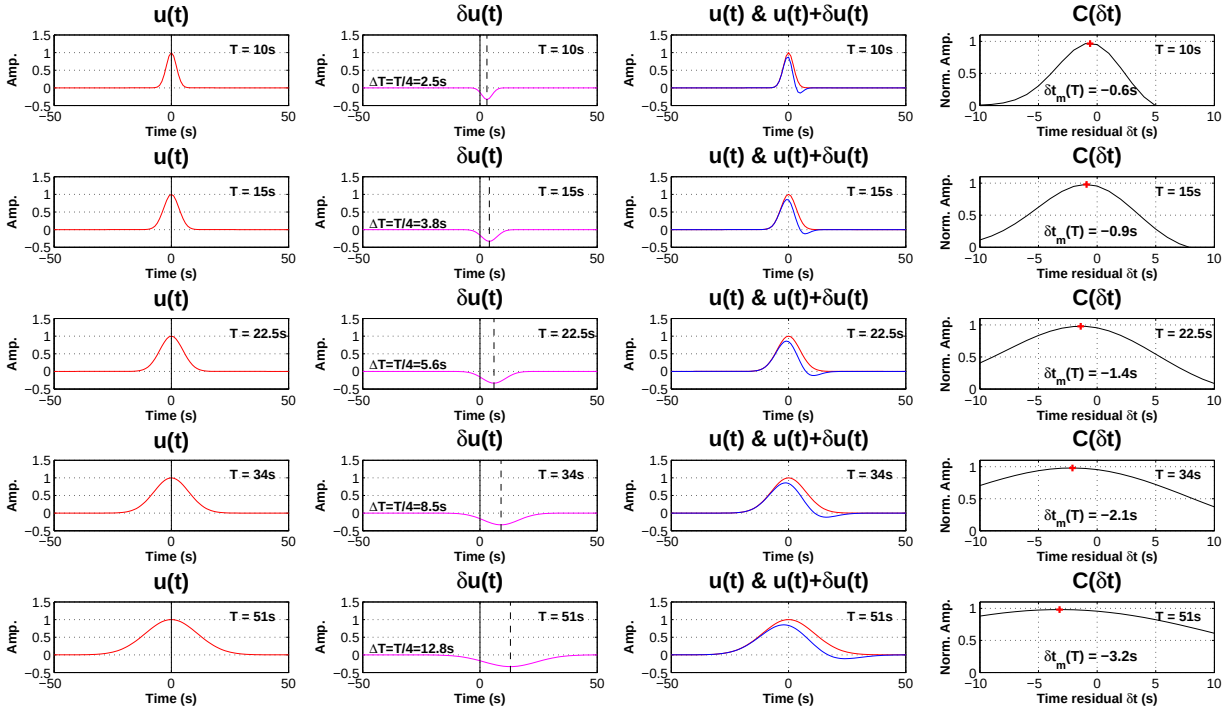


Figure A.1: Each raw corresponds to a different dominant period T of the signal. First column: synthetic waveform $u(t)$ (in red). Second column: scattered waveform $\delta u(t)$ (in magenta). Third column: synthetic waveform $u(t)$ (in red) and observed waveform $u(t) + \delta u(t)$ (in blue). Fourth column: the time residual, $\delta t_m(T)$, maximizes the cross-correlation between the synthetic and observed waveforms. Here, $u(t)$ is approximated by a Gaussian source time function (Komatitsch & Tromp 2002), simulating the ground displacement, with a half duration T. In the Born approximation, we consider that $\delta u(t) = -sign(\delta V_S/V_S) \times 1/3 \times u(t-\Delta T)$. Finally, we see that, at first order, high velocity (i.e. $\delta V_S/V_S > 0$) scatterers should lead to a decreasing dispersion curve $\delta t_m(T)$.

In section 2.3.2 (chapter 2), we have seen that the observed[1] waveform, $u(t)+\delta u(t)$, is expected to be dominated by arrivals of scattered waves, $\delta u(t)$, with detour times $\Delta T(T) \simeq T/4$, where T is the dominant period of the wave. If there is no phase shift, one may assume that the scattered wave, $\delta u(t)$, preserves the shape of the synthetic[2] waveform, $u(t)$, so that they will only differ by their amplitudes. In the Born approximation, one may also consider

[1]We assume that the observed signal at the receiver, $d(t)$, consists of the direct, $u(t)$, and scattered, $\delta u(t)$, wave arrivals, so that $d(t) = u(t) + \delta u(t)$.

[2]As the synthetic signal, $s(t)$, is calculated with ray theory, in a smooth reference model, it does not account for wave diffraction, so that $s(t) = u(t)$.

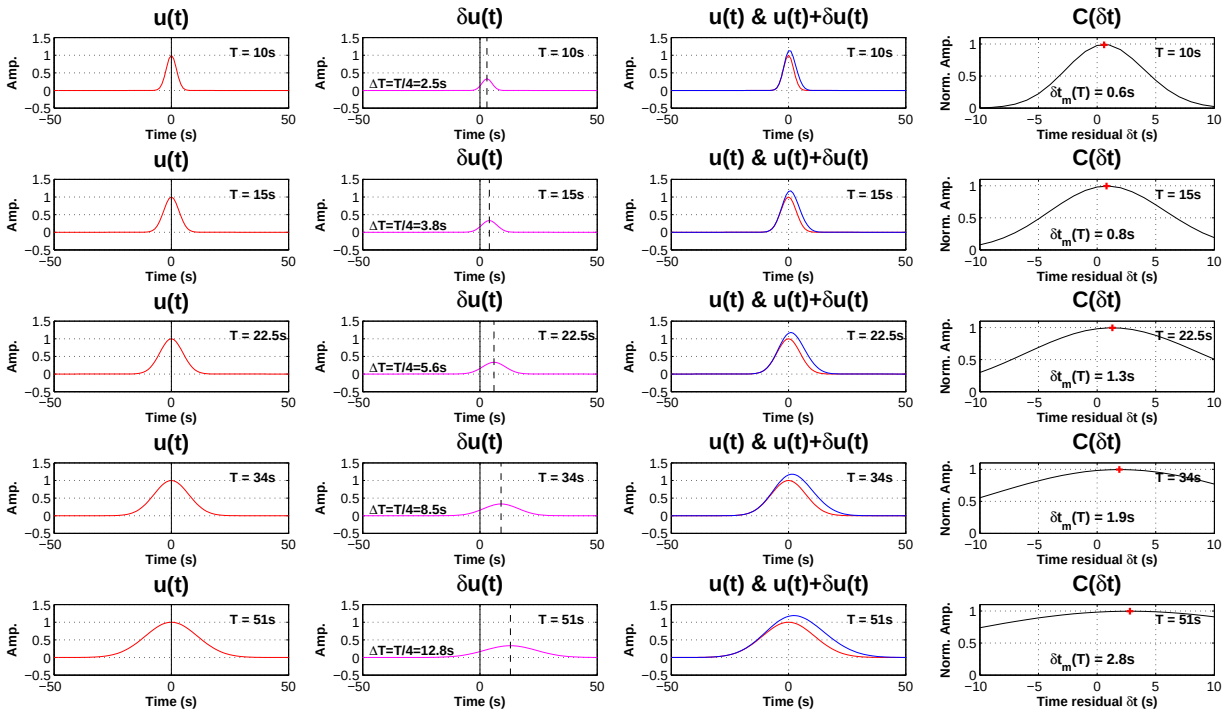


Figure A.2: Same as figure A.1, in the case of low-velocity scatterers (i.e. $\delta V_S/V_S < 0$) which should lead to an increasing dispersion curve $\delta t_m(T)$.

that the amplitude of $\delta u(t)$ is a small fraction of the amplitude of $u(t)$. Therefore, we may consider the two following cases, in order to characterize the frequency-dependent effects of single-scattering (cf. section 1.3.2, of chapter 1) as a function of the sign of the velocity anomaly $\delta V_S/V_s$ of the scatterer, in the 10–51 s period range.

1. Let the polarity of the scattered wave be negative, as for a high velocity anomaly. The contribution of $\delta u(t)$ is then to decrease the amplitude of $u(t) + \delta u(t)$ around the time $t \simeq \Delta T(T) \simeq T/4$, compared to $u(t)$. This leads to an "advancing" effect on the time residual δt_m (cf. figure 1.7, of chapter 1), which is expected to increase with the period T (cf. Figure A.1), and give a decreasing dispersion curve $\delta t_m(T)$.

2. Let the polarity of the scattered wave be positive, as for a low velocity anomaly. The contribution of $\delta u(t)$ is then to increase the amplitude of $u(t) + \delta u(t)$ around the time $t \simeq \Delta T(T) \simeq T/4$, compared to $u(t)$. This leads to a "delaying" effect on the time residual δt_m (cf. figure 1.7, of chapter 1), which is expected to increase with the period T (cf. Figure A.2), and give an increasing dispersion curve $\delta t_m(T)$. In such low velocity regions, we also expect that wavefront-healing (cf. section 2.3.1, of chapter 2) and scattering effects are competing.

APPENDIX B

Outline of the optimization of the parameterization

Here, we give the outline of the optimization of the nodes agency, which is based on the minimization, with a conjugate gradient, of the penalty function defined in section 3.2.3.a (chapter 3) - i.e. equation (3.5).

INPUT:

$\Rightarrow$ Resolution function $\ell(\mathbf{r})$ defined in each layer z;
$\Rightarrow$ Starting set of nodes $S_z^0(\theta, \phi)$ in each layer z.

OPTIMIZATION ALGORITHM:

For each layer z, do the following:

Set 'convergence' to 0 and let i the iteration number.

