Fast probing of many Earth models with full waveforms

Harriet Godwin

Thesis submitted to the University of Oxford
for the degree of Doctor of Philosophy



St Edmund Hall,
University of Oxford

Michaelmas 2019

Declaration

The contents of this thesis are all my own work, except where otherwise stated. The views and opinions expressed herein are mine and not necessarily those of any other person or body unless so attributed.

Citation: Harriet Godwin (2019), Fast probing of many Earth models with full waveforms, D.Phil. Thesis, University of Oxford, Department of Earth Sciences, Oxford, UK.

Keywords: Seismology

This thesis has been typeset using LaTeX 2$_\varepsilon$ and references were compiled using BibTeX.

Copyright ©, by Harriet Godwin, 2019.
All rights reserved.

No part of the material protected by this copyright notice may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

Printed in the United Kingdom.

Abstract

Fast probing of many Earth models with full waveforms

Harriet Godwin

Michaelmas 2019

Seismology is our primary method for inferring the structure and dynamics of Earth's interior. A region of particular importance is the lowermost mantle near the core-mantle boundary. Detailed information about lowermost mantle seismic structures can be inferred by comparing synthetic seismic waveforms to those observed at Earth's surface.

In this thesis, a forward modelling method is presented which is based on first order single scattering approximations. These include the well-known Born approximation (which involves linearly perturbing the amplitude of the waveform), as well as the Rytov approximation (which involves linearly perturbing the phase of the waveform) and a whole family of intermediate approximations termed Marks approximations. The computation of these first order waveforms is facilitated by time and space dependent waveform sensitivity kernels. Kernels are computed for each source-receiver pair, but they are independent of the perturbed model. Once the kernel is calculated, synthetic seismograms for many models are generated very quickly ($\approx 7\ s$ per model).

The applicabilities of the Born, Rytov and Marks approximations are tested by comparing to reference solutions computed with a fully numerical method. For small perturbations to the background model, an excellent agreement is found between the first order and reference solutions, providing validation of the method. The first order method breaks down with increasingly large and strong perturbations. The Rytov approximation consistently performs better than Born (the worst) or Marks(intermediate). The method is valid for P, S and converted phases.

The first order method is used to generate synthetic P/Pdiff waveforms under the Rytov approximation for 500 models representative of the large low velocity province under Africa. The database of synthetic data is probed in a statistical manner; sets of models which fit certain supposed observations are presented.

Acknowledgements

This thesis would not have been possible without all of the support and guidance of my primary supervisor, Tarje, over these four years. I am also very grateful to my second supervisor, Karin, for always offering her insight and advice during my DPhil.

I must also thank the colleagues who have helped me along the way - Paula, for her scientific and moral support during my DPhil, Kasra for offering lots of ideas and help, Simon for making the trip to Oxford last year, and Kuangdai for patiently explaining AxiSEM3D to me.

Thank you to the seismology group members and officemates who have been here for the ups and downs; to Marta for all the laughter you brought to the office, Afsaneh for always lending an ear or a shoulder, Eoghan for his patience and running analogies when I complained, and to Maria for her advice over the years. Also to Alex, Brook, Claudia, Ben x 2, Tom, Andy, Laura, Kara, Matthew and Lucia for all that you brought to the group meetings and pub trips.

Thank you to Theresa and to Becca for always being there through the past few years, sharing in the high points and giving support when it was needed most. To the friends in London including Grace, Helena, Dom and Shiv - for their visits to Oxford, and to my lifelong best friends Lily and Sarah who are always ready to listen and who offer the best advice.

Finally, thank you to my parents and my brothers Toby, Alistair and Jonathan for their unending support and encouragement, during this DPhil and always.

Contents

Declaration i

Abstract iii

Acknowledgments v

Contents x

1 Introduction **1**

1.1 History of global seismology 1

1.2 Seismic Tomography 2

1.3 Forward waveform modelling 4

1.4 Lower mantle seismic structures 5

1.4.1 Large low velocity provinces 5

1.4.2 Edges of the LLVP under Africa 8

1.5 Motivation and thesis outline 10

2 Theory of linearised wavefield scattering **13**

2.1 The seismic wave equation 13

2.1.1 Green's tensor 15

2.2 Wave scattering 17

2.3 The Born approximation 18

2.3.1 Derivation 20

2.3.2 Waveform kernels 21

2.3.3 Measurement kernels 23
2.4 Generalised first order solution 24
2.4.1 Background 24
2.4.2 Derivation 27
2.4.3 Discussion 36

3 Sensitivity kernel calculation 37
3.1 Introduction 37
3.1.1 Background: Sensitivity kernels for seismic tomography 38
3.2 Methodology 39
3.2.1 Wavefield Computation 39
3.2.2 Wavefield rotation 42
3.2.3 Projection onto 3-D meshes 44
3.3 Implementation 47
3.4 Kernel examples 50
3.4.1 Travel time kernels 50
3.4.2 Waveform kernels 53

4 Waveform modelling with Born, Rytov and Marks approximations 57
4.1 Introduction 57
4.2 Calculation of first-order seismograms 59
4.2.1 Kernel calculation 61
4.2.2 Model projection 61
4.2.3 Checks on the method 62
4.2.4 Note on the calculation of ϕ 64
4.2.5 Quantifying differences between first order and reference seismograms 65
4.3 1-D models 65

4.3.1 Perturbations in compressional wavespeed 67
4.3.2 Perturbations in shear wavespeed 70
4.4 Bubble models 73
4.4.1 Perturbations in compressional wavespeed 75
4.4.2 Perturbations in shear wavespeed 84
4.4.3 Converted phases and perturbations in both v_p and v_s 85
4.5 Discussion 87
4.5.1 1-D models 87
4.5.2 3-D models 88
4.6 Conclusions 91

5 Pathway towards statistical probing of lower mantle structures with waveforms 93

5.1 Introduction 93
5.2 Tomography model benchmark 94
5.3 Synthetic study of the LLVP under Africa 97
5.3.1 Model Parameterisation 97
5.3.2 AxiSEM3D synthetics 99
5.3.3 Benchmarks 104
5.3.4 Rytov seismograms for many models 105
5.4 Discussion 121

6 Conclusions 123

6.1 Summary 123
6.1.1 First order Approach 123
6.1.2 Applicability and limitations 125
6.1.3 Advantages and applications 126
6.2 Further work 126

6.2.1 First order method 126

6.2.2 Applications 127

Chapter 1

Introduction

1.1 History of global seismology

Global seismology is our primary method for investigating Earth's interior. The imprint of structures that seismic waves encounter along their path is encoded in recordings of ground motion at Earth's surface. The aim of global seismology is to interpret this signature and thereby to infer Earth's structure.

The first global studies in seismology used travel time curves, plots of arrival times of seismic phases against distance from an earthquake source.The first discontinuity to be discovered within the Earth is also the strongest; the core-mantle boundary (CMB) was found by Oldham in 1906 (*Oldham*, 1906). Shortly afterwards, in 1909, Mohorovicic found the seismic discontinuity separating the crust from the mantle (*Mohorovicic*, 1909). The fourth major region of the Earth to be revealed was the inner core. Its existence was inferred by Lehman in 1936 from the detection of faint P arrivals in the shadow zone from the outer core. The first 1-D radial model of Earth's structure to be able to accurately predict phase arrival times was arrived at by Jeffreys and Bullen (*Jeffreys and Bullen*, 1940) from their compilation of large numbers of travel time measurements for many seismic phases. By the 1970s, the 1-D structure of the Earth was well established, and there are now several 1-D models which are used a reference Earth models in global seismology including PREM (*Dziewonski and Anderson*, 1981),

IASP91 (*Kennett and Engdahl*, 1991) and AK135 (*Kennett et al.*, 1995).

While these 1-D models can explain a significant proportion of observed travel time data, it is clear that there are regional 3-D variations in seismic structure. Encoded in this 3-D global seismic structure is information about mantle dynamics, plate tectonics and the evolution of the planet. Therefore, it has since been the aim of global seismology to discover how the real Earth deviates from the well known 1-D structure.

Methods in global seismology can broadly be divided into seismic tomography and seismic waveform modelling.

1.2 Seismic Tomography

Seismic tomography is the method by which differences between observed and synthetic data are inverted to produce 3-D maps, mainly of seismic velocity but also of density, attenuation or anisotropy. The problem is linearised by inverting the system

$$\mathbf{d} = \mathbf{G}\mathbf{m}$$

for $\mathbf{m}$ as the model of Earth's structure. $\mathbf{d}$ is observed data and $\mathbf{G}$ is the sensitivity matrix. Constraints of smoothing and damping are applied, which are necessary to stabilise what is generally an ill-posed inverse problem. Damping ensures that the deviations from the 1-D reference Earth model are not too extreme, while smoothing means that there are no rapid changes in velocity in the model. The model,$\mathbf{m}$, is defined in terms of perturbations to a 1-D background model, and the observations from observed data, $\mathbf{d}$ are measured relative to the reference model (e.g. a travel time delay δt). For this reason, tomography therefore also requires a solution to the forward problem.

The first reliable global tomography model was produced by *Dziewonski* (1984) by inverting travel time residual data to produce a model for perturbations in v_p. Significant progress has since been made in terms of the volume and quality of seismological data, as well as inversion methods, resulting in the ever higher resolution of global

tomography models.

Each row of the sensitivity matrix, $\mathbf{G}$, defines how velocity perturbations within the Earth affect a particular measurement. In the case of single travel time measurement, the simplest way to define the sensitivity is using the ray-theoretical approximation of infinite frequency. Travel time delays are assumed to be accumulated along the geometrical ray path. This has been widely used to produce global tomography models in both v_p and v_s (e.g. *Grand et al.* 1997, *Ritsema et al.* 2011, *Koelemeijer et al.* 2016). Although computationally efficient, ray based inversion is only valid where the typical length-scale of structure is larger than the wavelength of the seismic waves. Another limitation of such inversions is that onset times are hand-picked single measurements with an associated uncertainty that is difficult to quantify.

Two new types of tomographic inversion have been developed, which are based on comparing waveforms of observed to synthetic data. These are finite frequency tomography (e.g. *Dahlen et al.* 2000, *Montelli et al.* 2004, *Hosseini et al.* 2019) and full-waveform inversion (e.g. *Tarantola and Valette* 1982, *Tromp et al.* 2005, *Fichtner et al.* 2009, *Tape et al.* 2009). Observables are extracted from the waveforms, including cross-correlation travel time differences and amplitudes. These measurements can be automated, the error quantified, and measurements can be made in multiple frequency bands. The sensitivity of such measurements to Earth structure is calculated by modelling wave-structure interaction under the first order Born approximation. These sensitivity kernels are frequency-dependent and converge to rays with increasing frequency.

Both finite frequency tomography and full waveform inversion attempt to minimise a misfit functional between observed and synthetic data, but they differ in their method of inversion. Finite frequency tomography employs a linearised inversion scheme with a single iteration. Full waveform inversion, on the other hand, uses an iterative method to update the preferred model based on the gradient of the misfit function with respect to model parameters.

We now have a catalogue of tomography models available to us. These models

differ in their choice of observable, the parameterisation of the resulting model and the inversion technique.

1.3 Forward waveform modelling

Forward waveform modelling is a complementary method to seismic tomography, which is limited by incomplete data coverage due to the irregular spacing of sources and receivers. On top of this, regularisation in the form of damping and smoothing results in models which reveal only large scale features. On a much smaller scale, waveform modelling uses high frequency waves to determine the seismic structure in localised areas. This method involves computing synthetic waveforms from models of perturbations in elastic parameters and comparing them to observed data. Various input models are tested to find the one that most closely matches the observed data.

Several methods for computing synthetic seismograms exist. These may be categorised into transform, ray-based, modal or numerical methods. Transform methods are applicable to plane-layered media or radially symmetric Earth models, once an Earth flattening approximation has been applied. The most common transform method is reflectivity (*Fuchs and Müller*, 1971; *Kind*, 1985; *Kennett*, 1983, 2001), which finds plane wave solutions to the wave equation in ray parameter-frequency space (hence the term transform method.)

Ray-based (or asymptotic) methods are based on high-frequency ray theory; seismograms are produced by summing over rays. Methods for summation include WKBJ (*Chapman*, 1978), Maslov (*Chapman and Drummond*, 1982) and Gaussian beam summation (*Červený et al.*, 1982). Asymptotic methods are valid where properties of the medium change over lengthscales longer than the seismic wavelength. Synthetic seismograms can also be calculated from a summation of modes (*Dahlen*, 1987; *Li and Tanimoto*, 1993). Such a summation can describe completely the wavefield at any depth within the Earth.

Finally, numerical methods provide a complete solution to the equations of seismic

wave propagation in arbitrarily complex 3-D media. There are many examples of numerical methods that can be used for this purpose including finite difference (e.g. *Chaljub and Tarantola* 1997; *Igel et al.* 2002), spectral element (e.g. *Chaljub et al.* 2007; *Komatitsch and Tromp* 2002a,b) and discontinuous Galerkin (e.g. *Dumbser and Käser* 2006). The community code SPECFEM (*Komatitsch and Tromp*, 2002a,b) is a widely used spectral element method designed to handle fully 3-D Earth models containing velocity variations over a range of lengthscales. AxiSEM3D (*Leng et al.*, 2016, 2019) is another such method, a hybrid between the pseudo-spectral and spectral element methods with increased computational efficiency compared to SPECFEM. Although numerical methods take into account the full interaction between wavefield and structure, their practical limitation is that they are computationally expensive.

1.4 Lower mantle seismic structures

An area of particular interest to seismologists is the region of the core-mantle boundary (CMB). The CMB lies at a depth of nearly 2900 km below Earth's surface. This is both a thermal and compositional boundary between the metallic, liquid outer core and the rocky, solid mantle. The lowermost 200 - 300 km of the mantle is therefore expected to be a region of chemical heterogeneity as well as vertical and lateral movement. Understanding the dynamics in the lowermost mantle is crucial for gaining a complete picture of mantle convection, mantle chemistry, and the thermal structure and evolution of the Earth. However, seismic structures in the lower mantle are difficult to observe because their signature in seismic observables can be lost or distorted as the waves travel to the surface.

1.4.1 Large low velocity provinces

The large scale structure of the lowermost mantle is revealed by seismic tomography. Tomography models typically show two large areas of slow seismic velocity surrounded by fast material (Figure 1.1.) The fast material is thought to be the remnants of

ancient subducted slabs, which have sunk to the bottom of the mantle. The locations of faster velocity in the lower mantle geographically align with the locations of ancient subduction zones.

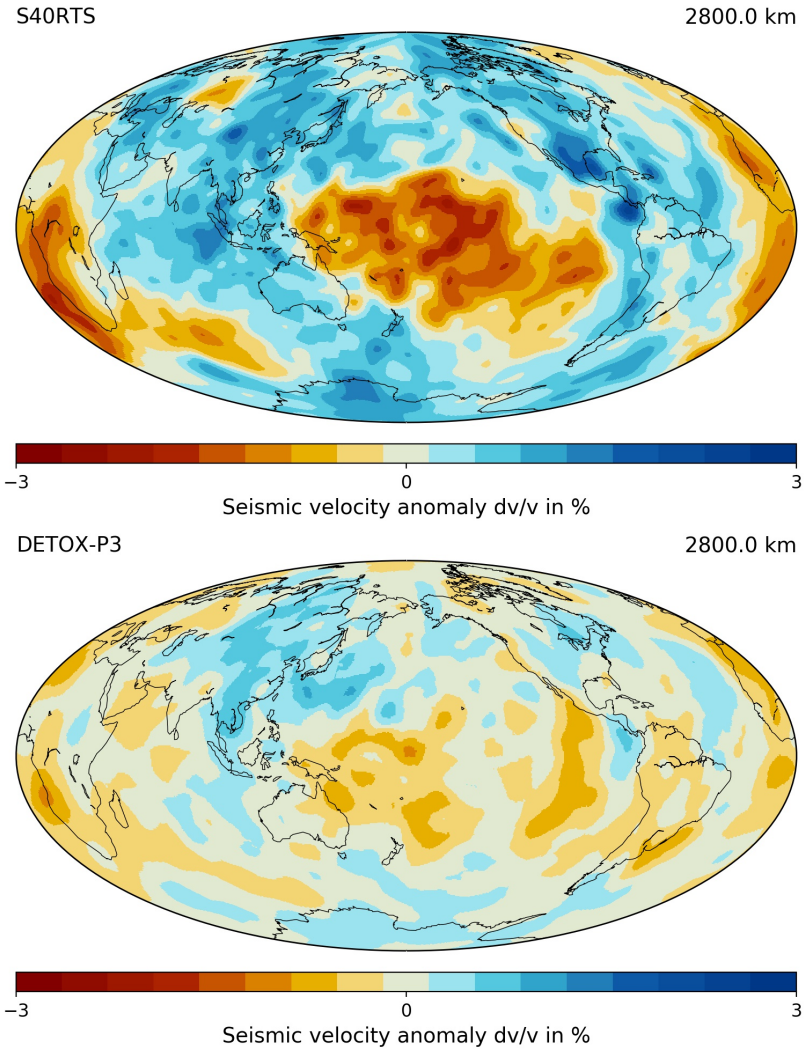


Figure 1.1: Tomography model slices 100 *km* above the CMB. The upper figure is a slice from S40RTS (*Ritsema et al.*, 2011) showing perturbations in v_s. The lower figure is from DETOX-P3 (*Hosseini et al.*, 2019) and shows perturbations in v_p. The LLVPs can be seen in both models under Africa and the Pacific, surrounded by fast material. Figure created using SubMachine (*Hosseini et al.*, 2018).

Termed large low velocity provinces (LLVPs), the two areas of slow seismic wavespeed they in two almost antipodal locations; one is under the Pacific Ocean and the other is beneath Africa and the Atlantic Ocean. They are structures of significant size, covering around 30% of the CMB by area (*Garnero et al.*, 2016) and extending thousands of

kilometres vertically into the mantle from the CMB. They are observed in models of both v_p and v_s, with average perturbation of around -1% and -3% respectively.

Different tomography models are mostly in agreement on the large scale structure of LLVPs. Further observations of fine scale structure of the LLVPs come from waveform modelling. Waves which interact with LLVP edges display abrupt changes in arrival time as well as in waveform shape (e.g. *He and Wen* 2012, *Zhao et al.* 2015, *Sun and Miller* 2013, *Frost and Rost* 2014). A broadening of the waveform is associated with a sharp change in velocity as a seismic wave is split into energy that passes through the LLVP and energy that refracts around it (multipathing). Past studies have predominantly focussed on shear waves, although some also use data from P phases (*Frost and Rost*, 2014; *Ni and Helmberger*, 2003a; *Wang and Wen*, 2007a). Waveform modelling studies have also found internal boundaries within the LLVPs (*He and Wen*, 2009), while layered structures have also been suggested (*Lay et al.*, 2006; *Schumacher et al.*, 2018). Studies examining shear wave splitting have also found evidence for increased anisotropy in the regions near the edges of the LLVPs (*Lynner and Long*, 2014; *Ford et al.*, 2015; *Wang and Wen*, 2007b).

The origin of the LLVPs has proved difficult to determine. It is widely agreed that the LLVPs must be hotter than their surroundings, and the locations of the LLVPs have been associated with the presence of volcanic hotspots and large igneous provinces on Earth's surface (*Torsvik et al.*, 2010). However, the question of whether or not they are chemically distinct remains yet to be answered. If the LLVPs are purely thermal structures, they are likely to be clusters of plumes which are blurred together by tomography. This is because individual upwellings are likely to be much narrower than the LLVPs themselves (*Schubert et al.*, 2004; *Bull et al.*, 2009).

The observation of sharp sides of the LLVPs has been used to support the hypothesis that the structures are chemically distinct from their surroundings. These thermochemical piles could be composed of material with an intrinsically elevated density that collects at the bottom of the mantle (*McNamara and Zhong*, 2005; *Bull et al.*, 2009; *Mulyukova et al.*, 2015). Such thermochemical piles could be primordial, the

remnants of differentiation in the early Earth. They they could also have accumulated in the lower mantle from subducted oceanic crust (*Brandenburg and van Keken*, 2007; *Nakagawa et al.*, 2009; *Mulyukova et al.*, 2015; *Hirose et al.*, 1999) or from core-mantle interactions (*Knittle and Jeanloz*, 1989; *Dubrovinsky et al.*, 2003). Mantle convection simulations which use reconstructed plate motions as boundary conditions show how dense material could be swept up into piles in the same locations as the LLVPs today (*Brandenburg and van Keken*, 2007; *McNamara and Zhong*, 2005; *Bull et al.*, 2009), making this a feasible scenario. However, so far measurements of the density of the LLVPs using different techniques have arrived at different conclusions. Earlier normal mode studies suggested an increased density (*Trampert et al.*, 2004), but a subsequent inversion of Stoneley mode data (*Koelemeijer et al.*, 2017) suggested that the LLVPs have decreased density relative to their surroundings (although they may still be denser that material above them). Recently however, *Lau et al.* (2017) use constraints from tidal tomography to suggest that the LLVPs are denser than their surroundings, thus keeping the debate open.

1.4.2 Edges of the LLVP under Africa

The edges of the LLVPs appear to be of particular importance in determining the nature of the LLVPs and how they fit into mantle dynamics as a whole. The edges of the African LLVP (ALLVP) are less studied than the Pacific LLVP, so there is potential for far more information to be extracted from observed data.

Waveform modelling studies are in agreement that the edges of the ALLVP are sharp, but the shapes and structures as derived by different authors are not consistent. Some models suggest a layer at the bottom of the ALLVP with a stronger velocity reduction compared to the rest of the structure (e.g. *Wang and Wen* 2004, 2007a), whereas others find a constant reduction in velocity (e.g. *Ni et al.* 2002). The most studied 2-D cross section is that from the East Pacific Rise, through the Drake Passage, the South Sandwich Islands, Southern Africa and Iran to the Hindu Kush. Three competing models have been proposed for the shape of the ALLVP in this cross-section

(*Ni et al.*, 2002; *Ni and Helmberger*, 2003b,c,a; *Wang and Wen*, 2007a; *Kästle et al.*, 2017), which are depicted in Figure 1.2. Data from earthquakes located to the West of Africa constrain the characteristics of the southwestern flank of the structure, which is in agreement between the models and is found to be sloping towards Africa. However, the northeastern side is disputed and models differ greatly. *Ritsema et al.* (1998), *Ni et al.* (2002) and *Ni and Helmberger* (2003b,a) propose a model where both flanks of the structure dip in the same direction (red in Figure 1.2). However *Wang and Wen* (2007a) suggest that the northeastern flank dips in the opposite direction to the southwestern flank producing a bell-like structure (blue in Figure 1.2). *Kästle et al.* (2017) have suggested that the LLVP along this corridor is actually composed of two parts. They favour the same structure as *Ni et al.* (2002) to the west, and propose a further parallel structure to the east (green in Figure 1.2).

Earthquakes in other locations provide further constraints on the shape of the AL-LVP beyond this 2-D cross-section. For example, waves from events in the South Atlantic Ridge recorded in the European-Mediterranean region traverse a roughly perpendicular path (*Wen et al.*, 2001; *Sun and Miller*, 2013). *Wang and Wen* (2004) use a wide range of azimuths to obtain good coverage of all but the Northern edge of the structure.

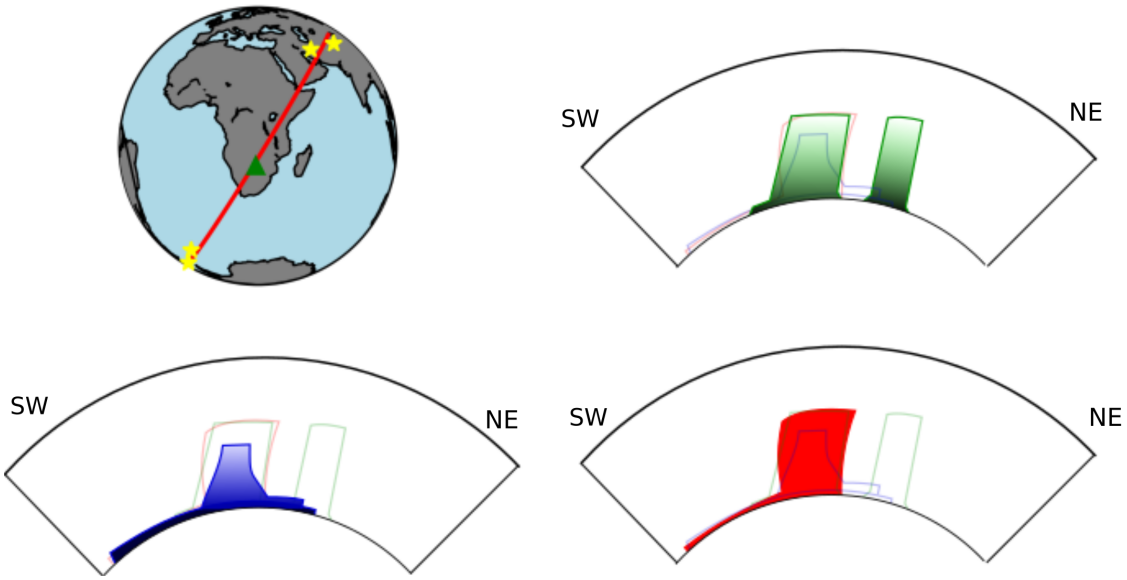


Figure 1.2: Summary of three competing models in $\delta ln(v_s)$ of the African LLVP in a 2-D cross-section. The cross-section is shown in the upper left figure, represented by the red line. Stations are typically located in the South of the African continent (represented by the green triangle), and earthquakes are located around Drake Passage, South Sandwich, Iran and the Hindu Kush (represented by yellow stars). The upper right shows the model suggested by *Kästle et al.* (2017) in green. Darker colours represent stronger velocity reductions; this model has a linearly increasing reduction in $\delta ln(v_s)$ with depth. The bottom right shows the model suggested by *Ni et al.* (2002) in red. This has a constant velocity reduction. The bottom left shows the model favoured by *Wang and Wen* (2004) and *Wang and Wen* (2007a) in blue. This includes a very low velocity layer at the bottom of the mantle, and a weaker bell-shaped structure above.

1.5 Motivation and thesis outline

Investigations into the edges of the LLVP are limited by their chosen method for forward modelling. Many of them focus on source-receiver geometries with small variations in azimuth, providing the justification for 2-D modelling to be used. However, this neglects 3-D effects on the waveform which could be important. Although the final suggested structures may be defined in 3-D, these are built up from multiple 2-D cross-sections. In addition, studies are often based on ray theory, which neglects wave effects (such as diffraction) and is only valid for large structures compared to the wavelength. A ray based method also does not take into account the full sensitivity of the wavefield to the structure. Where the full 3-D wave-structure interaction is taken into account

using numerical methods, forward modelling is very expensive and limits the number of models that can be tested.

There is inherent non-uniqueness in global seismology, with many models capable of matching the observed data. This is particularly the case in the ALLVP region, where different authors in some cases use the same data as evidence to support different models. Therefore, the most complete and desirable solution is a set of models that fit our observed data to within error. In order to do this, a forward modelling scheme is required that is fast enough to allow for the generation of synthetic seismograms for many models, while steering clear of asymptotic approximations or limiting the modelling to 2-D.

The aim of this thesis is to provide such a modelling scheme using first order, single scattering approximations. Chapter 2 details the theory behind this approach, which is dependent on calculating waveform sensitivity kernels under the Born approximation. Chapter 3 presents MC Kernel, which is a method to calculate sensitivity kernels in radially symmetric Earth models. Chapter 4 describes the implementation and benchmarking of our first order method. The domain of validity of the approximations is determined by testing them on models with spherical perturbations in the lower mantle. Chapter 5 is a demonstration of how this method can be applied to lower mantle-type structures. The method is tested on a tomography model, and then on a a simple ALLVP-like structure defined in v_p. Finally, the conclusions of this work are presented.

Chapter 2

Theory of linearised wavefield scattering

2.1 The seismic wave equation

The governing equation in seismology is the seismic wave equation, which describes how seismic waves propagate through the Earth. The wave equation is derived by applying Newton's second law ($F = ma$) to an infinitesimal cube in a continuous elastic medium. We do not reproduce the derivation completely here (see e.g. *Shearer* 2009), but we define the important physical quantities and concepts.

Internal forces in an elastic medium are fully described by the stress tensor, σ. Each element σ_{ij} is the force per unit area in the i direction, on the surface perpendicular to $\boldsymbol{x_j}$. Another quantity, the traction vector, is the total stress acting on a single surface dS with unit normal $\mathbf{n}$. Each component of the traction vector T_i, is given by $T_i = \sigma_{ij} n_j$.

As the result of an applied stress, the medium is displaced. This displacement may be decomposed into rigid rotation and distortion. The distortion is described by the symmetric strain tensor

$$\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i). \tag{2.1}$$

Stress and strain are linearly related by the elastic tensor, c_{ijkl}, according to Hooke's law,

$$\epsilon_{ij} = c_{ijkl}\sigma_{kl}. \tag{2.2}$$

The elastic tensor has 81 elements, 21 of which are independent in an anisotropic medium, where the material properties vary depending on directions. The simpler case is that of an isotropic medium, which is considered in this thesis. The subsequent extension to anisotropy would be straightforward. In the isotropic case, there are only two independent parameters (λ and μ) and the elastic tensor is defined as

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{il}\delta_{jk} + \delta_{ik}\delta jl). \tag{2.3}$$

In the above equation, λ and μ are the Lamé parameters. μ is the shear modulus, which quantifies the resistance to shearing while λ does not have a straightforward physical interpretation.

Applying Newton's second law, combined with some algebraic manipulation, takes us to the well-known seismic wave equation,

$$\rho\ddot{\mathbf{u}} = \nabla\lambda(\nabla\cdot\mathbf{u}) + \nabla\mu\cdot[\nabla\mathbf{u} + (\nabla\mathbf{u})^T] + (\lambda+\mu)\nabla\nabla\cdot\mathbf{u} + \mu\nabla^2\mathbf{u} + \mathbf{f}, \tag{2.4}$$

where ρ is density, $\mathbf{u}$ is the displacement vector, and $\ddot{\mathbf{u}}$ denotes its second derivative in time. $\mathbf{f}$ represents external body forces acting on the medium.

Equation 2.4 can be expressed more succinctly in terms of the linear operator $\mathscr{L}$, as:

$$\mathscr{L}\mathbf{u} = \mathbf{f}. \tag{2.5}$$

where:

$$\mathscr{L}\mathbf{u} = \rho\ddot{\mathbf{u}} - \nabla\lambda(\nabla\cdot\mathbf{u}) - \nabla\mu\cdot[\nabla\mathbf{u} - (\nabla\mathbf{u})^T] - (\lambda+\mu)\nabla\nabla\cdot\mathbf{u} - \mu\nabla^2\mathbf{u} \tag{2.6}$$

in the time domain. Switching to the frequency domain, $\mathscr{L}$ is defined by

$$\mathscr{L}\mathbf{u} = -\rho\omega^2\mathbf{u} - \nabla\lambda(\nabla\cdot\mathbf{u}) - \nabla\mu\cdot[\nabla\mathbf{u} - (\nabla\mathbf{u})^T] - (\lambda+\mu)\nabla\nabla\cdot\mathbf{u} - \mu\nabla^2\mathbf{u}, \tag{2.7}$$

where ω is frequency.

The free surface (zero traction) boundary condition requires that $\boldsymbol{u}$ must also satisfy

$$[\mathscr{B}\boldsymbol{u}]^+_- = [\hat{\boldsymbol{n}}(\lambda\nabla\cdot\boldsymbol{u}) + \hat{\boldsymbol{n}}\cdot(\mu[\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T])]^+_- = \mathbf{0}. \tag{2.8}$$

2.1.1 Green's tensor

The Green's tensor $\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}')$ is the displacement at receiver location $\boldsymbol{r}$ due to an instantaneous point force at $\boldsymbol{r}'$. Each element G_{ij} is the displacement in the i direction due to a point force in the j direction. $\boldsymbol{G}$ is defined as the solution to the equation:

$$\mathscr{L}\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}') = \boldsymbol{I}\delta(\boldsymbol{r}-\boldsymbol{r}',\omega), \tag{2.9}$$

where $\boldsymbol{I}$ is the identity matrix, and δ represents the Kronecker delta function in space. In the time domain, this would be written as:

$$\mathscr{L}\boldsymbol{G}(\boldsymbol{r},t;\boldsymbol{r}') = \boldsymbol{I}\delta(\boldsymbol{r}-\boldsymbol{r}')\boldsymbol{I}\delta(t-t'), \tag{2.10}$$

for a source at time t'. Since the Green's functions give the response to a point source, where there is a single force distribution (as opposed to a double couple), $\mathbf{f}$, the solution for $\boldsymbol{u}$ is

$$\mathbf{u} = \int \boldsymbol{G}(\boldsymbol{r}, \omega; \mathbf{r}')\mathbf{f}(\boldsymbol{r}', \omega)d^3\mathbf{r}'. \tag{2.11}$$

In the above equation the integral is over space, and multiplication in the frequency domain corresponds to a convolution in time. The force distribution at an earthquake source is represented by the moment tensor. Each component M_{ij} represents a pair of forces (a double couple) acting in opposite directions along the i axis, separated by distance in the j direction. The component M_{ij} is given by the magnitude of the force multiplied by the distance by which the forces are separated. the displacement at a location $\boldsymbol{r}$ due to a moment tensor source at $\boldsymbol{r}'$ is obtained using the representation theorem (*Dahlen and Tromp*, 1998), and is expressed as:

$$\boldsymbol{u}(\boldsymbol{r}, \omega) = (i\omega)^{-1}\boldsymbol{M}(\omega) : \nabla_s \boldsymbol{G}^T(\boldsymbol{r}, \omega; \boldsymbol{r}') \tag{2.12}$$

The seismic wavefield may be split into compressional (P) wave and shear (S) wave motion under the assumption that the Lamé parameters are constant. This may be assumed when the wavelength is considerably shorter than the wavelength over which the Lamé parameters change, or when the Earth is stratified into layers with constant Lamé parameters. The familiar expressions for compressional and shear wavespeed, v_p and v_s respectively are shown below.

$$\begin{aligned} v_p &= \sqrt{\frac{\lambda + 2\mu}{\rho}} \\ v_s &= \sqrt{\frac{\mu}{\rho}} \end{aligned} \tag{2.13}$$

The most convenient parameterisation of Earth's interior is not in the fundamental parameters ρ, λ, μ, but in ρ, v_p, v_s.

2.2 Wave scattering

The 1-D structure of the Earth has been well constrained in ρ, v_p, v_s. In global seismology, these 1-D models are used as reference or background models, on top of which 3-D heterogeneities are imposed. The seismic wavefield is analogously separated into a background wavefield and a scattered wavefield. Straightforwardly, the background wavefield, $\mathbf{u^0}$, is the solution to the wave equation in a 1-D background medium defined by some reference Earth model. The scattered wavefield, $\mathbf{u^s}$, is the addition to the background wavefield which results from seismic waves scattering off 3-D perturbations to the background medium. The interactions between seismic waves and scatterers within the Earth are depicted in Figure 2.1. It may be formulated as

$$\mathbf{u} = \mathbf{u}^0 + \int_V [\sum_i^\infty K_i \delta m^i] dV. \tag{2.14}$$

Here, $\mathbf{u^0}$ is the background wavefield, $\mathbf{u}$ is the total wavefield and $\mathbf{u_s}$ is given by the integral expression in the second term. δm is the perturbation to the background model, while V is the global volume over which the integral is calculated. K_i are scattering coefficients, and the index i denotes the scattering order. i^{th} order scattering means that i scattering events take place between the source and the receiver. The strength of the i^{th} order scattered wave is proportional to δm^i, and so for small perturbations scattering becomes weaker the higher the order becomes. This thesis is concerned with first order approximations, i.e. ones which truncate the scattering integral at first order. Physically, this means that only single scattering interactions are taken into account.

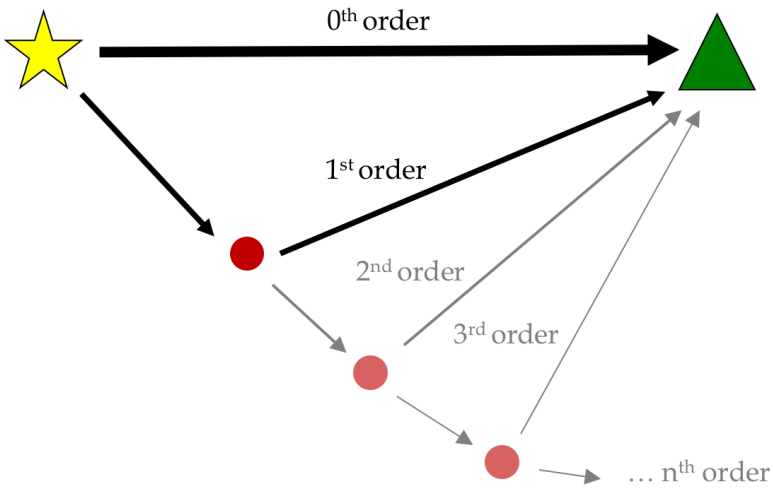


Figure 2.1: Schematic depicting scattering of seismic waves within the Earth. Waves travel from the source (yellow star) to the receiver (green triangle). The 0^{th} order wave through the background medium is represented by the thick black arrow. The scattered wave induced by a single encounter with a heterogeneity arrives at the station as the first order perturbation. Some of the scattered waves go on to encounter a second scatterer, and so on. The black arrows indicate the interactions which are taken into account by first order approximations, and the grey arrows represent higher order scattering which is ignored.

2.3 The Born approximation

The most commonly applied first order approximation in global seismology is the Born approximation. Neglecting multiple scattering, we may write Equation 2.14 simply as $\mathbf{u} = \mathbf{u^0} + \boldsymbol{\delta}\mathbf{u}$, i.e. $\mathbf{u^s} = \boldsymbol{\delta}\mathbf{u}$.

In global seismology, the Born approximation is most often applied to the computation of measurement sensitivity kernels used in finite frequency tomography, or to the calculation of Frechét derivatives defining gradients in full waveform inversion. Finite frequency tomography relies on the linearity of cross-correlation travel time measurements, that is the linear dependence of a delay time measured by cross-correlation on the perturbation to a velocity model. This linear relationship has been demonstrated to hold with a smooth velocity model for deviations of up to 20%, and with a sharp model for deviations up to 10% (*Diego Mercerat and Nolet*, 2013). Therefore, for this derived observable (the cross-correlation travel time measurement), the first order Born approximation is valid for realistic Earth models. However, the domain of validity for cross-correlation measurements is considerably larger than that of direct

waveform differences (*Diego Mercerat and Nolet*, 2013).

Previous work on the applicability of first order approximations to forward modelling of seismic waveforms has focused on the Born approximation as applied to waveform modelling of long period ($> 30s$) surface waves, using a normal mode framework (*Capdeville et al.*, 2000, 2002; *Capdeville*, 2005; *Romanowicz et al.*, 2008; *Panning et al.*, 2009; *Parisi et al.*, 2015). *Capdeville* (2005) compared the performance of the Born approximation to a coupled spectral element - normal mode approach. They computed synthetic seismograms for major and minor arc Rayleigh waves at a period of 100 s using both methods in models based on a tomography model, one with amplitudes divided by 10, and the other with the original amplitudes of the tomography model. Good agreement was found for the first model, but significant differences arising from non-linear effects were found between the two approaches for the realistic model. Much better agreement was found between Frechét derivatives calculated with the two methods.

Romanowicz et al. (2008) calculate long-period seismograms (> 100 s) using the Born approximation (as well as two other asymptotic approximations) and compare them to 3-D spectral-element solutions, computed for simple 3-D blob-like models. They found that the performance of the Born approximation decreases with increasing size of anomaly as it is not able to account for the continuously increasing phase delay as waves travel through an extended structure.

Panning et al. (2009) tested the applicability of the Born approximation for seismic waveforms down to 100 s, both in simple models (blobs) and in random models with a similar structural wavelength to models from global tomography. They determine that the Born approximation performs poorly wherever the accumulated phase delay is greater than a quarter of the period. More accurate synthetic seismograms may be obtained by including information on accumulated phase. They also find that phase delay kernels calculated using the Born approximation agree with measurements from synthetic data. *Parisi et al.* (2015) also find that for fundamental mode Rayleigh and Love waves in realistic global 3-D models based on tomography the performance of the

Born approximation is correlated with phase delay. They propose that once the phase delay exceeds 15% the Born approximation does not sufficiently model phase delay. In order to sufficiently model the amplitude of the seismogram the phase delay should not exceed 5%.

The results discussed above highlight the ability of the Born approximation to predict phase more accurately than amplitude when compared to more exact methods, and this is exploited in tomography. However, waveform modelling of the seismogram's amplitude appears to be more difficult under the Born approximation.