While 'convergence' is equal to 0
$\cdots \Rightarrow$ Compute the Delaunay mesh of the set of nodes S_z^i
$\cdots \Rightarrow$ Calculate the energy value E_z^i of S_z^i - equation (3.5)
$\cdots$**if** $|E_z^i - E_z^{i-1}| < \eta$ (i.e. the refining process is good enough, given the precision η)
$\cdots\cdots \Rightarrow$ Set 'convergence' to 1 (i.e. stop the refining process)
$\cdots$**otherwise**
$\cdots\cdots \Rightarrow$ Calculate the derivatives $(\frac{\partial E_z^i}{\partial \theta}, \frac{\partial E_z^i}{\partial \phi})$
$\cdots\cdots \Rightarrow$ Use conjugate gradient to find a new set of nodes S_z^{i+1} with lower energy
$\cdots\cdots \Rightarrow$ Set i to $i+1$
$\cdots$**end**
end

OUTPUT:

$\Rightarrow$ A globally irregular parameterization of the Earth's mantle, i.e. a set of nodes $S_z(\theta, \phi)$ for each layer z, that is optimally driven by the expected "resolution length" of the data.

APPENDIX C

The "Where am I" algorithm

Our global Spherical Triangular Prism (STP) parameterization of the mantle consists of 18 constant-depth spherical layers, ranging from the surface to the CMB, filled in with irregularly spaced spherical triangular prisms. Though it is straightforward to find in which layer z is contained a given point $\mathbf{p}$, we need an efficient algorithm to find which is the prism that encloses $\mathbf{p}$.

C.1 Locating a point in a spherical Delaunay mesh

Consider the Delaunay mesh of one layer z of our model parameterization, which consists of spherical triangles. Here, we aim to find in which cell of this mesh is located the query point $\mathbf{p}$. Barycentric coordinates (cf. section C.5) are useful for solving this point location problem. That is, if one computes the barycentric coordinates (α, β, γ) of $\mathbf{p}$ with respect to the triangle f, we can test if f does contain the query point; this is the case if $\alpha > 0$, $\beta > 0$ and $\gamma > 0$. The brutal method, for finding in which triangle sits the query point, is then to apply this test to all triangles of the mesh. However, in large scale problem, this is too much time consuming and thus prohibited.

In this thesis we have used an alternative method, from Wu *et al.* (2005), for solving the spherical point location problem. This approach heavily relies on the *walking triangle algorithm* from Sambridge *et al.* (1995), whose the outline is:

1. Compute the neighbor triangles N_f of each triangle f.
2. Select a start triangle f_i.
3. Compute the barycentric coordinates (α, β, γ) of $\mathbf{p}$ with respect to f_i (cf. sections C.2 & C.5).
4. According to (α, β, γ), choose the next triangle $f_{i+1} \in N_{f_i}$ which is the "closest" to the query point $\mathbf{p}$ (cf. section C.4).
5. Steps '3' and '4' are executed iteratively, until the current triangle does contain the query point $\mathbf{p}$, i.e. $\alpha > 0$, $\beta > 0$ and $\gamma > 0$.

Therefore, the main idea is to "walk" from the starting triangle towards the triangle containing the query point, so that we only test a limited number of triangles, which drastically reduces the computing time. In the approach[1] of Wu *et al.* (2005), such "walking" is done

[1]The *spherical point location* algorithm of Wu *et. al* (2005) was developed in the field of "virtual reality". That is, spherical parameterization based 3-D geometry applications have recently received much interest. The basic idea is to transform original irregular meshes into regular ones by spherical remeshing methods.

on a sphere, which is briefly presented in the following. In section C.3, we show how to select an optimal start triangle to walk (i.e. step '2'), in the case of our tomographic study.

C.2 Computing the barycentric coordinates on a sphere

If the query point $\mathbf{p}$ was contained in the plane defined by the triangle $f(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$, the barycentric coordinates (α, β, γ) of $\mathbf{p}$ with respect to f would be defined as

$$\begin{cases} \mathbf{p} = \alpha\mathbf{p}_0 + \beta\mathbf{p}_1 + \gamma\mathbf{p}_2 \\ \alpha + \beta + \gamma = 1 \end{cases} \tag{C.1}$$

where p_0, p_1 and p_2 are the three vertices of f. However, if $f(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$ is a spherical triangle, we cannot compute the barycentric coordinates with equation (C.1). One should rather use the corresponding planar triangle $f'(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$ (cf. figure C.1). Because $\mathbf{p}$ may be located everywhere on the sphere, it may be far away from the plane defined by f', denoted $P_{f'}$ in the following. Thus, one should also consider the projection **p'** of $\mathbf{p}$ onto $P_{f'}$ (cf. figure C.1). Therefore, we can compute the barycentric coordinates of **p'** with respect to f' using equation (C.1).

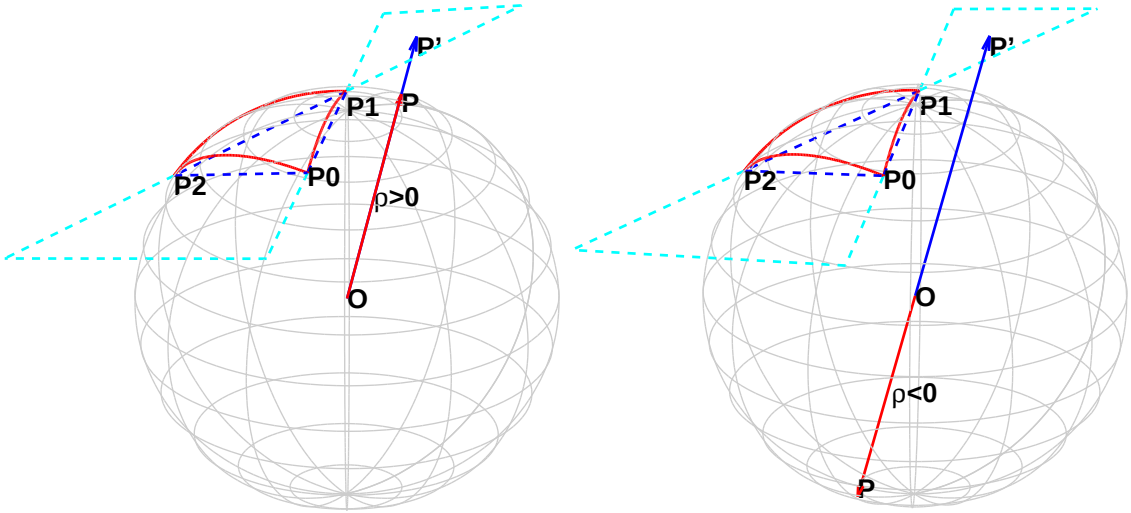


Figure C.1: Projection of the query point $\boldsymbol{p}$ *onto the plane* $P_{f'}$ *(in cyan), which gives the point* **p'***. (Left) Case* $\rho > 0$*. (Right) Case* $\rho < 0$*. Modified from Wu et.al (2005).*

Considering that the general equation of the plane $P_{f'}$ is $Ax + By + Cz + D = 0$ (cf. section C.6), the intersection point **p'** can be computed as

$$\mathbf{p'} = \rho\frac{\overrightarrow{op}}{\|\overrightarrow{op}\|} = \rho(l, m, n)^t \quad , \quad \text{with} \quad \rho = -\frac{D}{Al + Bm + Cn}. \tag{C.2}$$

C.3 Choosing the next triangle to walk

At this stage, we need to choose the next triangle to walk towards the query point. Here, we briefly explain how to do it, according to the value of ρ - equation (C.2) - and the barycentric coordinates (α, β, γ) of **p'** with respect to f'.

1. $\underline{\rho > 0}$
   In this case, $\overrightarrow{op'}$ has the same direction as $\overrightarrow{op}$ (cf. figure C.1).

   (a) All components of (α, β, γ) are positive (cf. figure C.2.a).
   In this case, **p'** lies in f', so that the corresponding spherical triangle f of f' is the target triangle containing the query point **p**.

   (b) One component of (α, β, γ) such as α is negative (cf. figure C.2.b).
   In this case, **p'** lies in the right of the opposite edge $\{\mathbf{p}_1, \mathbf{p}_2\}$ of $\mathbf{p}_0$, so that we select the triangle which is linked with f by the edge $\{\mathbf{p}_1, \mathbf{p}_2\}$ as the next one to walk.

   (c) Two components of (α, β, γ) such as α and β are negative (cf. figure C.2.c).
   In this case, **p'** lies in the intersection region of the two right hemispheres defined by the edges $\{\mathbf{p}_1, \mathbf{p}_2\}$ and $\{\mathbf{p}_2, \mathbf{p}_0\}$, respectively. We find the unvisited 1-ring neighbor[2] triangles set $T^*(\mathbf{p}_2)$ of the vertex $\mathbf{p}_2$, and sort them by anticlockwise order[3]. We then select the $\lfloor \frac{N^*(\mathbf{p}_2)}{2} \rfloor$ *-th* triangle of $T^*(\mathbf{p}_2)$ as the next one to walk, where $N^*(\mathbf{p}_2)$ is the number of triangles in $T^*(\mathbf{p}_2)$, and $\lfloor x \rfloor$ are Gaussian brackets denoting the integer value of x.

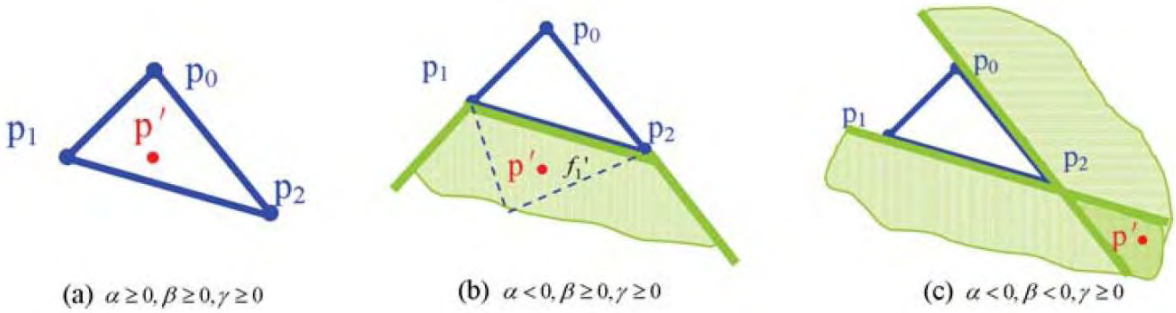


Figure C.2: Three kind of relationships between p' and $f'(p_0, p_1, p_2)$ for selecting the next triangle to walk. From Wu et.al (2005).

[2]The 1-ring neighbor triangles set $T^*(\mathbf{p}_2)$ comprises all the triangles that have the vertex $\mathbf{p}_2$ in common.
[3]Here, one assumes that the vertices of each triangle are anticlockwise ordered.

2. $\underline{\rho < 0}$
   It means that the vector direction of $\overrightarrow{op'}$ is contrary to that of $\overrightarrow{op}$ (cf. figure C.1), so that the strategy for choosing the next triangle is contrary to the case $\rho > 0$.

   (a) All components of (α, β, γ) are positive.
   In this case, p lies in the opposite hemisphere surface of f. We randomly select one unvisited neighbor triangle of f as the next one to walk.

   (b) One component of (α, β, γ) such as α is negative.
   In this case, we make the supposition that $\alpha \geq 0$, $\beta < 0$, $\gamma < 0$ and use the same approach as illustrated in figure C.2.c.

   (c) Two components of (α, β, γ) such as α and β are negative.
   In this case, we make the supposition that $\alpha \geq 0$, $\beta \geq 0$, $\gamma < 0$ and use the same approach as illustrated in figure C.2.b.

Finally, the figure C.3 shows an example of the "walk" from the start triangle towards the target triangle containing the spherical point $\mathbf{p}$ to be located.

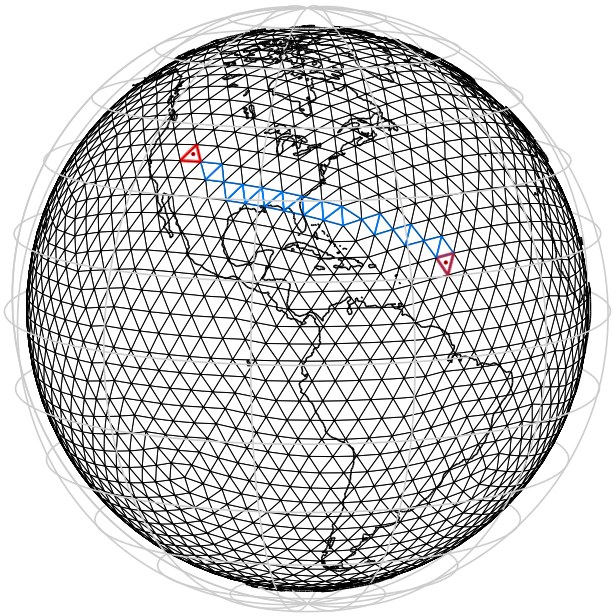


Figure C.3: Locating a point in a spherical Delaunay mesh Red triangle on the left/right: start/final triangle of the "walk". Blue triangles: all triangles tested by the algorithm during the 'walk' to the query point. In black: Delaunay mesh corresponding to an icosahedron iterated four times.

C.4 Choosing the start triangle

If one aims to minimize the computing time for locating the query point, we clearly need to choose a start triangle that is close enough to the target triangle. In the following, we explain how to choose such a good start triangle in the context of this tomographic study. Our strategy differs from Wu *et al.* (2005), who construct a new spherical mesh with subdivision connectivity, for partitioning the spherical domain into some subdivision regions. Here, we use a simpler subdivision, which is still very efficient in term of computing time.

First, we construct a regular $10^\circ \times 10^\circ$ mesh, which partitions each spherical layer z into $36 \times 18 = 648$ subdivision regions. We compute the centroid point $\mathbf{q}$ of each $(10^\circ \times 10^\circ)$ subdivision region. We use the barycentric coordinates approach to find the spherical triangle, of our STP parameterization, containing $\mathbf{q}$, which is named the "representation triangle" of the associated subdivision region.

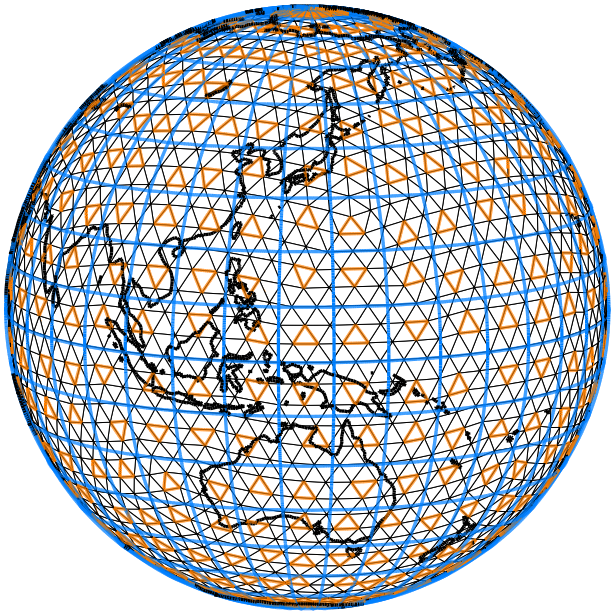


Figure C.4: Representation triangles (in orange) in $10^\circ \times 10^\circ$ *cells (in blue). Black triangles represent our STP parameterization, but in fact they correspond to an icosahedron iterated 4 times.*

In this thesis, we aim to locate a large amount of points $(\mathbf{p}_1, \cdots, \mathbf{p}_n)$ corresponding to 3-D sensitivity kernels calculated on a very fine 'kernel grid" (cf. section 3.4.2, of chapter 3). Therefore, a large majority of these kernel points are adjacent or close enough, so that one may use the triangle containing the point $\mathbf{p}_{i-1}$ as the start triangle for locating the

point $\mathbf{p}_i$.

However, this interesting property is lost if two triangles, though they may be very close, belong to two different layers of our STP parameterization. In this case, we choose, as the start triangle, the "representation triangle" of the subdivision region containing the point to be located. In this study, this case often happens, so that we have preffered to use a "representation triangle", rather than a randomly chosen triangle, as a start triangle. Though this choice should reduce the computing time, it is also likely not to make a large difference.

C.5 Calculating barycentric coordinates in 3-D

Here, we aim to show how to calculate the barycentric coordinates of a point $\mathbf{p}$, with respect to the triangle $f(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$. We suppose that the labeling of the $\mathbf{p}_i$ is *counterclockwise*. Figure C.5 shows the situation. The barycentric coordinates (α, β, γ) are defined by the equation (C.1), and represent the "signed" sub-areas A_1, A_2, A_3, respectively, as illustrated in figure C.5.a.

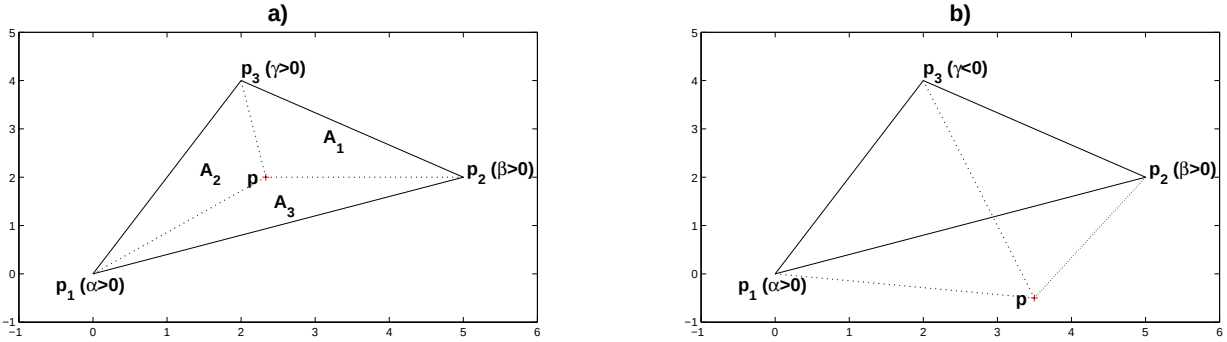


*Figure C.5: The barycentric coordinates of a point $\boldsymbol{p}$, with respect to the triangle $f(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$, are defined as ratios of triangle areas. **a)** Case where all the barycentric coordinates are positive, i.e. the point $\boldsymbol{p}$ is inside the triangle. **b)** Case where at least one barycentric coordinate is negative, i.e. the point $\boldsymbol{p}$ is outside the triangle.*

First, we form the vectors $\mathbf{v} = \mathbf{p}_2 - \mathbf{p}_1$ and $\mathbf{w} = \mathbf{p}_3 - \mathbf{p}_1$. The cross-product of these two vectors is $\mathbf{u} = \mathbf{v} \times \mathbf{w}$. The area of the parallelogram formed by $\mathbf{v}$ and $\mathbf{w}$ is then $A = \|\mathbf{u}\|$. For the sub-areas calculation A_i, we construct $\mathbf{u}_i$ as

$$\begin{cases} \mathbf{u} = (\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_3 - \mathbf{p}_1) & \Rightarrow \quad A = \|\mathbf{u}\| \\ \mathbf{u}_1 = (\mathbf{p}_2 - \mathbf{p}) \times (\mathbf{p}_3 - \mathbf{p}) & \Rightarrow \quad A_1 = \|\mathbf{u}_1\| \\ \mathbf{u}_2 = (\mathbf{p} - \mathbf{p}_1) \times (\mathbf{p}_3 - \mathbf{p}_1) & \Rightarrow \quad A_2 = \|\mathbf{u}_2\| \\ \mathbf{u}_3 = (\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p} - \mathbf{p}_1) & \Rightarrow \quad A_3 = \|\mathbf{u}_3\| \end{cases} \tag{C.3}$$