2.3.1 Derivation

Below, the well-known application of the Born approximation to the seismic wave equation is recounted. This derivation may be found in e.g. *Nolet* (2008); *Dahlen et al.* (2000). The elastic parameters are linearly perturbed:

$$\begin{aligned} \rho &\to \rho + \delta\rho, \\ \lambda &\to \lambda + \delta\lambda, \\ \mu &\to \mu + \delta\mu. \end{aligned} \tag{2.15}$$

The linear operator $\mathscr{L}$ as defined in Equation 2.7 is correspondingly perturbed by $\delta\mathscr{L}$. The perturbed wavefield must also satisfy the seismic wave equation and therefore we may write:

$$(\mathscr{L} + \delta\mathscr{L})(\boldsymbol{u} + \delta\boldsymbol{u}) = \boldsymbol{f} \tag{2.16}$$

Multiplying out the above equation, we obtain both first order and second order terms. Neglecting second order terms and using Equation 2.5 means we have

$$\mathscr{L}\delta\boldsymbol{u} = -\delta\mathscr{L}\boldsymbol{u}. \tag{2.17}$$

Applying the boundary conditions in the perturbed medium means that the follow-

ing equation must also be satisfied:

$$[\mathscr{B}\delta\boldsymbol{u}]^+_- = -[\delta\mathscr{B}\boldsymbol{u}]^+_- \tag{2.18}$$

$-\delta\mathscr{L}\boldsymbol{u}$ acts as a body force and we can solve for $\delta\boldsymbol{u}$ using Equation 2.11. The total perturbation to the waveform, including the boundary conditions is:

$$\delta\boldsymbol{u} = -\int \boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}')\cdot\delta\mathscr{L}\boldsymbol{u}d^3\boldsymbol{r}' + \int_\Sigma [\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}')\cdot\delta\mathscr{B}\boldsymbol{u}]^+_- d^2\mathbf{r}' \tag{2.19}$$

Applying Gauss' theorem and integration by parts eliminates derivatives of the model parameters, and the surface integrals are cancelled by the the boundary conditions. Therefore the perturbation to the waveform under the Born approximation can be written as:

$$\begin{aligned}\delta\boldsymbol{u} = &\int \delta\rho\omega^2\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}')\cdot\boldsymbol{u}(\boldsymbol{r}')d^3\boldsymbol{r}' \\ &-\int \delta\lambda(\nabla'\cdot\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}'))(\nabla'\cdot\boldsymbol{u}(\boldsymbol{r}'))d^3\boldsymbol{r}' \\ &-\int \delta\mu(\nabla'\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}') : [\nabla'\boldsymbol{u}(\boldsymbol{r}') + \nabla'\boldsymbol{u}(\boldsymbol{r}')^T])d^3\boldsymbol{r}'\end{aligned} \tag{2.20}$$

2.3.2 Waveform kernels

Equation 2.20 may be written as

$$\boldsymbol{\delta u} = \int [K^0_\rho(\omega,\mathbf{r}')\delta\rho + K^0_\lambda(\omega,\mathbf{r}')\delta\lambda + K^0_\mu(\omega,\mathbf{r}')\delta\mu)]d^3\mathbf{r}', \tag{2.21}$$

which defines the waveform kernels $K^0_X(\omega,\mathbf{r}')$. These quantify the sensitivity of $\boldsymbol{\delta}\mathbf{u}$ to perturbations in an elastic parameter X, where the superscript 0 indicates that X is a fundamental parameter. The fundamental parameters are one of ρ,λ,μ in this parameterisation. The waveform kernels describe how a scatterer at location $\mathbf{r}'$ perturbs the waveform at a frequency ω. Using the reciprocity of Green's functions ($\boldsymbol{G}(\boldsymbol{r},\omega;\boldsymbol{r}') = \boldsymbol{G}(\boldsymbol{r}',\omega;\boldsymbol{r})$) we may write the waveform kernels as

$$K^0_\rho(\omega, \mathbf{r'}) = \omega^2 \boldsymbol{G}(\boldsymbol{r'}, \omega; \boldsymbol{r}) \cdot \boldsymbol{u}(\boldsymbol{r'}) \tag{2.22}$$

$$K^0_\lambda(\omega, \mathbf{r'}) = -\nabla' \cdot \boldsymbol{G}(\boldsymbol{r'}, \omega; \boldsymbol{r})(\nabla' \cdot \boldsymbol{u}(\boldsymbol{r'})) \tag{2.23}$$

$$K^0_\mu(\omega, \mathbf{r'}) = -\nabla' \boldsymbol{G}(\boldsymbol{r'}, \omega; \boldsymbol{r}) : [\nabla' \boldsymbol{u}(\boldsymbol{r'}) + \nabla' \boldsymbol{u}(\boldsymbol{r'})^T]. \tag{2.24}$$

where $\mathbf{r'}$ is the location of the scatterer and $\mathbf{r}$ is the location of the receiver. $\boldsymbol{G}(\boldsymbol{r'}, \omega; \boldsymbol{r})$ is called the backward wavefield, as it propagates from the receiver to the scatterer. Its source is a delta function at the receiver. $\boldsymbol{u}(\boldsymbol{r'})$ is called the forward wavefield as it propagates from the source to the scatterer. In realistic scenarios its source is an earthquake, represented by a moment tensor.

We can rewrite the expressions for waveform kernels in terms of the forward strain tensor $\overrightarrow{\boldsymbol{\epsilon}} = \frac{1}{2}[(\nabla \mathbf{u})(\nabla \mathbf{u})^T]$ and the backward strain tensor $\overleftarrow{\boldsymbol{\epsilon}} = \frac{1}{2}[(\nabla \mathbf{G})(\nabla \mathbf{G})^T]$:

$$\begin{aligned} K^0_\rho(\omega, \mathbf{r'}) &= \omega^2 \boldsymbol{G} \cdot \boldsymbol{u} \\ K^0_\lambda(\omega, \mathbf{r'}) &= -(tr(\overrightarrow{\boldsymbol{\epsilon}}))(tr(\overleftarrow{\boldsymbol{\epsilon}})) \\ K^0_\mu(\omega, \mathbf{r'}) &= -2\overleftarrow{\boldsymbol{\epsilon}} : \overrightarrow{\boldsymbol{\epsilon}} \end{aligned} \tag{2.25}$$

Since shear and compressional wavespeed (v_p and v_s respectively) are generally more useful parameters than λ and μ, we reparameterise into v_p, v_s, ρ, in which case the waveform kernels are:

$$K_{v_s}(\omega, \mathbf{r'}) = 2\rho v_s K^0_\mu(\omega, \mathbf{r'}) - 4\rho v_s K^0_\lambda(\omega, \mathbf{r'}) \tag{2.26}$$

$$K_{v_p}(\omega, \mathbf{r'}) = 2\rho v_p K^0_\lambda(\omega, \mathbf{r'}) \tag{2.27}$$

$$K_\rho(\omega, \mathbf{r'}) = K^0_\rho(\omega, \mathbf{r'}) + (v_p^2 - 2v_s^2) K^0_\lambda(\omega, \mathbf{r'}) + v_s^2 K^0_\mu(\omega, \mathbf{r'}) \tag{2.28}$$

2.3.3 Measurement kernels

Usually, measurements taken from seismograms are used when analysing data. In this case, it is useful to quantify the effect that a scatterer has on the measurement rather than the seismogram itself. Most common is the measurement of a travel time delay of a particular phase as determined by cross-correlation. Below, we follow the derivation of *Marquering et al.* (1999) to find the sensitivity of cross-correlation measurements to structure within the Earth.

The method of measuring a travel time delay using cross-correlation is as follows. An observed seismogram u^{obs} is recorded during the time interval $[t_1, t_2]$. We define u^0 as the corresponding reference synthetic seismogram, computed using a 1-D reference Earth model. The cross-correlagram, $C(\tau)$, is the inner product of the observed signal with the reference signal that has been shifted by a time τ. It is defined by $C(\tau) = \int_{t_1}^{t_2} u^0(t-\tau)u^{obs}(t)dt$. The larger the value of $C(\tau)$, the better the fit between the two seismograms. The recorded travel time delay is the value of τ which corresponds to the maximum value of $C(\tau)$.

We may find the sensitivity of such a measurement within the Born approximation, assuming single scattering to be representative of the full wavefield at the point of measurement. We replace $u^{obs}(t)$ with $u(t) = u^0(t) + \delta u(t)$. The resulting travel time shift, δT is found to be:

$$\delta T = -\frac{\int_{t_1}^{t_2} \dot{u^0}(t)\delta u(t)dt}{\int_{t_1}^{t_2} \dot{u^0}(t)^2 dt} \tag{2.29}$$

We know that under the Born approximation δu can be expressed in terms of the waveform kernels (Equation 2.21). A straightforward substitution reveals that δT can be expressed as:

$$\delta T = \int_V K_\rho^{0,T}(\boldsymbol{r'})\delta\rho d^3\boldsymbol{r'} + \int_V K_\lambda^{0,T}(\boldsymbol{r'})\delta\lambda d^3\boldsymbol{r'} + \int_V K_\mu^{0,T}(\boldsymbol{r'})\delta\mu d^3\boldsymbol{r'} \tag{2.30}$$

where $K_X^T(\boldsymbol{r})$ denotes the travel time kernel and the superscript 0 denotes a fun-

damental elastic parameter. These can be expressed in terms of the time dependent waveform kernels, $K_X(\boldsymbol{r}, t)$:

$$K_\rho^{0,T}(\boldsymbol{r'}) = -\frac{\int_{t_1}^{t_2} \dot{u^0}(t) K_\rho^0(\boldsymbol{r'}, t) dt}{\int_{t_1}^{t_2} \ddot{u^0}(t) u(t) dt} \tag{2.31}$$

$$K_\mu^{0,T}(\boldsymbol{r'}) = -\frac{\int_{t_1}^{t_2} \dot{u^0}(t) K_\mu^0(\boldsymbol{r'}, t) dt}{\int_{t_1}^{t_2} \ddot{u^0}(t) u(t) dt} \tag{2.32}$$

$$K_\lambda^{0,T}(\boldsymbol{r'}) = -\frac{\int_{t_1}^{t_2} \dot{u^0}(t) K_\lambda^T(\boldsymbol{r'}, t) dt}{\int_{t_1}^{t_2} \ddot{u^0}(t) u(t) dt} \tag{2.33}$$

As with waveform kernels, the cross-correlation kernels can be reparameterised into ρ, v_s, v_p:

$$K_{v_s}^T(\boldsymbol{r'}) = 2\rho v_s K_\mu^{0,T}(\boldsymbol{r'}) - 4\rho v_s K_\lambda^{0,T}(\boldsymbol{r'}) \tag{2.34}$$

$$K_{v_p}^T(\boldsymbol{r'}) = 2\rho v_p K_\lambda^{0,T}(\boldsymbol{r'}) \tag{2.35}$$

$$K_\rho^T(\boldsymbol{r'}) = K_\rho^{0,T}(\boldsymbol{r'}) + (v_p^2 - 2v_s^2) K_\lambda^{0,T}(\boldsymbol{r'}) + v_s^2 K_\mu^{0,T}(\boldsymbol{r'}) \tag{2.36}$$

This formulation of the Born approximation as applied to seismic waves is commonly used in finite frequency and full waveform tomography and a summary may be found in *Fichtner* (2011).

2.4 Generalised first order solution

2.4.1 Background

We now extend our discussion from the Born approximation to other first order approximations. The Rytov approximation is another such approximation. Where the Born approximation is derived from a linear perturbation to the displacement field, **u**, the Rytov approximation is derived from a first order perturbation to the phase

of the displacement wavefield, or $ln(\mathbf{u})$. *Beydoun and Tarantola* (1988) compared the applicability of the Born and Rytov approximations for acoustic waves and a simple 1-D slab model. This scenario also offers an analytical solution, to which the first order seismograms are compared. They showed that the Born approximation performs better in back-scattering scenarios, while the Rytov approximation is more applicable to forward scattering. The Rytov approximation has also been used to compute Frechét derivatives to predict arrival times (*Spetzler and Snieder*, 2001).

Marks (2006) proposed that the Born approximation is one end member of a family of first order approximations, in which the Rytov approximation is the other end member. In that paper, the approximations are applied to to the Helmholtz equation, $\nabla^2 u + k^2(\boldsymbol{r})u = 0$.

As we have seen above, under the Born approximation the perturbed wavefield is written as $u = u^0 + \delta u$. Conversely, the Rytov approximation involves a perturbation to the phase of the wavefield. The perturbed wavefield is written as $u = u^0 exp(\phi)$, where ϕ is the complex phase of the perturbed wavefield. Marks showed that these two formulations of the perturbed field u could both be expressed as $u = u^0(1 + \frac{\phi}{n})^n$ with different values of $n = 1, 2, 3, ...$ In the Born case, $n = 1$ and so $u = u^0(1 + \phi)$. The perturbation to the wavefield $\delta u = u^0\phi$. Under the Rytov approximation, the perturbed wavefield may be expressed as

$$u = u^0 \lim_{n\to\infty} \left(1 + \frac{\phi}{n}\right)^n = u^0 e^{\phi}. \tag{2.37}$$

There is a whole family of intermediate approximations corresponding to intermediate values of n.

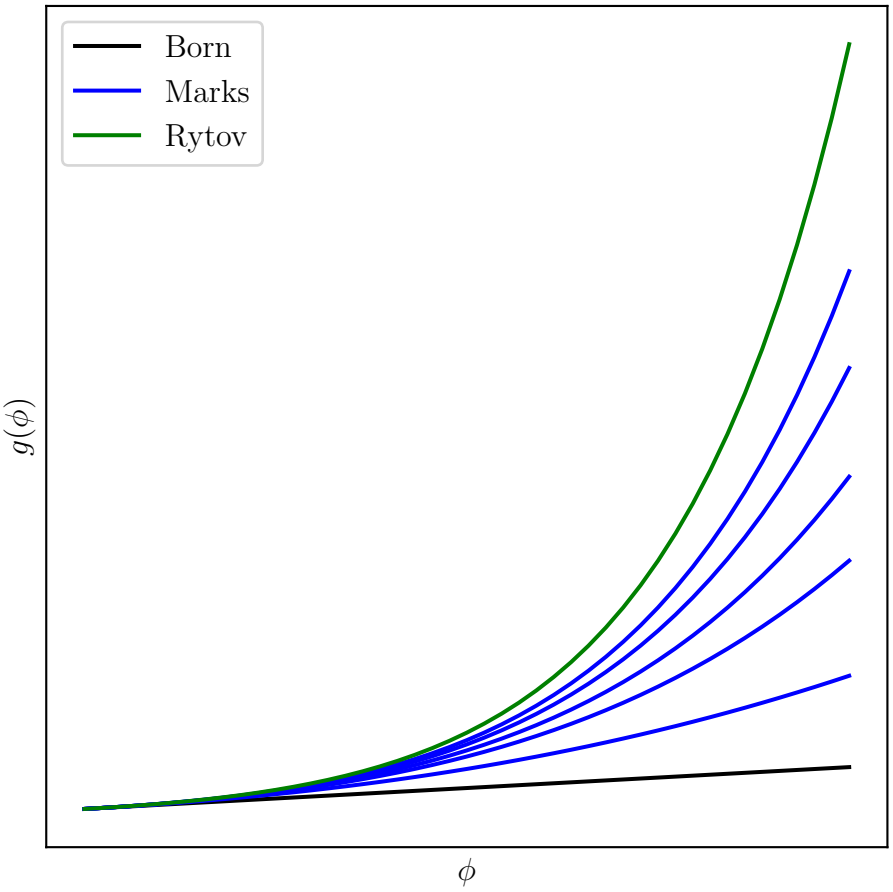


Figure 2.2: Figure showing the dependence of $g = (1 + \phi/n)^n$ on ϕ for different first order approximations. Each coloured line corresponds to a different value of n. The Born approximation ($n = 1$) is shown in black, the Rytov approximation $n \to \infty$ is shown in Green. A selection of intermediate values of $n = 3, 6, 9, 15, 25$ are shown in blue, all of which we term the Marks approximation.

2.4.2 Derivation

In this section,the approach of*Marks* (2006) is extended from the Helmholtz equation to the case of a generalised vector solution of the seismic wave equation (Equation 2.4). There is now a vector displacement wavefield $\boldsymbol{u}$, and a vector complex phase $\boldsymbol{\phi}$. This is the first time that the Rytov approximation has been applied in such a case, and furthermore the first time the Marks approach to derive a family of first order equations has been applied to a vector wavefield. It is supposed that the perturbed wavefield can be written as

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} u_1^0 g(\phi_1) \\ u_2^0 g(\phi_2) \\ u_3^0 g(\phi_3), \end{pmatrix} \tag{2.38}$$

where $g(\phi_i) = (1 + \frac{\phi_i(\boldsymbol{r})}{n})^n$

The Born approximation corresponds to the case where $n = 1$ and therefore $\mathbf{u} = \mathbf{u}^0 + \boldsymbol{\delta}\mathbf{u}$. The Rytov approximation corresponds to the case where $n \to \infty$, such that $u_i = u^0 e^{\phi_i}$ and the phase of the wavefield is linearly perturbed.

We may find the linear dependence of ϕ on changes to the elastic parameters using perturbation theory, with order parameter $\epsilon \ll 1$:

$$\begin{aligned} \mu &\to \mu^0 + \epsilon\mu^1 \\ \lambda &\to \lambda^0 + \epsilon\lambda^1 \\ \rho &\to \rho^0 + \epsilon\rho^1 \end{aligned} \tag{2.39}$$

ϕ_i can also be expanded as a perturbation series, $\phi_i = \phi_i^0 + \epsilon\phi_i^1 + \epsilon^2\phi 2_i + ...$ The first non-zero term is $\epsilon\phi^1$, since the zero order terms satisfies the unperturbed seismic wave equation. We now substitute this solution into the seismic wave equation, and neglect terms higher than first order in ϵ.

For ease of reading, we divide the linear operator $\mathscr{L}$ into its five terms $(\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3}, \mathbf{t_4}, \mathbf{t_5})$, and attend to each individually.

$$\begin{aligned}
\mathbf{t_1} &= \nabla\lambda(\nabla\cdot\boldsymbol{u}) \\
\mathbf{t_2} &= \nabla\mu\cdot[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^T] \\
\mathbf{t_3} &= (\lambda+\mu)\nabla\nabla\cdot\boldsymbol{u} \\
\mathbf{t_4} &= \mu\nabla^2\boldsymbol{u} \\
\mathbf{t_5} &= \rho\omega^2\mathbf{u}
\end{aligned} \tag{2.40}$$

Since all terms involve derivatives of $\boldsymbol{u}$, and therefore $g(\phi_i)$ it is useful to first expand $g(\phi_i)$ and $g'(\phi_i)$ to first order in a Taylor series. This gives

$$g(\phi_i) \approx 1+\epsilon\phi_i^1 \tag{2.41}$$

$$g'(\phi_i) \approx 1+\frac{n-1}{n}\epsilon\phi_i^1. \tag{2.42}$$

We start with $\mathbf{t_1}$. Firstly, we expand $\nabla\cdot\mathbf{u}$ to first order:

$$
\begin{aligned}
\nabla \cdot \boldsymbol{u} &= \frac{\partial}{\partial x_1}\left(u_1^0 g\left(\phi_1\right)\right) + \frac{\partial}{\partial x_2}\left(u_2^0 g\left(\phi_2\right)\right) + \frac{\partial}{\partial x_3}\left(u_3^0 g\left(\phi_3\right)\right) \\
&= \frac{\partial u_1^0}{\partial x_1} g\left(\phi_1\right) + u_1^0 \frac{\partial g\left(\phi_1\right)}{\partial x_1} + \frac{\partial u_2^0}{\partial x_2} g\left(\phi_2\right) + u_2^0 \frac{\partial g\left(\phi_2\right)}{\partial x_2} + \frac{\partial u_3^0}{\partial x_3} g\left(\phi_3\right) + u_3^0 \frac{\partial g\left(\phi_3\right)}{\partial x_3} \\
&\approx \frac{\partial u_1^0}{\partial x_1}\left(1+\epsilon\phi_1^1\right) + u_1^0 \frac{\partial\left(\epsilon\phi_1^1\right)}{\partial x_1}\left(1+\epsilon\phi_1^1\left(\frac{n-1}{n}\right)\right) \\
&\quad + \frac{\partial u_2^0}{\partial x_2}\left(1+\epsilon\phi_2^1\right) + u_2^0 \frac{\partial \epsilon\phi_2^1}{\partial x_2}\left(1+\epsilon\phi_2^1\left(\frac{n-1}{n}\right)\right) \\
&\quad + \frac{\partial u_3^0}{\partial x_3}\left(1+\epsilon\phi_3^1\right) + u_3^0 \frac{\partial \epsilon\phi_3^1}{\partial x_3}\left(1+\epsilon\phi_3^1\left(\frac{n-1}{n}\right)\right) \\
&= \frac{\partial u_1^0}{\partial x_1} + \frac{\partial u_1^0}{\partial x_1}\epsilon\phi_1^1 + u_1^0 \frac{\partial \epsilon\phi_1^1}{\partial x_1} + \frac{n-1}{n}\phi_1 u_1^0 \frac{\partial \epsilon\phi_1^1}{\partial x_1} \\
&\quad + \frac{\partial u_2^0}{\partial x_2} + \frac{\partial u_2^0}{\partial x_2}\epsilon\phi_2^1 + u_2^0 \frac{\partial \epsilon\phi_2^1}{\partial x_2} + \frac{n-1}{n}\epsilon\phi_2^1 u_2^0 \frac{\partial \epsilon\phi_2^1}{\partial x_2} \\
&\quad + \frac{\partial u_3^0}{\partial x_3} + \frac{\partial u_3^0}{\partial x_3}\epsilon\phi_3^1 + u_3^0 \frac{\partial \epsilon\phi_3^1}{\partial x_3} + \frac{n-1}{n}\epsilon\phi_3^1 u_3^0 \frac{\partial \epsilon\phi_3^1}{\partial x_3} \\
&\approx \frac{\partial u_1^0}{\partial x_1} + \epsilon\left(\frac{\partial u_1^0}{\partial x_1}\phi_1^1 + u_1^0 \frac{\partial \phi_1^1}{\partial x_1}\right) \\
&\quad + \frac{\partial u_2^0}{\partial x_2} + \epsilon\left(\frac{\partial u_2^0}{\partial x_2}\phi_2^1 + u_2^0 \frac{\partial \phi_2^1}{\partial x_2}\right) \\
&\quad + \frac{\partial u_3^0}{\partial x_3} + \epsilon\left(\frac{\partial u_3^0}{\partial x_3}\phi_3^1 + u_3^0 \frac{\partial \phi_3^1}{\partial x_3}\right) \\
&= \frac{\partial u_1^0}{\partial x_1} + \frac{\partial u_2^0}{\partial x_2} + \frac{\partial u_3^0}{\partial x_3} + \epsilon\left(\frac{\partial}{\partial x_1}\left(u_1^0\phi_1^1\right) + \frac{\partial}{\partial x_2}\left(u_2^0\phi_2^1\right) + \frac{\partial}{\partial x_3}\left(u_3^0\phi_3^1\right)\right)
\end{aligned}
\tag{2.43}
$$

Secondly, we define $\nabla\lambda$ to first order as:

$$
\nabla\lambda = \begin{pmatrix} \frac{\partial\lambda^0}{\partial x_1} + \epsilon\frac{\partial\lambda^1}{\partial x_1} \\ \frac{\partial\lambda^0}{\partial x_2} + \epsilon\frac{\partial\lambda^1}{\partial x_2} \\ \frac{\partial\lambda^0}{\partial x_3} + \epsilon\frac{\partial\lambda^1}{\partial x_3} \end{pmatrix} \tag{2.44}
$$

For brevity, we now consider the first element of the vector $\mathbf{t_1}$. We do not present the analogous derivation for the other two elements because they have the same form, but with reordered subscripts.

$$
\begin{aligned}
[\mathbf{t_1}]_1 &= \left(\frac{\partial\lambda^0}{\partial x_1} + \epsilon\frac{\partial\lambda^1}{\partial x_1}\right)\left(\frac{\partial u_1^0}{\partial x_1} + \frac{\partial u_2^0}{\partial x_2} + \frac{\partial u_3^0}{\partial x_3} + \epsilon\left(\frac{\partial}{\partial x_1}\left(u_1^0\phi_1^1\right) + \frac{\partial}{\partial x_2}\left(u_2^0\phi_2^1\right) + + \frac{\partial}{\partial x_3}\left(u_3^0\phi_3^1\right)\right)\right) \\
&= \frac{\partial\lambda^0}{\partial x_1}\frac{\partial u_1^0}{\partial x_1} + \epsilon\frac{\partial\lambda^0}{\partial x_1}\frac{\partial}{\partial x_1}\left(u_1^0\phi_1^1\right) + \epsilon\frac{\partial\lambda^1}{\partial x_1}\frac{\partial u_1^0}{\partial x_1} + 2^{nd}order \\
&\quad + \frac{\partial\lambda^0}{\partial x_1}\frac{\partial u_2^0}{\partial x_2} + \epsilon\frac{\partial\lambda^0}{\partial x_1}\frac{\partial}{\partial x_2}\left(u_2^0\phi_2^1\right) + \epsilon\frac{\partial\lambda^1}{\partial x_1}\frac{\partial u_2^0}{\partial x_2} + 2^{nd}order \\
&\quad + \frac{\partial\lambda^0}{\partial x_1}\frac{\partial u_3^0}{\partial x_3} + \epsilon\frac{\partial\lambda^0}{\partial x_1}\frac{\partial}{\partial x_3}\left(u_3^0\phi_3^1\right) + \epsilon\frac{\partial\lambda^1}{\partial x_1}\frac{\partial u_3^0}{\partial x_3} + 2^{nd}order
\end{aligned}
\tag{2.45}
$$

It can be seen from the above equation that there are terms corresponding to the 0^{th} order $\nabla\lambda_0(\nabla\cdot\boldsymbol{u}^0)$ as well as first order terms. Combining with $[\mathbf{t_1}]_2$ and $[\mathbf{t_1}]_3$, we see that:

$$
\nabla\lambda\left(\nabla\cdot\boldsymbol{u}\right) \approx \nabla\lambda_0\left(\nabla\cdot\boldsymbol{u}^0\right) + \nabla\lambda^0\left(\nabla\cdot\begin{pmatrix} u_1^0\epsilon\phi_1^1 \\ u_2^0\epsilon\phi_2^1 \\ u_3^0\epsilon\phi_3^1 \end{pmatrix}\right) + \nabla\epsilon\lambda^1\left(\nabla\cdot\boldsymbol{u}^0\right) \tag{2.46}
$$

to first order in ϵ.

For the second term, we first calculate the elements of $\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T$, which are:

$$
\begin{pmatrix}
2\frac{\partial}{\partial x_1}\left(u_1^0 g\left(\phi_1\right)\right) & \frac{\partial}{\partial x_2}\left(u_1^0 g\left(\phi_1\right)\right) + \frac{\partial}{\partial x_1}\left(u_2^0 g\left(\phi_2\right)\right) & \frac{\partial}{\partial x_3}\left(u_1^0 g\left(\phi_1\right)\right) + \frac{\partial}{\partial x_1}\left(u_3^0 g\left(\phi_3\right)\right) \\
\frac{\partial}{\partial x_1}\left(u_2^0 g\left(\phi_2\right)\right) + \frac{\partial}{\partial x_2}\left(u_1^0 g\left(\phi_1\right)\right) & 2\frac{\partial}{\partial x_2}\left(u_2^0 g\left(\phi_2\right)\right) & \frac{\partial}{\partial x_3}\left(u_2^0 g\left(\phi_2\right)\right) + \frac{\partial}{\partial x_2}\left(u_3^0 g\left(\phi_3\right)\right) \\
\frac{\partial}{\partial x_1}\left(u_3^0 g\left(\phi_3\right)\right) + \frac{\partial}{\partial x_3}\left(u_1^0 g\left(\phi_1\right)\right) & \frac{\partial}{\partial x_2}\left(u_3^0 g\left(\phi_3\right)\right) + \frac{\partial}{\partial x_3}\left(u_2^0 g\left(\phi_2\right)\right) & 2\frac{\partial}{\partial x_3}\left(u_3^0 g\left(\phi_3\right)\right)
\end{pmatrix}
\tag{2.47}
$$

We may also write:

$$
\nabla\mu = \begin{pmatrix} \frac{\partial\mu^0}{\partial x_1} + \epsilon\frac{\partial\mu^1}{\partial x_1} \\ \frac{\partial\mu^0}{\partial x_2} + \epsilon\frac{\partial\mu^1}{\partial x_2} \\ \frac{\partial\mu^0}{\partial x_3} + \epsilon\frac{\partial\mu^1}{\partial x_3} \end{pmatrix} \tag{2.48}
$$

Taking the dot product of 2.48 and 2.47, and Taylor expanding $g(\phi_i)$ and its derivatives gives:

$$\begin{aligned}
[\mathbf{t_2}]_1 = 2&\left(\frac{\partial\mu^0}{\partial x_1}+\epsilon\frac{\partial\mu^1}{\partial x_1}\right)\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_1}\left(1+\frac{n-1}{n}\epsilon\phi_1^1\right)+\left(1+\epsilon\phi_1^1\right)\frac{\partial u_1^0}{\partial x_1}\right)\\
&+\left(\frac{\partial\mu^0}{\partial x_2}+\epsilon\frac{\partial\mu^1}{\partial x_2}\right)\left(u_2^0\frac{\partial\epsilon\phi_2^1}{\partial x_1}\left(1+\frac{n-1}{n}\epsilon\phi_2^1\right)+\left(1+\epsilon\phi_2^1\right)\frac{\partial u_2^0}{\partial x_1}\right)\\
&+\left(\frac{\partial\mu^0}{\partial x_2}+\epsilon\frac{\partial\mu^1}{\partial x_2}\right)\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_2}\left(1+\frac{n-1}{n}\epsilon\phi_1^1\right)+\left(1+\epsilon\phi_1^1\right)\frac{\partial u_1^0}{\partial x_2}\right)\\
&+\left(\frac{\partial\mu^0}{\partial x_3}+\epsilon\frac{\partial\mu^1}{\partial x_3}\right)\left(u_3^0\frac{\partial\epsilon\phi_3^1}{\partial x_1}\left(1+\frac{n-1}{n}\epsilon\phi_3^1\right)+\left(1+\epsilon\phi_3^1\right)\frac{\partial u_3^0}{\partial x_1}\right)\\
&+\left(\frac{\partial\mu^0}{\partial x_3}+\epsilon\frac{\partial\mu^1}{\partial x_3}\right)\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_3}\left(1+\frac{n-1}{n}\epsilon\phi_1^1\right)+\left(1+\epsilon\phi_1^1\right)\frac{\partial u_1^0}{\partial x_3}\right)
\end{aligned} \tag{2.49}$$

Expanding the brackets and keeping terms only to first order in ϵ gives:

$$\begin{aligned}
[\mathbf{t_2}]_1 = 2&\frac{\partial\mu^0}{\partial x_1}\frac{\partial u_1^0}{\partial x_1}+\frac{\partial\mu^0}{\partial x_2}\left(\frac{\partial u_2^0}{\partial x_1}+\frac{\partial u_1^0}{\partial x_2}\right)+\frac{\partial\mu^0}{\partial x_3}\left(\frac{\partial u_3^0}{\partial x_1}+\frac{\partial u_1^0}{\partial x_3}\right)\\
&+2\frac{\partial\mu^0}{\partial x_1}\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_1}+\epsilon\phi_1^1\frac{\partial u_1^0}{\partial x_1}\right)\\
&+\frac{\partial\mu^0}{\partial x_2}\left(u_2^0\frac{\partial\epsilon\phi_2^1}{\partial x_1}+\epsilon\phi_2^1\frac{\partial u_2^0}{\partial x_1}\right)+\frac{\partial\mu^0}{\partial x_2}\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_2}+\epsilon\phi_1^1\frac{\partial u_1^0}{\partial x_2}\right)\\
&+\frac{\partial\mu^0}{\partial x_3}\left(u_3^0\frac{\partial\epsilon\phi_3^1}{\partial x_1}+\epsilon\phi_3^1\frac{\partial u_3^0}{\partial x_1}\right)+\frac{\partial\mu^0}{\partial x_3}\left(u_1^0\frac{\partial\epsilon\phi_1^1}{\partial x_3}+\epsilon\phi_1^1\frac{\partial u_1^0}{\partial x_3}\right)\\
&+2\epsilon\frac{\partial\mu^1}{\partial x_1}\frac{\partial u_1^0}{\partial x_1}\\
&+\epsilon\frac{\partial\mu^1}{\partial x_2}\frac{\partial u_2^0}{\partial x_1}+\epsilon\frac{\partial\mu^1}{\partial x_2}\frac{\partial u_1^0}{\partial x_2}\\
&+\epsilon\frac{\partial\mu^1}{\partial x_3}\frac{\partial u_3^0}{\partial x_1}+\epsilon\frac{\partial\mu^1}{\partial x_3}\frac{\partial u_1^0}{\partial x_3}
\end{aligned} \tag{2.50}$$

This can be rewritten as:

$$
\begin{aligned}
[\mathbf{t_2}]_1 = & 2\frac{\partial \mu^0}{\partial x_1}\frac{\partial u_1^0}{\partial x_1} + \frac{\partial \mu^0}{\partial x_2}\left(\frac{\partial u_2^0}{\partial x_1} + \frac{\partial u_1^0}{\partial x_2}\right) + \frac{\partial \mu^0}{\partial x_3}\left(\frac{\partial u_3^0}{\partial x_1} + \frac{\partial u_1^0}{\partial x_3}\right) \\
& + 2\frac{\partial \mu^0}{\partial x_1}\frac{\partial}{\partial x_1}\left(u_1^0 \epsilon \phi_1^1\right) \\
& + \frac{\partial \mu^0}{\partial x_2}\left(\frac{\partial}{\partial x_1}\left(u_2^0 \epsilon \phi_2^1\right) + \frac{\partial}{\partial x_2}\left(u_1^0 \epsilon \phi_1^1\right)\right) \\
& + \frac{\partial \mu^0}{\partial x_3}\left(\frac{\partial}{\partial x_1}\left(u_3^0 \epsilon \phi_3^1\right) + \frac{\partial}{\partial x_3}\left(u_1^0 \epsilon \phi_1^1\right)\right) \\
& + 2\epsilon\frac{\partial \mu^1}{\partial x_1}\frac{\partial u_1^0}{\partial x_1} \\
& + \epsilon\frac{\partial \mu^1}{\partial x_2}\frac{\partial u_2^0}{\partial x_1} + \epsilon\frac{\partial \mu^1}{\partial x_2}\frac{\partial u_1^0}{\partial x_2} \\
& + \epsilon\frac{\partial \mu^1}{\partial x_3}\frac{\partial u_3^0}{\partial x_1} + \epsilon\frac{\partial \mu^1}{\partial x_3}\frac{\partial u_1^0}{\partial x_3}
\end{aligned}
\tag{2.51}
$$

In a similar way to $\mathbf{t_1}$, $\mathbf{t_2}$ may be written to first order as:

$$
\begin{aligned}
\nabla\mu \cdot [\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T] \approx & \nabla\mu^0 \cdot [\nabla \boldsymbol{u^0} + \left(\nabla \boldsymbol{u^0}\right)^T] + \epsilon \nabla \mu^1 [\nabla \boldsymbol{u}^0 + \left(\nabla \boldsymbol{u}^0\right)^T] \\
& + \nabla\mu^0 \cdot \left[\nabla \begin{pmatrix} u_1^0 \epsilon \phi_1^1 \\ u_2^0 \epsilon \phi_2^1 \\ u_3^0 \epsilon \phi_3^1 \end{pmatrix} + \left(\nabla \begin{pmatrix} u_1^0 \epsilon \phi_1^1 \\ u_2^0 \epsilon \phi_2^1 \\ u_3^0 \epsilon \phi_3^1 \end{pmatrix}\right)^T\right]
\end{aligned}
\tag{2.52}
$$

We now turn to $\mathbf{t_3}$, firstly defining the elements of $\nabla\nabla \cdot \boldsymbol{u}$:

$$
\nabla\nabla \cdot \boldsymbol{u} = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2}(u_1^0 g(\phi_1)) + \frac{\partial^2}{\partial x_1 \partial x_2}(u_2^0 g(\phi_2)) + \frac{\partial^2}{\partial x_1 \partial x_3}(u_3^0 g(\phi_3)) \\ \frac{\partial^2}{\partial x_2 \partial x_1}(u_1^0 g(\phi_1)) + \frac{\partial^2}{\partial x_2^2}(u_2^0 g(\phi_2)) + \frac{\partial^2}{\partial x_2 \partial x_3}(u_3^0 g(\phi_3)) \\ \frac{\partial^2}{\partial x_3 \partial x_1}(u_1^0 g(\phi_1)) + \frac{\partial^2}{\partial x_3 \partial x_2}(u_2^0 g(\phi_2)) + \frac{\partial^2}{\partial x_3^2}(u_3^0 g(\phi_3)) \end{pmatrix}
\tag{2.53}
$$

Expanding $[\nabla\nabla \cdot \boldsymbol{u}]_{11}$ to first order requires us to expand $\frac{\partial^2}{\partial x_1^2}(u_1^0 g(\phi_1))$, $\frac{\partial^2}{\partial x_1 \partial x_2}(u_2^0 g(\phi_2))$ and $\frac{\partial^2}{\partial x_1 \partial x_3}(u_3^0 g(\phi_3))$. This is done below:

$$\begin{aligned}\frac{\partial^2}{\partial x_1^2}\left(u_1^0 g\left(\phi_1\right)\right) &= \frac{\partial^2 u_1^0}{\partial x_1^2} g\left(\phi_1\right) + 2\frac{\partial u_1^0}{\partial x_1} g'\left(\phi_1\right)\frac{\partial \phi_1}{\partial x_1} + u_1^0\left(g'\left(\phi_1\right)\frac{\partial^2 \phi_1}{\partial x_1^2} + \left(\frac{\partial \phi_1}{\partial x_1}\right)^2 g''\left(\phi_1\right)\right) \\ &\approx \frac{\partial^2 u_1^0}{\partial x_1^2}\left(1+\epsilon\phi_1^1\right) + 2\frac{\partial u_1^0}{\partial x_1}\frac{\partial \epsilon\phi_1^1}{\partial x_1}\left(1+\frac{n-1}{n}\epsilon\phi_1^1\right) + u_1^0\frac{\partial^2 \epsilon\phi_1^1}{\partial x_1^2}\left(1+\frac{n-1}{n}\epsilon\phi_1^1\right) \\ &\quad + 2^{nd}order \\ &\approx \frac{\partial^2 u_1^0}{\partial x_1^2} + \epsilon\phi_1^1\frac{\partial^2 u_1^0}{\partial x_1^2} + 2\frac{\partial u_1^0}{\partial x_1}\frac{\partial \epsilon\phi_1^1}{\partial x_1} + u_1^0\frac{\partial^2 \epsilon\phi_1^1}{\partial x_1^2} + 2^{nd}order\end{aligned} \tag{2.54}$$

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2}\left(u_2^0 g\left(\phi_2\right)\right) &= \frac{\partial^2 u_2^0}{\partial x_1 \partial x_2} g\left(\phi_2\right) + \frac{\partial u_2^0}{\partial x_2} g'\left(\phi_2\right)\frac{\partial \phi_2}{\partial x_1} + \frac{\partial u_2^0}{\partial x_1} g'\left(\phi_2\right)\frac{\partial \phi_2}{\partial x_2} \\ &\quad + u_2^0\left(g'\left(\phi_2\right)\frac{\partial^2 \phi_2}{\partial x_1 \partial x_2} + \frac{\partial \phi_2}{\partial x_1}\frac{\partial \phi_2}{\partial x_2} g''\left(\phi_2\right)\right) \\ &\approx \frac{\partial^2 u_2^0}{\partial x_1 \partial x_2}\left(1+\epsilon\phi_2^1\right) \\ &\quad + \left(1+\frac{n-1}{n}\epsilon\phi_2^1\right)\left(\frac{\partial u_2^0}{\partial x_2}\frac{\partial \epsilon\phi_2^1}{\partial x_1} + \frac{\partial u_2^0}{\partial x_1}\frac{\partial \epsilon\phi_2^1}{\partial x_2} + u_2^0\frac{\partial^2 \epsilon\phi_2^1}{\partial x_1 \partial x_2}\right) + 2^{nd}order \\ &\approx \frac{\partial^2 u_2^0}{\partial x_1 \partial x_2} + \epsilon\phi_2^1\frac{\partial^2 u_2^0}{\partial x_1 \partial x_2} + \frac{\partial u_2^0}{\partial x_2}\frac{\partial \epsilon\phi_2^1}{\partial x_1} + \frac{\partial u_2^0}{\partial x_1}\frac{\partial \epsilon\phi_2^1}{\partial x_2} + u_2^0\frac{\partial^2 \epsilon\phi_2^1}{\partial x_1 \partial x_2}\end{aligned} \tag{2.55}$$

$$\frac{\partial^2}{\partial x_1 \partial x_3}\left(u_3^0 g\left(\phi_3\right)\right) \approx \frac{\partial^2 u_3^0}{\partial x_1 \partial x_3} + \epsilon\phi_3^1\frac{\partial^2 u_3^0}{\partial x_1 \partial x_3} + \frac{\partial u_3^0}{\partial x_3}\frac{\partial \epsilon\phi_3^1}{\partial x_1} + \frac{\partial u_3^0}{\partial x_1}\frac{\partial \epsilon\phi_3^1}{\partial x_3} + u_3^0\frac{\partial^2 \epsilon\phi_3^1}{\partial x_1 \partial x_3} \tag{2.56}$$