Taking the dot-product, $s_i = \mathbf{u} \cdot \mathbf{u}_i$, results in a value that determines the sign of the area A_i. If $s_i > 0$, then the area A_i is positive; if $s_i < 0$, then the area A_i is negative. Therefore,

$$\alpha = \text{sign}(\mathbf{u} \cdot \mathbf{u}_1) \cdot \frac{A_1}{A}, \quad \beta = \text{sign}(\mathbf{u} \cdot \mathbf{u}_2) \cdot \frac{A_2}{A}, \quad \gamma = \text{sign}(\mathbf{u} \cdot \mathbf{u}_3) \cdot \frac{A_3}{A}. \tag{C.4}$$

C.6 Calculating the general equation of the plane $P_{f'}$

We have previously seen that the general equation of the plane $P_{f'}$, containing the planar triangle $f'(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$, is given by: $A \cdot x + B \cdot y + C \cdot z + D = 0$. Here, we aim to show how to calculate A, B, C, and D. We note:

$$\begin{cases} \mathbf{p}_0 = (x_{p_0}, y_{p_0}, z_{p_0})^t \\ \mathbf{p}_1 = (x_{p_1}, y_{p_1}, z_{p_1})^t \\ \mathbf{p}_2 = (x_{p_2}, y_{p_2}, z_{p_2})^t. \end{cases} \tag{C.5}$$

Let us consider the two co-planar vectors: $\overrightarrow{p_1p_0}$ and $\overrightarrow{p_1p_2}$. We have $\overrightarrow{p_1p_0} \in P_{f'}$ and $\overrightarrow{p_1p_2} \in P_{f'}$. We note:

$$\begin{cases} \overrightarrow{p_1p_0} = (x_{p_0} - x_{p_1}, y_{p_0} - y_{p_1}, z_{p_0} - z_{p_1})^t \\ \overrightarrow{p_1p_2} = (x_{p_2} - x_{p_1}, y_{p_2} - y_{p_1}, z_{p_2} - z_{p_1})^t. \end{cases} \tag{C.6}$$

For a given point $M(x, y, z) \in \Re^3$, we consider the vector:

$$\overrightarrow{p_1M} = (x - x_{p_1}, y - y_{p_1}, z - z_{p_1})^t. \tag{C.7}$$

Let's assume that $M \in P_{f'}$, this implies that the three vectors $\overrightarrow{p_1p_0}$, $\overrightarrow{p_1p_2}$ and $\overrightarrow{p_1M}$ are co-planar, which is equivalent to the following condition:

$$(\overrightarrow{p_1M} \times \overrightarrow{p_1p_0}) \cdot \overrightarrow{p_1p_2} = \det \begin{pmatrix} x - x_{p_1} & x_{p_0} - x_{p_1} & x_{p_2} - x_{p_1} \\ y - y_{p_1} & y_{p_0} - y_{p_1} & y_{p_2} - y_{p_1} \\ z - z_{p_1} & z_{p_0} - z_{p_1} & z_{p_2} - z_{p_1} \end{pmatrix} = 0. \tag{C.8}$$

Finally, this leads to:

$$\begin{cases} A = (y_{p_0} - y_{p_1}) \cdot (z_{p_2} - z_{p_1}) - (z_{p_0} - z_{p_1}) \cdot (y_{p_2} - y_{p_1}) \\ B = (z_{p_0} - z_{p_1}) \cdot (x_{p_2} - x_{p_1}) - (x_{p_0} - x_{p_1}) \cdot (z_{p_2} - z_{p_1}) \\ C = (x_{p_0} - x_{p_1}) \cdot (y_{p_2} - y_{p_1}) - (y_{p_0} - y_{p_1}) \cdot (x_{p_2} - x_{p_1}) \\ D = -(A \cdot x_{p_1} + B \cdot y_{p_1} + C \cdot z_{p_1}). \end{cases} \tag{C.9}$$

APPENDIX D

Integrals calculation for analytical kernels

Here, we derive analytical expressions of some complicated integrals, that show up in the calculation of analytical "finite-frequency" kernels (Dahlen *et al.* 2000).

A.1 Calculation of $\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \sin[\omega\Delta T]d\omega$

We have (Gradshteyn & Ryzhik 1965):

$$\int_0^\infty x^{2n+1} e^{-\beta^2 x^2} \sin(ax)dx = (-1)^n \sqrt{\pi}/(2\beta)^{2n+2} e^{-a^2/(4\beta^2)} H_{2n+1}(a/2\beta) \tag{D.1}$$

with H_m the Hermite polynomial of order m. The 'physicists' Hermite polynomials are defined by: $H_m(x) = (-1)^m e^{x^2} d^m e^{-x^2}/dx^m$. For m=5, we have $H_5(x) = 32x^5 - 160x^3 + 120x$. We then use the previous formula - equation (A.1) - with: $n = 2$, $\beta = T/2\pi$ and $a = \Delta T$, and we obtain:

$$\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \sin[\omega\Delta T]d\omega = (-1)^2 \sqrt{\pi}/(T/\pi)^6 e^{-\Delta T^2/4(T/2\pi)^2} H_5(\Delta T\pi/T). \tag{D.2}$$

Finally, by using the previous formula for $H_5(x)$, we obtain the analytical expression of the integral $\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \sin[\omega\Delta T]d\omega$.

A.2 Calculation of $\int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2} \cos[\omega\Delta T]d\omega$

By using the software *Mathematica* (http://integrals.wolfram.com), we have:

$$\int x^5 e^{-\beta^2 x^2} \cos[ax]dx = 1/(128\beta^{11}) e^{-a^2/4\beta^2 - iax - \beta^2 x^2} [\cdots]_1 \tag{D.3}$$

with $\beta = T/2\pi$, $a = \Delta T$, and:

$$\begin{cases} [\cdots]_1 = -2\beta e^{a^2/4\beta^2} [\cdots]_2 + [\cdots]_3 \\ [\cdots]_2 = (1 + e^{2iax})a^4 - 2i\beta^2(-1 + e^{2iax})xa^3 - 2\beta^2(1 + e^{2iax})(2\beta^2 x^2 + 9)a^2 \\ +4i\beta^4(-1 + e^{2iax})x(2\beta^2 x^2 + 7) + 16\beta^4(1 + e^{2iax})(\beta^4 x^4 + 2\beta^2 x^2 + 2) \\ [\cdots]_3 = a(a^4 - 20\beta^2 a^2 + 60\beta^4) e^{x(x\beta^2 + ia)} \sqrt{\pi}(-i\mathrm{erf}(x\beta + ia/(2\beta)) + \mathrm{erfi}(a/2\beta + i\beta x)). \end{cases} \tag{D.4}$$

Then, we take the limit of the previous expression in $+\infty$ minus the limit in 0 in order to obtain the integral value of $\int_0^{+\infty} x^5 e^{-\beta^2 x^2} \cos(ax)dx$.

1. Limit in $+\infty$.
   We have $\mathrm{erf}(z) = -i\mathrm{erfi}(iz)$, where erf is the error function and erfi is the imaginary

error function, that is $\text{erf}(z) = 2/\sqrt{\pi}\int_0^z e^{-t^2}dt$ and $\text{erfi}(z) = 2/\sqrt{\pi}\int_0^z e^{+t^2}dt$. We have:

$$\begin{cases} \lim\limits_{z\to+\infty} \text{erf}(z) = 1 \\ \lim\limits_{z\to-\infty} \text{erf}(z) = -1 \\ \lim\limits_{z\to 0} \text{erf}(z) = 0. \end{cases} \tag{D.5}$$

We have $\lim\limits_{x\to+\infty} e^{-\beta^2x^2}x^n = 0$ when $n = [1, 2, 3, 4, \cdots]$, so that, as $\beta > 0$,:

$$\begin{cases} \lim\limits_{x\to+\infty} -i\text{erf}(x\beta + ia/(2\beta)) = -i \\ \lim\limits_{x\to+\infty} \text{erfi}(a/(2\beta) + i\beta x) = +i. \end{cases} \tag{D.6}$$

The sum of these two last terms is zero when $x \to +\infty$. We conclude that the limit in $+\infty$ is zero.

2. Limit in 0.
We have:

$$\begin{cases} \lim\limits_{x\to 0} -i\text{erf}(x\beta + ia/(2\beta)) = i\text{erf}(ia/(2\beta)) = \text{erfi}(a/(2\beta)) \\ \lim\limits_{x\to 0} \text{erfi}(a/(2\beta) + i\beta x) = \text{erfi}(a/(2\beta)) \end{cases} \tag{D.7}$$

The sum of these two last terms is then $2\text{erfi}(a/(2\beta))$. We then have:

$$\begin{cases} \lim\limits_{x\to 0}[\cdots]_2 = 2a^4 - 36\beta^2a^2 + 64\beta^4 \\ \lim\limits_{x\to 0}[\cdots]_3 = a(a^4 - 20\beta^2a^2 + 60\beta^4)2\sqrt{\pi}\text{erfi}(a/(2\beta)) \end{cases} \tag{D.8}$$

We conclude that the limit in 0 is:
$-1/(a(2\beta)^5)[32(a/(2\beta))^5 - 144(a/(2\beta))^3 + 64(a/(2\beta))] + e^{-a^2/(4\beta^2)}\sqrt{\pi}/(2\beta)^6[32(a/(2\beta))^5] - 160(a/(2\beta))^3 + 120(a/(2\beta))]\text{erfi}(a/(2\beta))$.