Therefore, the first element of $\mathbf{t_3}$ to first order is:

$$
\begin{aligned}
[\mathbf{t_3}]_1 \approx & \left(\lambda^0+\mu^0\right)\left(\frac{\partial^2 u_1^0}{\partial x_1^2}+\epsilon\phi_1^1\frac{\partial^2 u_1^0}{\partial x_1^2}+2\frac{\partial u_1^0}{\partial x_1}\frac{\partial \epsilon\phi_1^1}{\partial x_1}+u_1^0\frac{\partial^2 \epsilon\phi_1^1}{\partial x_1^2}\right) \\
& \left(\lambda^0+\mu^0\right)\left(+\frac{\partial^2 u_2^0}{\partial x_1\partial x_2}+\epsilon\phi_2^1\frac{\partial^2 u_2^0}{\partial x_1\partial x_2}+\frac{\partial u_2^0}{\partial x_2}\frac{\partial \epsilon\phi_2^1}{\partial x_1}+\frac{\partial u_2^0}{\partial x_1}\frac{\partial \epsilon\phi_2^1}{\partial x_2}\ +u_2^0\frac{\partial^2 \epsilon\phi_2^1}{\partial x_1\partial x_2}\right) \\
& +\left(\lambda^0+\mu^0\right)\left(\frac{\partial^2 u_3^0}{\partial x_1\partial x_3}+\epsilon\phi_3^1\frac{\partial^2 u_3^0}{\partial x_1\partial x_3}+\frac{\partial u_3^0}{\partial x_3}\frac{\partial \epsilon\phi_3^1}{\partial x_1}+\frac{\partial u_3^0}{\partial x_1}\frac{\partial \epsilon\phi_3^1}{\partial x_3}\ +u_3^0\frac{\partial^2 \epsilon\phi_3^1}{\partial x_1\partial x_3}\right) \\
& +\left(\epsilon\lambda^1+\epsilon\mu^1\right)\left(\frac{\partial^2 u_1^0}{\partial x_1^2}+\frac{\partial^2 u_2^0}{\partial x_1\partial x_2}+\frac{\partial^2 u_3^0}{\partial x_1\partial x_3}\right) \\
= & \left(\lambda^0+\mu^0\right)\left(\frac{\partial^2 u_1^0}{\partial x_1^2}+\frac{\partial^2 u_2^0}{\partial x_1\partial x_2}+\frac{\partial^2 u_3^0}{\partial x_1\partial x_3}\right) \\
& +\left(\lambda^0+\mu^0\right)\left(\frac{\partial^2}{\partial x_1^2}\left(u_1^0\epsilon\phi_1^1\right)+\frac{\partial^2}{\partial x_1\partial x_2}\left(u_2^0\epsilon\phi_2^1\right)+\frac{\partial^2}{\partial x_1\partial x_3}\left(u_3^0\epsilon\phi_3^1\right)\right) \\
& +\left(\epsilon\lambda^1+\epsilon\mu^1\right)\left(\frac{\partial^2 u_1^0}{\partial x_1^2}+\frac{\partial^2 u_2^0}{\partial x_1\partial x_2}+\frac{\partial^2 u_3^0}{\partial x_1\partial x_3}\right)
\end{aligned}
\tag{2.57}
$$

In vector notation, $\mathbf{t_3}$ can be written to first order as:

$$
(\lambda+\mu)\nabla\nabla\cdot\boldsymbol{u}\approx(\lambda^0+\mu^0)\nabla\nabla\cdot\boldsymbol{u^0}+(\lambda^0+\mu^0)\nabla\nabla\cdot\begin{pmatrix}u_1^0\epsilon\phi_1^1\\u_2^0\epsilon\phi_2^1\\u_3^0\epsilon\phi_3^1\end{pmatrix}+(\epsilon\lambda^1+\epsilon\mu^1)\nabla\cdot\boldsymbol{u^0} \tag{2.58}
$$

Taking now $\mathbf{t_4}$ we may write $\nabla^2\boldsymbol{u}$ as:

$$
\nabla^2\boldsymbol{u}=\begin{pmatrix}\frac{\partial^2}{\partial x_1^2}(u_1^0g(\phi_1))+\frac{\partial^2}{\partial x_2^2}(u_1^0g(\phi_1))+\frac{\partial^2}{\partial x_1^2}(u_1^0g(\phi_1))\\\frac{\partial^2}{\partial x_1^2}(u_2^0g(\phi_2))+\frac{\partial^2}{\partial x_2^2}(u_2^0g(\phi_2))+\frac{\partial^2}{\partial x_3^2}(u_2^0g(\phi_2))\\\frac{\partial^2}{\partial x_1^2}(u_3^0g(\phi_3))+\frac{\partial^2}{\partial x_2^2}(u_3^0g(\phi_3))+\frac{\partial^2}{\partial x_3^2}(u_3^0g(\phi_3)\end{pmatrix} \tag{2.59}
$$

We expanded the second derivative in Equation 2.54, and so for the first element of $\mathbf{t_4}$ we may write:

$$[\mathbf{t_4}]_1 \approx \left(\mu^0 + \epsilon\mu^1\right)\left(\frac{\partial^2 u_1^0}{\partial x_1^2} + \frac{\partial}{\partial x_1^2}\left(u_1^0\epsilon\phi_1^1\right)\right) \tag{2.60}$$

$$\approx \mu^0\frac{\partial^2 u_1^0}{\partial x_1^2} + \mu^0\frac{\partial}{\partial x_1^2}\left(u_1^0\epsilon\phi_1^1\right) + \epsilon\mu^1\frac{\partial^2 u_1^0}{\partial x_1^2} \tag{2.61}$$

$\mathbf{t_4}$ to first order is therefore:

$$\mu\nabla^2\boldsymbol{u} \approx \mu^0\nabla^2\boldsymbol{u^0} + \mu^0\nabla^2\begin{pmatrix} u_1^0\epsilon\phi_1^1 \\ u_2^0\epsilon\phi_2^1 \\ u_3^0\epsilon\phi_3^1 \end{pmatrix} + \epsilon\mu^1\nabla^2\mathbf{u^0} \tag{2.62}$$

The term $\mathbf{t_5}$ is the most straightforward, and to first order may be written as:

$$\mathbf{t_5} \approx \omega^2\left(\rho^0\mathbf{u^0} + \epsilon\rho^1\mathbf{u^0} + \rho^0\begin{pmatrix} u_1^0\epsilon\phi_1^1 \\ u_2^0\epsilon\phi_2^1 \\ u_3^0\epsilon\phi_3^1 \end{pmatrix}\right) \tag{2.63}$$

Collecting the first order terms (and again using that $\mathscr{L}^0\mathbf{u^0} = \mathbf{f}$), the order parameter ϵ cancels out, and we are left with:

$$\mathscr{L}^0\begin{pmatrix} u_1^0\phi_1^1 \\ u_2^0\phi_2^1 \\ u_3^0\phi_3^1 \end{pmatrix} = -\delta\mathscr{L}\mathbf{u^0}. \tag{2.64}$$

Relabelling μ^1, λ^1, ρ^1 to $\delta\mu, \delta\lambda, \delta\rho$ makes this exactly the same equation that was solved above in order to find $\boldsymbol{\delta}\mathbf{u}$ under the Born approximation. This means that the solution for ϕ to first order is:

$$\phi_i = \delta u_i/u_i^0. \tag{2.65}$$

This expression is independent of n, and therefore ϕ is the same for all values of n. This is analogous to the results for the Helmholtz equation presented in *Marks* (2006). Here, the results have been extended to a vector wave with an independent ϕ_i for each

of the components of the displacement wavefield. Once ϕ is known, it can be used to calculate a seismogram under the Born, Rytov or Marks approximations, something which has not previously been done:

$$u_i^{1^{st}} = (1 + \phi_i/n)^n \,. \tag{2.66}$$

2.4.3 Discussion

In the above derivation ϕ_i is independent of n, and we may calculate ϕ_i using $\boldsymbol{\delta}\mathbf{u}_i$ as determined under the Born approximate. Once ϕ_i is found, the difference between the three first order approximations is in the definition of $g(\phi_i)$, or the value of n. We see that when ϕ_i is small, the approximations converge, and it should be noted that ϵ (not ϕ_i) is small in the previous section.

ϕ_i is a complex phase, which perturbs the seismogram in different ways depending on the value of n. Perturbations in any of the elastic parameters (v_s, v_p, ρ) contribute to the value of ϕ_i additively via the waveform kernels. Each component of the seismogram has a different value of ϕ_i, which is independent of the other components as terms mixed in ϕ_i and ϕ_j are second order and above. Contained in ϕ_i is therefore information on how any perturbation to the background model affects a particular component of the seismogram instantaneously at a time t. It is indiscriminate of which phase the time corresponds to and therefore is equally applicable to direct, reflected or converted phases. It may also be applied to models with any combination of perturbations in both v_p and v_s.

Crucial to the determination of ϕ_i are the waveform kernels derived under the Born approximation. The method for kernel calculations is described in the next chapter. In the subsequent chapter, we apply the method derived above to compute first order seismograms under the Born, Marks and Rytov approximations.

Chapter 3

Sensitivity kernel calculation

3.1 Introduction

In this chapter, we detail the method used to compute sensitivity kernels in this thesis, called MC (Monte Carlo) Kernel. The publication of the method is currently in preparation:

M.C. Kernel: Global broadband seismic sensitivity kernels on heterogeneous meshes calculated from a wavefield database, *S. C. Stähler, M. van Driel, H. Godwin, E. Totten, K. Hosseini, L. Auer, K. Sigloch and T. Nissen Meyer*

MC Kernel has primarily been developed as a tool to compute sensitivity kernels for tomographic inversion, and therefore it is geared towards the calculation of cross-correlation travel time kernels or similar. These are computed by integrating the time dependent waveform kernels over the selected measurement window. For the subsequent applications of MC Kernel in the following chapters, the waveform kernel output is used.

When work for this thesis began, the MC Kernel method had been implemented but not thoroughly tested. Testing and significant debugging of the software was carried out as part of this work.

3.1.1 Background: Sensitivity kernels for seismic tomography

In finite frequency tomography and full waveform inversion, measurements of e.g. cross-correlation travel time differences are related back to model parameters using sensitivity kernels calculated under the Born approximation, as we derived in the previous chapter.

Travel time tomography has traditionally been ray-based, with a travel time delay accumulated along the theoretical ray path. However, taking into account the finite frequency of seismic waves means that cross-correlation travel time measurements become sensitive to velocity perturbations located away from the ray path. The shape of finite frequency kernels may be described as a "banana-doughnut" (*Marquering et al.*, 1999). The "banana" part refers to the kernel's shape parallel to the geometrical ray. The width of the kernel is dependent on frequency; the higher the frequency, the narrower the kernel. The "doughnut" part refers to the fact that the kernel is hollow; counter-intuitively there is no sensitivity along the geometrical ray-path. This may be understood by considering that waves which are scattered by material along the geometrical ray path arrive at the same time as the direct wave. Therefore, they can only affect the amplitude of the waveform, but not the distribution of energy in time.

Sensitivity kernels have been computed using semi analytical techniques (*Tian et al.*, 2007) and mode coupling (*Li and Romanowicz*, 1996) as well as fully numerical methods (*Nissen-meyer et al.*, 2007; *Liu and Tromp*, 2008). MC Kernel is one such numerical approach, using forward and backward wavefields computed in the background medium using the axisymmetric spectral element method, AxiSEM (*Nissen-Meyer et al.*, 2014). Kernels are calculated on irregular, adapted grids, akin to those that have been employed in recent tomographic inversions where the element size may be dictated by data coverage e.g. (*Zaroli et al.*, 2015). The kernel value on a particular basis function is calculated using Monte Carlo integration.

3.2 Methodology

3.2.1 Wavefield Computation

As described in the previous chapter, the building blocks of sensitivity kernels are the forward and backward wavefields, $\overrightarrow{\mathbf{u}}$ and $\overleftarrow{\mathbf{G}}$, computed in a 1-D background medium. The computation of these wavefields is facilitated by the method AxiSEM.

AxiSEM is a spectral element method for solving the wave equation that takes advantage of a short-cut which may be used in cases where the Earth model is axisymmetric (*Nissen-Meyer et al.*, 2007, 2014). This means that wavefields can be computed as if the source were on the North Pole. The wave equation is decomposed into a series of 2-D equations, where the azimuthal dependence of the wavefield can be found analytically. Therefore, the numerical solution may be computed on a 2-D rather than a 3-D mesh, leading to a signficant speed-up in computing time. A source may be represented as a sum of monopole, dipole and quadrupole sources. This classification is defined by the dependence on ϕ. A monopole source is independent of ϕ while dipole sources are dependent on $cos\phi$ or $sin\phi$ and quadrupole sources are dependent on $cos(2\phi)$ or $sin(2\phi)$. The forward and backward wavefields are assembled from linear combinations of these three different types of sources.

The response to a monopole, dipole or quadrupole source may be written in cylindrical polar coordinates, (s, z, ϕ) as

$$\begin{aligned}\mathbf{u}^m &= \hat{\mathbf{s}}u_s(s,z,\phi) + \hat{\mathbf{z}}u_z(s,z,\phi) + \hat{\boldsymbol{\phi}}u_\phi(s,z,\phi) \\ &= (\hat{\mathbf{s}}\bar{u}_s(s,z) + \hat{\mathbf{z}}\bar{u}_z(s,z))A_1^m(\phi) + \hat{\boldsymbol{\phi}}\bar{u}_\phi(s,z)A_2^m(\phi),\end{aligned} \tag{3.1}$$

where $A_1^m(\phi)$ and $A_2^m(\phi)$ are azimuthal factors. $m = 0, 1, 2$ is the order of the multi-pole expansion and corresponds to a monopole, dipole or quadrupole source respectively. The wavefields $\bar{u}_s, \bar{u}_z, \bar{u}_\phi$ are found numerically by solving the wave equation on a 2-D mesh. A multi-pole expansion of the source allows the azimuthal dependence to be tackled analytically via the azimuthal factors $A_1^m(\phi)$ and $A_2^m(\phi)$.

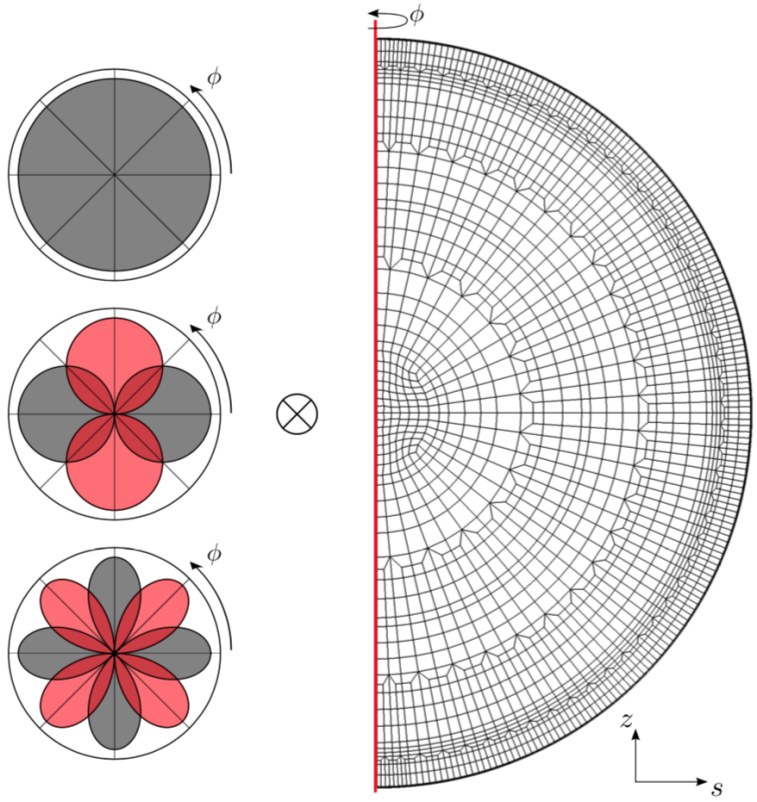


Figure 3.1: From (*Van Driel*, 2015). On the left are radiation patterns for monopole (top), dipole (middle) and quadrupole (bottom) sources, viewed down the z axis and showing their dependence on ϕ. On the right is an example of a 2-D mesh on which s and z dependence of the wavefield is computed numerically.

Forward wavefield

Four separate computations in the 2-D domain are required to find the response to each of the six independent moment tensor elements $(M_{xx}, M_{yy}, M_{zz}, M_{xy}, M_{xz}, M_{yz})$ before being combined with the azimuthal factors. These are:

1. M_{zz}: This is a monopole source $(m = 0)$ and therefore the corresponding wavefields are indepedent of azimuth.

$$A_1^0(\phi) = 1, \qquad A_2^0(\phi) = 0. \tag{3.2}$$

2. $M_{xx} + M_{yy}$: Also a monopole source with $A_1^0(\phi) = 1, A_2^0(\phi) = 0$.

3. M_{xz}: This is a dipole source $(m = 1)$ where the displacements are functions of

ϕ. It is equivalent to M_{yz} with rotation by $\pi/2$ in azimuth.

$$A_1^1(\phi) = \begin{cases} cos(\phi), & \text{for } M_{xz} \\ sin(\phi) & \text{for} M_{yz} \end{cases}, \qquad A_2^1(\phi) = \begin{cases} -sin(\phi), & \text{for } M_{xz} \\ cos(\phi) & \text{for } M_{yz} \end{cases} \tag{3.3}$$

4. $M_{xx} - M_{yy}$: This is a quadrupole source ($m = 2$) where the displacements are functions of 2ϕ, equivalent to M_{xy} with a $\pi/4$ rotation in azimuth.

$$A_1^2(\phi) = \begin{cases} cos(2\phi), & \text{for } M_{xx} - M_{yy} \\ sin(2\phi) & \text{for } M_{xy} \end{cases}, \qquad A_2^2(\phi) = \begin{cases} -sin(2\phi), & \text{for } M_{xx} - M_{yy} \\ cos(2\phi) & \text{for } M_{xy} \end{cases} \tag{3.4}$$

Since the solution for the full moment tensor source is a linear combination of each of the moment tensor elements, we may use the same forward AxiSEM run for any moment tensor source at a single depth.

Backward wavefield

The backward wavefield is computed with a point force source and it may be calculated with two AxiSEM simulations in the 2-D domain:

1. P_z: This is a monopole source with $m = 0$ and therefore the same azimuthal factors as in Equation 3.2.

2. P_x: This is a dipole source with $m = 1$, which is equivalent to P_y with a $\pi/2$ rotation in azimuth.

$$A_1^1(\phi) = \begin{cases} cos(\phi), & \text{for } P_x \\ sin(\phi) & \text{for} P_y \end{cases}, \qquad A_2^1(\phi) = \begin{cases} -sin(\phi), & \text{for } P_x \\ cos(\phi) & \text{for} P_y \end{cases} \tag{3.5}$$

Since receivers are located on Earth's surface, only one backward wavefield is required for any source-receiver geometry.

3.2.2 Wavefield rotation

Wavefields are computed using AxiSEM with the source at the North Pole, and defined in source-centered cylindrical coordinates (s_s, ϕ_s, z_s). Before the wavefields can be convolved, they must be rotated so that the source of the forward wavefield is at the earthquake location and the source of the backward wavefield is at the receiver location. Both wavefields are also transformed into global cartesian coordinates before they are convolved. The same steps are carried out for the forward and backward wavefields, therefore what is referred to below as the source could either be the earthquake event or the receiver. Below are the steps required to find the wavefield in global cartesian coordinates at a particular scatterer location, as derived in *Nissen-Meyer et al.* (2007).

Converting from source-centered cylindrical coordinates takes place via the source-centered cartesian coordinate system. This is a cartesian coordinate system with axes pointing in the radial, azimuthal and polar directions when transported to the surface. The relationship between the source-centered and global cartesian coordinate systems are shown in Figure 3.2.

First we find the wavefield at the scatterer location in source-centered cylindrical coordinates. The first step is to convert to source-centered cartesian coordinates, (x_s, y_s, z_s). The z_s component is unchanged in this transformation, so we convert ϕ and s to x and y. This requires the azimuthal angle between the scatterer and the source, ϕ_{sd} to be determined (Figure 3.3). The coordinate transform is performed using the matrix

$$\boldsymbol{T} = \begin{pmatrix} cos\phi_{sd} & sin\phi_{sd} & 0 \\ -sin\phi_{sd} & cos\phi_{sd} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.6}$$

The second step is to convert from source-centered cartesian coordinates into global cartesian coordinates. This is a rotation of the wavefield from the North Pole to the source location, with the rotation matrix. This is a function of θ_s and ϕ_s which are the colatitude and longitude of the source in the global frame.

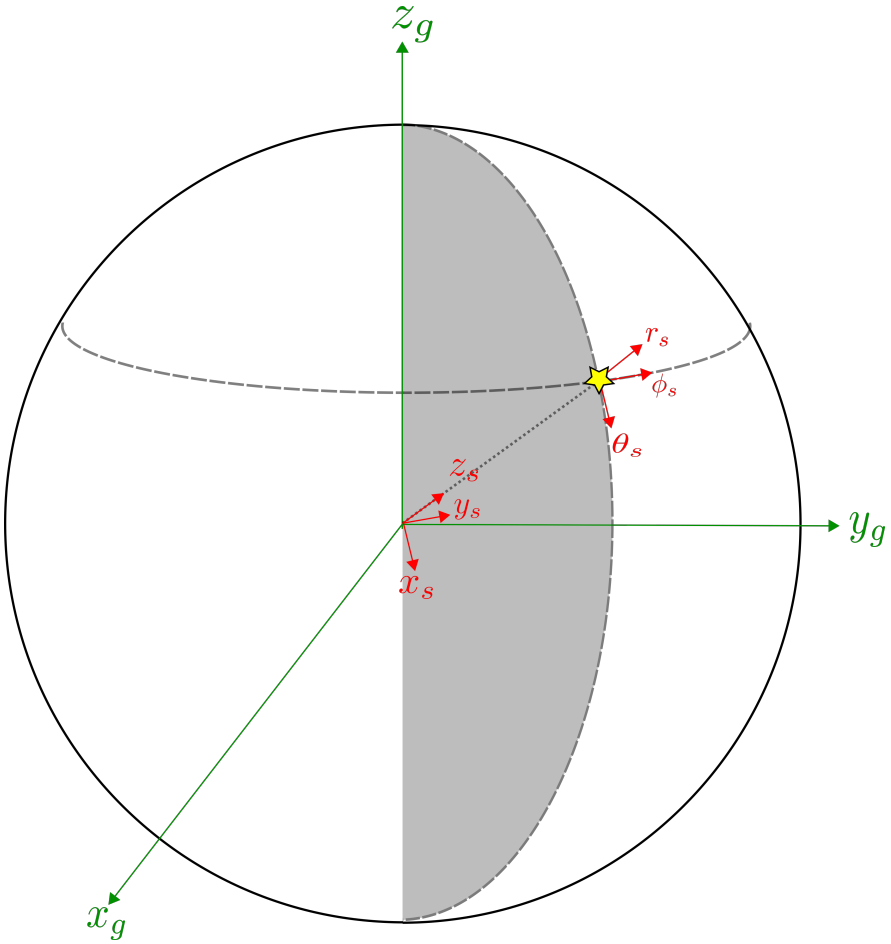


Figure 3.2: Source-centered and global cartesian coordinate systems, (x_g, y_g, z_g) and x_s, y_s, z_s respectively. When transported to the surface, the source-centered cartesian axes point in the (r, θ, ϕ) directions of spherical polar coordinates. The source location in global coordinates is shown with a yellow star.

$$\boldsymbol{R}(\theta_{gs}, \phi_{gs}) = \begin{pmatrix} cos\theta_{gs}cos\phi_{gs} & cos\theta_{gs}sin\phi_{gs} & -sin\theta_{gs} \\ -sin\phi_{gs} & cos\phi_{gs} & 0 \\ sin\theta_{gs}cos\phi_{gs} & sin\theta_{gs}sin\phi_{gs} & cos\theta_{gs}. \end{pmatrix} \tag{3.7}$$

Wavefield rotation is only necessary for v_s kernels, which require the full strain tensor (see equation something in theory). The computation of kernels in v_p uses the trace of the strain tensor, which is invariant under rotation.The transformed strain tensor, $\boldsymbol{\epsilon}$, is found from

$$\boldsymbol{\epsilon}(x_g, y_g, z_g) = \mathbf{R}(\mathbf{T}\mathbf{A}\mathbf{T}^T)\mathbf{R}^T. \tag{3.8}$$

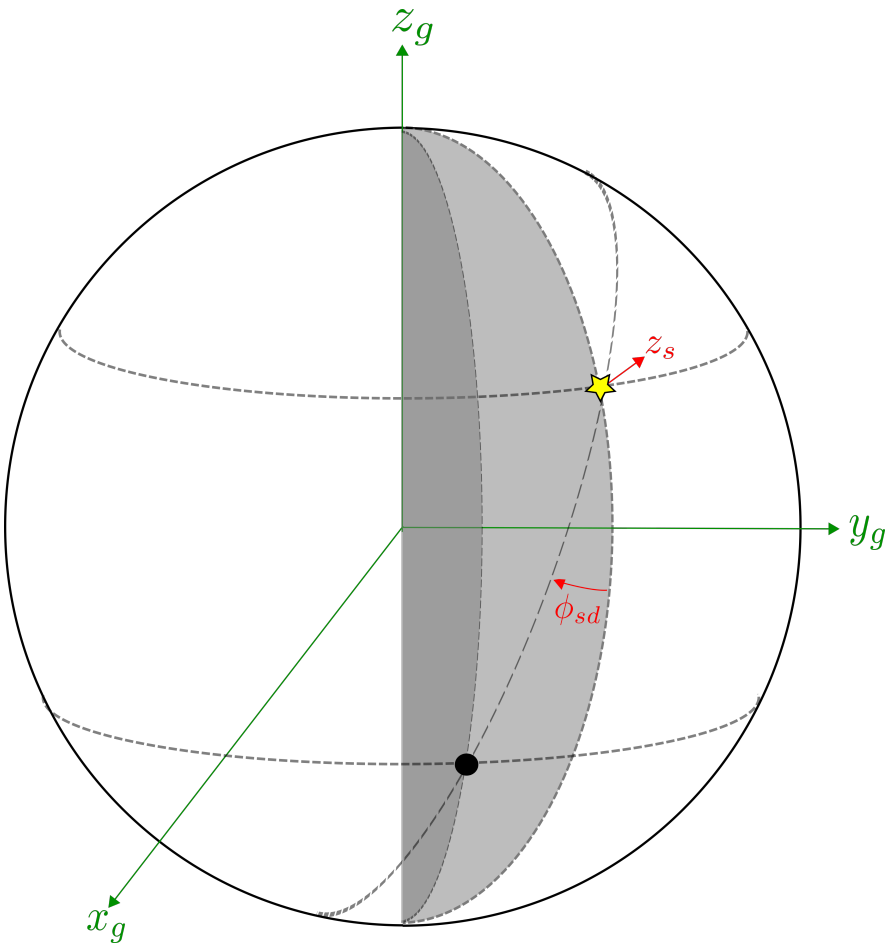
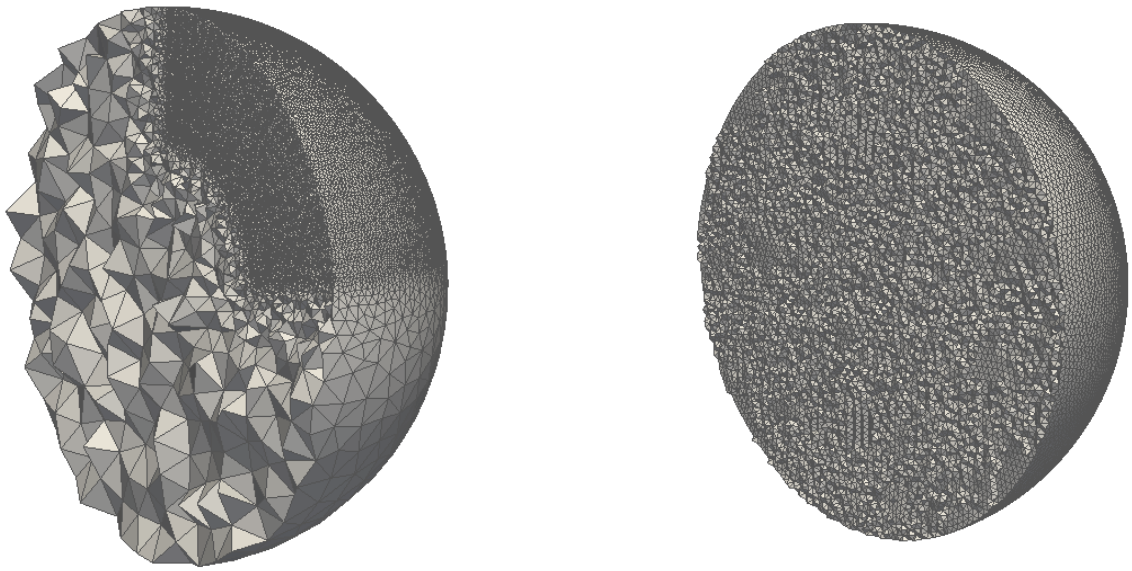


Figure 3.3: Definition of the angle ϕ_{sd}, which is the angle between the source (yellow star) and a scatterer(black dot) in source-centered cylindrical coordinates. Rotating by the angle ϕ_{sd} about the z_s axis is equivalent to converting to source-centered cartesian coordinates.

3.2.3 Projection onto 3-D meshes

AxiSEM wavefields are computed on a 2-D mesh but kernels must be defined on a 3-D mesh. This is facilitated by using Monte Carlo integration to project the kernels onto such a 3-D mesh. Monte Carlo integration is based on evaluating the kernel value at randomly chosen points across the domain. Additional points are added until the error falls below a certain threshold. This approach means that a slightly different result is obtained after each computation, but it is chosen for its quantifiable error.



(a) **(b)**

Figure 3.4: Two tetrahedral meshes on which kernels are computed in this thesis. (a) is computed with a target edge length of 30 km in one quarter of the Earth and 800 km elsewhere. (b) is computed with a uniform target edge length of 100 km throughout Earth's interior.

Mesh generation

The MC Kernel code is capable of handling different mesh types, including 3-D meshes composed of tetrahedrons or curvilinear voxels.

In this thesis, we use 3-D tetrahedral meshes generated using the software Distmesh (*Persson and Strang*, 2004). Distmesh allows meshes with variable edge-lengths to be created. The user defines a target edge-length at various locations within the Earth. Two meshes are used in this work. One is generated with a uniform target edge-length of 100 km. The other is generated with a target edge length of 30 km in one quarter of the Earth down to the core-mantle boundary to provide higher resolution in the region of interest. Both meshes are displayed in Figure 3.4.

Monte Carlo projection method

Within the code MC Kernel, measurement kernels employed in seismic tomography (that are integrated over time) are projected onto the 3-D mesh. Prior to projection, the measurement kernels are defined on the 2D AxiSEM mesh, with the azimuthal dependence calculated analytically. Projection onto the 3D mesh includes integration over volume, so that the volume of the mesh element does not need to be taken into account at later stages during tomographic inversion. This integration over volume is carried out using Monte Carlo (MC) integration. The algorithm for carrying out the MC integration of the measurement kernel for a mesh element b_l occupying a volume V_l is as follows:

1. Find N random points, $\mathbf{x_i}$ where $i = 1, 2, ..., N$, within the volume V_l
2. Find the kernel value $K(\mathbf{x_i})$ on each of the N points
3. Update the integral estimator in this element, Q_l:

$$Q_l = V_l \langle K(\mathbf{x_i}) \rangle \tag{3.9}$$

4. Update the error estimator in this element, δQ_l

$$\delta Q_l = V_l \frac{\sqrt{\langle K(\mathbf{x_i}) \rangle^2 - \langle K(\mathbf{x_i})^2 \rangle}}{\sqrt{N}} \tag{3.10}$$

5. Repeat steps 1 to 4 until the absolute error, Q_l or the relative error, $\delta Q_l / Q_l$, becomes lower than a threshold value.

The parameters used in the Monte Carlo integration are defined by the user, who chooses the threshold error, relative error, number of points per step and the maximum number of iterations. The choice of these parameters can be a trade-off between accuracy and time. Most cells converge quickly (in fewer than 10 iterations), but near the source and receiver thousands of iterations may be reached. The kernels in these

regions are more complex and have very large magnitudes. For this reason, kernels are often damped in the vicinity of the source and receiver (*Zhu et al.*, 2015; *Tao et al.*, 2018).

If an unsuitable mesh is chosen for the kernel frequency, convergence may not be reached as the values of the kernel change too much over the size of each measurement.

Since the MC Kernel code is geared towards applications in seismic tomography, it currently checks the convergence of measurement kernels rather than waveform kernels. This is discussed further in the following chapter, where we 'manually' check the convergence of the waveform kernels. The choice of time window over which the measurement is made can affect the convergence since waveform kernels are far more coherent around phase arrivals.

3.3 Implementation

The MC Kernel code is written in Fortran90, and is released under the General Purpose License v3 (`https://github.com/seismology/mc_kernel`). It is organised into a master-worker architecture, in which the inversion mesh is read by the master instance. MC Kernel is embarassingly parallel; the master sends groups of elements to the worker ranks. Each worker calculates the kernel projected onto the mesh in its allocated mesh elements and sends the results back to the master. The workflow of the program is visualised in Figure 3.5.

The generation of a seismic tomography model will typically involve the inversion of data from several thousand earthquakes. Each earthquake is in turn recorded by hundreds of receivers, and measurements are taken from the resulting seismograms. These measurements are made on a particular component (R,T,Z) in a particular time window and in a particular frequency band arrived at through filtering. A corresponding kernel is required for each of the many measurements and so the MC Kernel code is ordered for efficiency. The order of loops is as follows:

1. **Earthquake** One run of MC Kernel is required for each earthquake, since little

computation can be shared between two sources.

2. **Mesh elements** The main loop of the code is over mesh elements. Since the kernel value in one basis element is computed independently of other elements, the code is very easily parallelised at this level. The forward wavefield need only be read in once at each point and is used for all receivers.

3. **Receivers** Since the location of each receiver is different, the backward wavefield must be read in once for each receiever.

4. **Base kernels** Time dependent kernels are calcualted in the fundamental elastic parameters (μ, λ, ρ)

5. **Measurement kernels** These are assembled from the base kernels after filtering, applying the earthquake source-time function (STF) and integrating over time. The convergence is checked for measurement kernels.

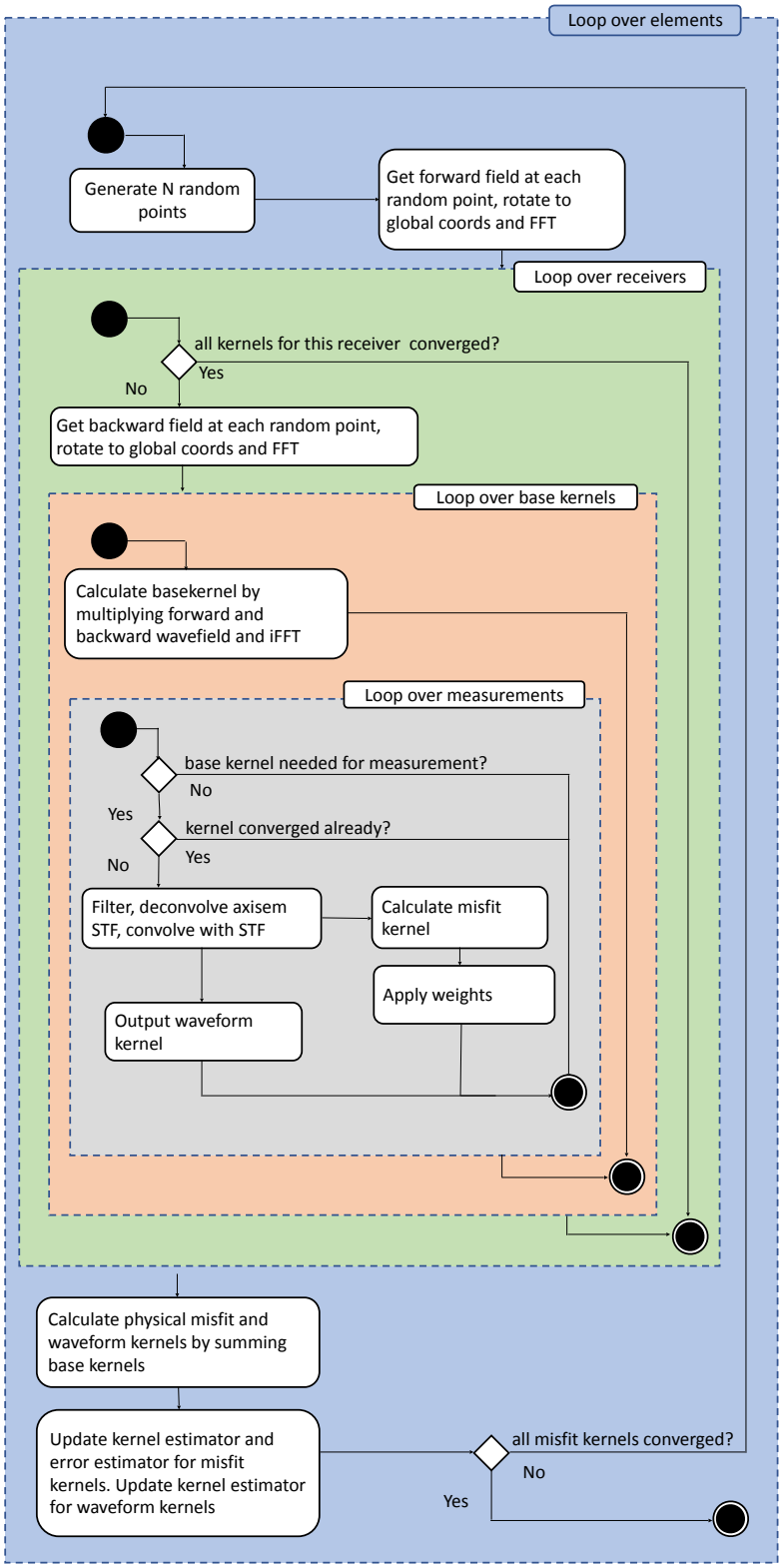


Figure 3.5: Flowchart showing the workflow of MC Kernel, adapted from Staehler et al.

3.4 Kernel examples

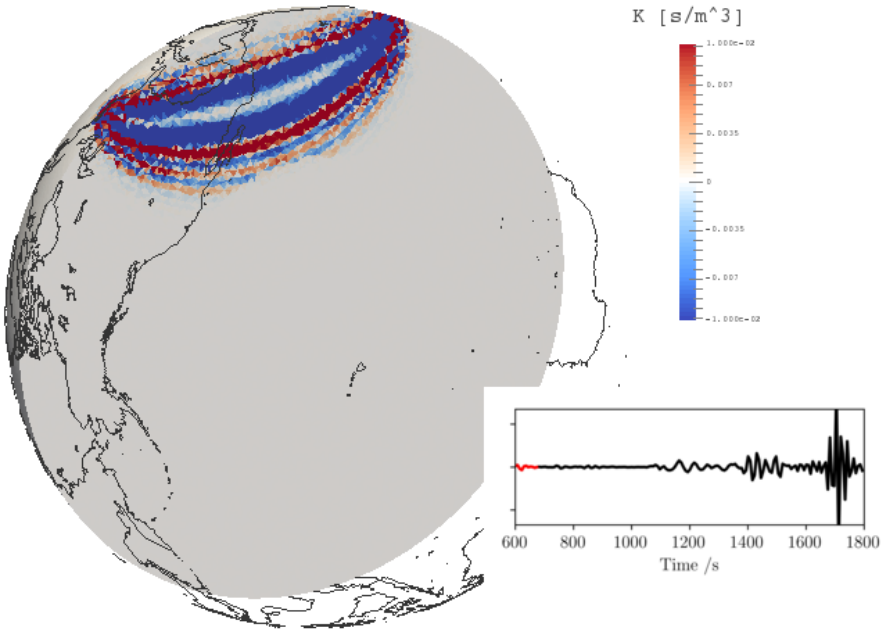
3.4.1 Travel time kernels

First, we present examples of travel time kernels computed using MC Kernel. We use a 60 s dominant period, a moment tensor source and a 60 s Gabor filter. v_p kernels are calculated for the Z component of the seismogram, while v_s kernels are calculated for the T component. They are projected on a tetrahedral mesh with a target edge length of 100 km throughout the Earth. Figure 3.6 shows kernels for perturbations in v_p. Figure 3.6a is a 60° kernel in a time window corresponding to the P arrival. Figure 3.6b is a 90° kernel in a time window where P and PcP morph into a single phase.

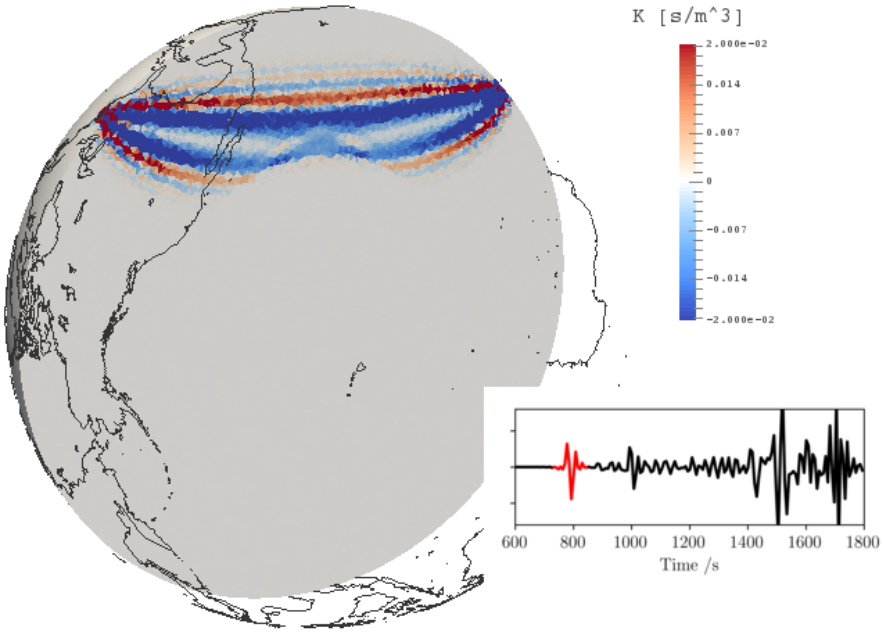
The 60° kernel for the direct P phase shows the typical 'banana-doughnut' shape of finite frequency sensitivity kernels for cross-correlation measurements. There is no sensitivity in the middle of the 'banana' kernel, corresponding to the geometrical ray path, showing how material along the ray path only affects the amplitude of the phase, but not its arrival time. The first Fresnel zone has negative sensitivity to travel time, and the subsequent Fresnel zones have alternating positive and negative frequency.

The 90° kernel corresponds to the P/PcP arrival, and here we see added complexity in the kernel around the core-mantle boundary where the reflection takes place.

Similar figures for kernels in v_s are shown in Figures 3.7. Again, we plot kernels at 60° and 90°, corresponding to the direct S phase and the S/ScS phase respectively. These kernels have similar properties to the v_p kernels described above.

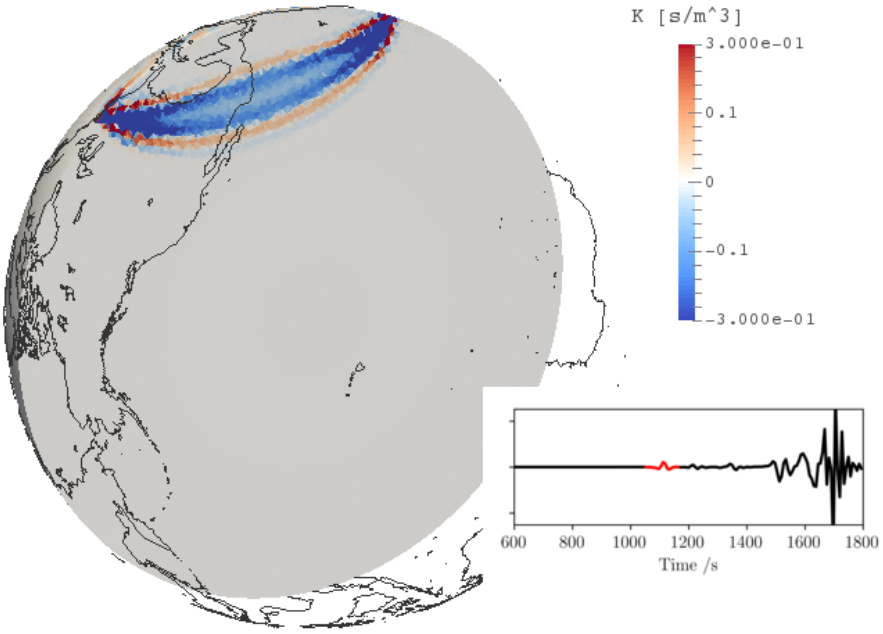


(a)

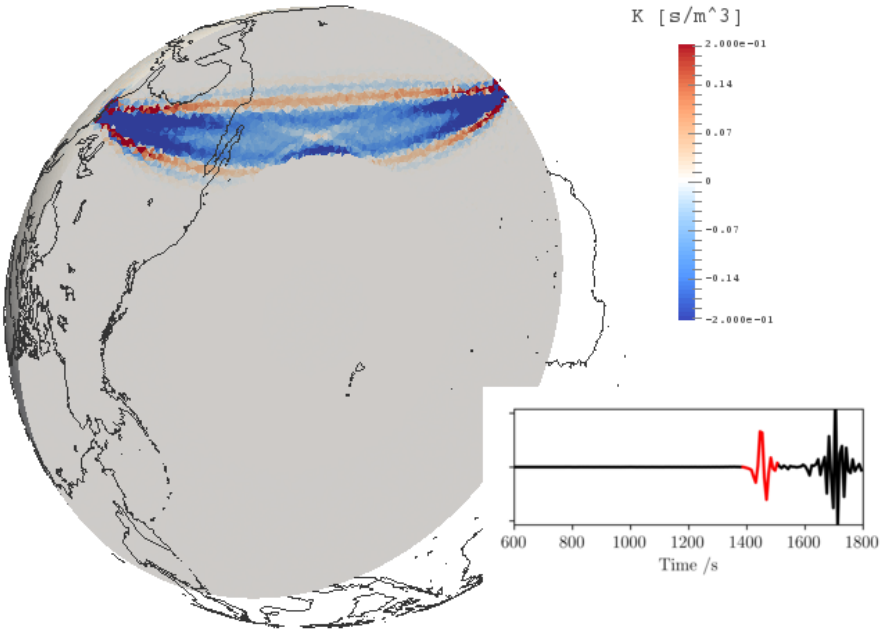


(b)

Figure 3.6: Travel time kernels showing sensitivity to v_p. The kernels are calculated in the background model IASP91, at a dominant period of 60 s and filtered with a 60 s Gabor filter. In the receivers are 60° and 90° away from the source in a) and b) respectively. The inset shows the recorded Z component seismogram at the receiver, with the red section corresponding to the cross-correlation time window. The kernels are projected onto a tetrahedral mesh with a target edge length of 100 km.



(a)



(b)

Figure 3.7: Travel time kernels showing sensitivity to v_s. The kernels are calculated in the background model IASP91, at a dominant period of 60 s and filtered with a 60 s Gabor filter. In the receivers are 60° and 90° away from the source in a) and b) respectively. The inset shows the recorded T component seismogram at the receiver, with the red section corresponding to the cross-correlation time window. The kernels are projected onto a tetrahedral mesh with a target edge length of 100 km.

3.4.2 Waveform kernels

In this section, we present the results of computing time dependent waveform kernels in v_s and v_p using MC Kernel. These are the kernels which can then be integrated over time and normalised to compute the more familiar travel time kernels such as those displayed in the previous section. The kernels in this section are computed with a 20 s dominant period, and low pass filtered at 20 s. Different sources and components (Z and T) are used for P and S computations, but both are projected onto a tetrahedral mesh with a target edge length of 30 km in the kernel region. Figure 3.8 shows a waveform kernel in v_p at several points in time around the arrival of the direct P and the core reflected PcP phases. Figure 3.9 shows a waveform in v_s at several points in time around the arrival of the direct S and the core-reflected ScS phases.

We note that around the arrivals of the direct P and S waves, the waveform kernels have the characteristic banana shape, with very little sensitivity away from this region surrounding the ray path. However, we do not observe the doughnut shaped 'hole' in the way that we do for travel time kernels. This is because the waveform kernel defines how the value of the seismogram (the displacement) changes, rather than a travel time delay measured by cross-correlation.

The shape of the waveform kernel is not so straightforward at times apart from the arrivals of distinct phases. The sensitivity covers a larger range, and its values are far smaller.

In the next chapter, we use waveform kernels calculated using MC Kernel to compute perturbed seismograms from including spherical heterogeneities in the lower mantle. For small perturbations, these seismograms match those calculated using a fully numerical method, providing quantitative validation of the MC Kernel method.

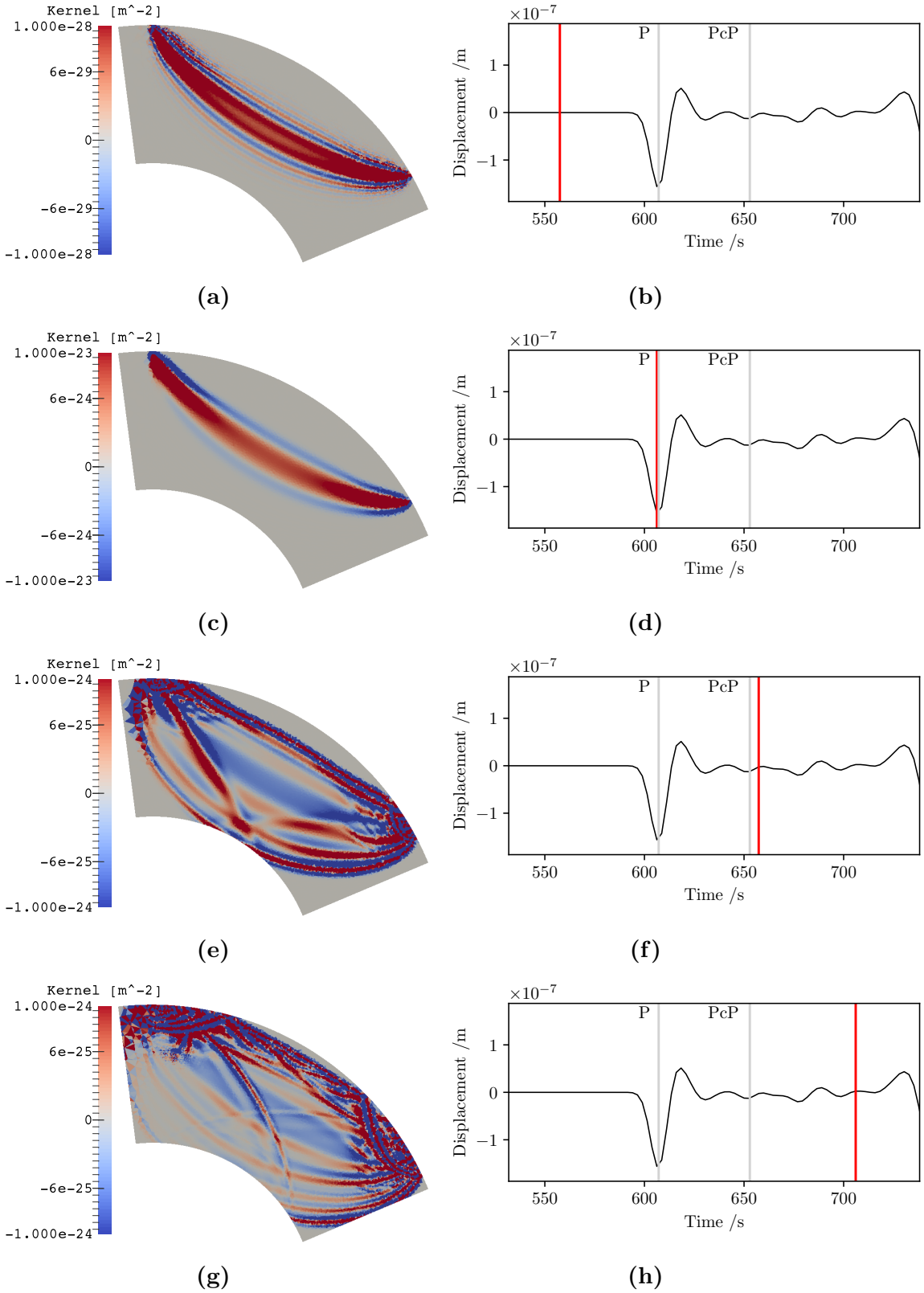


Figure 3.8: Waveform kernels in v_p for an explosion source at the North Pole, and receiver 60° away. The kernels are computed with a dominant period of 20 s. In the right column, the Z component seismograms are shown. The red line indicates the time that the snapshot of the waveform kernels in the left column are taken. We also indicate with grey lines the TauP predicted arrival times of the P and PcP phases.

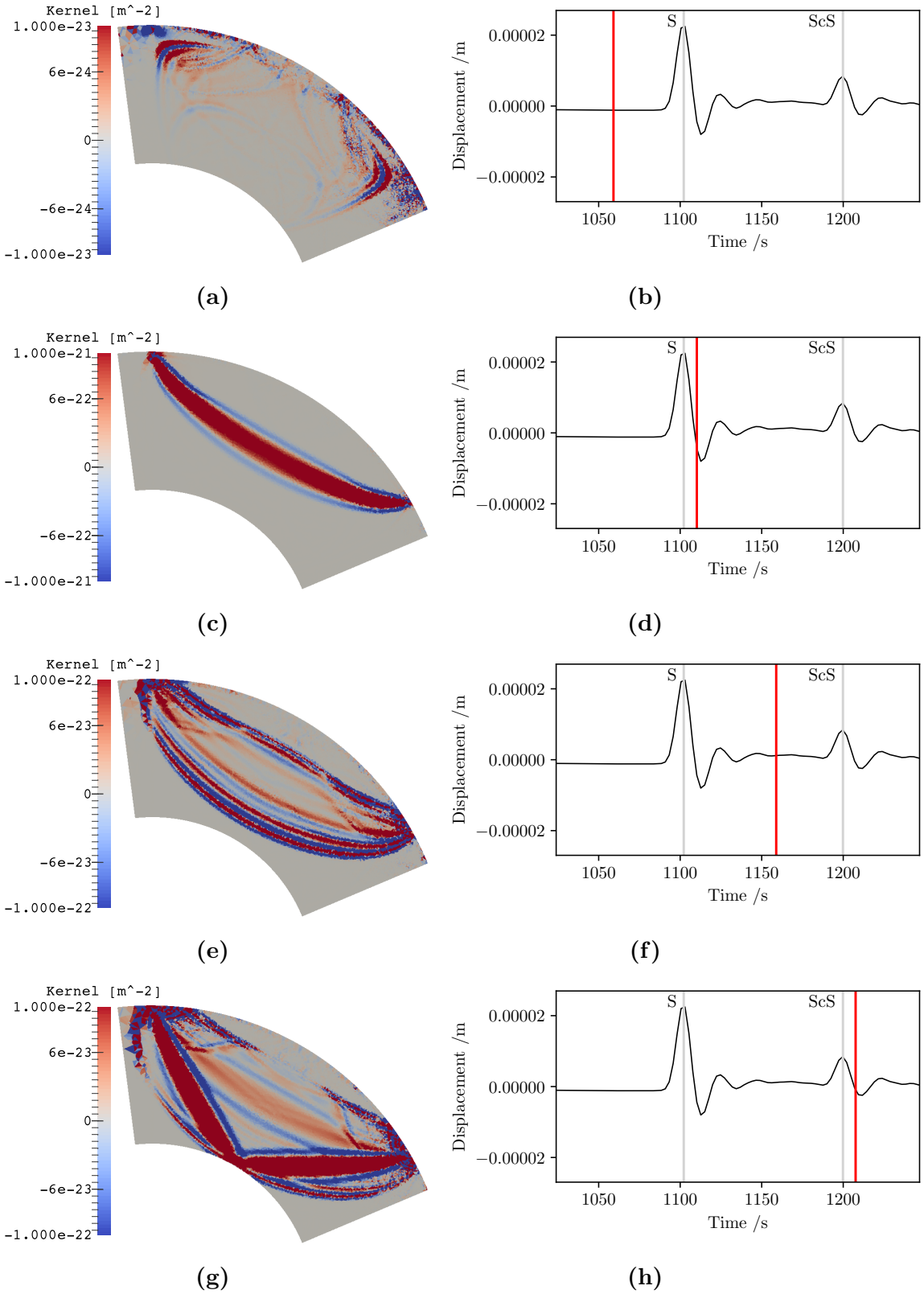


Figure 3.9: Waveform kernels in v_s for a moment tensor source at the North Pole, and receiver 60° away. The kernels are computed with a dominant period of 20 s. In the right column, the T component seismograms are shown. The red line indicates the time that the snapshot of the waveform kernels in the left column are taken. We also indicate with grey lines the TauP predicted arrival times of the S and ScS phases.

Chapter 4

Waveform modelling with Born, Rytov and Marks approximations

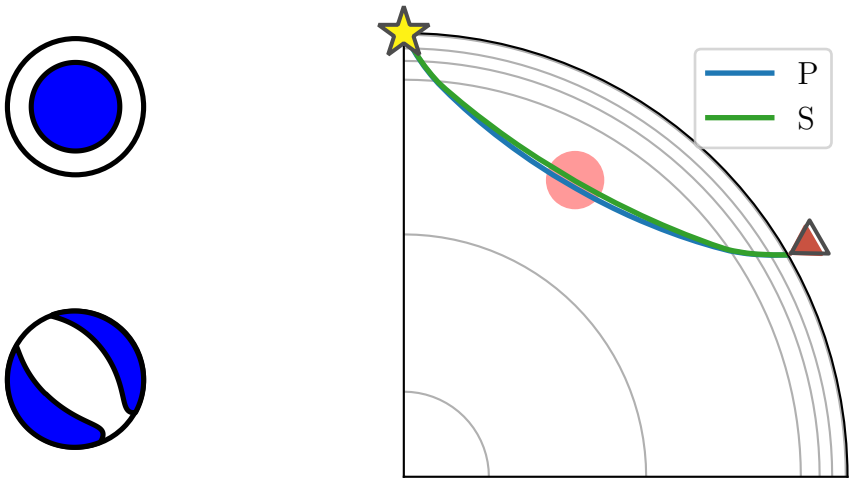
4.1 Introduction

In this chapter, I use waveform kernels calculated using MC Kernel to compute first order seismograms. These first order seismograms are compared to reference solutions which are found from the spectral element methods, AxiSEM (*Nissen-Meyer et al.*, 2014) and AxiSEM3D (*Leng et al.*, 2019). For small perturbations to the background model, comparison to AxiSEM/AxiSEM3D synthetics provides validation of the method as I benchmark the two set of seismograms against each other. With further deviation from the background model, the first order seismograms start to diverge from the reference solutions and I compare the two to determine where the applications are valid.

I first describe how the first order seismograms are computed from waveform kernels. These seismograms are compared to reference SEM solutions for two sets of perturbed models. The first set of models is defined by perturbing the 1-D Earth model PREM in a spherically symmetric manner such that the wavespeed below the discontinuity at 660 km depth is altered by up to $\pm 1\%$. The second set of models have spherical heterogeneities in the lower mantle, termed 'bubbles'. Bubble size ranges from 200 km

to 800 km in diameter, and the strength of the perturbation ranges from $\pm 1\%$ to $\pm 10\%$.

For the majority of this chapter, I examine the seismogram in windows around the direct P and S arrivals, and address perturbations in v_p and v_s separately. All seismograms are calculated using a Gaussian source-time function with a dominant period of 20 s, and filtered using a Butterworth bandpass filter from 20 - 100 s. I position an earthquake source on Earth's surface at the North Pole, and a receiver at 0°E, 30°N (Figure 4.1a). Compressional waves are studied using a simple moment tensor source with only an M_{zz} component, and I examine the vertical (Z) component of the seismogram. On the other hand shear waves are studied using a full moment tensor source based on a real earthquake, and I examine the transverse (T) component of the seismogram (Figure 4.1b).Displacement seismograms are studied in all cases.



(a) **(b)**

Figure 4.1: Set-up for waveform modelling used in this chapter. (a) shows beachball representations of the M_{zz} source for P waves (upper) and the full moment tensor source for S waves (lower). (b) shows the source (yellow star) at the North Pole and receiver (green triangle). The ray paths of the direct P and S arrivals are plotted in blue and green respectively.

4.2 Calculation of first-order seismograms

I describe the method for computing first order seismograms using, as an example, a bubble model of 800 km diameter with a perturbation in P wavespeed of -5% .

As shown by the theory presented in Chapter 2, the perturbation to the seismogram under the Born approximation, $\delta u(t)$, is found by integrating a waveform kernel over a perturbed model. In our discretised Earth model, this is carried out by multiplying the model perturbation in every mesh element by the waveform kernel, and then summing over mesh elements.

I also showed that we can extend from the the Born approximation to the Rytov and Marks approximations via the parameter $\phi = \delta u/u^0$. In the frequency domain, the first order seismograms are found using $u^{1^{st}} = (1+\phi/n)^n$, where n is dependent on the approximation being applied. In the case of the Rytov approximation, ϕ is added to the phase of the seismogram, or $ln(u)$.

Figure 4.2 shows the perturbations to the seismograms for an 800 km bubble with $\delta ln(v_p) = -5\%$. In the top panel, it can be seen that δu computed under the Born approximation compared to the unperturbed seismogram. The two are added together to find the Born seismogram in the time domain. The second panel shows $\phi(f)$ and the natural logarithm of the unperturbed seismogram in the frequency domain. The two are added together to find the phase of the Rytov seismogram. The bottom two panels show the unperturbed seismogram along with the Born and Rytov seismograms in the frequency and time domains, showing the differences between the two approximations. Although not shown in Figure 4.2, the intermediate Marks approximation falls between the two, and in this section two values are chosen with $n = 3$ and $n = 6$.

A summary of the computational procedure for obtaining first order seismograms:

1. Calculate waveform kernel using MC Kernel facilitated by AxiSEM. Extract unperturbed seismogram from AxiSEM run.
2. Project velocity model onto the same mesh as the kernel
3. Calculate δu under the Born approximation by integrating the waveform kernels

over the perturbed model

4. Find ϕ in the frequency domain from $\delta u/u^0$. Use ϕ to compute first order seismograms with any value of n in the frequency domain.
5. Apply filter and transform back to time domain.

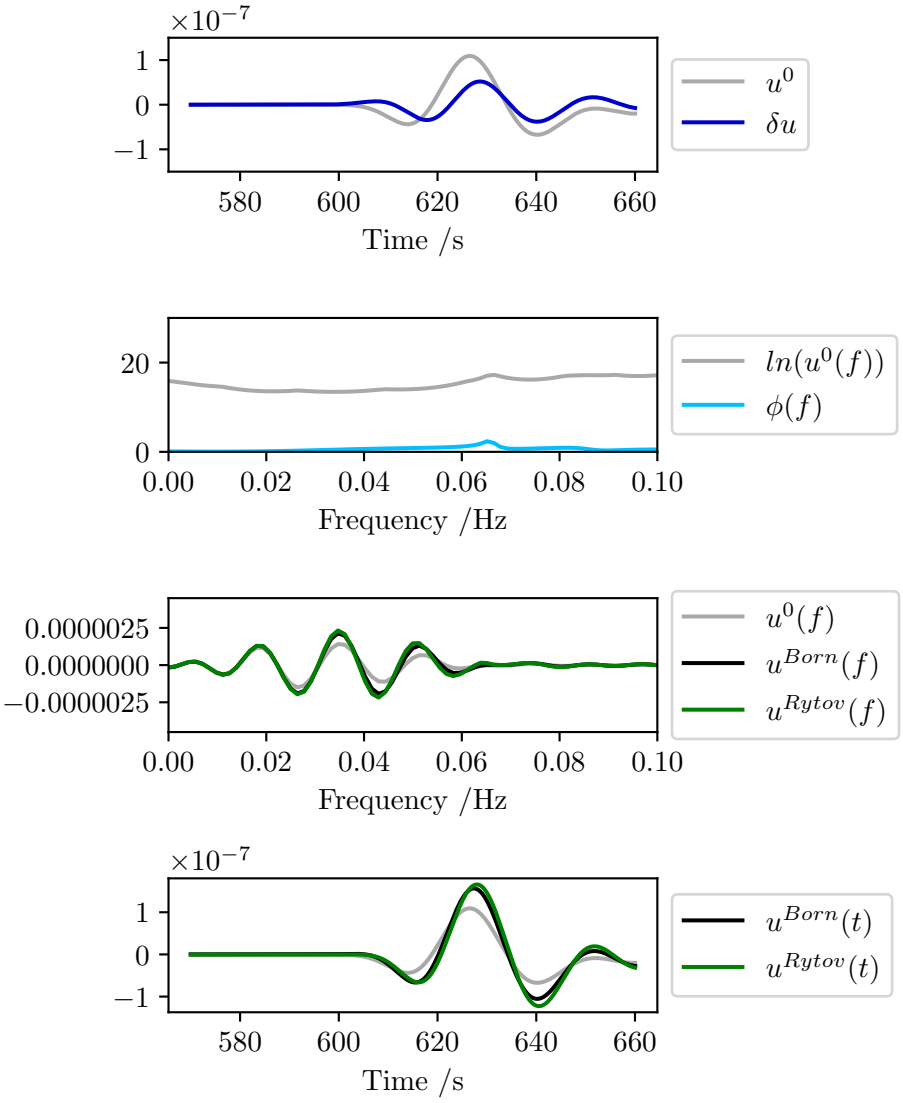


Figure 4.2: Steps in computing Born and Rytov seismograms. The top panel shows the unperturbed seismogram $u^0(t)$ and the perturbation under the Born approximation $\delta u(t)$. In the frequency domain, the Rytov approximation is applied by the addition $\phi(f)$ to the natural logarithm of $u^0(f)$, as shown in the second panel. The third and fourth panels show the Born, Rytov and unperturbed seismograms in the frequency and time domains respectively.

4.2.1 Kernel calculation

The kernels are calculated in PREM using AxiSEM wavefields with a dominant period of 20 s, and a sampling rate of 8 samples per period. I use the 30 km tetrahedral mesh shown in Figure 3.4a. The kernels are computed with with a low pass Butterworth filter at 20 s. A snapshot of the P waveform kernel at 60° is shown in Figure 4.3.

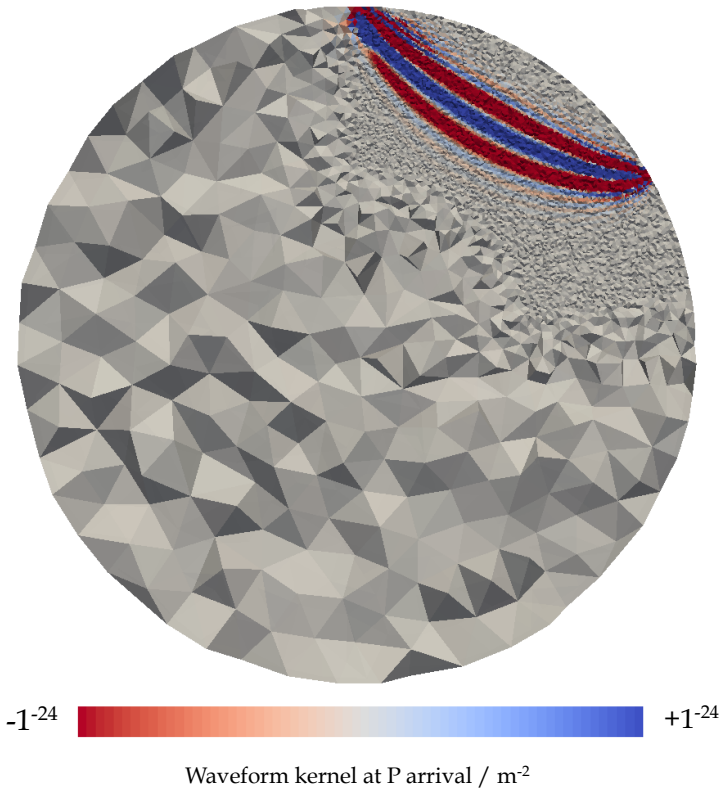


Figure 4.3: Slice of the waveform kernel with respect to compressional wavespeed at 607 s (the TauP predicted arrival time of the P wave) for a source at the North pole and a station 60° away.

4.2.2 Model projection

The perturbed model is initially defined on its own grid, and projected onto the same tetrahedral mesh as the waveform kernel. The centre point of each tetrahedron, $\overline{\mathbf{x}}$, is calculated from the mean coordinates of the vertices. The four nearest points on the

model grid are found using a KD tree (*Maneewongvatana and D. M. Mount*, 1999). The velocity perturbation v_n inside the n^{th} element is found from the average perturbation of the four nearest model points, v_i, weighted by their distance to the centre of the tetrahedron, r_i:

$$v_n = \frac{\sum_{i=1}^{i=4} v_i r_i}{\sum_{i=1}^{i=4} r_i}, \tag{4.1}$$

where $r_i = |\mathbf{x_i} - \overline{\mathbf{x}}|$. The calculation of the midpoints is independent of the model and is carried out separately.

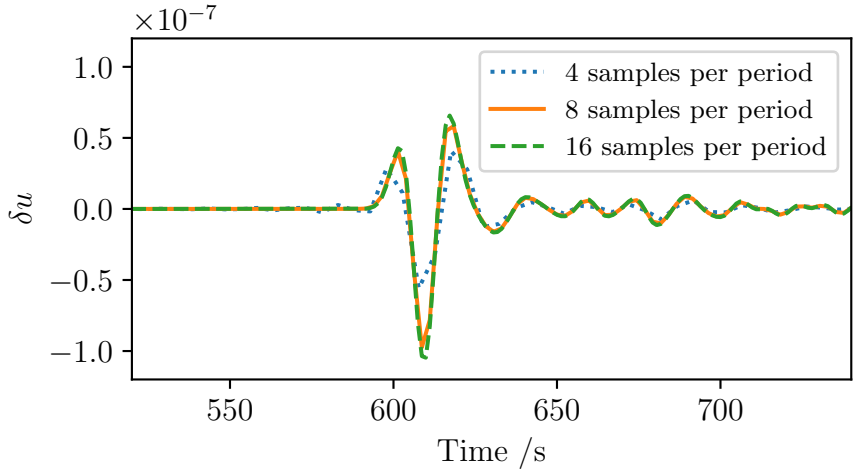
4.2.3 Checks on the method

I test the methodology by examining $\delta u(t)$ in the following ways:

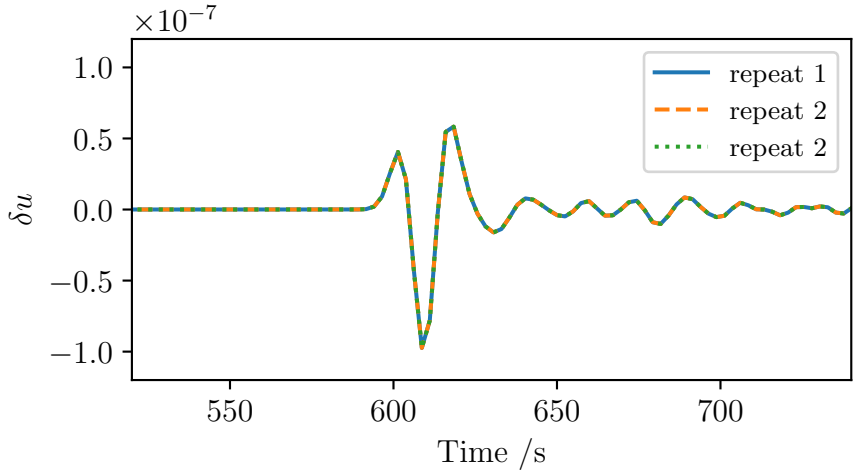
1. I determine the minimum number of samples per period (spp) needed to capture the effects of the heterogeneity on the waveform. I compare $\delta u(t)$ calculated using 4, 8 and 16 samples per period and find that at least 8 are required (Figure 4.4a).

2. I manually check the convergence of the waveform kernels by comparing δu from three repeats of the same run. Figure 4.4*b* shows perfect agreements between the three repeats for the P kernel used in this chapter. This test is carried out on every kernel used in this thesis until a more suitable method is found.

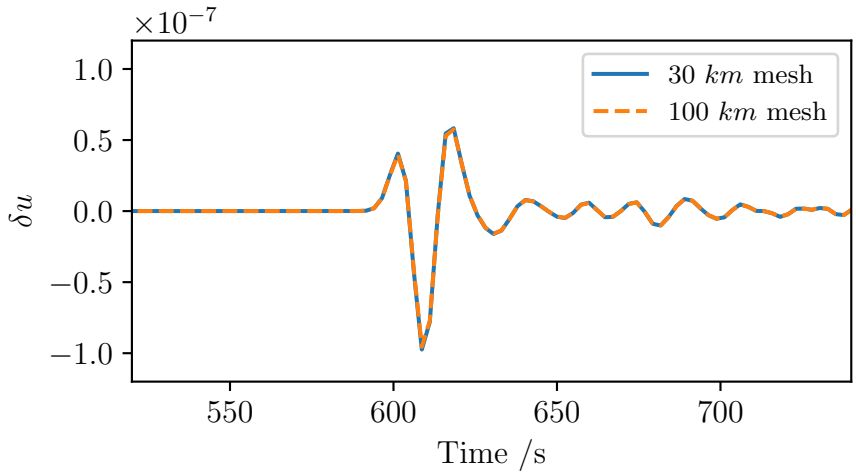
3. I confirm that the results are independent of the tetrahedral mesh by comparing results from our 30 km and 100 km meshes (Figure 4.4c).



(a)



(b)



(c)

Figure 4.4: A summary of the checks performed on the method, showing how results are independent of (a) samples per period above 8 and (c) the specific mesh used. (b) is a manual convergence check on the waveform kernels.

4.2.4 Note on the calculation of ϕ

A well-established problem with the deconvolution of one signal from another (division in the frequency domain) is that of encountering a value of zero, or close to zero, in the demoninator.

Because ϕ is computed by deconvolving u_0 from δu there is the potential for very large values of ϕ to be obtained. After exponentiating ϕ to compute the Rytov seismogram in the frequency domain, these values become even larger and could lead to a single frequency dominating the signal in the Rytov seismogram. This means that the Rytov seismogram computed using the method does not accurately reflect the approximation, but rather a problem with the technical implementation.

Various methods have been applied in the context of receiever functions to alleviate this problem. These include the water level method (e.g. *Ammon* 1991), where small values of the denominator are replaced with a constant value. However, when applying such methods in this context, the question of whether the resulting seismograms represent the first order approximations is raised.

Taking the Born approximation, the 'true' Born seismogram is simply obtained in the time domain by the addition $u^{Born}(t) = u^0(t) + \delta u(t)$. Consider performing a calculation in the frequency domain, in which $\phi(f) = \delta u/u^0$ is computed, and then $u^{Born}(f) = u^0(f)(1 + \phi(f))$, before Fourier transforming back into the time domain. In this case, if $\phi(f)$ were to be altered such that very large values are not incurred, the resulting $u^{Born}(t)$ will not be the same as the 'true' Born seismogram.

For these reasons, I refrain from applying any correctional approaches to the deconvolution problem, in order to assess the approximations in their raw form. Instead, I apply an automated approach to flag when very small values of $u_0(f)$ are encountered, and the Rytov and Marks seismograms are not used in such cases. This has not occurred for any of the seismograms computed for this thesis.

4.2.5 Quantifying differences between first order and reference seismograms

Differences in the amplitude and phase of the first order seismograms compared to the reference AxiSEM3D solutions are independently evaluated using phase and envelope misfits as defined by *Kristeková et al.* (2006).

These are based on the time-frequency representation of the signal, which shows the time evolution of spectral content. Phase misfit (PM) and envelope misfit (EM) are single values obtained by integrating over time and frequency. For each AxiSEM3D-first order seismogram pair, EM and PM are calculated. The values of EM and PM may be interpreted as percentage misfit in phase and amplitude with reference to the AxiSEM or AxiSEM3D seismogram.

Based on the analysis of *Kristeková et al.* (2009), where EM and PM values are associated with verbal values for goodness-of-fit, a threshold EM and PM is defined of 10% below which we do not consider the approximations to be applicable. This corresponds to a verbal goodness-of-fit of 'excellent'.

4.3 1-D models

In this section, first order seismograms are compared to reference solutions calculated using AxiSEM for 1-D models based on perturbing PREM below the 660 km discontinuity. An example model is shown in Figure 4.5. In total, perturbations to v_p and v_s of $\pm 0.1\%, \pm 0.2\%, \pm 0.5\%, \pm 1.0\%$ are considered .

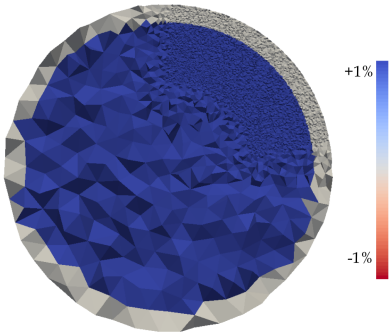


Figure 4.5: A 1-D perturbed model projected onto the 30 *km* tetrahedral mesh.

4.3.1 Perturbations in compressional wavespeed

The effect of the 1-D models on the waveform around the P arrival is shown in Figure 4.6. The effect is a simple advance or delay in arrival time of the P phase with very small changes in amplitude and without any significant distortion of the waveform.

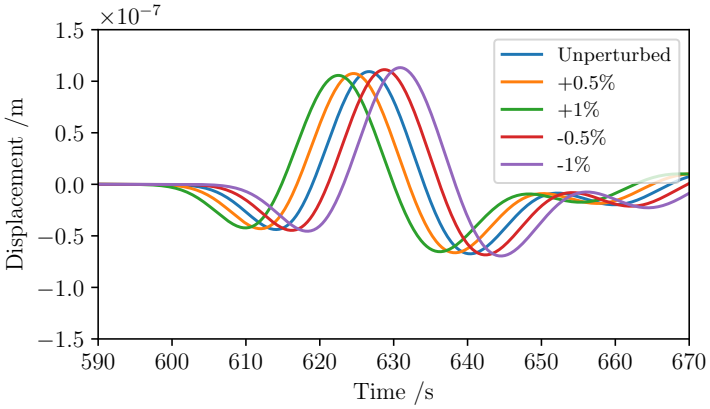


Figure 4.6: The effect of modifying v_p below the 660 km discontinuity. The Z component seismogram is shown in a window around the P arrival.

Figure 4.7 shows the P seismograms for all percentage perturbations except ±0.1% where the perturbation is too small to be seen clearly. Envelope and phase misfits are calculated for each pair of a first order seismogram with a reference seismogram. the envelope and phase misfits are plotted in Figure 4.8. Envelope and phase misfits are consistently highest in the Born case, lowest in the Rytov case and somewhere in between for examples of the Marks approximation. Marks with $n = 6$ is closer to the Rytov case, while Marks with $n = 3$ is closer to the Born case. The misfits for positive perturbations are lower than negative perturbations. I also calculate the cross-correlation time delay of the first-order seismogram relative to the PREM seismogram, since this is a common measurement in both tomography and waveform modelling. This is compared to the measured cross-correlation time delay between the true solution and the PREM seismogram, as shown in Figure 4.8c. At larger magnitudes of $\delta ln(v_p)$, the measurements on Born seismograms diverge from measurements made from the true solution. Measurements on the Rytov seismograms are very close to the reference values, while Marks is in between Born and Rytov.