Finally, we obtain :

$$\begin{array}{l} \int_0^\infty \omega^5 e^{-\omega^2(T/2\pi)^2}\cos[\omega\Delta T]d\omega = 2^5(\frac{\pi}{T})^{10}[\Delta T^4 - \frac{9}{2}\Delta T^2T^2/\pi^2 + 2T^4/\pi^4] \\ -\sqrt{\pi}2^5(\frac{\pi}{T})^{11}\Delta T e^{-(\Delta T\pi/T)^2}[\Delta T^4 - 5\Delta T^2T^2/\pi^2 + \frac{15}{4}T^4/\pi^4]\text{erfi}(\Delta T\pi/T) \end{array} \tag{D.9}$$

A.3 Calculation of $\int_0^\infty \omega^4 e^{-\omega^2(T/(2\pi))^2} d\omega$

By using the software *Mathematica* (http://integrals.wolfram.com), we have:

$$\int x^4 e^{-\beta^2 x^2} dx = \frac{3\sqrt{\pi}\text{erf}(\beta x) - 2\beta e^{-\beta^2 x^2} x(2\beta^2 x^2 + 3)}{8\beta^5} \tag{D.10}$$

The limit of this integral is 0 when $x \to 0$ and $\frac{3\sqrt{\pi}}{8\beta^5}$ when $x \to +\infty$ and $\beta > 0$. Therefore, we have: $\int_0^\infty x^4 e^{-\beta^2 x^2} dx = \frac{3\sqrt{\pi}}{8\beta^5}$. With $\beta = T/(2\pi)$, we then obtain:

$$\int_0^\infty \omega^4 e^{-\omega^2(T/(2\pi))^2} d\omega = \frac{3\sqrt{\pi}}{8}(\pi/T)^5 2^5 \tag{D.11}$$

A.4 Calculation of $\int_0^\infty \omega^4 e^{-(b\omega)^2} \cos(a\omega) d\omega$

We calculate the following integral by using the software Mathematica:

$$\begin{cases} \int x^4 e^{-(bx)^2} \cos(ax) dx = \frac{1}{64b^9} e^{-(a/2b)^2} [\cdots] \\ [\cdots] = -2be^{(a-2ib^2x)^2/(4b^2)} \{\cdots\} + (a^4 - 12b^2a^2 + 12b^4)\sqrt{\pi}\{\text{erf}(\frac{2xb^2+ia}{2b}) - i\text{erfi}(\frac{a}{2b} + ibx)\} \\ \{\cdots\} = 4(1 + e^{2iax})x(2b^2x^2 + 3)b^4 - 2a^2(1 + e^{2iax})xb^2 + \\ 2ia(-1 + e^{2iax})(2b^2x^2 + 5)b^2 - ia^3(-1 + e^{2iax}) \end{cases} \tag{D.12}$$

1. Limit in $+\infty$.
   We have: $\lim_{x\to+\infty} \text{erf}(xb + ia/(2b)) = 1$, with $b > 0$, and $\text{erfi}(a/(2b) + ibx) = \text{erf}(bx - ia/(2b))$, because of $\text{erf}(z) = -i\text{erfi}(iz)$. So: $\lim_{x\to+\infty} -i\text{erfi}(a/(2b)+ibx) = 1$, with $b > 0$. Then, the sum of these two last terms is 2 when $x \to +\infty$. Finally, $\lim_{x\to+\infty} \int \cdots = \frac{1}{64b^9} e^{-(a/2b)^2}(a^4 - 12b^2a^2 + 12b^4)\sqrt{\pi}2$.

2. Limit in 0.
   We have to calculate : $[\cdots]_{x\to0} = -2be^{(a/2b)^2}\{\cdots\}_{x\to0} + (a^4 - 12b^2a^2 + 12b^4)\sqrt{\pi}\{\text{erf}(\frac{ia}{2b}) - i\text{erfi}(\frac{a}{2b})\}$. We have: $\text{erf}(\frac{ia}{2b}) = -i\text{erfi}(i\frac{ia}{2b}) = -i\text{erfi}(-\frac{a}{2b})$, and, as $\text{erfi}(-z) = -\text{erfi}(z)$ (parity), we obtain $\text{erf}(\frac{ia}{2b}) = -i\text{erfi}(-\frac{a}{2b}) = -i(-\text{erfi}(\frac{a}{2b})) = i\text{erfi}(\frac{a}{2b})$. As a result: $\{\text{erf}(\frac{ia}{2b}) - i\text{erfi}(\frac{a}{2b})\} = i\text{erfi}(\frac{a}{2b}) - i\text{erfi}(\frac{a}{2b}) = 0$. Moreover, we have: $\{\cdots\}_{x\to0} = 0$. Finally: $\lim_{x\to0} \int \cdots = 0$.

Therefore, we end up with:

$$\int_0^\infty \omega^4 e^{-(b\omega)^2} \cos(a\omega) d\omega = \frac{e^{-(a/2b)^2}\sqrt{\pi}}{2^5 b^9}(a^4 - 12b^2a^2 + 12b^4) \tag{D.13}$$

A.5 Calculation of erfi$(z) = 2/\sqrt{\pi}\int_0^z e^{+t^2}dt$

The imaginary error function (erfi) is an exponential function with its argument squared. Hence, this function is unbounded and rapidly increasing, so that computing tabulated values of erfi(z) is not a good option. Following Simpson *et. al* (2003), we use the derived Maclaurin series expansion for the evaluation of erfi(z). That is:

$$\text{erfi}(z) = \frac{2}{\sqrt{\pi}}\{\frac{z^1}{10!} + \frac{z^3}{31!} + \frac{z^5}{52!} + \cdots\} = \frac{2}{\sqrt{\pi}}\sum_{n=1}^{+\infty}\frac{z^{2n-1}}{[2n-1][(n-1)!]}. \tag{D.14}$$

As this series converges for any argument z (Simpson *et. al* 2003), it may be truncated after a finite number of terms. Figure D.1 shows the relationship between the argument of the series, z, and the number of terms, n, in the truncated series, such as the magnitude of the $(n+1)$-th term is smaller than $\epsilon = 10^{-7}$. Here, we need to calculate erfi(z), with

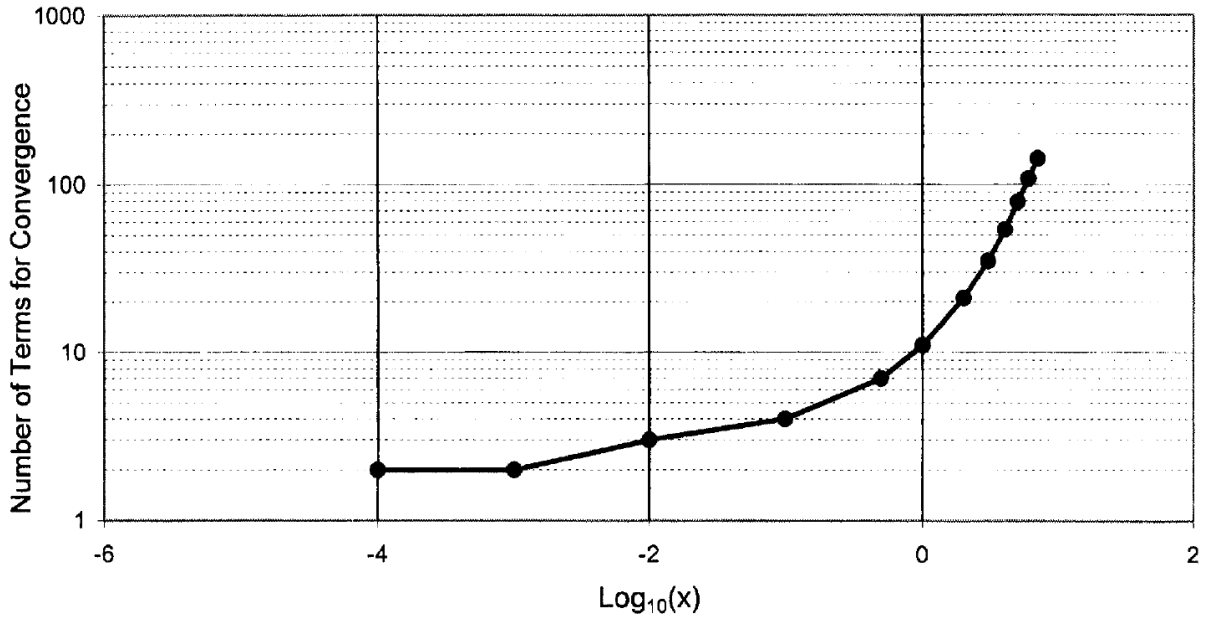


Figure D.1: Convergence behaviour of the Maclaurin series expansion of the imaginary error function for $\epsilon = 10^{-7}$. From Simpson et.al (2003).

$z = \pi\Delta T/T$, where T is the wave period, and ΔT is the scattered wave detour time. The maximum detour time may be set to $\Delta T_{max} = T/4$ (section 3.3.6, of chapter 3). Therefore, the condition $\Delta T \leq \Delta T_{max}$ leads to $|z| \leq \pi/4$. Using the fact that $\log_{10}(\pi/4) \simeq -0.1049$, the figure D.1 gives us an upper limit of the number of terms,n, required for the series to converge, that is: $n = 10$.