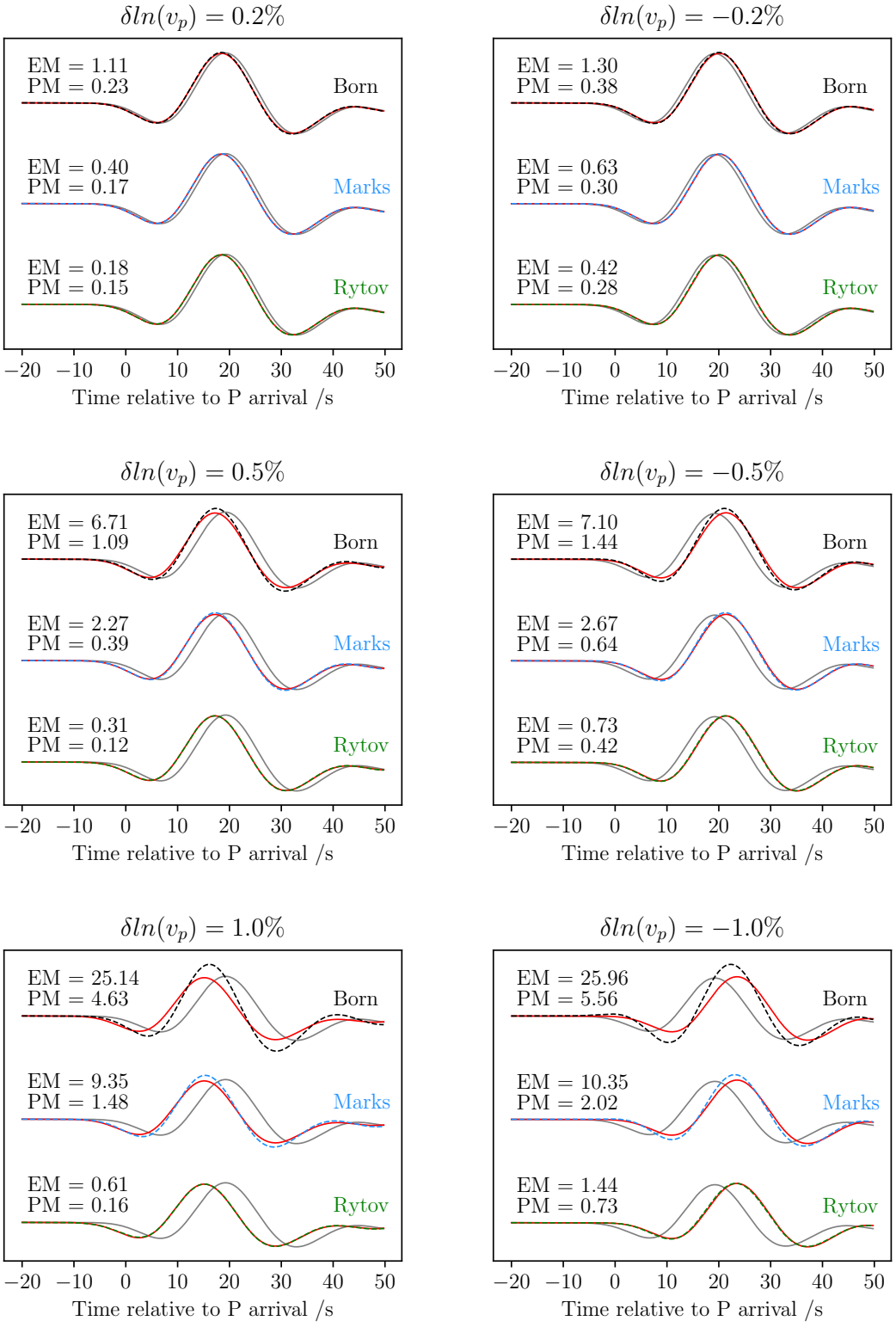
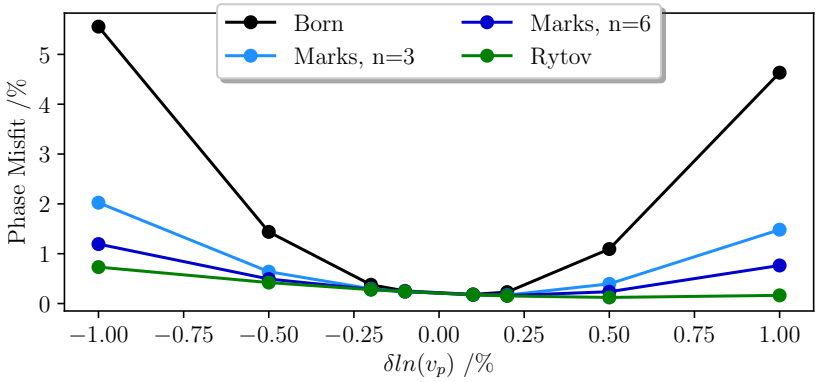
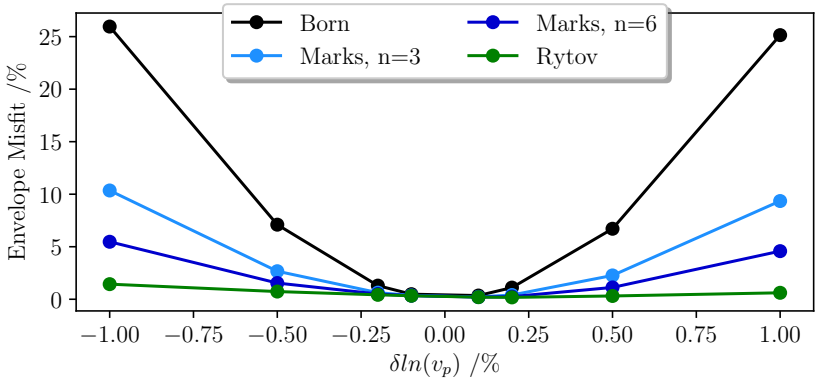


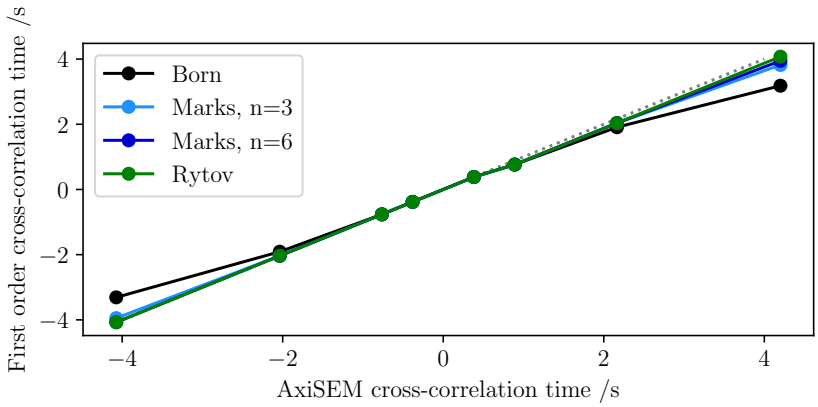
Figure 4.7: Plots comparing the PREM seismogram (grey), true solution (red), and the first order seismograms (Born in black, Marks with n=3 in blue and Rytov in green) for a selection of perturbations to PREM in compressional wavespeed below the 660 *km* discontinuity. Envelope misfit (EM) and phase misfit (PM) between the reference seismogram and the first order seismogram are displayed in percent.



(a)



(b)



(c)

Figure 4.8: Phase misfits (a) and envelope misfits (b) against percentage perturbation in $\delta ln(v_p)$. (c) shows AxiSEM cross-correlation travel time against first order cross-correlation time. AxiSEM cross-correlation time is the difference in arrival time of the P wave in PREM compared to the modified 1-D model, as measured by cross-correlation between the PREM seismogram and the true solution computed using AxiSEM. The first order cross-correlation time is the difference in arrival time as measured by cross-correlation between the PREM seismogram and the first order seismogram. The grey dotted line is the line of parity between the two measurements.

1-D perturbations throughout the whole Earth

We deviate slightly here to present the results of perturbing v_p by -0.2% throughout the whole Earth, (as opposed to only in the lower mantle). Figure 4.9 compares the first order approximations to reference solutions calculated using AxiSEM. The fit between the two sets of seismograms is poor, despite it being a small perturbation. Our first order method is not applicable, and this is due to the complexity in the kernel around the source and receiver regions.

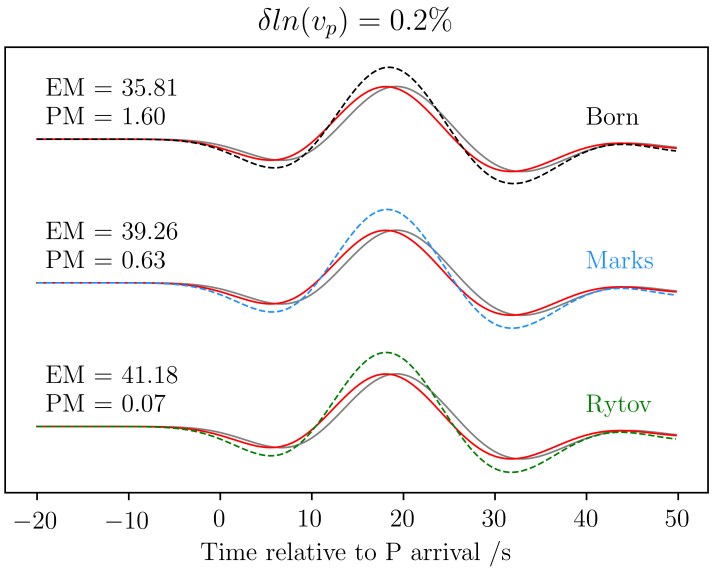


Figure 4.9: Comparison of first order seismograms to reference AxiSEM solutions for a 1-D model in which PREM is perturbed uniformly by -0.2%.

4.3.2 Perturbations in shear wavespeed

The effects of modifying v_s below the 660 km discontinuity on the waveform around the S arrival are very similar to those of perturbing v_p on the P arrival.

In an analogous manner to Section 4.3.1, we compare the first order seismograms to the true solution computed using AxiSEM. The pattern in misfit values is very similar to the results shown above for P waves (Figures 4.7 and 4.8). However, the misfit values are larger in the S case.

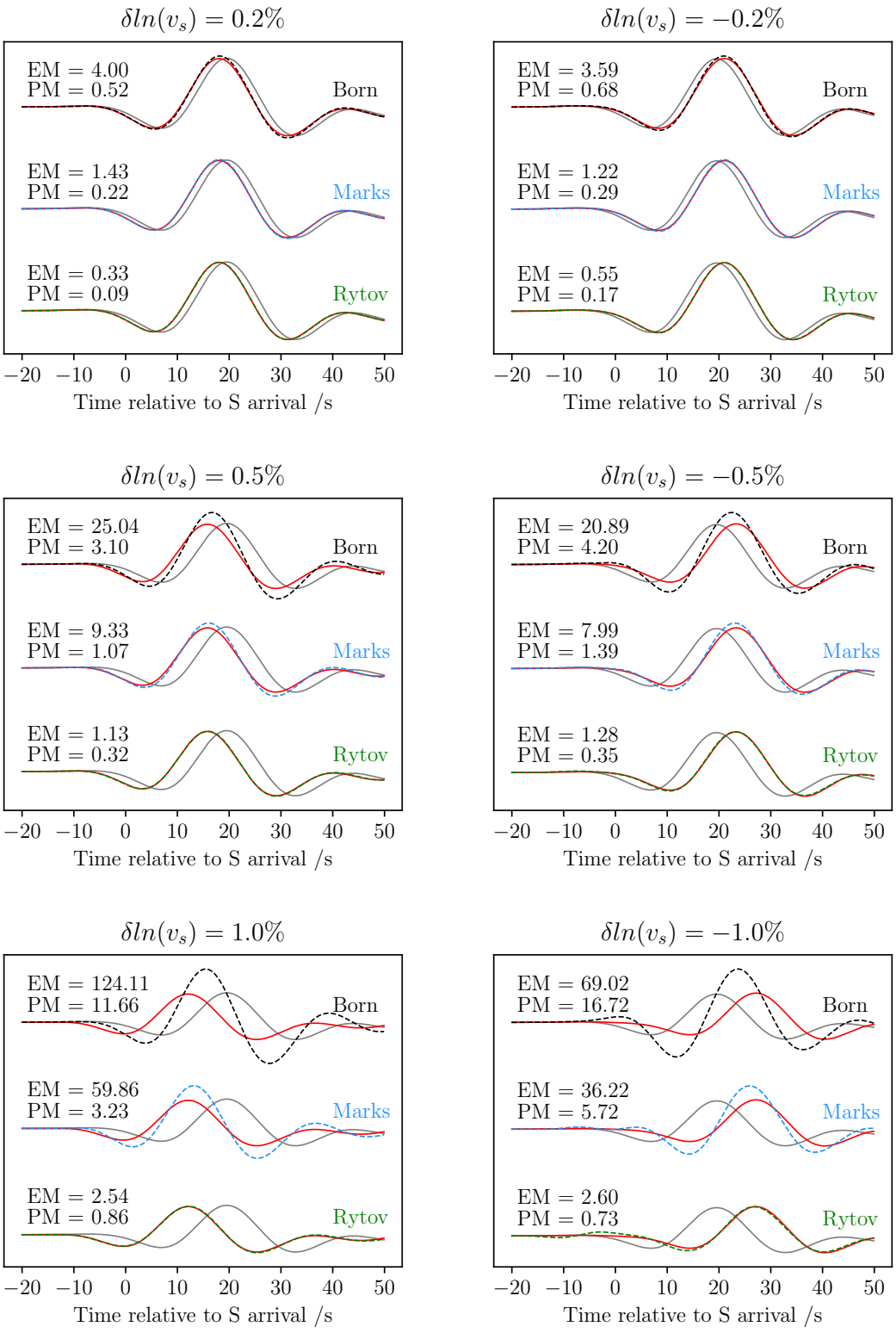
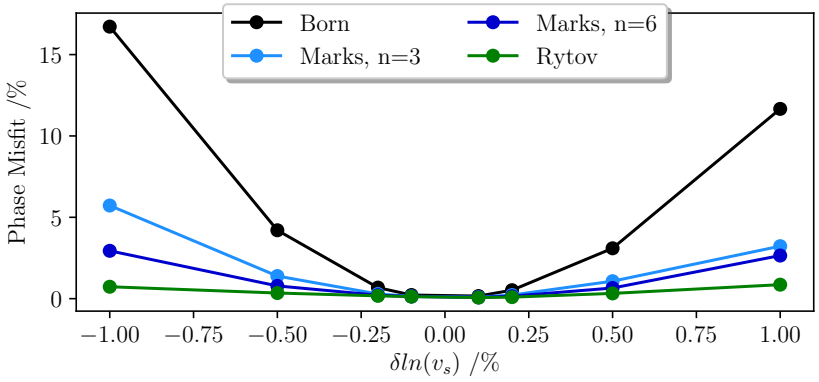
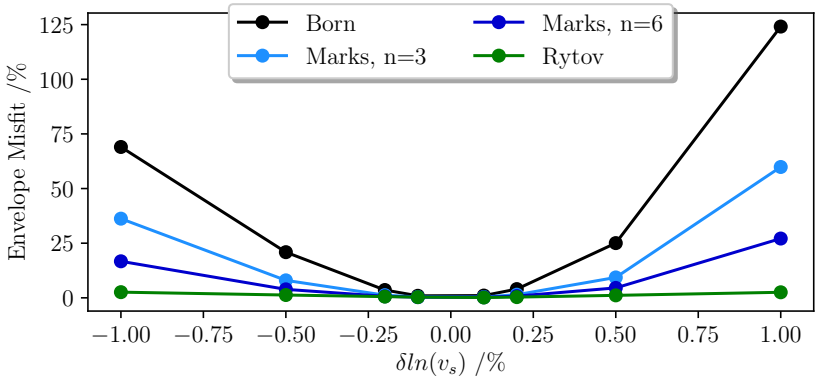


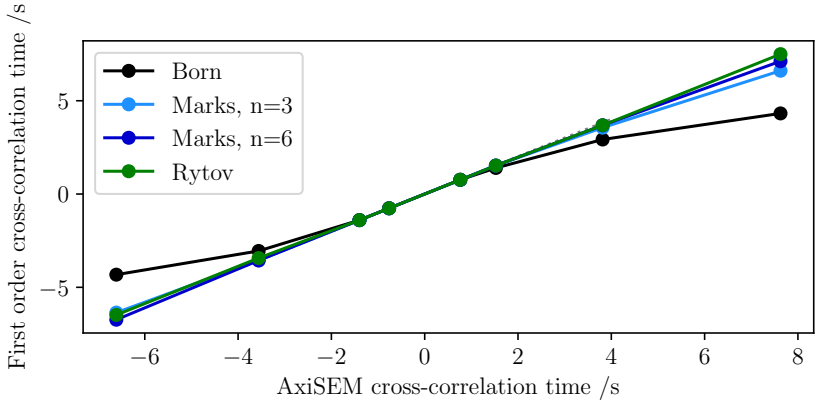
Figure 4.10: Plots comparing the PREM seismogram (grey), true solution (red), and the first order seismograms (Born in black, Marks with n=3 in blue and Rytov in green) for a selection of perturbations to PREM below the 660 km discontinuity. Envelope misfit (EM) and phase misfit (PM) between the reference seismogram and the first order seismogram are displayed in percent.



(a)



(b)



(c)

Figure 4.11: Phase misfits (a) and envelope misfits (b) against percentage perturbation in $\delta ln(v_s)$. (c) shows AxiSEM cross-correlation travel time against first order cross-correlation time. AxiSEM cross-correlation time is the difference in arrival time of the S wave in PREM compared to the modified 1-D model, as measured by cross-correlation between the PREM seismogram and the true solution computed using AxiSEM. The first order cross-correlation time is the difference in arrival time as measured by cross-correlation between the PREM seismogram and the first order seismogram. The grey dotted line is the line of parity between the two measurements.

4.4 Bubble models

The performance of the first order approximations is now examined with the more complicated set of models consisting of spherical heterogeneities in the lower mantle. Again, we first analyse the results from perturbing v_p and subsequently turn to v_s. A full analysis is carried out on the v_p perturbations, while fewer scenarios are examined in the case of v_s as we observe similar results.

In the case of v_p, we consider bubbles with strengths of $\delta ln(v_p) = \pm 1\%, \pm 5\%, \pm 10\%$. The bubble diameter, a, ranges from 200 km to 800km . In the case of v_s, we select four bubbles with sizes between 200 km and 500 km and strengths between -1% and -5%. The bubbles are positioned at the turning point of the TauP (*Crotwell et al.*, 1999) predicted turning point of the direct P ray. Therefore, the P waves measured at 60° pass straight through the bubble. The turning point of the theoretical S ray is approximately 90 km shallower than the turning point of the P ray.

Period estimation

It becomes important at this stage to consider the size of the perturbation relative to the wavelength of the seismic wave as it encounters the heterogeneity. An estimate of the wavelength is obtained by multiplying the period of the wave by the PREM wavespeed at the depth at the centre of the bubble. In turn, the period is determined from the power spectrum of the PREM seismogram. Determination of the period from a seismogram is not exact, and the number obtained depends on subjective decisions. It is typical in computational global seismology to refer to the period of a specific simulation, as determined by spatial and temporal discretisation as well as the dominant period of the source. In this case the dominant period is estimated by a Fourier transform of the whole length of the seismogram. However, different phases in the seismogram have different periods, the main difference being between surface and body waves. In this case we are specifically interested in the body waves which sample the lower mantle. The seismogram is therefore trimmed around the P and S arrivals and calculate the dominant frequency in this window. The spectra of the whole seismogram and the

trimmed seismogram for the P set-up are shown in Figure 4.12. Using this method, the periods of the P and S arrivals are determined to be 28.17 s and 35.02 s. This corresponds to a wavelength of 377 km at the centre of each bubble for the P wave and 298.5 km for the S wave.

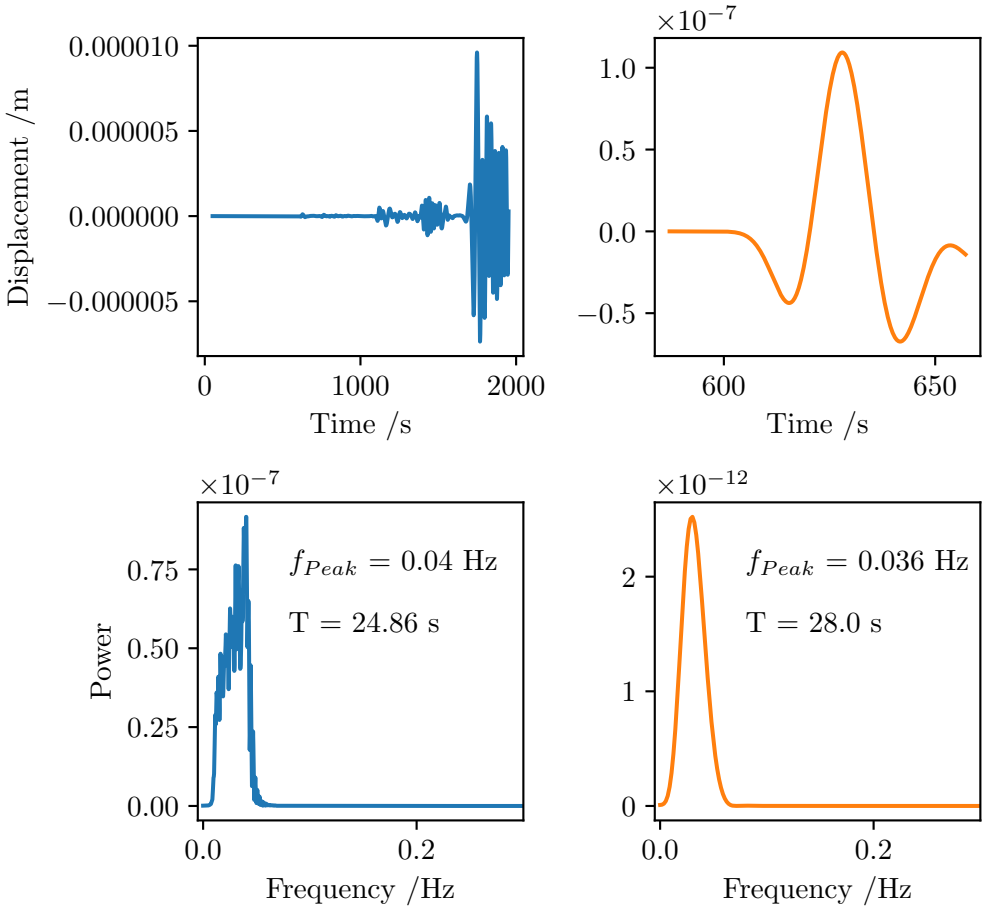


Figure 4.12: Frequency content of the Z component PREM seismogram for the P set-up. Power is equivalent to $|\mathscr{F}(u)|^2$, where $\mathscr{F}$ represents the Fourier transform. On the left (blue), the dominant period of the whole seismogram is shown to be 24 s. On the right, we see the dominant period of the seismogram trimmed around the P arrival is 28s.

4.4.1 Perturbations in compressional wavespeed

AxiSEM3D record sections

The effects of a selection of spherical heterogeneities with perturbations in v_p on the AxiSEM3D seismograms (taken to be the true solution) are shown in Figures 4.14 and 4.15. These are for bubbles with diameters, a, of 200 km, 400 km and 800 km corresponding to structure to wavelength ratios, a/λ, of 0.53, 1.06 and 2.12. Figure 4.14a shows how a 200 km bubble with a -10% perturbation effects the seismogram in a manner that is barely perceptible. In the zoomed section the effect can be seen to be very small effects on the amplitude. Comparatively, the perturbation to the seismogram from a 400 km bubble (Figures 4.14b - 4.14c) is far more significant. The dominant effect is that of an increase or decrease in amplitude for the negative and positive perturbation respectively. This change in amplitude is most significant at an epicentral distance of 60°, since the bubble is positioned at the turning point of the theoretical P ray. There is a minimal change in arrival time of the phase, as is expected for a heterogeneity that lies within the first Fresnel zone of the travel time kernel. Bubbles of 800 km diameter (Figure 4.15)affect the travel time of the P phase, with a delay in arrival time for a slow perturbation, and an advance in arrival time for a fast perturbation.

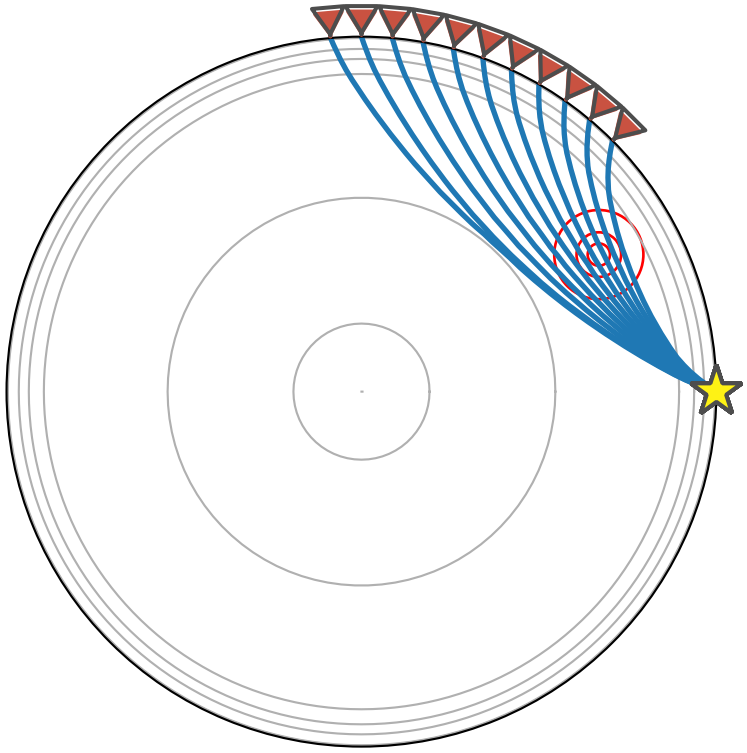


Figure 4.13: Ray geometry for Figures 4.14 and 4.15. The source is shown as a yellow star, and receivers with red triangles. Three different sizes of bubble heterogeneity (a = 200, 400, 800 km)are shown in red in the lower mantle.

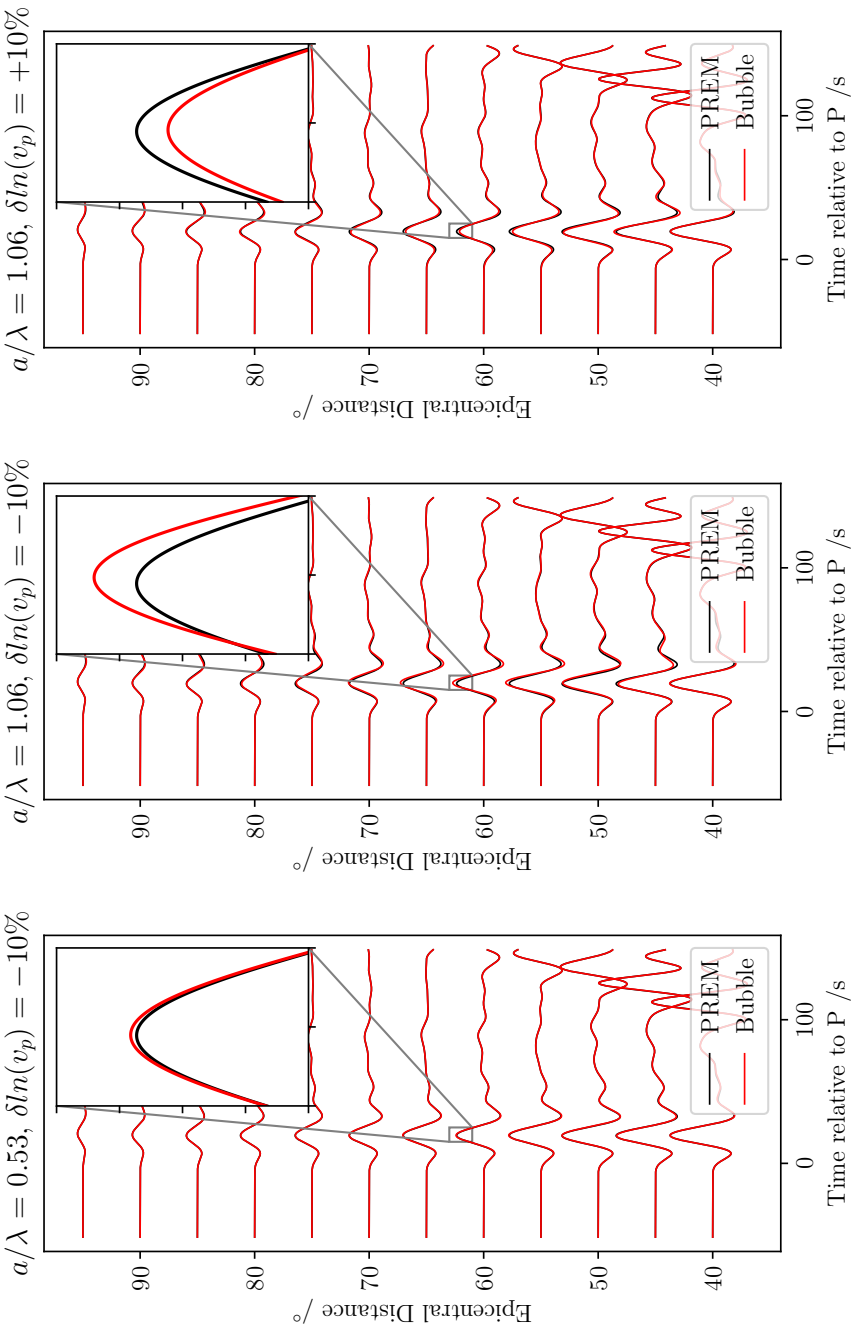


Figure 4.14: The effects of a 200 km bubble ($a/\lambda = 0.53$) and two 400 km ($a/\lambda = 1.06$) bubbles with perturbations in compressional wavespeed as described in each title.

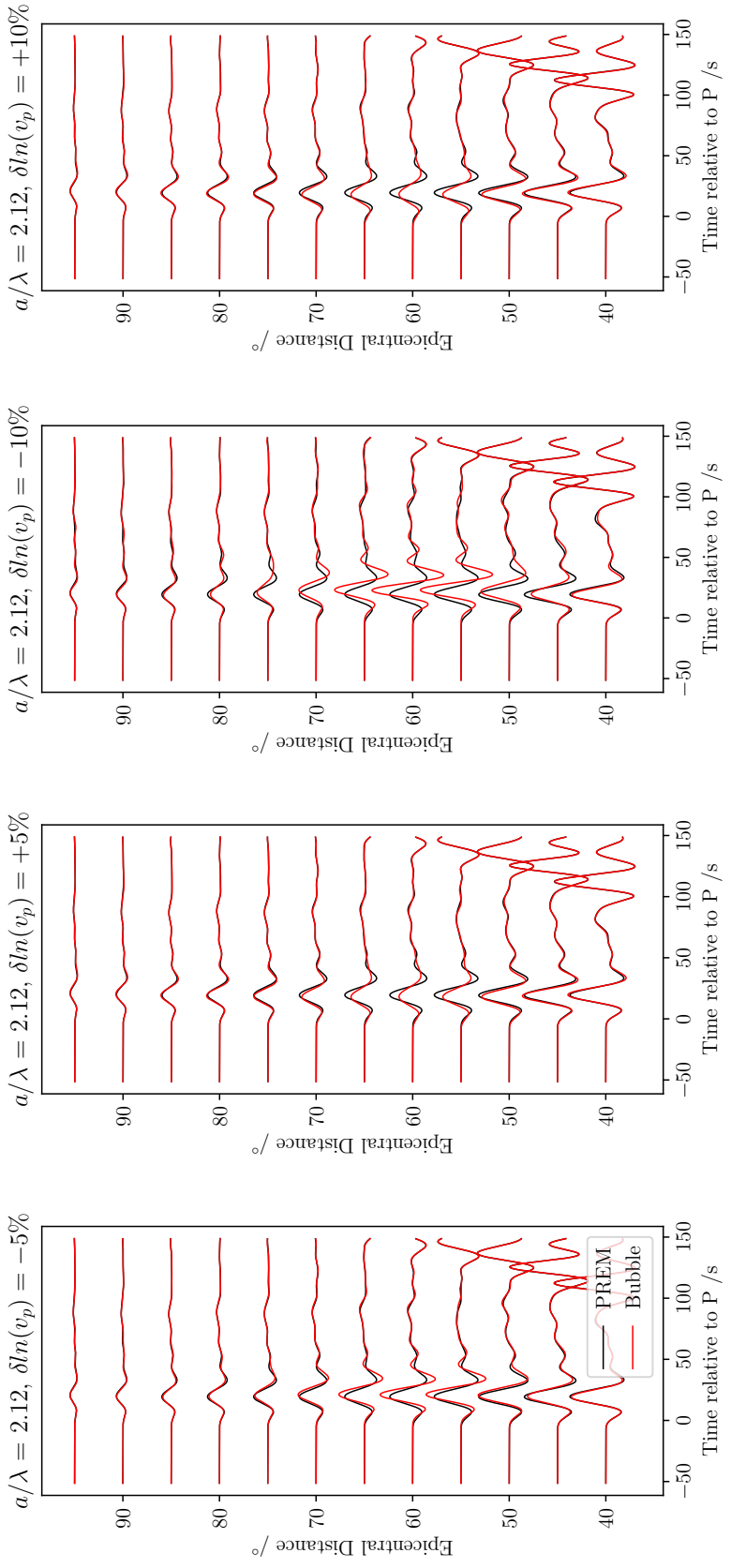


Figure 4.15: The effects of 800 km bubbles ($a/\lambda = 2.12$) with perturbations in compressional wavespeed as described in each title.

First order results

In this section, the results of applying the first order method to the bubble heterogeneities are presented. All of the seismograms shown are for an epicentral distance of 60°. Figures 4.16 and 4.17 are comparisons between first order seismograms and reference AxiSEM3D solutions for bubbles of 400 km and 800 km diameter. Equivalent plots for 200 km bubbles are not shown since the perturbation is very small.

Figures 4.18 and 4.19 are summaries of PM and EM for negative and positive perturbations respectively. EM and PM are plotted against the structure-wavelength ratio (a/λ) in a separate panel for each velocity perturbation. I determine a good fit to be a percentage misfit below 10%. As the heterogeneities become stronger or larger, corresponding to greater perturbations to the waveforms, the first order approximations do not perform as well. This means that the EM and PM misfit values become larger as is obvious, but they also become less reliable as they are not applicable to large waveform differences (*Kristeková et al.*, 2009).

In the figures below, the Rytov approximation is always more accurate in predicting the phase of a perturbed waveform as demonstrated by the consistently lower PM values. The Rytov approximation is also most often the best at estimating the amplitude as shown by EM values. However there are a few examples where the envelope misfit is lowest for the Marks or Born approximations.

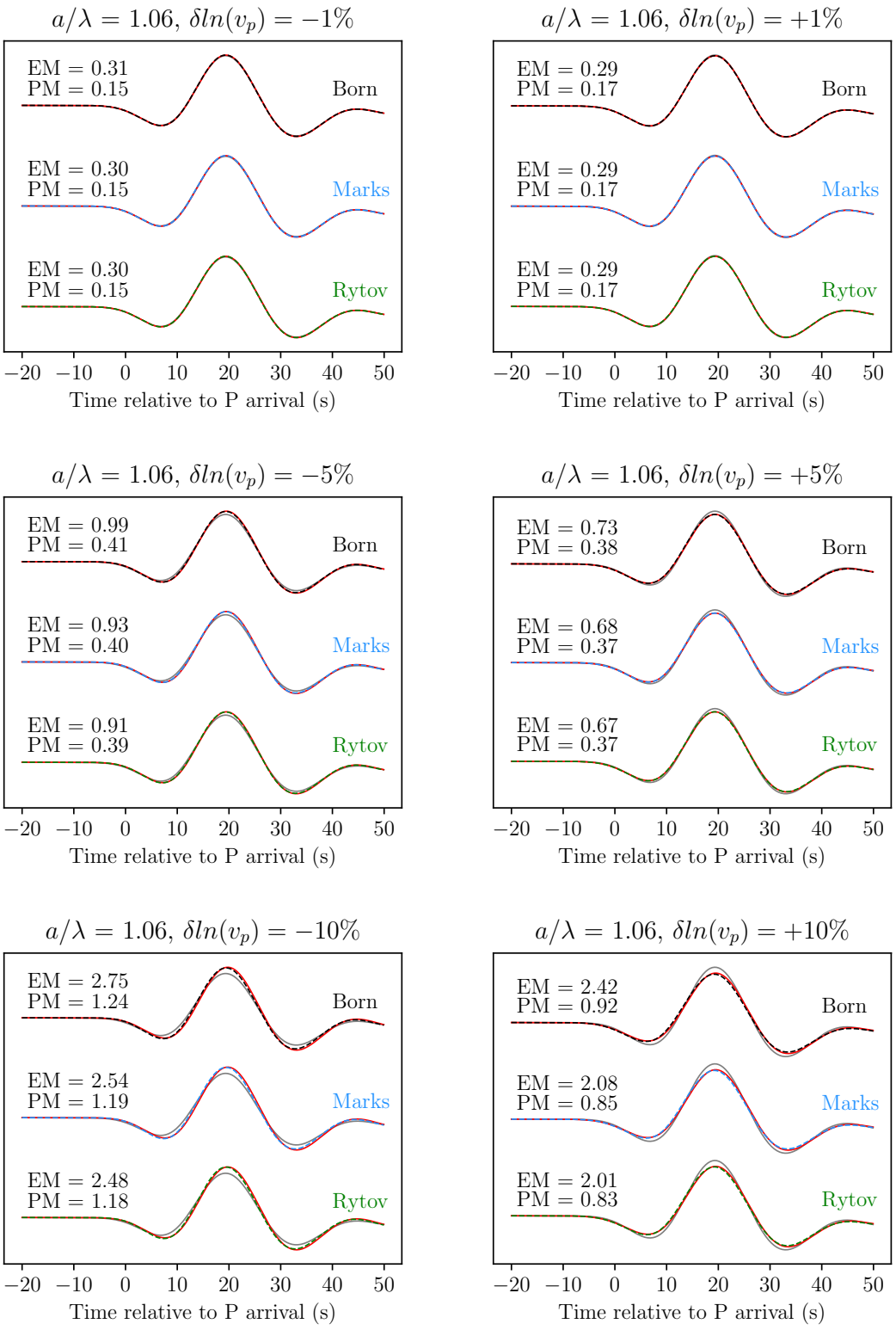


Figure 4.16: Plots comparing the PREM seismogram (grey), reference solution (red), and the first order seismograms (Born in black, Rytov in green, and Marks with n=3 in blue) for bubbles with structure to wavelength ratio of ≈1.06. All seismograms are recorded at an epicentral distance of 60° as per the geometry in Figure 4.1a. Envelope and phase misfit (EM and PM) are displayed in percent.

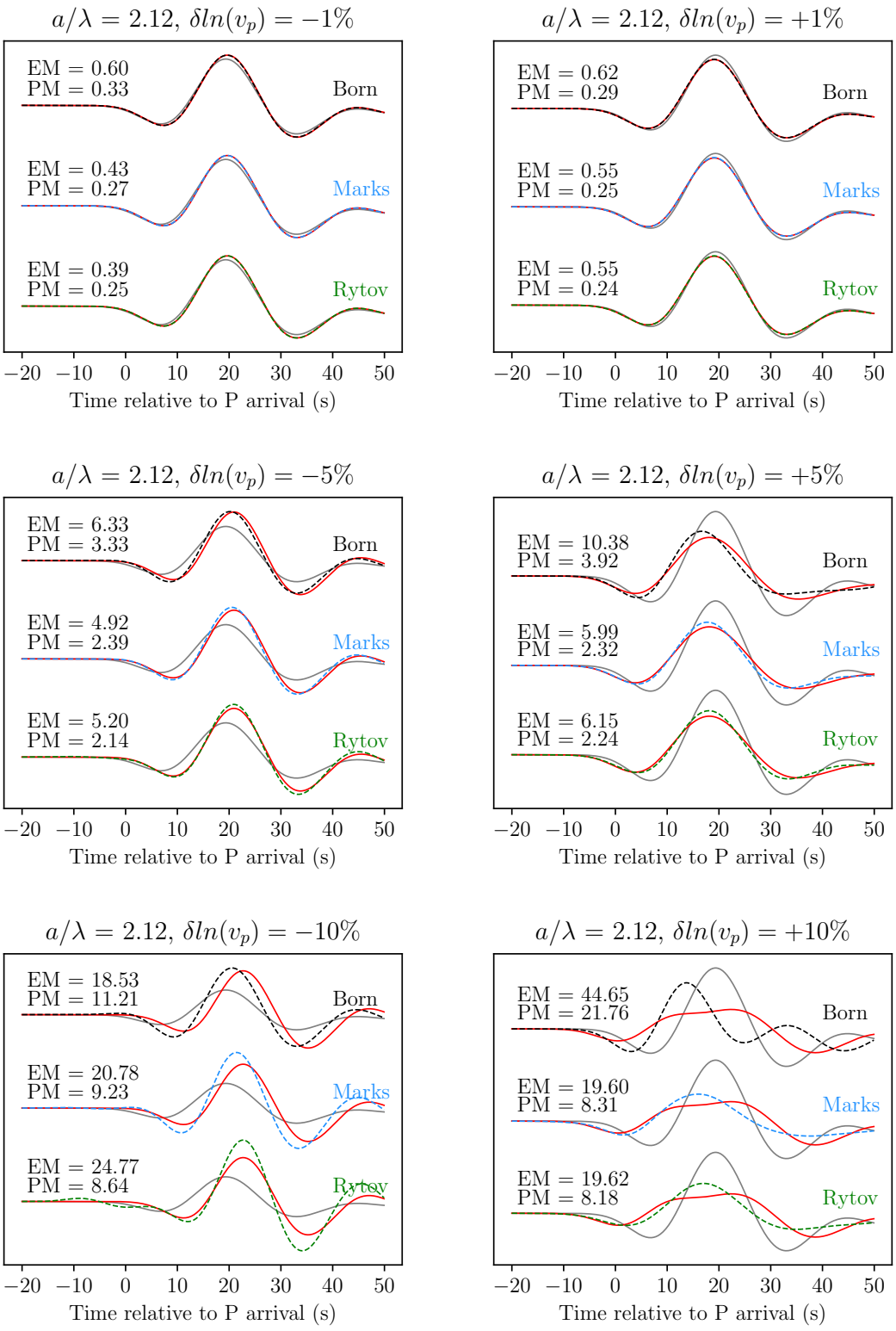


Figure 4.17: Plots comparing the PREM seismogram (grey), reference solution (red), and the first order seismograms (Born in black, Rytov in green, and Marks with n=3 in blue) for bubbles with structure to wavelength ratio of ≈2.12. All seismograms are recorded at an epicentral distance of 60° as per the geometry in Figure 4.1a. Envelope and phase misfit (EM and PM) are displayed in percent.

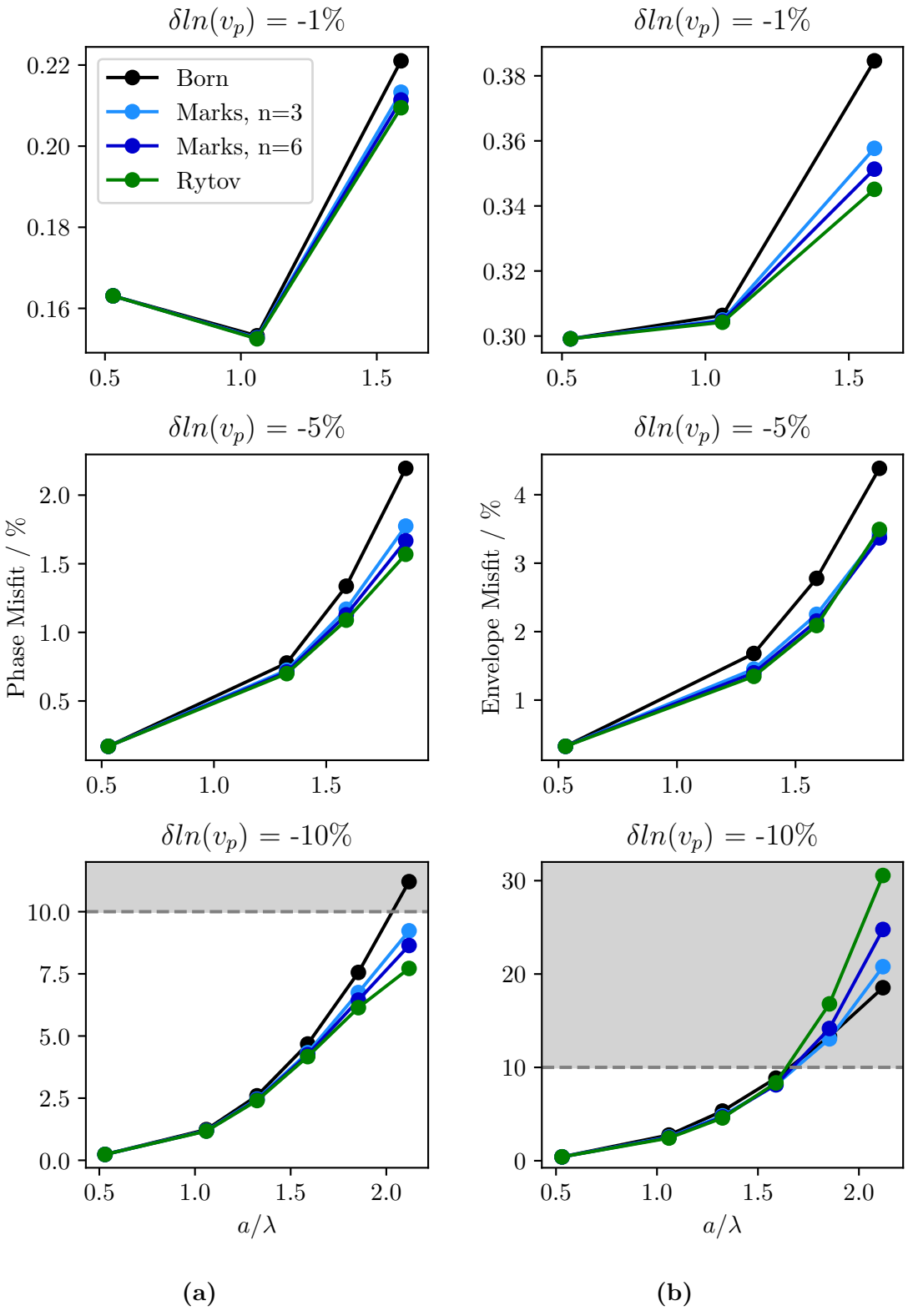


Figure 4.18: A summary of phase and envelope misfits between the reference AxiSEM3D solutions and the first order seismograms, for bubbles with negative perturbations in v_p. Misfits for Born seismograms are shown in black and Rytov in green. The misfits for two types of Marks seismograms corresponding to $n = 3$ and $n = 6$ are shown in light blue and dark blue respectively. Phase misfit is shown on the left in subfigure (a) and envelope misfit is shown on the right in subfigure (b). The grey shaded areas represent regions where the misfit is above 10%, where we consider the approximations to be invalid.

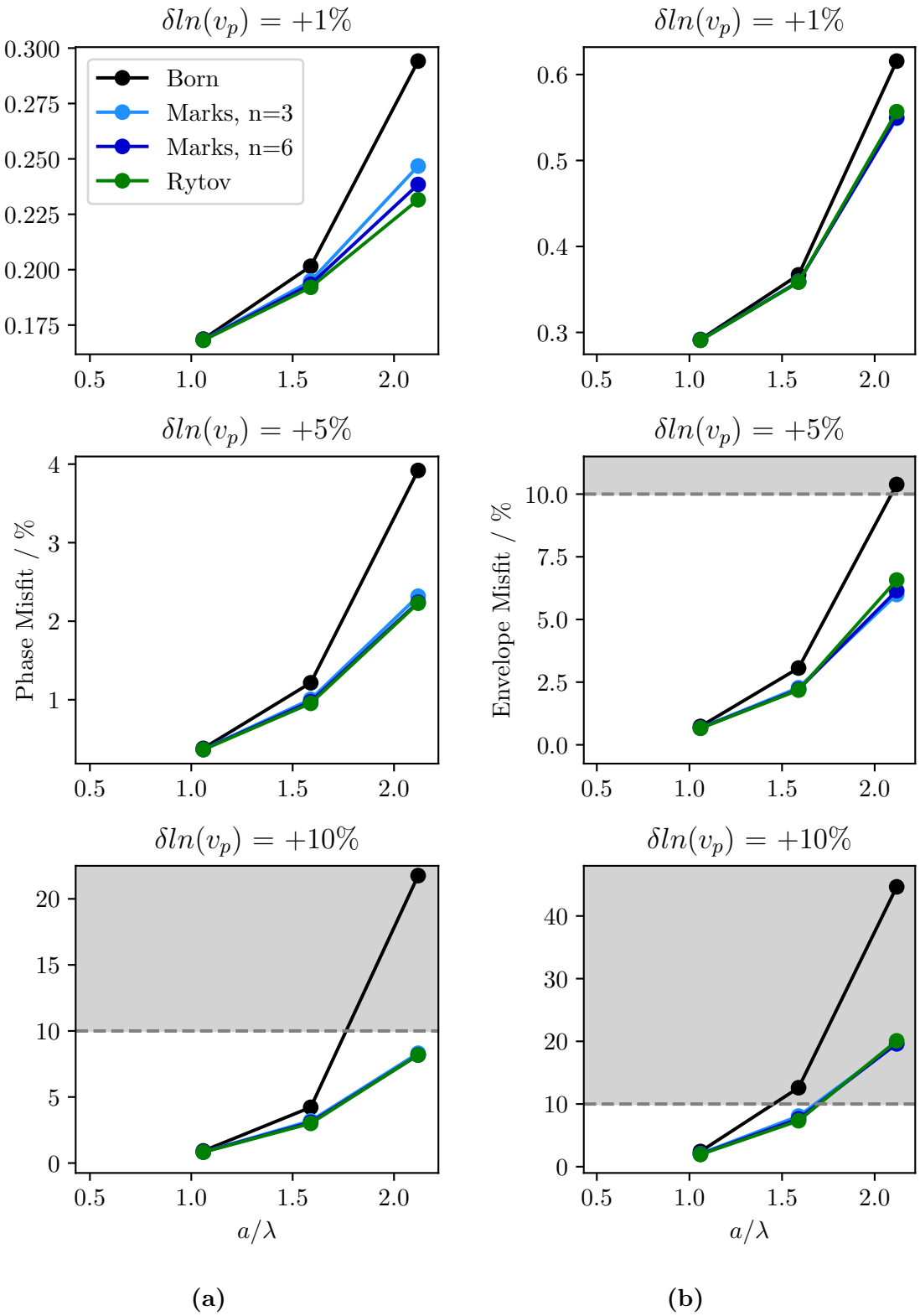


Figure 4.19: A summary of phase and envelope misfits between the reference AxiSEM3D solutions and the first order seismograms, for bubbles with positive perturbations in v_p. Misfits for Born seismograms are shown in black and Rytov in green. The misfits for two types of Marks seismograms corresponding to $n = 3$ and $n = 6$ are shown in light blue and dark blue respectively. Phase misfit is shown on the left in subfigure a) and envelope misfit is shown on the right in subfigure b). The grey shaded areas represent regions where the misfit is above 10% in which we consider the approximations to be invalid.

4.4.2 Perturbations in shear wavespeed

We now turn to Earth models in which the shear wavespeed, v_s, is perturbed inside spherical heterogeneities. The more comprehensive analysis performed above on perturbations in v_p are not reproduced, but four bubble models are selected with structure to wavelength ratios within the same range, corresponding to bubble diameters of 200 km ($a/\lambda = 0.67$), 400 km ($a/\lambda = 1.34$) and 500 km($a/\lambda = 1.67$). The results are shown in Figures 4.20 and 4.21.

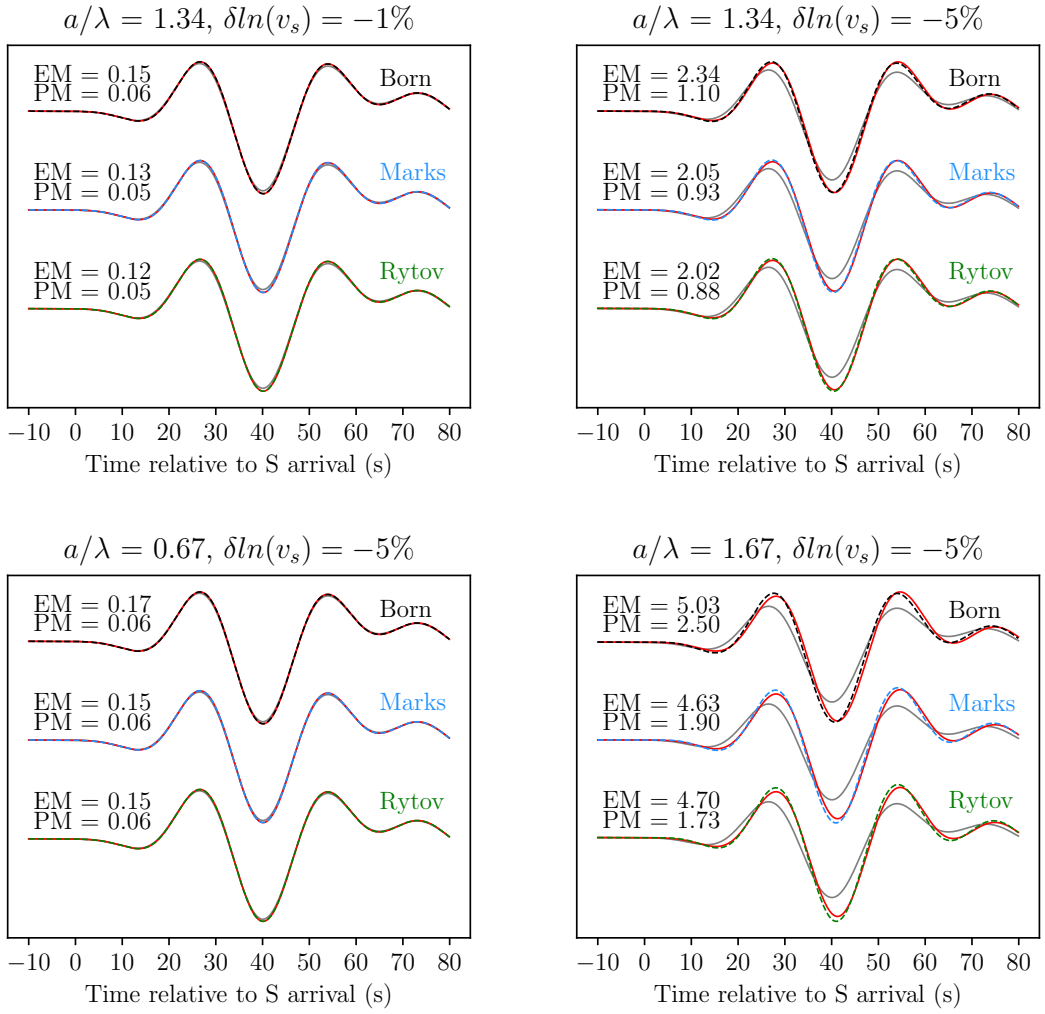


Figure 4.20: Comparison of AxiSEM3D seismograms (red) and three first order seismograms. The PREM waveform is shown in grey. Phase misfits (PM) and envelope misfits (EM) between the first order seismograms and the true solutions are presented as percentages. All seismograms are recorded at an epicentral distance of 60° as per the geometry in Figure 4.1a.

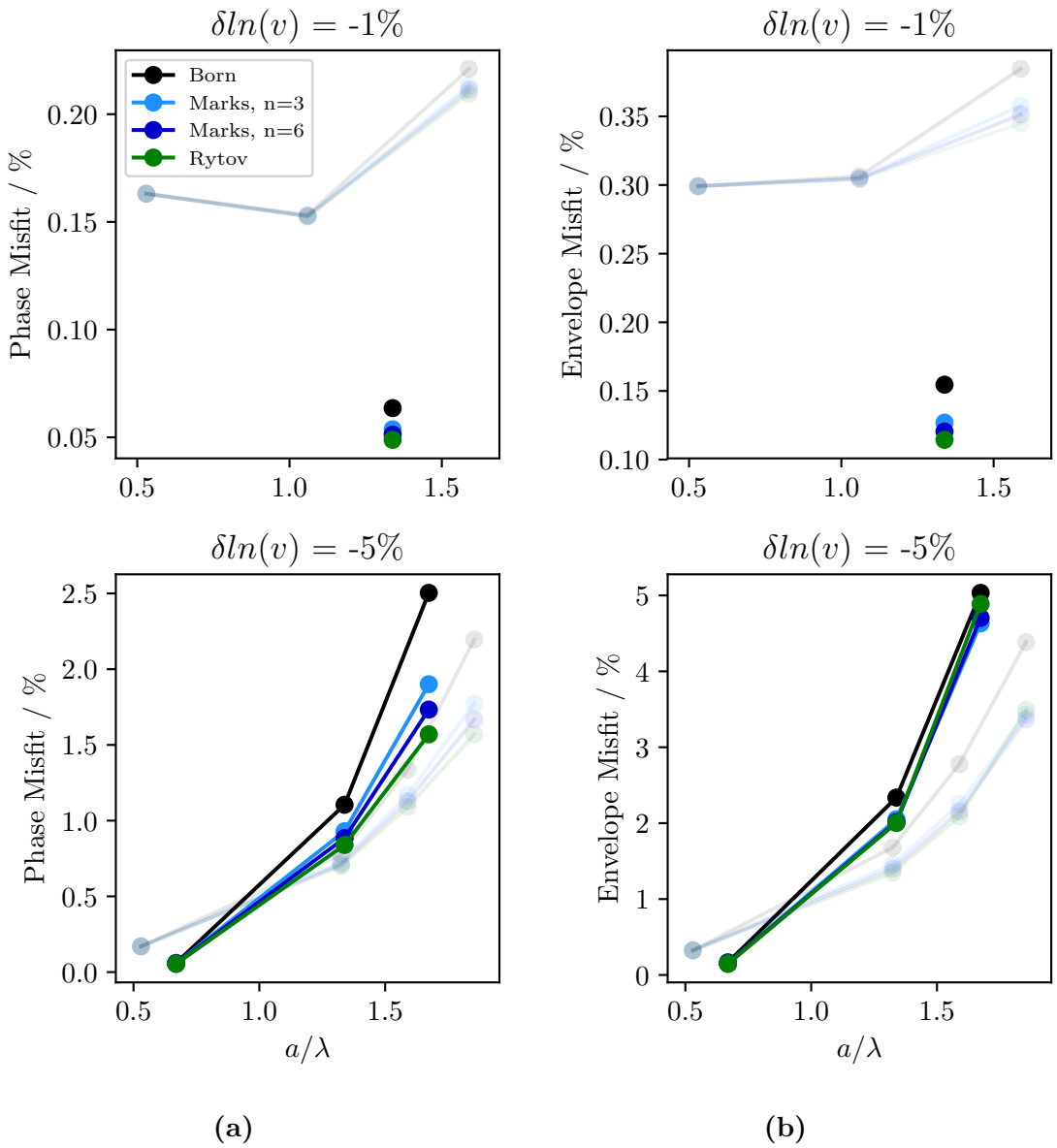
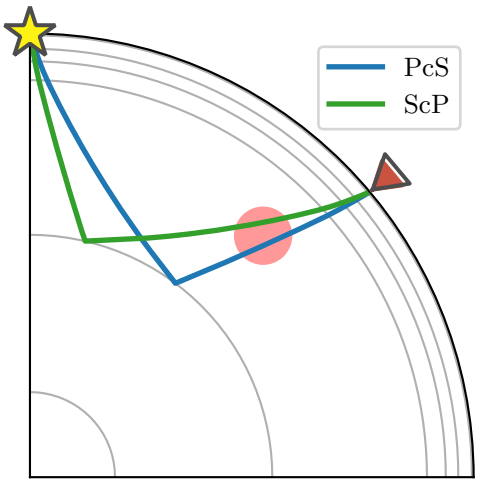


Figure 4.21: Plots showing envelope and phase misfits for bubbles with perturbations in v_s. The results of the analysis of P waves is shown faintly in the background for comparison. Each window has a fixed perturbation with structure to wavelength ratio on the x axis. The misfits for S waves are generally comparable to the misfits for P waves for the same perturbation.

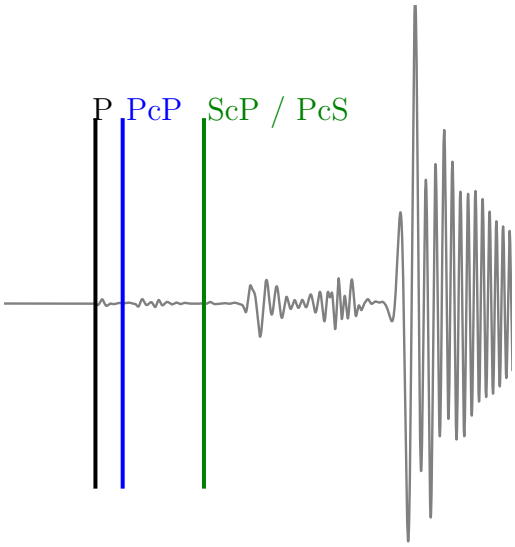
4.4.3 Converted phases and perturbations in both v_p and v_s

Until this point perturbations in only v_s or v_p have been considered, but realistic structures within the Earth display perturbations in both parameters. At the same time, we consider what happens when a wave is converted from a P to an S wave or vice versa at a discontinuity in the Earth. We therefore examine the arrival of

the phases PcS and ScP, which arrive at the same time, at an epicentral distance of 50°. A monopole source which does not generate S waves is used, meaning that the phase is dominantly PcS. The bubble model is located along the path of the PcS phase, has a diameter of 800 km and a perturbation in v_s and v_p of -1%. Figure 4.22a shows the set-up for this section. Figure 4.22b shows the Z component PREM seismogram, identifying the PcS/ScP phase. The comparison between first order and reference seismograms is shown in Figure 4.23 and there is good agreement, with Rytov performing best out of the three first order approximations.



(a) Geometry for testing the first order method on the PcS/ScP arrival. The source is an M_{zz} moment tensor (represented by the beach ball) and the receiver (red triangle) is 50° away. The bubble heterogeneity is plotted in red, and is 800 km in diameter and perturbed by -1% in v_p and v_s.



(b) Z component PREM seismogram identifying the ScP/PcS arrival.

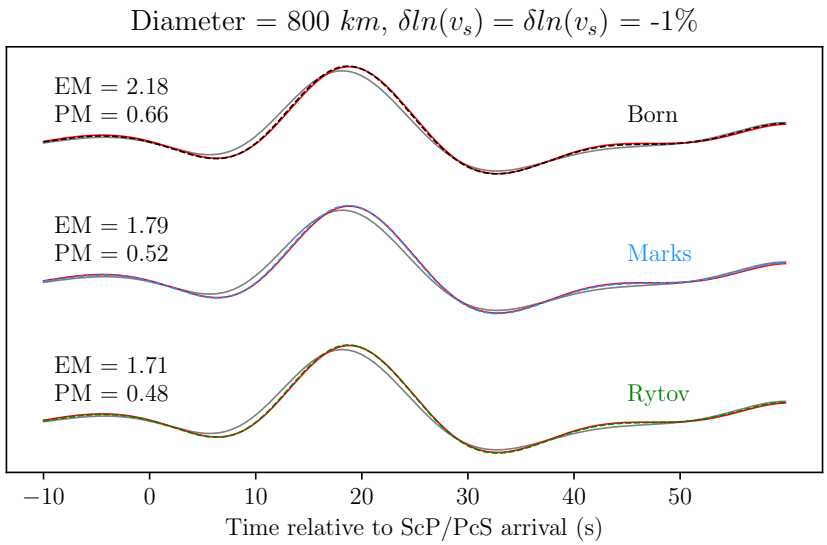


Figure 4.23: Comparison of first order and reference waveforms around the PcS/ScP arrival. The PREM seismogram is shown in grey for reference, and the AxiSEM3D seismogram which is taken as the true solution is shown in red. The Born, Rytov and Marks (with n=6) seismograms are plotted in black, green and blue respectively. Phase misfit (PM) and envelope misfit (EM) are displayed as percentages.

4.5 Discussion

4.5.1 1-D models

In the case of 1-D models, the effect of the perturbation is of a straightforward shift in arrival time of the P or S phase. This lends itself well to the linear perturbation in phase of the Rytov approximation, and indeed we see the Rytov seismograms matching the true solutions very well in this case. Referring to Figures 4.8 and 4.11, as the magnitude of the perturbation increases, EM and PM increase for all of the approximations, as expected. At the smallest value of $\delta ln(v_p)$=-0.1%, all three of the approximations have very similar values of EM and PM, reflecting how all of the approximations converge in the limit of very small ϕ. As the perturbations get larger, EM and PM of the three approximations start to diverge. Rytov consistently performs better, Born the worst and the Marks approximations between the two.

Figures 4.8 and 4.11 show that the first order seismograms match the true solutions more closely for positive perturbations. This could be due to Snell's law; the angle of

refraction at the 660 discontinuity is changed. By increasing the velocity below 660, the path in the lower mantle becomes shorter, as the angle of refraction increases. When the velocity below 660 is decreased however, the path becomes longer and the effect of the perturbation is therefore more significant. The first order approximations cannot take into account the effects of a heterogeneity on the path of a wave, as they are based on wave propagation through a PREM Earth. Also, decreasing the wavespeed has the effect of shortening the wavelength (if the period stays the same). Therefore for negative perturbations there are more wave cycles inside the perturbed material. This could also explain the misfit values for S being larger than P, as the S wave has a shorter wavelength in the lower mantle.

The first order method was tested on a 1-D model with a perturbation in v_p throughout the whole crust and mantle. The first order approximations gave inaccurate results in this case because of the kernel values in the region of the source and receiver. As discussed in Chapter 3, kernels used in seismic tomography are often masked in these regions, and I employ a similar strategy here by only considering perturbations below the 660 km discontinuity.

4.5.2 3-D models

Including a bubble perturbation in the lower mantle has a more significant effect on the shape of the waveform, to varying degrees depending on the strength of the bubble and its size relative to the wavelength. Negative bubbles delay and increase the amplitude of the arrival, while also shortening the period of the waveform. On the other hand positive bubbles advance the arrival of the phase, reduce its amplitude and broaden the waveform as if the wave were stretched by its interaction with the bubble.

Figures 4.19 and 4.18, which describe how perturbations in v_p affect the P phase, show that the PM values for every tested bubble model are lowest for the Rytov approximation, and the Born approximation has the highest PM values. The two Marks approximations ($n = 3$ and $n = 6$) have PM values in between Born and Rytov.

The measured EM values do not show such a consistent pattern as the PM values.

For weak perturbations of ±1%, the Rytov approximation consistently has the lowest EM values. However, for stronger perturbations and the largest bubbles either the Born or Marks approximations has the lowest EM values (for smaller bubble sizes, the Rytov approximation performs the best). It is not surprising that the Rytov approximation is outperformed by Born or Marks in these cases, as the Born approximation is based on perturbations to amplitude, and Marks is an intermediate approximation, they might be expected to match the amplitude better.

There are differences in the performance of the first order approximations between positive and negative bubbles. The negative bubbles seem to show a cross-over in the best performing model in terms of EM from Rytov to Born via the Marks approximation. Meanwhile for the positive bubbles, the Born approximation consistently predicts the amplitude of the waveform the worst, while the Marks approximation outperforms Rytov in some cases. I note that the measured misfit values are not absolute (for example, they are sensitive to the chosen time window), and so I do not put emphasis on very small differences such as those between the Marks and Rytov approximations for a +1% bubble. It is apparent that the performance of the approximations is dependent on the effect of the perturbation on the waveform, which differs significantly between positive and negative bubbles. For the 1-D models, positive perturbations consistently had lower misfits than negative perturbations, but that is not the case for the 3-D bubble models where the effect on the waveform is more profound.

Previous studies on the applicability of the Born approximation for waveform modelling have looked at surface waves. *Parisi et al.* (2015) perform the most comprehensive tests, and find that the Born approximation does not sufficiently model phase delay once the delay exceeds 15%, and that it does not sufficiently model the amplitude once the phase delay exceeds 5%. Although I do not analyse a high enough number of models to generalise to a rule of thumb in an analogous way, I compare our results for 3-D bubble models in $\delta ln(v_p)$. The phase misfit as measured between the reference AxiSEM3D seismogram and the PREM seismogram (PM_{Axisem3D}) is taken as a percentage measure of phase delay relative to period.

I first discuss the case of positive bubbles. Figure 4.24c shows that a good fit in amplitude (EM $< 10\%$) is obtained where $\mathrm{PM}_{\mathrm{AxiSEM3D}} < 4\%$ for the Born approximation and around 8% for the Rytov approximation. Figure 4.24d shows that good fit in phase (PM $< 10\%$) is obtained where $\mathrm{PM}_{\mathrm{AxiSEM3D}} < 8\%$ for the Born approximation. For the Rytov approximation, all of our tested bubble models fit the phase to within a PM of 10%.

For negative bubbles, the first order approximations perform better. A good fit in amplitude is obtained where $\mathrm{PM}_{\mathrm{AxiSEM3D}} < 8\%$ for both the Born and Rytov approximations (Figure 4.24a). Figure 4.24b shows how a good fit in phase is obtained where $\mathrm{PM}_{\mathrm{AxiSEM3D}} < 15\%$ of the period for the Born approximation. Again for the Rytov approximation, all of our tested bubble models fit the phase well. Therefore, for the Born approximation, our results are comparable to those of *Parisi et al.* (2015). I show that the Rytov approximation has a considerably larger domain of validity, particularly in terms of phase prediction.

Although this discussion has focussed on P waves, a similar pattern is observed for S waves. Figure 4.21 shows how the misfits for S waves are generally comparable to the misfits for P waves for the same perturbation. The small differences may be caused by the different paths of S and P waves through the bubble or inaccuracies in period determination as well as the fundamental differences between compressional and shear wave motion. In the case of a very weak perturbation at $\delta ln(v_p) = -1\%$, the misfits appear to be much lower for S waves. However all values are extremely small (less than 0.4%) so the differences are not significant. I have also shown that the first order method works just as well for the converted PcS phase as for direct P and S.

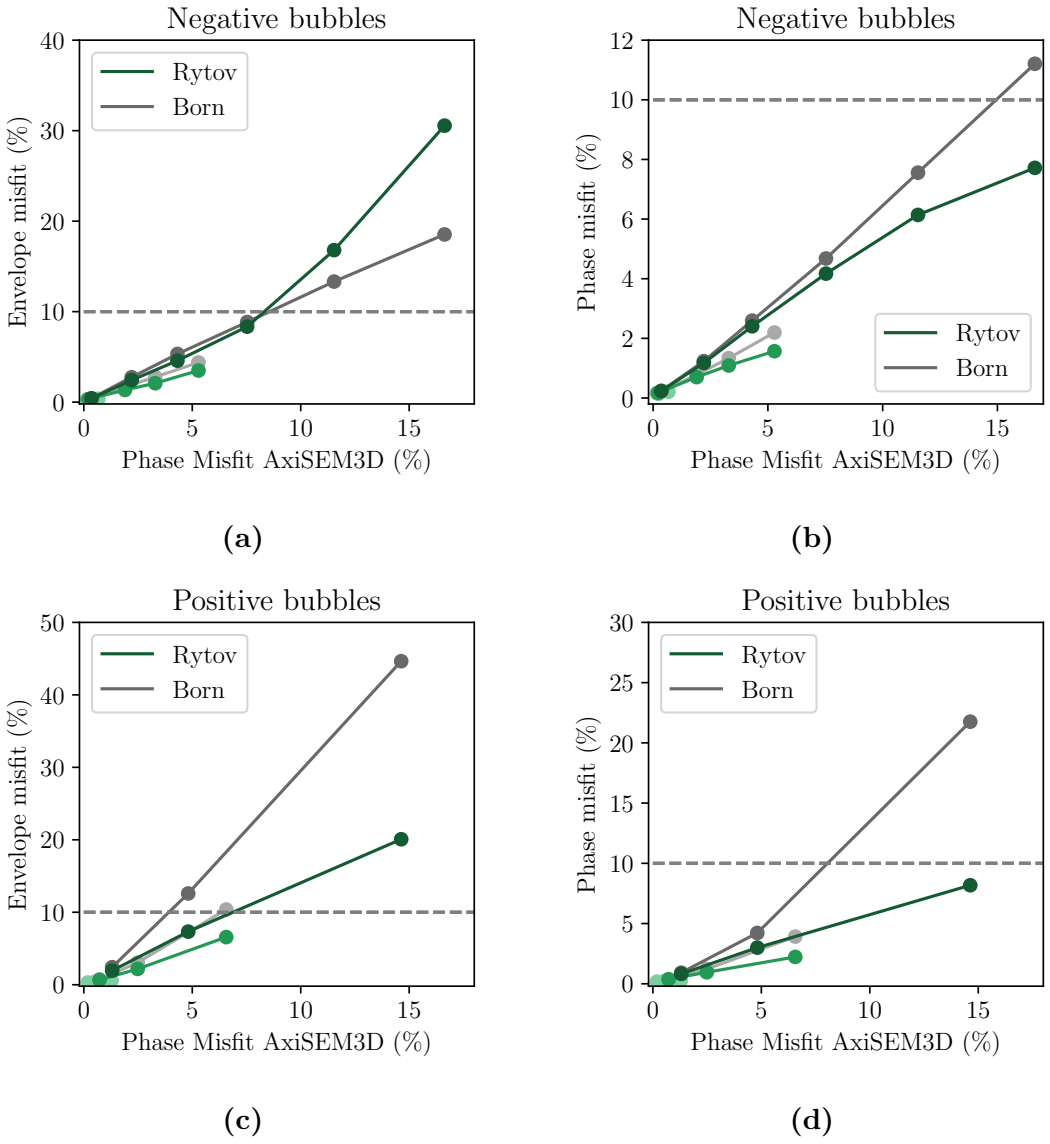


Figure 4.24: Envelope and phase misfits against AxiSEM3D phase misfits. Envelope and phase misfits are the percentage misfit between the Born or Rytov seismograms and the reference true solutions. AxiSEM3D phase misfit is the phase misfit between AxiSEM3D seismograms and the PREM seismogram, and is a percentage measure of the phase delay relative to the period. Values for Born seismograms are plotted in grey, with one different shaded line for each velocity perturbation and multiple sizes. Values for Rytov seismograms are plotted in green, with one different shaded line for each velocity perturbation and multiple sizes. Darker colours correspond to larger magnitudes of velocity perturbation.

4.6 Conclusions

In this chapter, the performance of first order approximations was tested for 1-D models based on perturbing PREM below the 660 *km* discontinuity, and 3-D models with

spherical perturbations in the lower mantle for direct P and S waves as well as converted phases. Previous studies have examined the performance of the Born approximation for waveform modelling of surface waves, but this is the first work to apply first order approximations to body waves. We obtain comparable results for the Born approximation, and furthermore we have extended beyond the Born approximation to the Rytov and Marks approximations.

The validity of first order approximations is dependent on the size and strength of the perturbation, and differs depending on the value of n (Born, Rytov or Marks). The approximations are considered to be valid where both phase and envelope misfit values are 10% or lower. I can conclude that in all cases the Rytov approximation most closely matches the reference SEM solutions, after balancing phase and envelope misfit values, and is therefore the most applicable to waveform modelling of body waves that sample the lower mantle. In our 3-D tests, where structure to wavelength ratios up to 2.12 and wavespeed perturbations up to $\pm 10\%$ are considered. The Rytov approximation is applicable for perturbations with strengths up to $\delta ln(v) = \pm 10\%$, where the size of the perturbation is below $\approx 1.6\lambda$. For larger sizes, the wavespeed perturbation should be below $\pm 5\%$. Where changes to the wavespeed are small (I have examined up to $\pm 1\%$), the whole of the lower mantle can be perturbed with almost perfect agreement between Rytov and SEM solutions.

In the next chapter, we move on to applying first order approximations to more realistic lower mantle structures including tomography models and LLVP-type structures.

Chapter 5

Pathway towards statistical probing of lower mantle structures with waveforms

5.1 Introduction

In this chapter, the applicability of the first-order method to more realistic lower mantle seismic structures is demonstrated.

Images produced from seismic tomography provide an approximation of Earth structure. These tomographic models differ in a significant way from the single spherical perturbations tested previously as there are many patches of slow and fast material next to each other. In the first section of this chapter, we benchmark our first order method against AxiSEM3D synthetics for the tomography model S40RTS (*Ritsema et al.*, 2011).

Where smaller scale seismic structure are investigated, we look for a level of detail which cannot be provided by seismic tomography. For example, the inherent damping and smoothing of seismic tomography means that a sharp edge of an LLVP will not be resolved. We therefore move away from seismic tomography models to a simple shape with some of the proposed features of the LLVP under Africa.

We use our first order method to produce synthetic seismograms for a large number (500) of variants on the LLVP model, demonstrating that the method can be used to produce synthetic seismograms for large numbers of models very quickly. This opens the possibility of finding a favoured model, or set of models, using a parameter space search. We create a database of synthetic waveforms which may then be probed, depending on features of the observed data. We limit our model to perturbations in v_p.

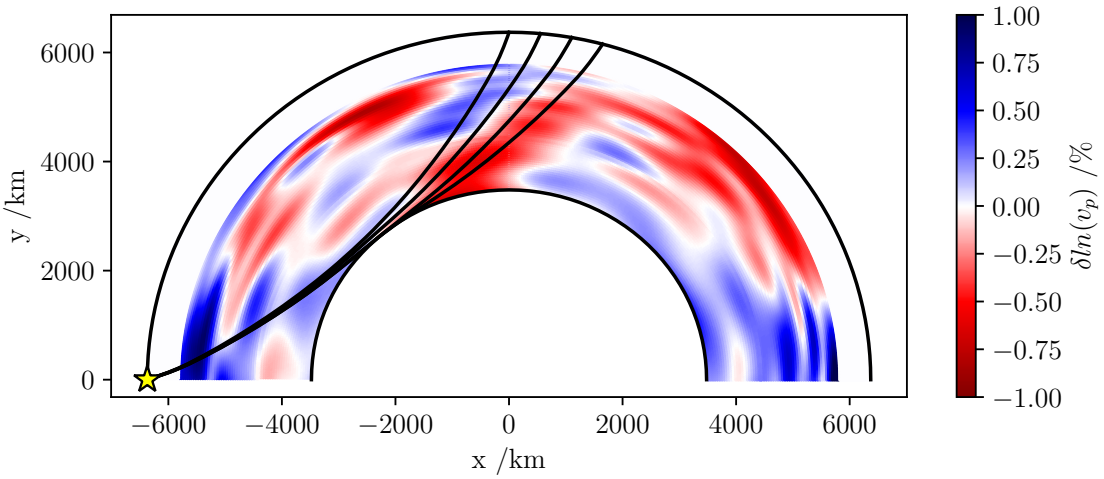
5.2 Tomography model benchmark

We generate synthetic seismograms for the tomography model S40RTS (*Ritsema et al.*, 2011) using AxiSEM3D and our first order method, again computing Born, Rytov and Marks ($n = 3$ and $n = 6$) seismograms. We make perturbations only in v_p, and examine the Z component seismogram around the P or Pdiff arrival. We use a moment tensor source with only an M_{zz} component. The waveforms are simulated with a Gaussian source-time function with a dominant period of 20 s, and they are filtered with a 20s-100s bandpass filter. The wavelength in the lower mantle is therefore approximately 377 km. The mesh used for this test is the 100 km mesh shown in Figure 3.4b. The geometry is based on real source-receiver configurations, data from which has been used to investigate the western edge of the LLVP. The source is in South America and stations are located in southern Africa. The four stations selected span the edge of the LLVP in the lower mantle.

The model S40RTS was produced using travel times of teleseismic phases which pass through the mantle as S waves, as well as normal mode data, and is defined in v_s. The v_p model is defined using a scaling of $\delta ln(v_s)/\delta ln(v_p)$ which increases linearly with depth from 2 at Earth's surface to 3 at the core-mantle boundary . Due to source and receiver effects, as with previous models, we define the model only below the 660 km discontinuity.

Figure 5.1a shows the source and receiver geometry, and the resulting waveforms

are displayed in Figures 5.1b - 5.1e. The waveform differences induced by the tomography model are very small, and all three first order approximations fit the reference AxiSEM3D synthetics well, though the Rytov approximation consistently performs best. At the shortest distance of 90° the waveform is shifted in time and arrives early, due to the predominantly fast material the wave travels through. At larger distances, the Pdiff phase samples the slow LLVP. However, its effect is barely perceptible as it is counteracted by fast material elsewhere.



(a) Cross section through S40RTS in the slice defined by the source (yellow star) and receivers. The TauP calculated ray paths of the P or Pdiff phase are shown as black lines.