APPENDIX E

Kernel to model grid: change of coordinates

In order to build the G matrix, corresponding to the multiple-frequency inverse problem, we aim to project each kernel value, defined in the "kernel grid" basis $F' = (O, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, onto the "model grid", defined in the basis $F = (O, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. The situation is sketched in figures 3.24 and 3.25, of chapter 3. Here, we show how to find the relationship between the two basis F and F'. We have:

$$\begin{cases} \text{E is the epicenter with coordinates } (x_E, y_E, z_E)_F \text{ and } (r_E, \theta_E, \phi_E)_F \\ \text{S is the station with coordinates } (x_S, y_S, z_S)_F \text{ and } (r_S, \theta_S, \phi_S)_F \\ \text{P, E, M, S, Q are on the same great-circle} \\ \text{P, O, Q are aligned} \\ \overrightarrow{OM}/\|\overrightarrow{OM} = (\overrightarrow{OE} + \overrightarrow{OS})/\|(\overrightarrow{OE} + \overrightarrow{OS})\| \\ \mathbf{Z}' = \overrightarrow{OM}/\|\overrightarrow{OM}\| \\ \mathbf{Y}' = \overrightarrow{OQ}/\|\overrightarrow{OQ}\| \\ \mathbf{X}' = \mathbf{Y}' \times \mathbf{Z}' \ \text{ (cross-product).} \end{cases} \tag{E.1}$$

Let $\mathbf{u} = \overrightarrow{OE} + \overrightarrow{OS}$ and $\mathbf{v} = \overrightarrow{ES}$. The coordinates of these two vectors are, in the basis F:

$$\mathbf{u} = (x_E + x_S, y_E + y_S, z_E + z_S)^t_F, \quad \mathbf{v} = (x_S - x_E, y_S - y_E, z_S - z_E)^t_F. \tag{E.2}$$

We also have $\mathbf{Z}' = \mathbf{u}/\|\mathbf{u}\|$ and $\mathbf{Y}' = \mathbf{v}/\|\mathbf{v}\|$. We denote the coordinates of the three unit vectors of F' in F as:

$$\mathbf{X}' = (X'_x, X'_y, X'_z)^t_F, \quad \mathbf{Y}' = (Y'_x, Y'_y, Y'_z)^t_F, \quad \mathbf{Z}' = (Z'_x, Z'_y, Z'_z)^t_F. \tag{E.3}$$

Finally, we can express the relationship between the unit vectors of F', expressed in F, as:

$$\begin{pmatrix} \mathbf{X}' \\ \mathbf{Y}' \\ \mathbf{Z}' \end{pmatrix} = \begin{pmatrix} X'_x & Y'_x & Z'_x \\ X'_y & Y'_y & Z'_y \\ X'_z & Y'_z & Z'_z \end{pmatrix} \cdot \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}. \tag{E.4}$$

The 3×3 matrix in the right term of equation (E.4) is called the "change-of-basis matrix" P between the two basis F and F'. Therefore, if the coordinates[1] of a point $\mathbf{r}_p$ in the "kernel grid" basis F' are $(x'_{r_p}, y'_{r_p}, z'_{r_p})^t_{F'}$, then its corresponding coordinates $(x_{r_p}, y_{r_p}, z_{r_p})^t_F$ in the "model grid" F are:

$$(x_{r_p}, y_{r_p}, z_{r_p})^t_F = P \cdot (x'_{r_p}, y'_{r_p}, z'_{r_p})^t_{F'}. \tag{E.5}$$

[1] In section 3.3.1.a (chapter 3), we show how to calculate the coordinates $(x'_{r_p}, y'_{r_p}, z'_{r_p})^t_{F'}$ of a point $\mathbf{r}_p$ in the "kernel grid" basis F'- equation (3.20).

APPENDIX F

Imaging the Earth's interior

Here, we show some nice features present in our global multiple-frequency tomographic model, *ZDS-S10*, of the Earth's mantle, such as sinking slabs, rising up low-velocity anomalies, etc.

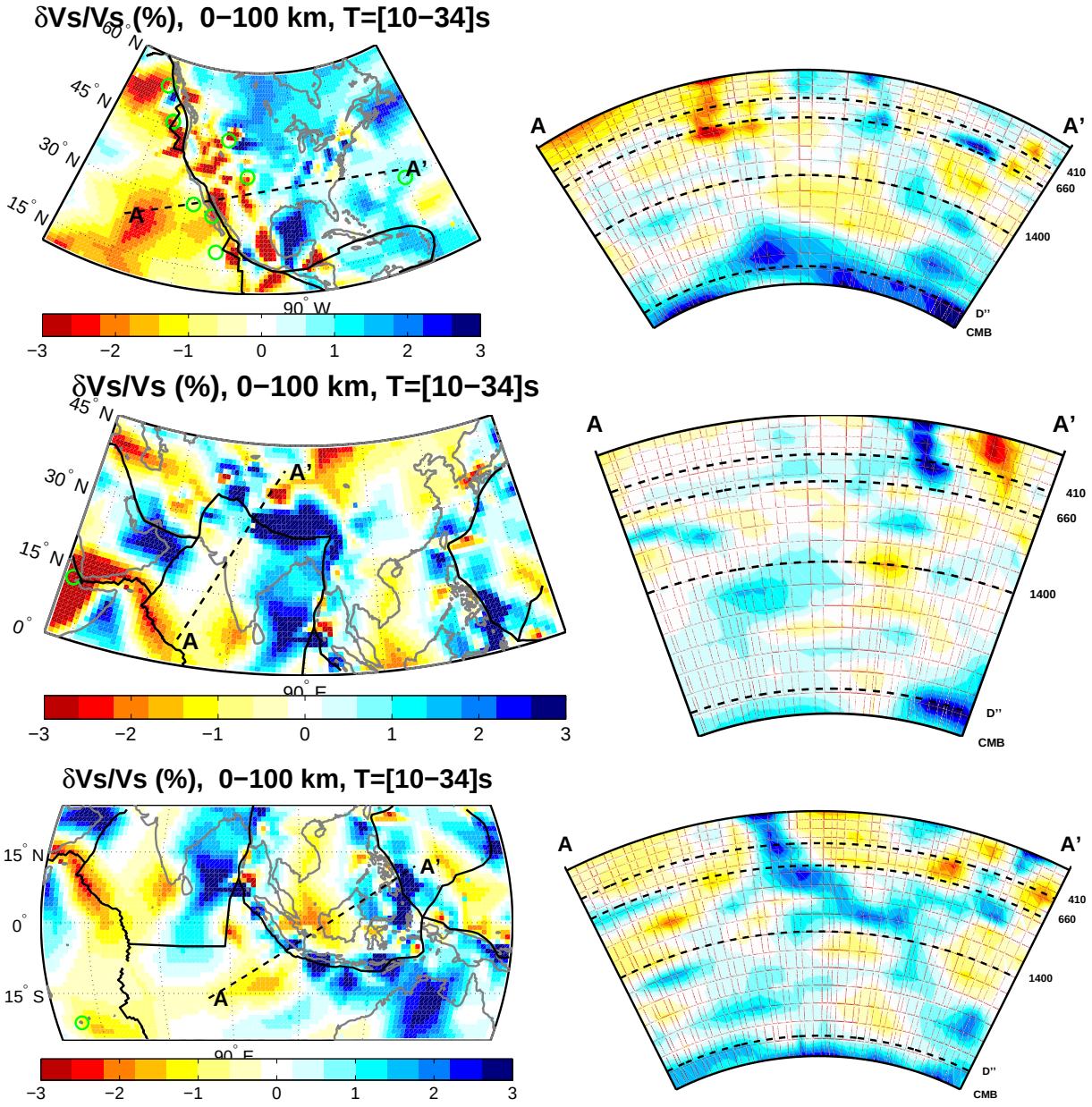


Figure F.1: Top row: ancient Farallon slab (e.g. Grand et al. 2002) beneath Northern America. Middle row: Tibet slab (e.g. Replumaz et al. 2004) related to the subduction of the India plate beneath Asia. Bottom row: Sumatra subducting slab (e.g. Replumaz et al. 2004), in Indonesia.

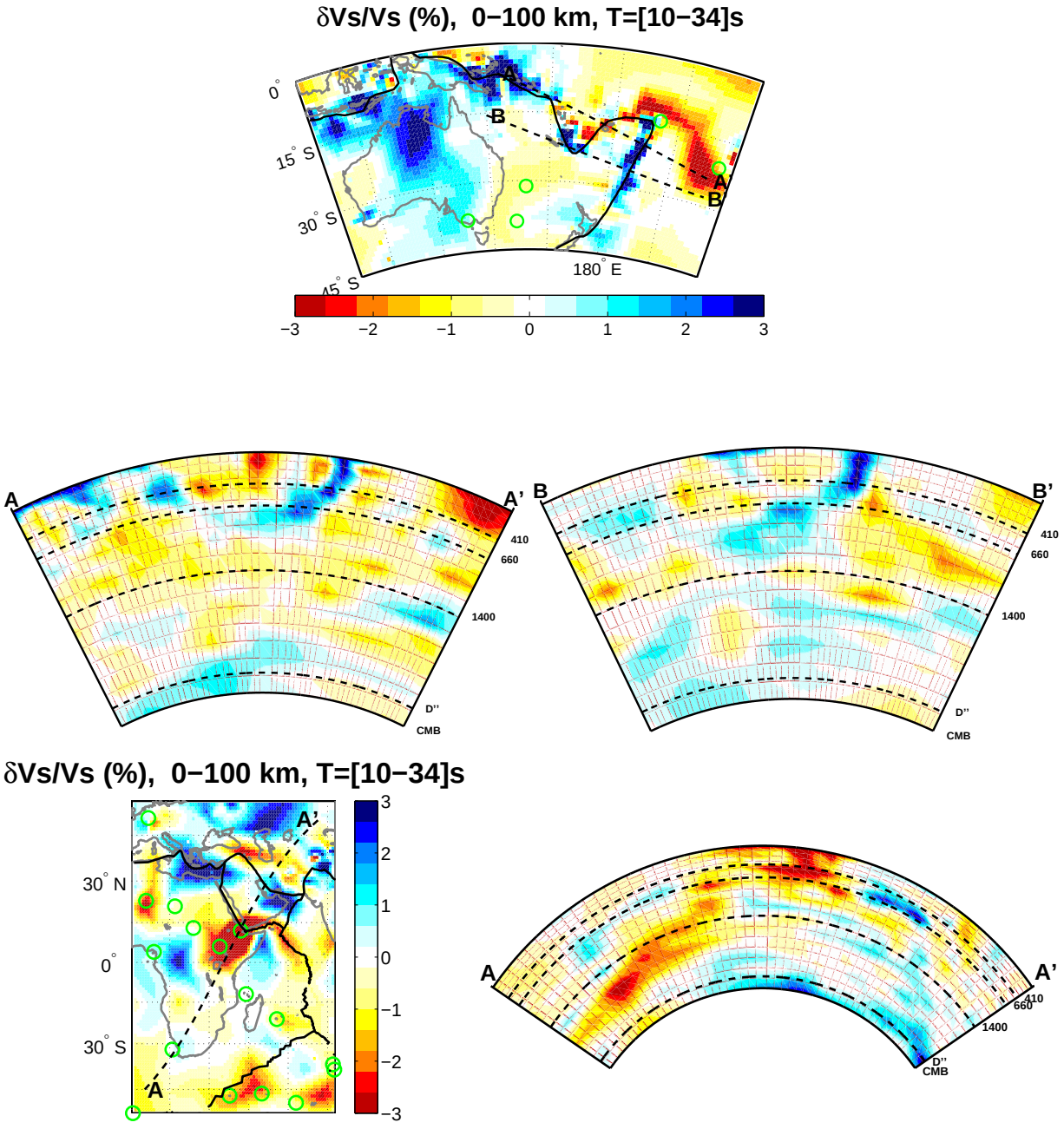


Figure F.2: Top and middle rows: Tonga and Fiji subducting slabs. Bottom row: tilted low-velocity anomaly extending from the CMB, beneath Southern Africa, to the bottom of the lithosphere beneath the Afar. A similar complex shear-wave velocity structure has been previously imaged by several authors (e.g. Ritsema et al. 1999, Montelli et al. 2006b). See section 4.4.2.c (chapter 4) for a comparison of single- versus multi-band tomography illustrated on this low-velocity anomaly beneath Africa.

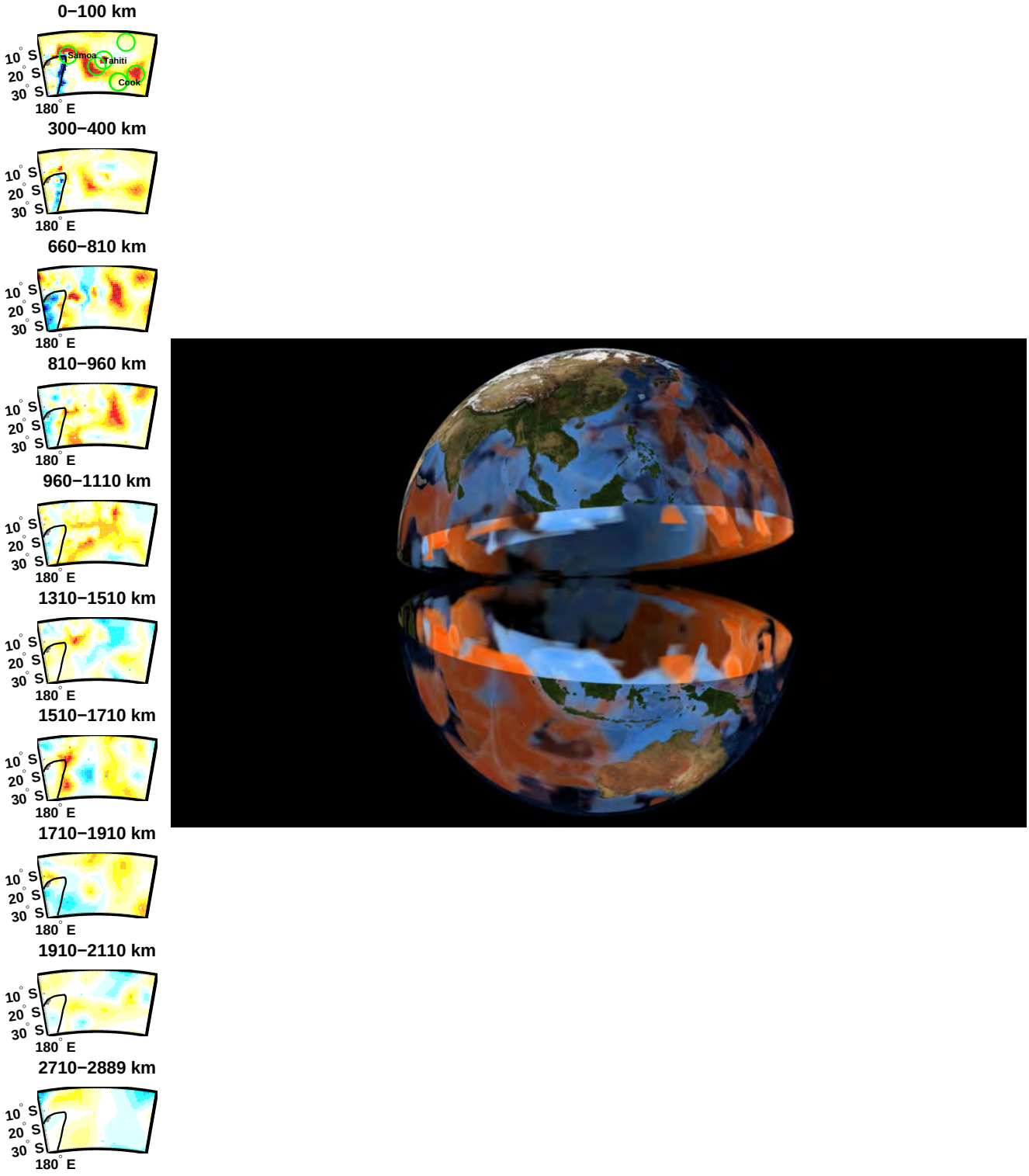


Figure F.3: (Left) Cross sections, in ZDS-S10, of the Samoa, Tahiti and Cook deep "plumes" (low velocity anomaly), which are suspected to extend from the CMB to the surface (e.g. Courtillot et al. 2003; Montelli et al. 2006b). The weaker "plume" signal near the CMB is likely due to a poor data sampling (section 4.2.3.c). Color scale is ±3%. (Right) Three-dimensional view of the Earth's mantle in a smoothed version of ZDS-S10 (Courtesy from Rhys Hawkins). One can "see-through" anomalies with negligible amplitude. Red/blue are slow/fast regions. First order convection occuring in the mantle is clearly observed.

References

Aki, K. & Richards, P. G., 1980. Quantitative Seismology, *W.H. Freeman*.

Albarède, F. & Van der Hilst, R. D., 1999. New mantle convection model may reconcile conflicting evidence, *Eos Transactions American Geophysical Union*, **45**, 535–539.

Anache-Ménier, D., 2008. Propagation des ondes en milieu désordonné: des ultrasons aux ondes sismiques. Etude de la phase. *PhD thesis*, Université de Grenoble.

Anderson, D., 1976. The Earth as a seismic absorption band, *Science*, **196**, 1104–1106.

Anderson, D. & Minster, J., 1979. The frequency dependence of Q in the Earth and implications for mantle rheology and Chandler wobble, *Geophys. J. R. Astron. Soc.*, **58**, 431–440.

Bai, C. Y. & Kennett, B.L.N., 2001. Phase identification and attribute analysis of broad-band seismograms at far regional distances, *J. Seismol.*, **5**, 217–231.

Baig, A. M. & Dahlen, F. A., 2004. Traveltimes biases in random media and the S-wave discrepancy, *Geophys. J. Int.*, **158**, 922–938.

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

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**, 13,527–13,540.

Boschi, L., Becker, T. W., Soldati, G. & Dziewonski, A. M., 2006. On the relevance

of Born theory in global seismic tomography, *Geophys. Res. Lett.*, **33**, L06302, doi: 10.1029/2005GL025063.

Burdick, S., Li, C., Martynov, V., Cox, T., Eakins J., Mulder, T., Astiz, L., Vernon, F., Pavlis, G. & Van der Hilst, R. D., 2008. Upper Mantle Heterogeneity beneath North America from Travel Time Tomography with Global and USArray Transportable Array Data, *Seismol. Res. Lett.*, **79**, 384–392.

Calvet, M. & Chevrot, S., 2005. Traveltime sensitivity kernels for PKP phases in the mantle, *Phys. Earth Planet. Inter.*, **153**, 21–31.

Chapman, C., 1978. A new method for computing synthetic seismograms, *Geophys. J. Roy. Astr. Soc.*, **54**, 481–518.

Cheng, H. X. & Kennett, B., 2002. Frequency dependence of seismic wave attenuation in the upper mantle beneath the Australian region, *Geophys. J. Int.*, **150**, 45–57.

Chevrot, S., 2002. Optimal measurement of relative and absolute delay-times by simulated annealing, *Geophys. J. Int.*, **151**, 164–171.

Choy, G.L. & Richards, P.G. 1975. Pulse distorsion and Hilbert transformation in multiply reflected and refracted body waves, *Bull. Seismol. Soc. Am.*, **65**, 55–70.

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

Dahlen, F. A., 2004. Resolution limit of traveltime tomography, *Geophys. J. Int.*, **157**, 315–331.

Dahlen, F. A. & Nolet, G., 2005. Comment on the paper "On sensitivity kernels for 'wave-equation' transmission tomography" by de Hoop and Van der Hilst, *Geophys. J. Int.*, **163**, 949–951.

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

Davaille, A., Stutzmann, E. Silveira G., Besse, J & Courtillot V. 2005. Convective patterns under the Indo-Atlantic box, *Earth Planet Sci. Lett.*, **239**, 233–252.