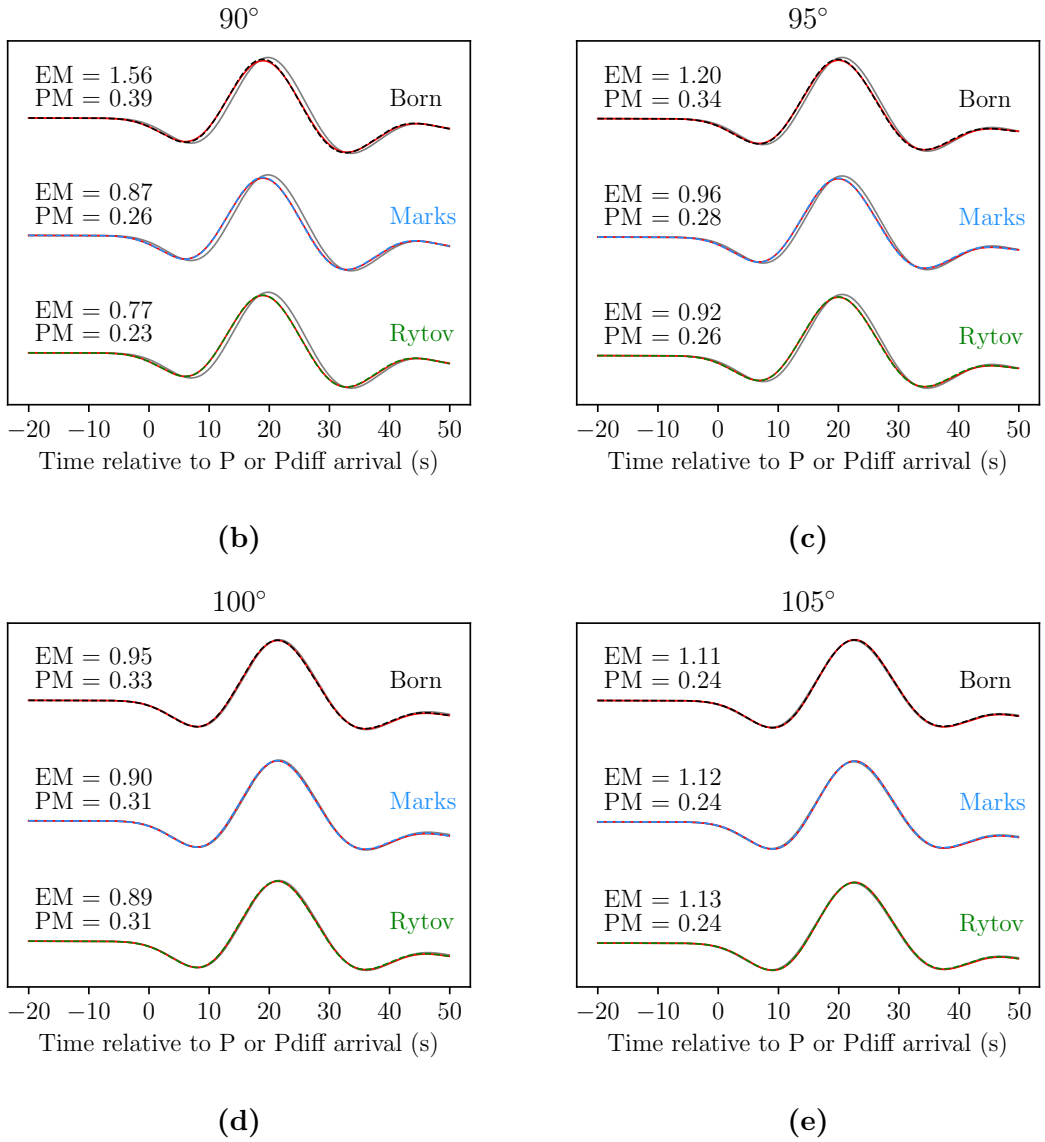


Figure 5.1: First order results for the tomography model S40RTS perturbed below 660 *km*. (a) shows the model and ray paths. (b) - (e) are P or Pdiff waveforms, calculated using the first order method and AxiSEM3D (red). PREM seismograms are shown in grey. Born is shown in black, Marks (n=3) in blue, and Rytov in green dashed lines. Envelope misfit (EM) and phase misfit (PM) measurements are shown to the left of each seismogram in percent.

5.3 Synthetic study of the LLVP under Africa

We produce synthetic seismograms for a set of models defined in $\delta ln(v_p)$. The shape of our model is based on that proposed by *Ni et al.* (2002), with its sloping parallel sides. The velocity perturbation is based on tomography models, in which $\delta ln(v_p)$ is typically between -1% and -2% (*Koelemeijer et al.*, 2016; *Hosseini et al.*, 2019; *Tesoniero et al.*, 2015). Waveform modelling studies also suggest wavespeed reductions of similar magnitide, for example *Wang and Wen* (2007a) use P and PcP travel time modelling to suggest $\delta ln(v_p)$ of between -0.7% and -1.7% in the African LLVP structure.

For simplicity, we position the source at the North pole and we use a simple M_{zz} component source. We use higher frequency waves here than we have done previously in order to capture more detail of the wave-structure interaction. The source-time function is Gaussian with a dominant period of 10 s, and we apply a 10 - 100 s bandpass filter to the seismograms. Frequency analysis around the first (P or Pdiff) arrival reveals the period to be around 17.7 s. This corresponds to a wavelength in the lower mantle of around 266 km.

We limit our parameter space by perturbing the model only in v_p, and we examine Z component seismograms. This set-up is a simplified version of one which may be used to directly compare to observed data. In that case, we would be required to use a more realistic moment tensor source and source-time function, as well as taking into account perturbations in v_s and more complexity in the structure. This is all possible with the first order approach, but a simple example is provided in this chapter.

5.3.1 Model Parameterisation

The LLVP structure is shown in Figures 5.2 and 5.3. We define four parameters, which are varied systematically. These are:

- **Height:** The radial distance in km from the core-mantle boundary to the top of the structure. Ranges from 1200 to 1700 km, or around 4.5 - 6.4 wavelengths.

- **Strength:** The largest reduction in $\delta ln(v_p)$ in the structure. Ranges from -1% to -2%.

- **x:** The distance in km from the edge of the structure, over which $\delta ln(v_p)$ changes linearly from 0 to its largest reduction in wavespeed (the strength). This is a proxy for the blurriness of the boundary of the structure. Ranges between 0 - 800 km, or 0 to approximately 3 wavelengths.

- **Extent in longitude (abbreviated to extent):** Angular extent of the structure in the azimuthal dimension. Ranges from 15 - 45°.

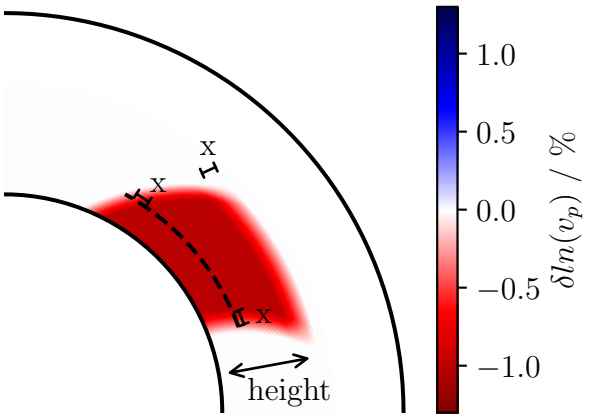


Figure 5.2: A slice through the LLVP model in the $x - z$ plane, defining the parameters height and x. In this example, x is 400 km and height is 1700 km.

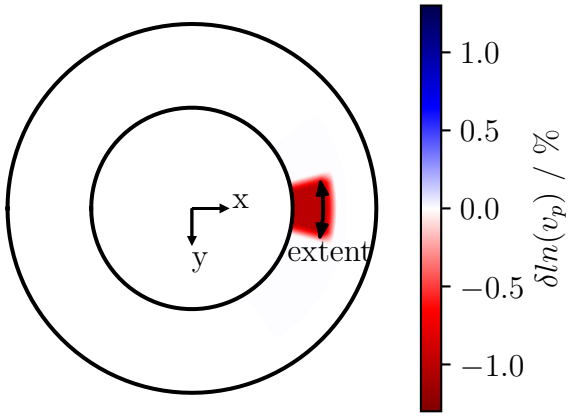


Figure 5.3: A slice through the LLVP model in the $x - y$ plane, viewed looking down the z axis. The parameter extent is defined. It can be seen that in the azimuthal direction, the edge is also blurred with the same linear gradient of strength/x (%/km).

5.3.2 AxiSEM3D synthetics

We begin by simulating synthetic seismograms using AxiSEM3D, and three variants of the LLVP model. This requires one wavefield calculation per model for many stations, enabling us to use a dense array of stations in this first instance in order to select stations for the subsequent first order study. We also benchmark the first order method against the AxiSEM3D seismograms to confirm that our method is applicable to this structure.

The source-receiver geometry for the AxiSEM3D simulations is shown in Figure 5.4. The selection of models for which AxiSEM3D synthetics are calculated are:

1. Height=1700 km, extent=30°, strength=-1%, x=0 km
2. Height=1700 km, extent=30°, strength=-1%, x=800 km
3. Height=1700 km, extent=30°, strength=-2%, x=0 km

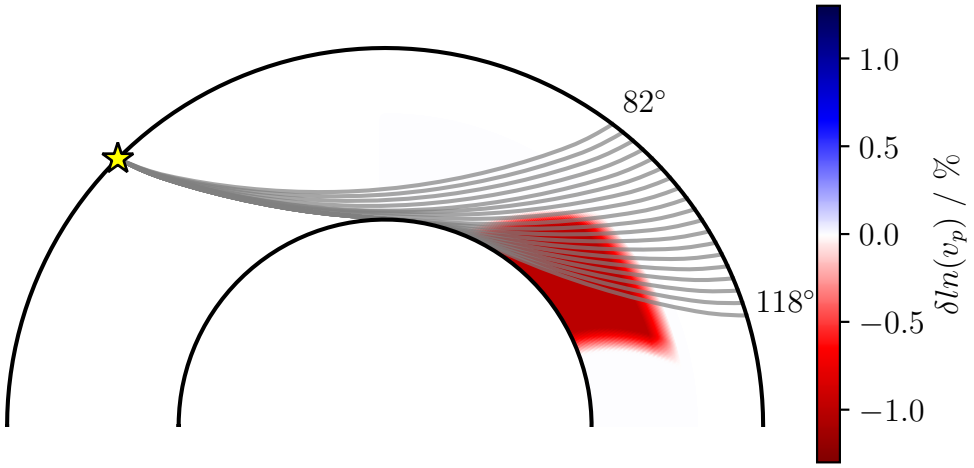


Figure 5.4: Source-receiver geometry for AxiSEM3D runs. This is a slice in the $x-z$ plane with the source represented by a yellow star on the North pole. All of the receivers are located at a longitude of 0° with epicentral distances from 82° to 118°, spaced every 2°. The TauP predicted ray paths for the P and then Pdiff phase are shown for each station.

In Figure 5.5, we present a record section highlighting the differences between PREM seismograms and those calculated using the first of the LLVP models listed

above. Differences between unperturbed and perturbed seismograms around the first P or Pdiff arrival start to become apparent at around 96°. At larger distances, differences around the PP arrival begin to appear. Both of these observations are consistent with the plotted ray paths for these phases in Figure 5.4. Differences also emerge at a later time for distances above 88°. This time period is near the arrival time of SKS; although there is not a clear phase arrival it is likely to be SPdKS or SKPdS.

We focus on the differences around the first P or Pdiff arrival. The seismograms in this time window for all three of the models listed above are plotted in Figures 5.6 and 5.7. The waveforms are plotted against both epicentral distance and longitude. We can see from these figures that the main difference between the perturbed and unperturbed seismograms for this range of models is the travel time of the phase. There are also some differences in amplitude and in the shape of the waveform, particularly at epicentral distances of around 98 - 106°, and at longitudes of around 8 - 12°. These stations correspond to interaction of the Pdiff phase with the edge of the LLVP. As expected, the effects on the seismogram are more pronounced in Figure 5.7, where the perturbation is stronger.

Figure 5.8 shows the travel time differences between the perturbed and unperturbed seismograms as measured by cross-correlation in the window around the P/Pdiff arrival. The window is selected to be 25 s before and 40 s after the TauP predicted P/Pdiff arrival. The cross-correlation time is typically a delay, with a positive value corresponding to a late arrival of the phase. Each of the travel time delay curves has a different shape, which is something that is noted here and investigated further in the subsequent sections.

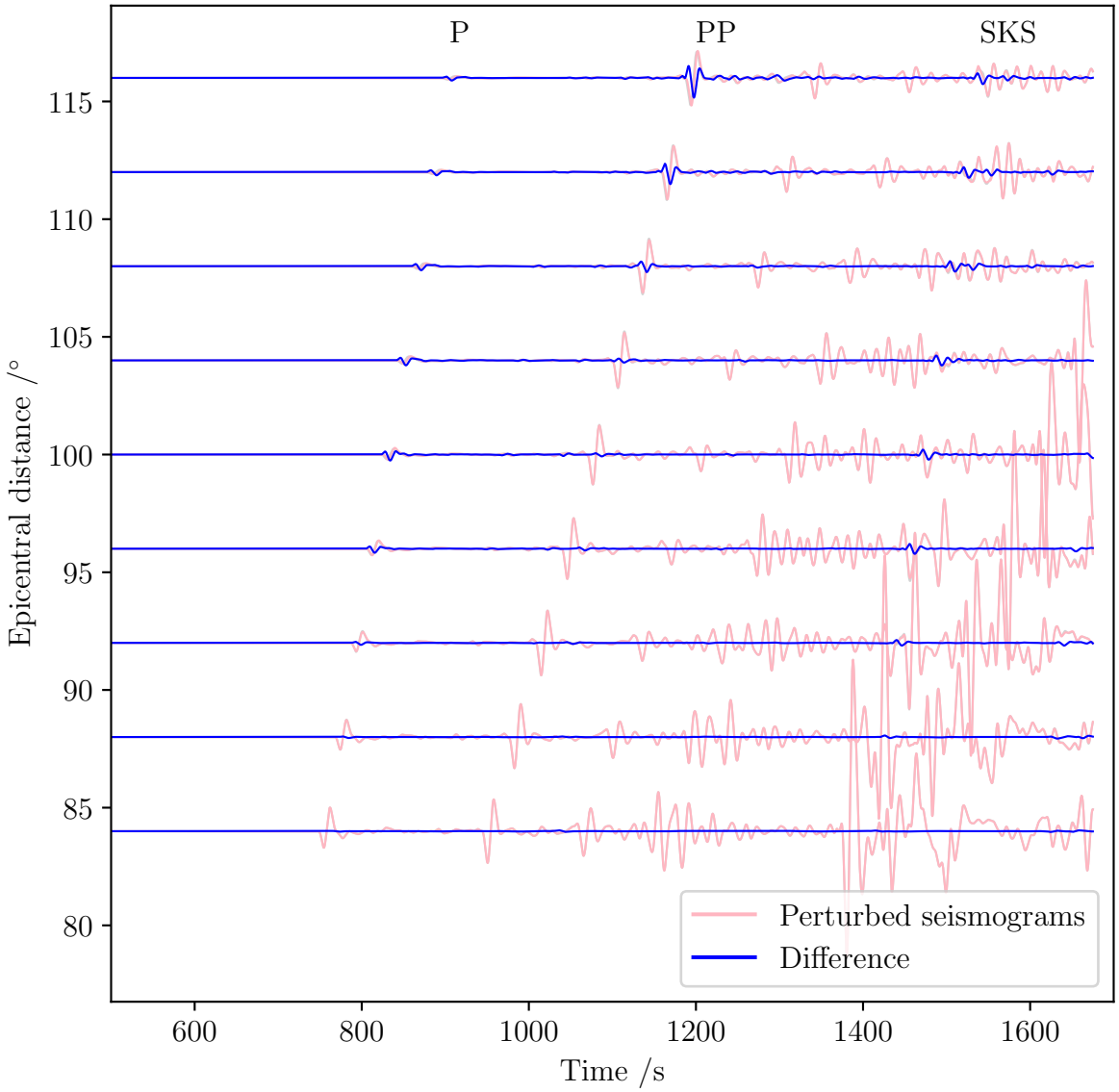


Figure 5.5: Record section showing the AxiSEM3D seismograms at alternate stations (every 4°) along the Greenwich meridian line of longitude=0°. Seismograms corresponding to an LLVP model with height=1700 *km*, extent=30°, strength=-1%, x=0 *km* are shown in light red. The direct waveform difference between the perturbed and unperturbed seismograms is plotted in blue. The same normalisation is used for each of the blue lines plotted, which is why the magnitude is reduced at larger distances.

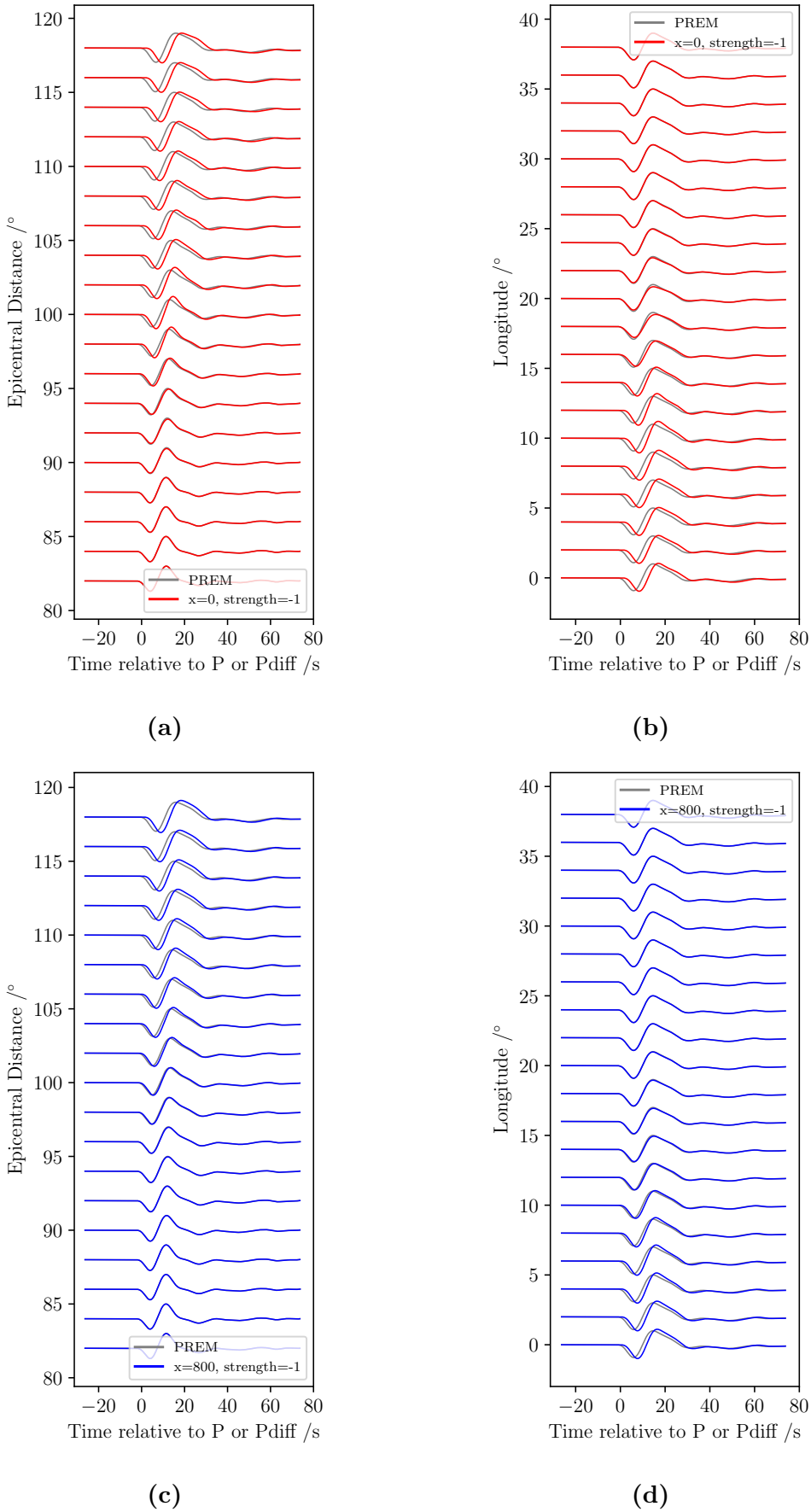


Figure 5.6: AxiSEM3D record sections, aligned on the P or Pdiff arrival for models with a strength of -1%, height of 1700 *km*, extent of 30°. x=0 *km* in (a) and (b) and x=800 *km* in (c) and (d). At each distance or longitude, the PREM and perturbed seismogram are both normalised by the absolute maximum displacement in the PREM seismogram, so that differences in amplitude are apparent.

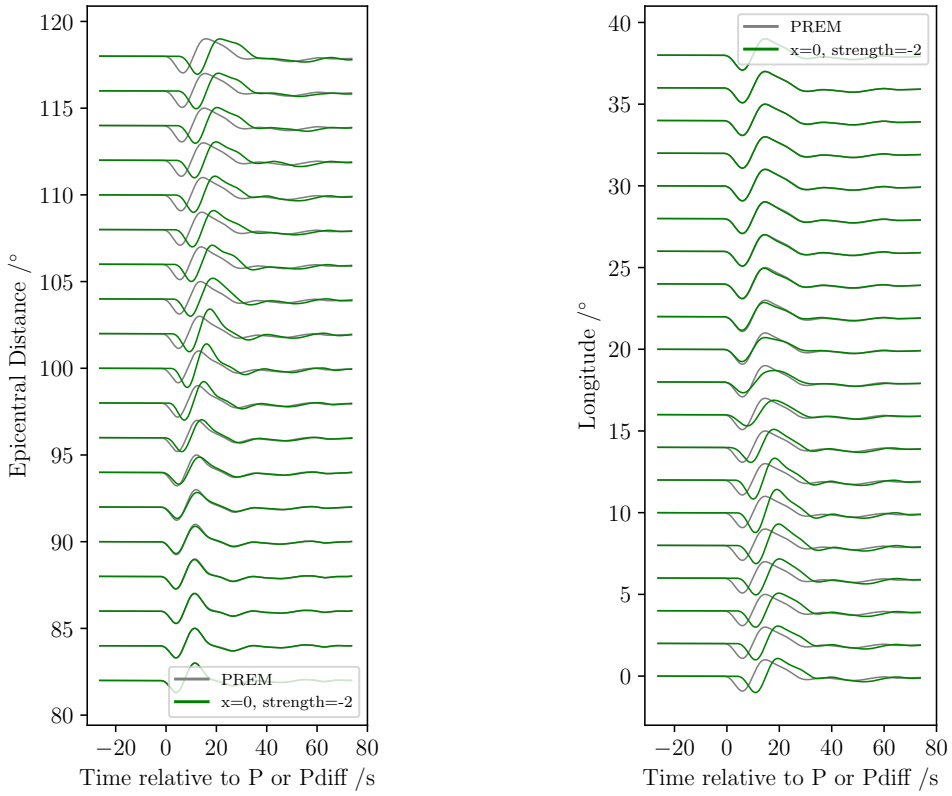


Figure 5.7: AxiSEM3D record sections, aligned on the P or Pdiff arrival for models with a strength of -2%, height of 1700 *km*, extent of 30° and x=0 *km*.

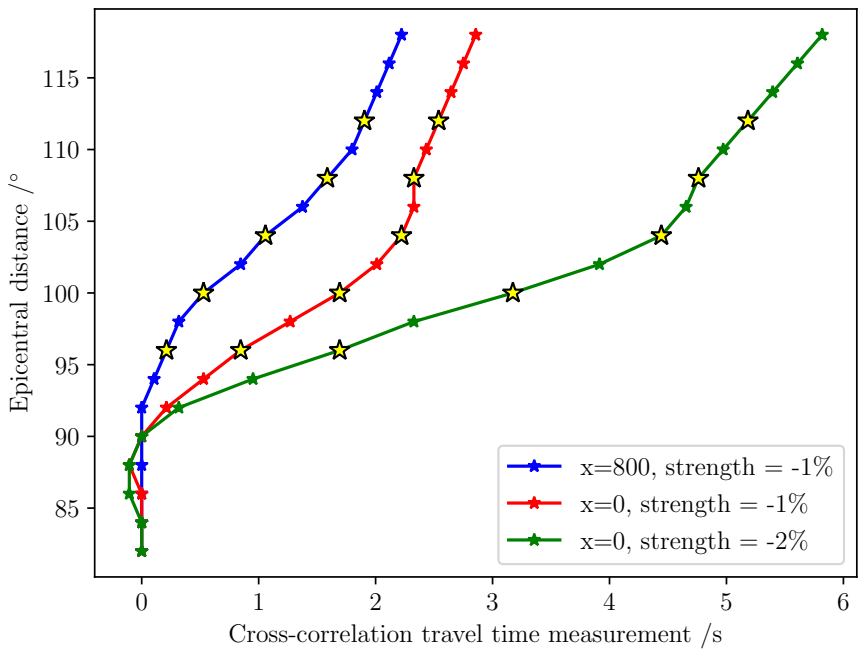


Figure 5.8: P or Pdiff travel time delay with respect to PREM as measured by cross-correlation, against epicentral distance for stations along the Greenwich meridian line at longitude=0°. Each line corresponds to one model and one run of AxiSEM3D. The measured delay at each station is represented by a star. The larger yellow stars correspond to stations which are selected for the first order study in the next section.

5.3.3 Benchmarks

We demonstrate here that the Rytov approximation is applicable to the LLVP structures by benchmarking against the AxiSEM3D seismograms of the previous section. Figure 5.9 compares AxiSEM3D and first order seismograms for the three models defined in Section 5.3.2, at a distance of 104°. For all of the models, the Rytov approximation performs the best, as expected given the previous results. For the model with the most significant heterogeneity of x=0 *km* and strength of -2%, the phase misfit value is still very low at 2% while the envelope misfit value is at the edge of what we have defined to be a good fit.

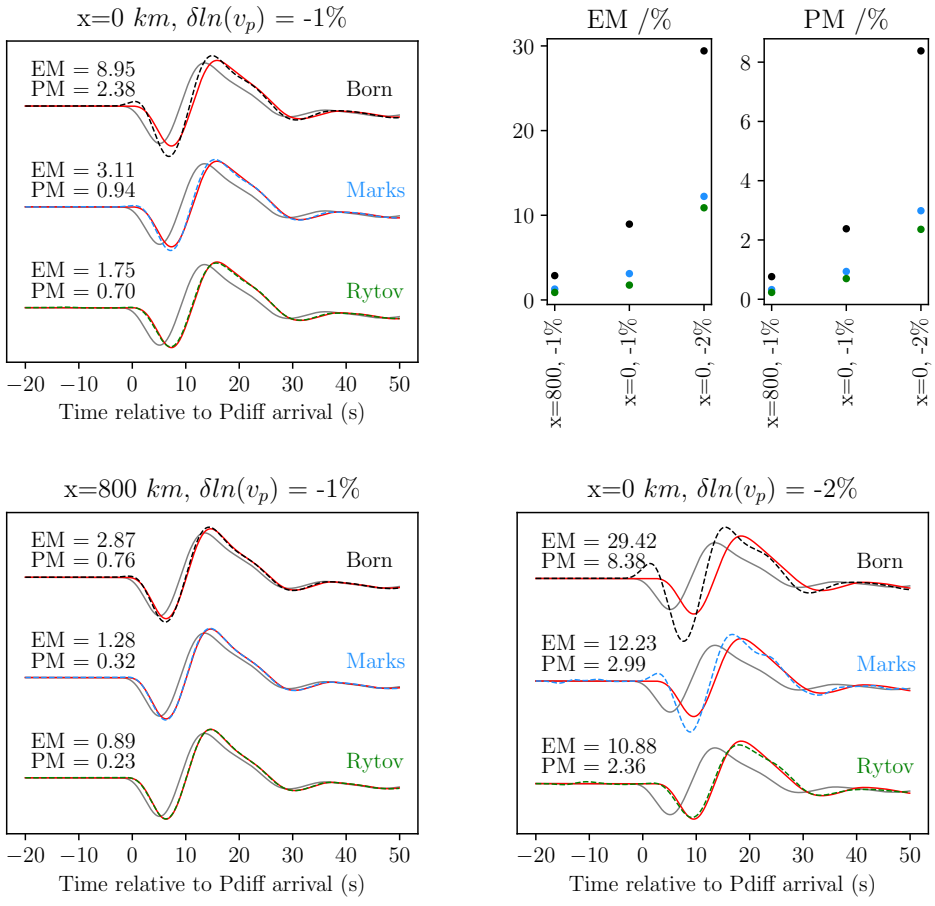


Figure 5.9: Pdiff waveforms at a distance of 104° comparing reference AxiSEM3D solutions (red lines) to the first order seismograms (Born in black, Marks(n=3) in blue, and Rytov in green). The PREM seismogram is shown in grey. Envelope misfit (EM) and phase misfit (PM) are given for each approximation in percent. A summary of EM and PM for each method is shown in the upper right panel with black (Rytov), blue(Marks) and green (Born).

5.3.4 Rytov seismograms for many models

Set-up

For the first order study, we select five epicentral distances based on Figures 5.8, 5.6 and 5.7. These are 96°, 100°, 104°, 108°, 112°. The Pdiff ray paths and source-receiver geometry are depicted in Figure 5.10. We also examine the changes to the seismogram with longitude. Since we use an axisymmetric M_{zz} source, the same waveform kernel can be used for any longitude with a simple rotation around the z axis of the model. We select longitudes of 0°E, 5°E, 10°E and 15°E. The interaction between the waveform kernel for a station at 0°E a distance of 104° from the source and an LLVP structure is shown in Figure 5.11.

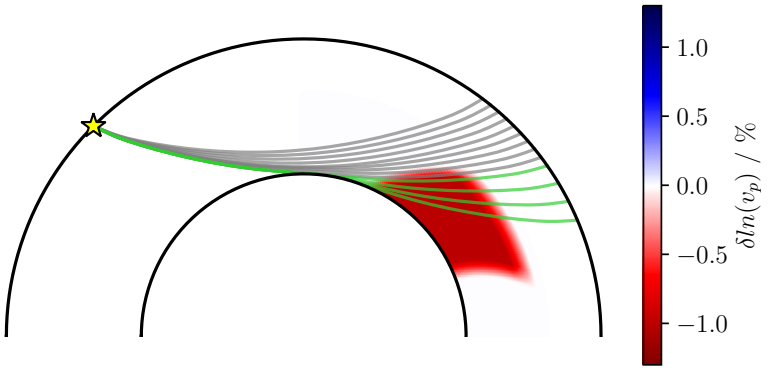


Figure 5.10: Geometry for the computation of first order seismograms showing the source (yellow star) and P/Pdiff ray paths in green to four stations along the Meridian line of 0°E.

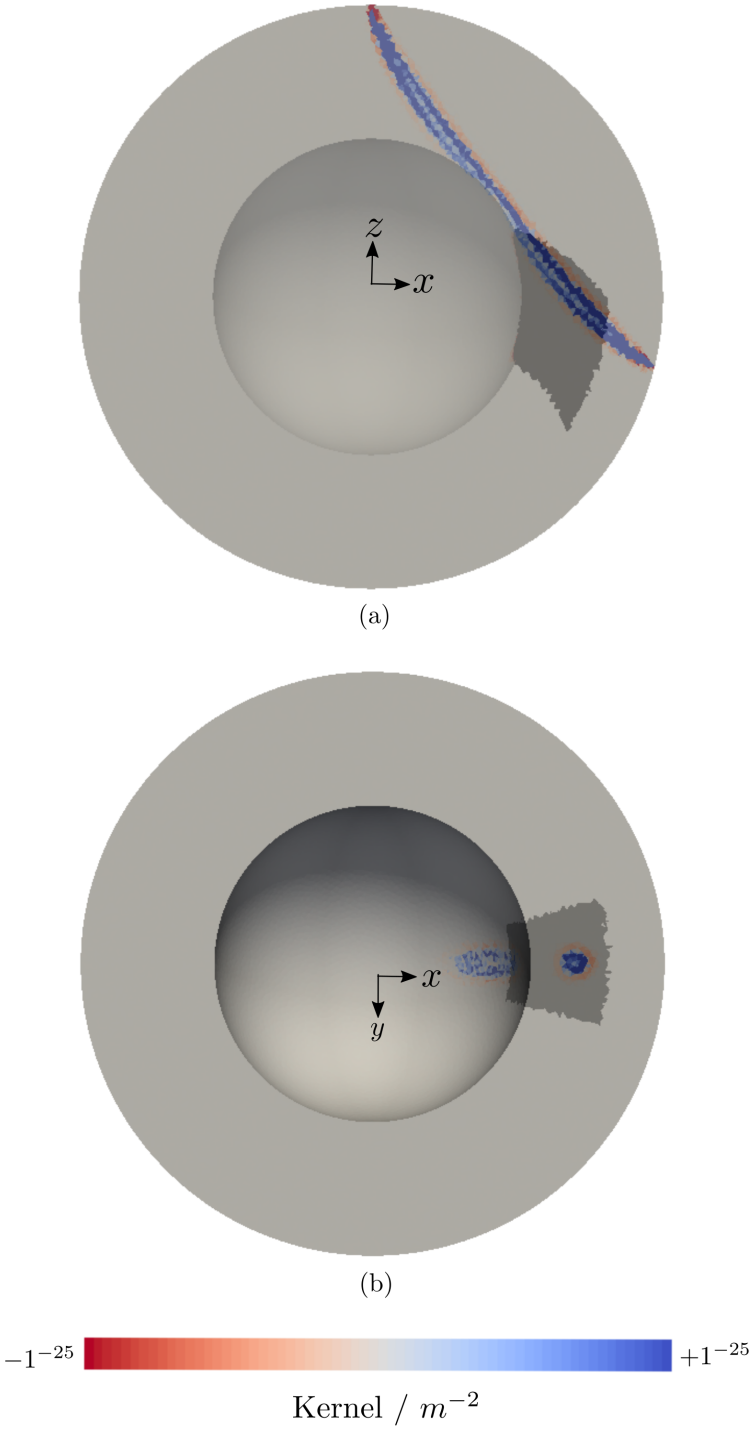


Figure 5.11: The interaction between waveform kernel and structure at a time of 842 s, at the onset of the Pdiff arrival. The source is at the North Pole and the station is located 104° away along the Greenwich meridian line. The LLVP structure with x=0 km, height=1700 km, strength=-1%, and extent=30° is shown in dark grey, projected onto the mesh. The waveform kernel is plotted on top in colour. (a) is a slice in the $x-z$ plane, viewed down the y axis. (b) is a slice in the $x-y$ plane.

Workflow

The first order method for computing Rytov seismograms is summarised in Figure 5.12, and is implemented in Python2. Outside of the first order calculation, we compute the AxiSEM wavefields (which give us the unperturbed seismogram) and the waveform kernel. Computation of Rytov seismograms involves first finding δu under the Born approximation by integrating the waveform kernel over the background model. We find the mesh elements where the model is non-zero and we only perform the integration over these elements. This is performed once for the model with the largest spatial extent. The waveform kernels on the non-zero elements are loaded in and we loop over models to compute the first order Rytov seismograms.

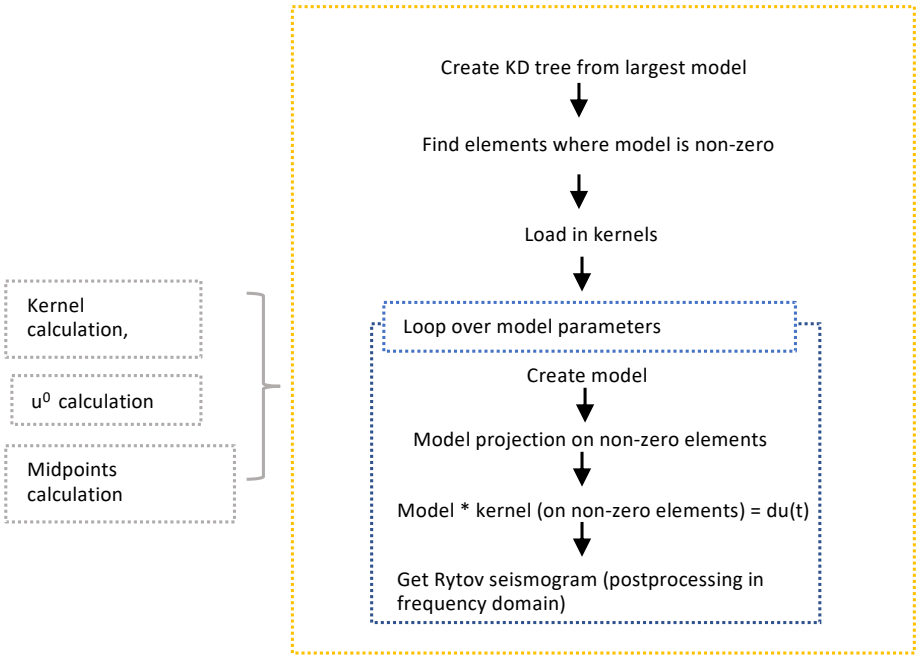


Figure 5.12: Flowchart showing the computation of many models at multiple stations. The overheads are carried out first and involve finding the elements in which the model is non-zero. Subsequently, the integration over volume is performed only over the non-zero elements.

Computational costs

The computational costs for the steps in generating synthetic seismograms under the Rytov approximation are shown in Table 5.1, for one source-receiver pair. The steps are subdivided into 'pre-computation', 'model-type specific' and 'model iterations'.

The pre-computation steps involve calculating the wavefield, the kernel, and the midpoints of the mesh. The wavefield calculation is specific to the earthquake depth and defines the time sampling of the kernel. The kernel calculation is specific to the mesh and earthquake source. The time given in Table 5.1 is for one receiver, for a seismogram length of 1800 s. The calculation of midpoints is currently performed outside of MC Kernel and is independent of the source-receiver geometry, so can be applied wherever the same mesh is used.

Building a KD-tree and using it to identify elements where the model is non-zero is a model-type specific procedure, which is applicable wherever the model is defined only in one region such as the LLVP. The integration over the perturbed model to find δu, and the post processing to find the Rytov seismograms are model specific.

The timings show that the added computational cost per model is very low, at ≈ 7 s per model on only one core. This is for changing the shape of the model within a certain size. In cases where the amplitude of $\delta ln(v_p)$ is scaled, only the post-processing step is required, and so the added computational time per model is even smaller.

Pre-computation steps	*CPU Hours*
Wavefield Calculation at 10s	1230
Kernel Calculation at 10s	690
Midpoint Computation	3
Model-type specific steps	*CPU Hours*
Building KD Tree	0.001
Finding non-zero elements	0.18
Model iteration steps	*CPU Hours*
Generating model, projection and multiplication (per model)	0.002
Post-processing (per model)	0.00001

Table 5.1: Computational costs for one source-receiver pair and many models

Seismogram Plots

In this section, record sections of Rytov seismograms for 18 models are presented, in order to demonstrate the effects of different parameters on the waveform. We plot against both longitude and epicentral distance. The models shown in each figure are summarised in Table 5.2. Height and extent are grouped into parameters relating to the size of the perturbation, and strength and x into parameters relating to the velocity perturbation. In Figures 5.14 and 5.15, height and extent are fixed, whereas in Figures 5.16 and 5.17 strength and x are fixed. Seismograms are plotted against either epicentral distance (in which case longitude is fixed to 0°) or longitude (in which case the distance is fixed to 104°).

Figure	Strength /%	x/ km	Height / km	Extent /°	Plotted against
5.14	-1, -1.5, -2	0, 400, 800	1700	30	Epicentral distance
5.15	-1, -1.5, -2	0, 400, 800	1700	30	Longitude
5.16	-1.5	400	1200, 1500, 1700	15, 30, 45	Epicentral distance
5.17	-1.5	400	1200, 1500, 1700	15, 30, 45	Longitude

Table 5.2: The model parameters corresponding to Figures 5.14 - 5.17

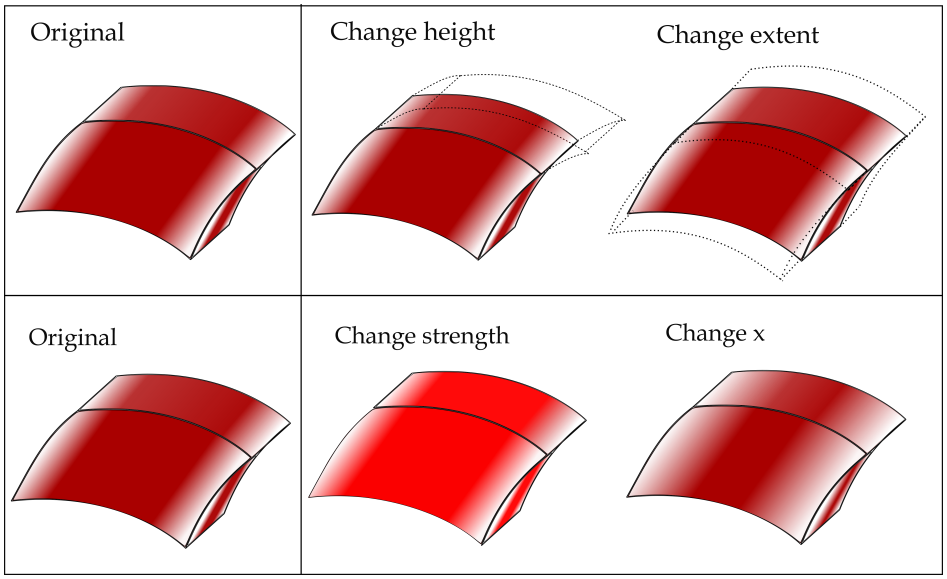


Figure 5.13: The top row shows changes to the model shape made in Figure 5.16 and 5.17. The bottom row shows changes made to the wavespeed perturbations within the model in Figures 5.14 and 5.15

Although results are not presented for a complete set of models, Figures 5.14 - 5.17

give an impression of the effects of these parameters on the waveform and travel time of the Pdiff phase. As noted in section 5.3.2, the main effect is on travel time, and for the remainder of this chapter we focus on travel time measurements.