Debayle, E., Lévêque, J.J & Cara, M., 2001. Seismic evidence for a deeply rooted low-velocity anomaly in the upper mantle beneath the northeastern afro/arabian continent, *Earth Planet. Sci. Lett.*, **193**, 423–436.

Debayle, E., Kennett, B.L.N. & Priestley, K., 2005. Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia, *Nature*, **433**, 509–512.

De Hoop, M. V. & Van der Hilst, R. D., 2005a. On sensitivity kernels for 'wave equation' transmission tomography, *Geophys. J. Int.*, **160**, 621–633.

De Hoop, M. V. & Van der Hilst, R. D., 2005b. Reply to comment by F.A. Dahlen and G. Nolet on "On sensitivity kernels for 'wave equation' transmission tomography", *Geophys. J. Int.*, **163**, 952–955.

Devilee, R. J. R., Trampert, J. & Paulssen, H., 2003. Dispersion measurements of P waves and their implications for mantle Qp, *Pure appl. geophys.*, **160**, 2223–2238.

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. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.

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

Favier, N., Chevrot, S. & Komatitsch, D., 2004. Near-field influence on shear wave splitting and traveltime sensitivity kernels, *Geophys. J. Int.*, **156**, 467–482.

Flanagan, M. P. & Wiens, D.A., 1998. Attenuation of broadband P and S waves in Tonga: observations of frequency dependent Q, *Pure appl. geophys.*, **153**, 345–375.

Fukao, Y., Widiyantoro, S. & M. Obayashi, 2001. Stagnant slabs in the upper and lower mantle transition region, *Rev. of Geophys.*, **39**, 291–323.

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**, 1–7.

Gudmundson, O. 1996. On th effect of diffraction on traveltime measurement, *Geophys. J. Int.*, **124**, 304–314.

Houser, C., Masters, G., Shearer, P. M. & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.*, **174**, 195–212.

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

Hung, S.-H, Dahlen, F. A. & Nolet, G., 2001. Wavefront-healing: a banana-doughnut perspective, *Geophys. J. Int.*, **146**, 289–312.

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

Julian, B. R., 2005. What can seismology say about hot spots?, in Foulger, G. R., J. H. Natland, D. C. Presnall and D. L. Anderson, ed., Plates, Plumes, and Paradigms, *Geological Society of America*, **388**, 155–170.

Kanamori, H. & Anderson, D. L., 1977. Importance of physical dispersion in surface wave and free oscillation problems, *Rev. Geophys. Space Phys.*, **15**, 105–12.

Kennett, B. & Engdahl, E., 1991. Travel times for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.

Komatitsch, D. & Tromp, J., 1995. Spectral-element simulations of global seismic wave propagation – I. Validation, *Geophys. J. Int.*, **122**, 108–124.

Komatitsch, D., Ritsema; J. & Tromp, J., 2002. The spectral-element method, Beowulf computing and global seismology, *Science*, **298**, 1737–1742.

Lekic, V., Matas, J., Panning, M. & Romanowicz, B., 2009. Measurement and implications of frequency dependence of attenuation, *Earth Planet. Sci. Lett.*, **282**, 285–293.

Liu, H.-P, Anderson, D. L. & Kanamori, H., 1976. Velocity dispersion due to anelasticity; implications for seismology and mantle composition, *Geophys. J. R. astr. Soc.*, **47**, 41–58.

Marquering, H., Nolet & 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. R. Soc. Lond.*, **A354**, 1385–1410.

Masters, G., Laske, G., Bolton, H. & Dziewonski, A. M., 2000. The relative behaviour of shear velocity, bulk sound speed and compressional velocity in the mantle: implications for chemical and thermal structure., *In S. Karato, A. Forte, R. C. Liebermann, G. Masters and L. Stixrude, eds., Earth's Deep Interior, AGU, Washington DC*, 63–88.

Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**, 257–281.

Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R. & Hung, S.-H., 2004a. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, **303**, 338–343.

Montelli, R., Nolet, G., Masters, G., Dahlen, F. A. & Hung, S.-H., 2004b. Global P and PP traveltime tomography: Rays versus waves, *Geophys. J. Int.*, **158**, 637–654.

Montelli, R., Nolet, G., Dahlen F. A. & Masters, G., 2006a. Comment on 'banana-doughnut kernels and mantle tomography' by van der Hilst and de Hoop, *Geophys. J. Int.*, **167**, 1204–1210.

Montelli, R., Nolet, G., Dahlen F. A. & Masters, G., 2006b. A catalogue of deep mantle plumes. New results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, **7**, Q11007, doi:10.1029/2006GC001248.

Mooney, W. D., Laske, G. & Masters, G., 1998. CRUST5.0: A global crustal model at 5°×5°, *J. Geophys. Res.*, **103**, 727–747.

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

Nissen-Meyer, T., Dahlen, F. A., & Fournier A. 2007. Spherical-earth Fréchet sensitivity kernels, *Geophys. J. Int.*, **168**, 1051–1066.

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

Nolet, G., Dahlen, F. A. & Montelli, R., 2005. Traveltimes and amplitudes of seismic waves: a re-assessment, *A. Levander, G. Nolet, Eds.. Array Analysis of Broadband Seismograms, AGU Monograph Series*, 37–48.

Nolet, G., 2008. A breviary of seismic tomography, *Cambridge University Press, Cambridge, UK.*

Nolet, G., 2009. Slabs Do Not Go Gently, *Science*, **324** no. 5931, pp. 1152–1153.

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

Pollitz, F. F., 2007. Finite-frequency travel time tomography of San Francisco Bay region crustal velocity structure, *Geophys. J. Int.*, **171**, 630–656.

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

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 travel times, *Geophys. J. Int.*, **149**, 482–489.

Ritsema, J., van Heijst, H.-J.,Woodhouse, J. H. & Deuss A., 2009. Long-period body wave traveltimes through the crust: implications for crustal corrections and seismic tomography, *Geophys. J. Int.*, **179**, 1255–1261.

Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years, *Ann. Rev. Earth Planet Sci.*, **31**, 303–328.

Romanowicz, B. & Mitchell, B., 2007. Deep Earth structure: Q of the Earth from crust to core, *Schubert, G. (Ed.), Treatise on Geophysics, 1. Elsevier*, pp. 731–774.

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

Sambridge, M. & Faletic, R., 2003. Adaptive whole Earth Tomography *Geochem; Geoph. Geosyst.*, **4**, No 3.

Shito, A., Karato, S. I. & Park, J., 2004. Frequency dependence of Q in Earth's upper mantle inferred from continuous spectra of body waves, *Geophys. Res. Lett.*, **31**, L24614, doi:10.1029/2004GL019582.

Sieminski, A., Leveque, J. J. & Debayle, E., 2004. Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.*, **31**, L24614, doi:10.1029/2004GL021402.

Sigloch, K. & Nolet, G., 2006. Measuring finite-frequency body wave amplitudes and travel times, *Geophys. J. Int.*, **167**, 271–287.

Simons, F.J., Nolet, G., Babcock, J.M., Davis, R.E., & Orcutt, J.A. 2006. A future for drifting seismic networks, *EOS Trans. AGU*, **31**, 305–307.

Sipkin, S. A. & Jordan, T. H., 1979. Frequency dependence of Q_{ScS}, *Bull. Seismol. Soc. Am.*, **69**. 1055–79, doi:10.1785/0120060038.

Sue, 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

Tian, Y., Nolet, G., Hung, S.-H., Montelli, R. & Dahlen, F, A., 2007a. Dynamic ray tracing and traveltime corrections for global seismic tomography, *J. Comput. Phys.*, **226**, 672–687.

Tian, Y., Montelli, R., Nolet, G. & Dahlen, F, A., 2007b. Computing traveltime and amplitude sensitivity kernels in finite-frequency tomography, *J. Comput. Phys.*, **226**, 2271–2288.

Tian, Y., Sigloch, K. & Nolet, G., 2009. Multiple-frequency SH-tomography of the western US upper mantle, *Geophys. J. Int.*, **178**, 1384–1402.

To, A. & Romanowicz, B., 2009. Finite frequency effects on global S diffracted traveltimes, *Geophys. J. Int.*, **179**, 1645–1657.

Trampert, J. & Spetzler, J., 2006. Surface wave tomography: finite frequency effects lost in the null space, *Geophys. J. Int.*, **164**, 394–400.

Trampert, J., Deschamps, F., Resovsky, J & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science*, **306**, 853–856.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.

Ulug, A. & Berckhemer, H., 1984. Frequency dependence of Q for seismic body waves in the Earth's mantle, *J. Geophys.*, **56**, 9–19.

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. & de Hoop, M. V., 2005. Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.*, **163**, 956–961.

Van der Hilst, R. D. & de Hoop, M. V., 2006. Reply to comment by R. Montelli, G. Nolet, and F. A. Dahlen on "Banana-doughnut kernels and mantle tomography", *Geophys. J. Int.*, **167**, 1211–1214.

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**, no 5409, pp. 1885–1888.

Vasco, D. W., Johnson, L.R. & Marques, O., 2003. Resolution, uncertainty, and whole Earth tomography, *J. Geophys. Res.*, **108**.

Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parameterization of three-dimensional Earth, *Geophys. Res. Lett.*, **22**, 3099–3102.

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.

Wu, Y., He Y. & Tian H., 2005. A spherical point location algorithm based on barycentric coordinates, *Gervasi et al. (Eds.): ICCSA 2005, LNCS 3482*, pp. 1099–1108.

Yang, T. & Shen, Y., 2006. Frequency-dependent crustal corrections for FF seismic tomography, *Bull. Seismol. Soc. Am.*, **96**6. 2441–2448, doi:10.1785/0120060038.

Zaroli C., Debayle E., & Sambridge M., 2010. Frequency-dependent effects on global S-wave traveltimes: wavefront-healing, scattering and attenuation, *Geophys. J. Int.*, **182**, 1025–1042.

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