The fourth column of Figures 5.14 - 5.17 shows how the travel time delay varies with epicentral distance and longitude for various models. We note from Figures 5.14 and 5.15 that, with constant height and extent the shape of the travel time delay curve is dependent mostly on the value of x, but the magnitude of the travel time delays is determined by the strength of the structure. In the fourth column of Figure 5.14, we may discern between the different values of x from the second derivative of travel time with respect to distance. In the case of x=400 km and x=800 km, the second derivative is negative, while in the case that x=0 km the second derivative is positive. Similarly, in the fourth column of Figure 5.15, we see that the double derivative of delay with respect to longitude between 5° and 15° is largest for x=0 km, and decreases with increasing x. Keeping other parameters constant, the strength determines the magnitude of the travel time delay.

The results of keeping x and strength constant, but changing height and extent are shown in Figures 5.16 and 5.17. We see from Figure 5.16 that even where stations are at a longitude of 0°, the azimuthal extent can affect the travel time, as demonstrated by the deviation of the extent=15° curve from the extent=30° extent=45° curves in the fourth column. This is explained by sensitivity outside of the LLVP structure in both perturbed and unperturbed material. The distortion of the shape of the seismogram with an extent of 15° is more significant than larger values due to the interaction with the edge of the structure. Measurements of travel time delay against longitude are of course more sensitive to the extent, as can be seen from the gradient of delays with respect to longitude in the fourth column of Figure 5.17. The height appears to change the magnitude of the travel time delays more than the shape of the delay curve, which is in accordance with the source-receiver geometry making these measurements sensitive to the northern edge of the LLVP.

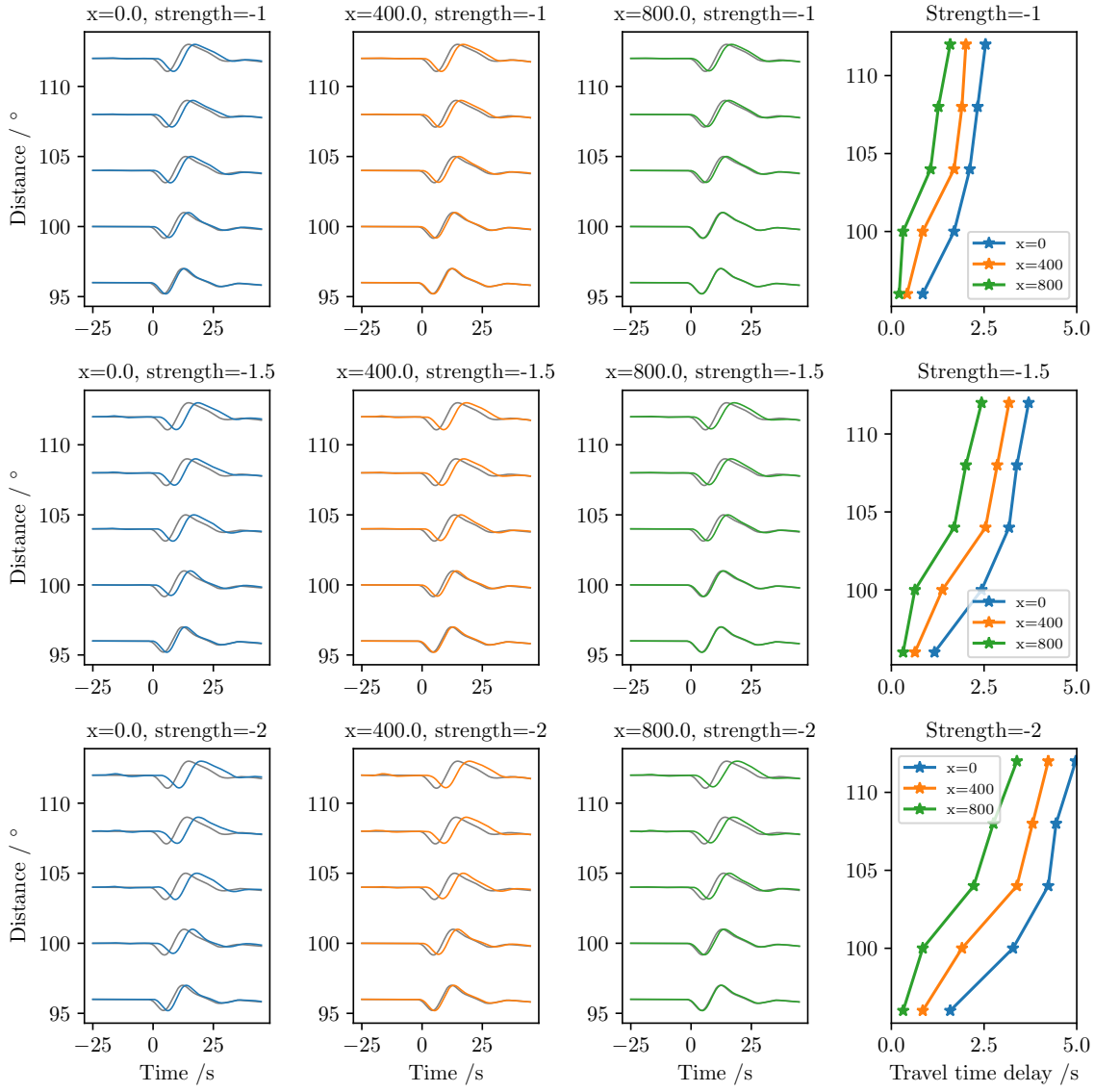


Figure 5.14: The first three columns are Rytov seismograms aligned on the Pdiff arrival plotted against epicentral distance on the y axis. The title of each panel defines the value of x in km and the strength in % of the model used to calculate the seismograms. Each row has a constant strength, while each column has a constant x value. The fourth column shows the cross-correlation measured travel time delays of the Pdiff phase, for the three models in the same row. Each panel corresponds to a fixed strength, height and extent, with a varying x.

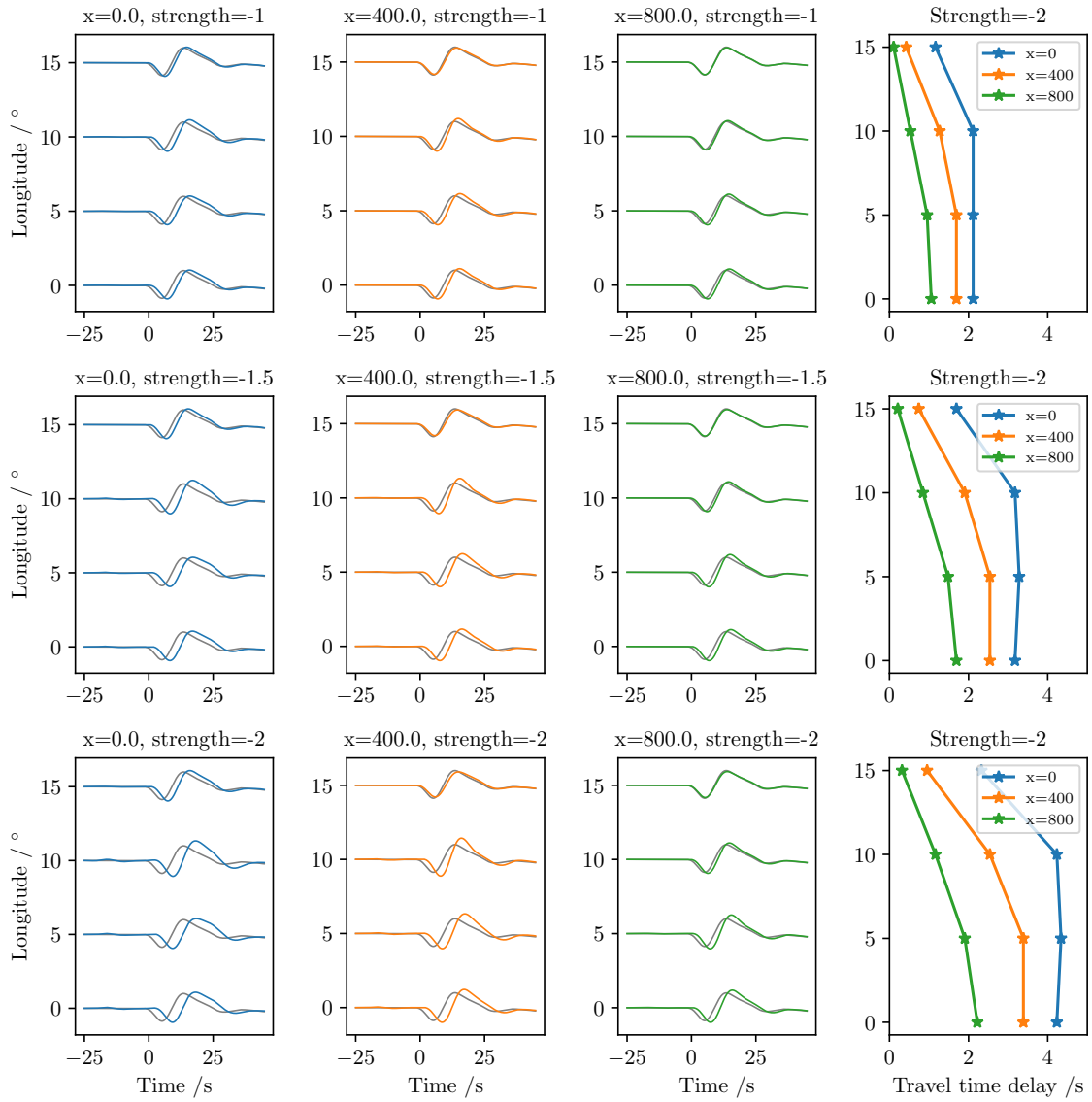


Figure 5.15: Same as Figure 5.14 but plotted against longitude rather than distance.

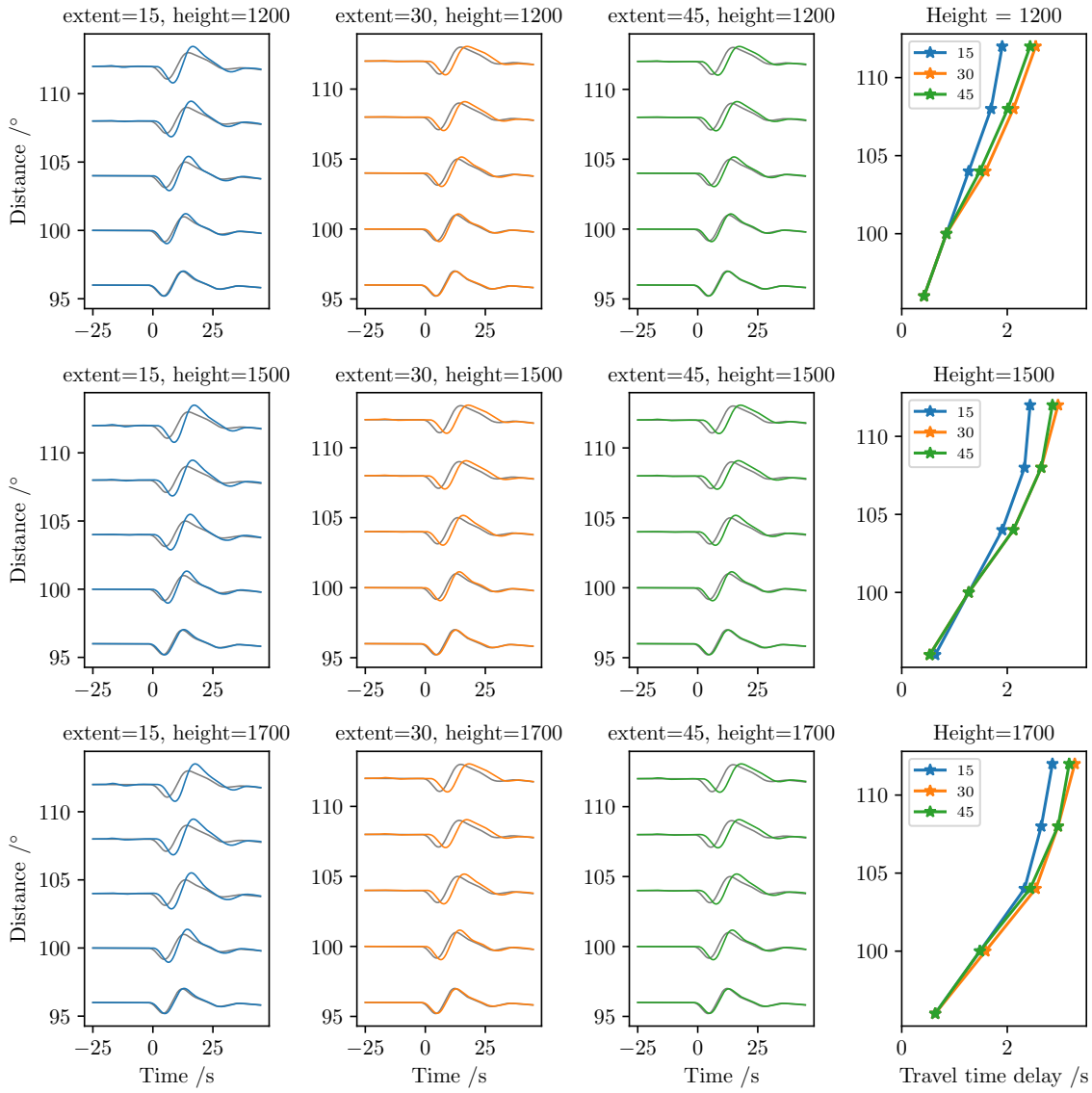


Figure 5.16: The first three columns are Rytov seismograms aligned on the Pdiff arrival plotted against epicentral distance on the y axis. The title of each panel defines the extent in degrees and the height in km of the model used to calculate the seismograms. Each row has a constant height, while each column has a constant extent. The fourth column shows the cross-correlation measured travel time delays of the Pdiff phase, for the three models in the same row. Each panel corresponds to a fixed strength, height and x, with a varying extent.

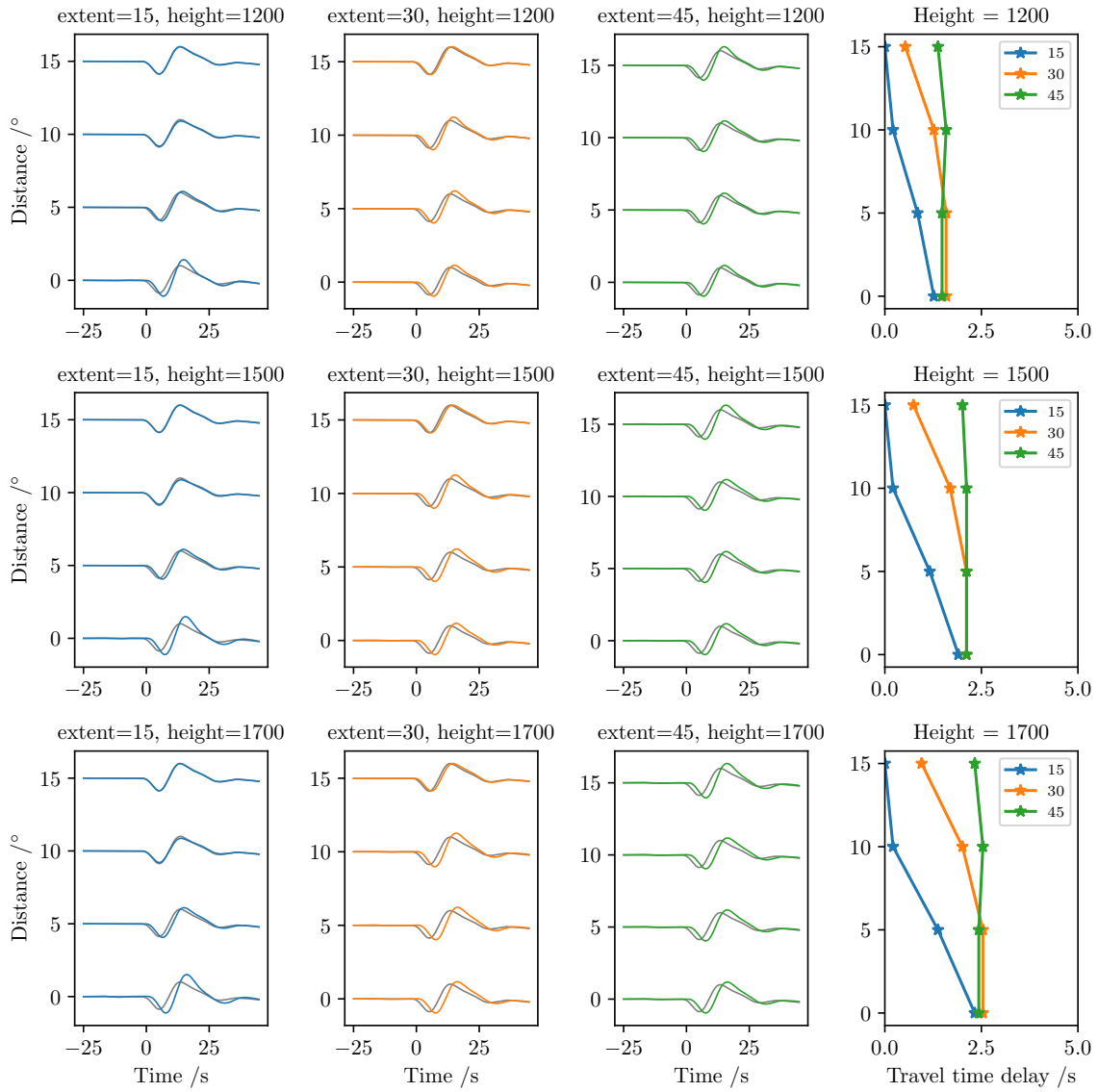


Figure 5.17: Same as Figure 5.16 but plotted against longitude rather than distance.

Travel time delay database

We expand on the previous section and produce synthetic data for 500 models. We generate synthetic waveforms and then measure the travel time delay relative to the PREM seismogram, creating a database of cross-correlation travel time delay measurements. We use the same epicentral distances as above (96-112°), but only examine stations at 0°E. The model parameters are shown below in Table 5.3. In this section, we suggest ways in which a database such as this could be analysed in a statistical manner in a waveform modelling study.

Parameter	Values
Height /km	900, 1100, 1300, 1500, 1700
Extent /°	15, 25, 35, 45
Strength /%	-1, -1.25, -1.5, -1.75, -2
x /km	0, 200, 400, 600, 800

Table 5.3: Model parameters used to compute database of synthetic seismograms. Every combination of the parameters listed is used.

We present the measured travel times of each model at each epicentral distance in Figure 5.18. In this figure, each model is represented by a rectangle with a hanging line. At each distance, the models with the smallest delay have smaller heights, strengths and extents and larger x values. There are a large number of models in the middle of the range of travel time delays, representing trade-offs between the different parameters. The models with the largest delay times have the largest strengths and heights, and smallest values of x. However, the values of extent are small for these models. The explanation for this is likely due to the more significant waveform distortion in models with an extent of 15°, making the cross-correlation measurement larger.

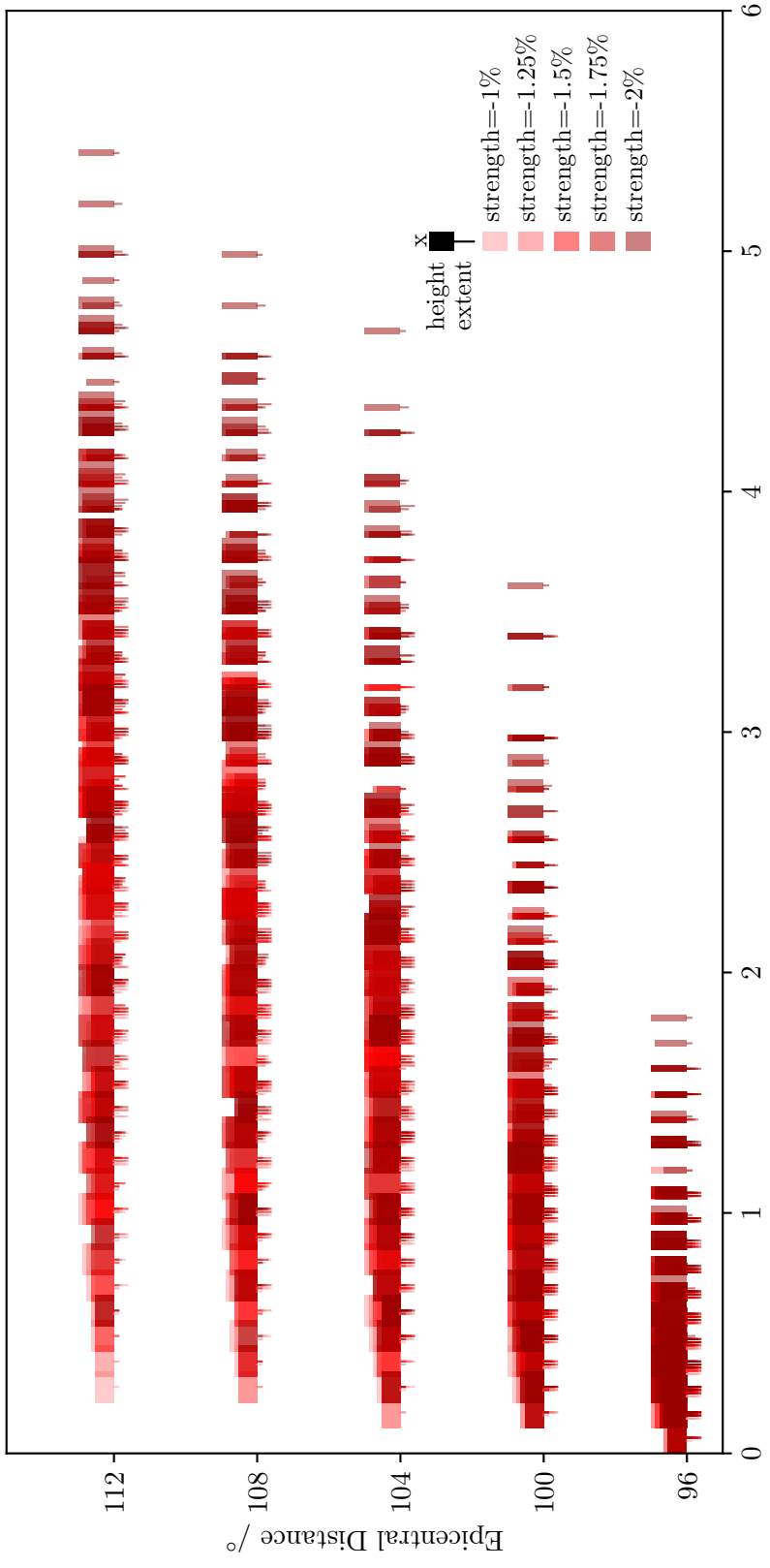


Figure 5.18: Figure representing travel time delay measurements (from cross-correlation) for 500 different LLVP models at 5 stations. All stations are at 0°E. Each model is represented by a rectangle. The height of the rectangle represents the height, the width represents x, and the colour represents the strength. The hanging line represents the extent of the structure.

Querying the travel time delay database

In this section, we give examples of ways in which the database might be queried to find a range of preferred models. First is a test of the hypothesis that models with a negative second derivative of delay time with respect to distance between the stations at 96 and 104° have sharp sides. The second derivative has no particular physical significance, nor has it been previously analysed in the context of waveform modelling for LLVPs. However, the second derivative is selected as a feature of the curves of travel time delay against epicentral distance. This feature is single valued and easily plucked from the delay curve and is an example of a way to quantify the shape of the curve when examining a large number of models. Figure 5.19 shows the parameters for models with a negative second derivative. In total there are 173/500 models which meet this criterion. We can see from this figure that, to a first approximation, any strength or extent is just as likely to result in a negative value of the second derivative. Larger heights are more likely to produce this result. However, the most striking feature is the high probability of a low x value. 130/173 (75.1%) of the models have an x value of 200 km or 0 km, with the number of models meeting the criterion tailing off with increasing x.

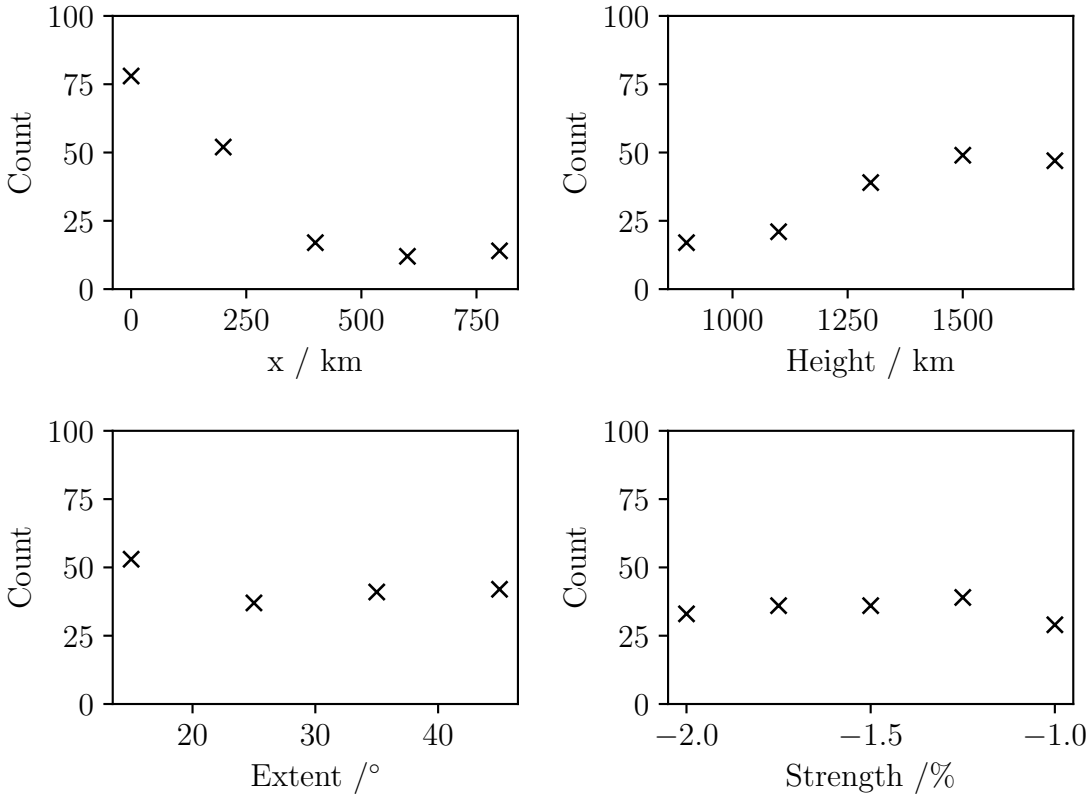


Figure 5.19: Model parameters for models which meet the criterion of a negative double derivative of travel time delay measurement with respect to epicentral distance between 96-104°. Each panel shows the number of models with each value of the parameters. In total there are 173 models.

Another example of how the database might be queried is by imagining that we have certain travel time delay measurements. This supposed data is shown in Figure 5.20. We search for models with these measured delay times within ± 0.3 s. In total, there are 10 models out of 500 which fit this criteria. The parameters of these models are shown in Figure 5.21. This shows how we can rule out certain scenarios. In particular, no models with x less than 400 km, an extent of 15°, a height of less than 1100 km,or a strength with magnitude less than 1.5% fit this criteria. The combinations of parameters are represented in Figure 5.22. This shows the trade-offs between different parameters. For example, there appears to be a trade-off between height and strength (stronger models tend to be shorter which weaker models taller). There is also a trade-off with the x value. For the tallest five models, a larger x corresponds to a smaller strength and vice versa. For the two shortest models, a larger height corresponds to a larger value of x.

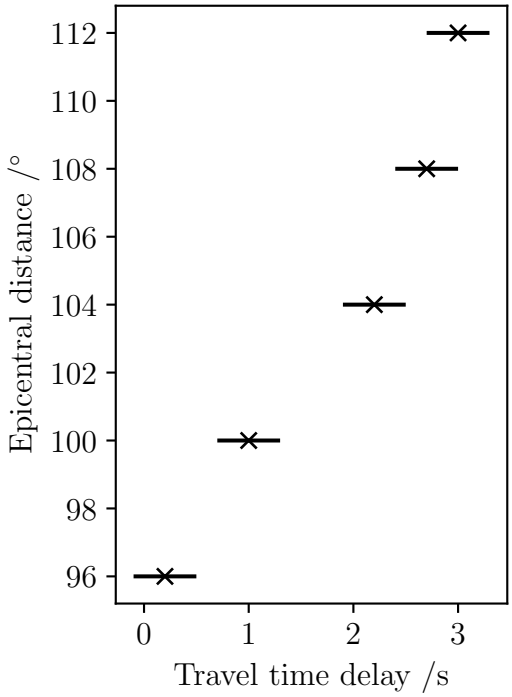


Figure 5.20: A supposed data set, as an example of the sort of measurement that could be taken from observed data. A travel time delay is measured at each station. We search for models which fit this data to within ± 0.3 s, as represented by the error bars.

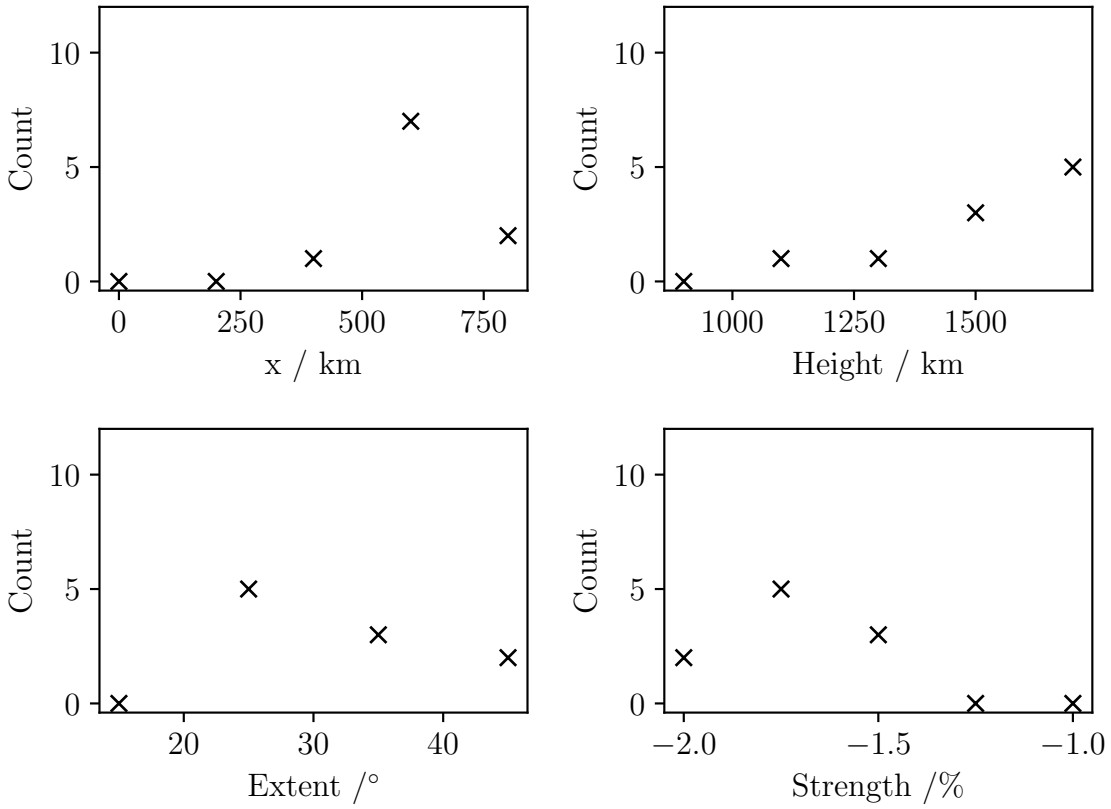


Figure 5.21: Model parameters for models which fit the supposed travel time delay measurements to within ± 0.3 s. Each panel shows the number of models with each value of the parameters. In total there are 10 models.

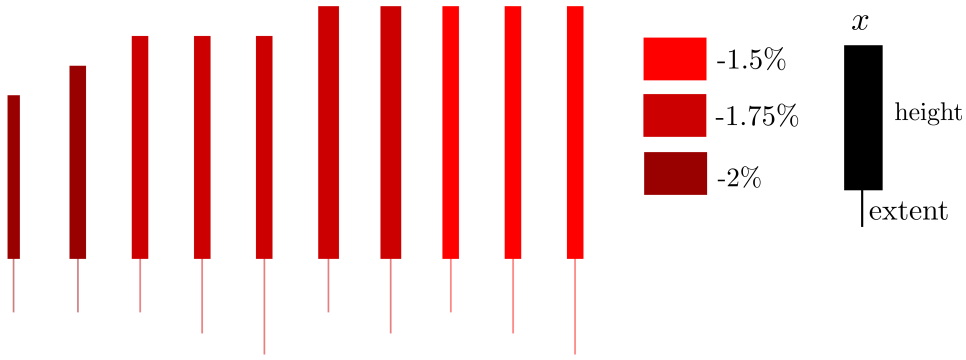


Figure 5.22: Representation of the 10 models which fit the supposed travel time delay measurements. Each model is depicted as a rectangle, in the same way as Figure 5.18. Note that the models are not plotted according to a travel time measurement, but are ordered with respect to strength for ease of viewing.

5.4 Discussion

The work in this chapter demonstrates how our first order method can be used to generate synthetic seismograms for the types of structures that may be found in the lower mantle. Although the kernel calculation is fairly expensive computationally, the added computational cost of an extra model is very small (in our example above, the added cost per model is ≈ 7 s on one core). Comparatively, to compute seismograms for a different model using AxiSEM3D requires over one thousand CPU hours.

The times stated are dependent on the particular mesh, the sampling in time (dependent on dominant period) and on the spatial extent of the model. Where the mesh is finer or the model defined over a greater extent, projection and multiplication is carried out over more elements. Here, we only perform the multiplication and projection on elements where the model is non-zero. In cases where the model is defined globally e.g. a tomography model, the method could be made more computationally efficient by instead identifying the elements where the kernel is of non-negligible magnitude. MC Kernel in its current state is geared towards the computation of measurement kernels for tomography, rather than waveform kernels. Therefore, there are possible improvements to the procedure for computing waveform kernels which would reduce the computational cost. For example, the waveform kernel could be calculated and written to disk only in the time window of interest instead of at each of the roughly 2000 time steps in the seismogram. A further modification would be to include the midpoint calculation within MC Kernel.

The first order method requires a waveform kernel for every source-receiver pair to be stored on disk. Each kernel used in this chapter has a size of around 12GB, and therefore the total hard drive storage requirement is close to 100 GB. (Significant less RAM is required as only a section in time and space of each waveform kernel is required). The disk space required could be significantly reduced by storing the waveform only in the time and space window required. For example, by selecting a time window of around 60 s, the size of each kernel would be reduced to around 0.4 GB. The storage requirements of the method could further be reduced for multiple

events by saving waveform kernels corresponding to a single component of the moment tensor. A linear combination of such kernels could be used for any event depth and station distance.

The very low added computational cost per model allows for synthetic data from hundreds or thousands of models to be compared with observations. This opens the possibility of approaching waveform modelling studies in a probabilistic way. In this chapter, we created a database of cross-correlation travel time delays for 500 different models based on the structure of the ALLVP, to give an example of an application of the method. There are many further applications of the method involving larger numbers of models in a higher dimensional parameter space. With large numbers of models, a Bayesian approach to waveform modelling of lower mantle structures could also be adopted.

Chapter 6

Conclusions

6.1 Summary

Comparing synthetic seismograms calculated using forward waveform modelling to observed data allows us to determine local scale structure within the Earth that is often blurred by seismic tomography. An important example are the edges of the LLVPs; their nature and location hold information about the composition and evolution of these enigmatic structures. Waveform modelling studies of the LLVP edges are limited by their choice of modelling technique. It was the aim of this thesis to develop a forward modelling approach to be used for studying the lower mantle, which is computationally fast while taking into account finite frequency as well as the effects of 3-D structure.

6.1.1 First order Approach

In this thesis, we have developed a method to generate synthetic seismograms under the first order Born, Rytov and Marks approximations. In Chapter 2 the theory behind our approach is presented. Under the first-order, linearised Born approximation, the perturbed seismogram is obtained from $u = u^0 + \delta u^{Born}$. $\delta u(t)^{Born}$ (The perturbation to the seismogram as a result of a perturbation to the seismic wavespeed $\delta v(x)$) can be obtained from waveform kernels $K(x,t)$:

$$\delta u(t)^{Born} = \int_V K(x,t)\delta v(x) d^3x \tag{6.1}$$

The three approximations may be summarised by the equation

$$u^{1^{st}} = u^0(1 + \phi/n)^n, \tag{6.2}$$

where u^0 is the unperturbed seismogram, $u^{1^{st}}$ is the first order seismogram, ϕ is a complex phase and $n = 1, 2, ..., \infty$. The Born approximation corresponds to $n = 1$, where a perturbation is added onto the seismogram. The Rytov approximation corresponds to the case where $n \to \infty$, where a perturbation is added onto the phase of the seismogram, and the Marks approximation is intermediate. We show how the parameter ϕ may be determined from $\phi = \delta u^{Born}/u^0$. Therefore the key to calculating first order seismograms is computation of the waveform kernels. Once these are calculated, an integration (Equation 6.1) and some manipulation (Equation 6.2) are all that is required to generate the perturbed seismograms under any approximation.

In Chapter 3 we describe the method developed to calculate sensitivity kernels, which is called MC (Monte Carlo) Kernel. Time and space dependent waveform kernels are computed using the convolution of a foward wavefield (which propagates from the source to the receiver) and a backward wavefield (which propagates from the receiver to the source). MC Kernel is facilitated by the software AxiSEM, which exploits a computational short-cut that may be used where the source and background model are axisymmetric. This means that the wave equation is solved numerically on a 2-D mesh which is D shaped, and the azimuthal dependence is calculated analytically. MC Kernel uses Monte Carlo integration to project the kernel values onto any desired 3-D grid such as those used by tomography. We use waveform kernels computed using MC Kernel in our subsequent computation of first order seismograms.

6.1.2 Applicability and limitations

We implement, test and apply our first order method for generating synthetic seismograms in Chapters 4 and 5. Since the Born, Rytov and Marks approximations are first order, they are only applicable to velocity models which do not deviate too significantly from the background model. In Chapter 4, we test the approximations by comparing first order seismograms to reference solutions computed using AxiSEM and AxiSEM3D. We used 1-D models where the velocity was perturbed uniformly below the 660 km discontinuity, and 3-D models with spherical perturbations in the lower mantle. Comparing the phase and amplitude of the first order seismograms with respect to the reference solutions, we found that the Rytov approximation matched most closely. This is because of its improved ability over the Born approximation to take into account phase shifts. From our tests on 1-D models, we see that the Rytov approximation matches the reference solution extremely well for models with velocity perturbations up to $\pm 1\%$. The tests on 3-D models (spherical perturbations with structure to wavelength ratios up to 2.12 and wavespeed perturbations up to $\pm 10\%$) show that the Rytov approximation is applicable for perturbations with strengths up to $\delta ln(v) = \pm 10\%$, where the size of the perturbation is below $\approx 1.6\lambda$. For larger sizes, the wavespeed perturbation should be lower than $\pm 5\%$. Similar results are found for P waves, S waves and converted phases. The first order method is also tested on a tomography model very good agreement with the reference seismograms was found. This is the first work on the applicability of the Born approximation to forward modelling of body waves in global seismology. It was shown that the domain of applicability of the Born approximation for body waves is similar to that of surface waves as determined by *Parisi et al.* (2015). The larger domain of applicability of the Rytov approximation suggests that it is a more suitable method for forward modelling than Born.

We have shown that our method is limited by the models for which the Rytov approximation is valid. Another limitation is that the method does not accurately take into account the effects of wavespeed perturbations near to the source and receiver, because of the complexity and large values of the kernel in these regions, and so we

only consider models where the wavespeed is perturbed below the 660 km discontinuity. Further work could investigate how close to the source and receiver the kernel should be masked. Since the kernel is frequency dependent, this distance likely also depends on the frequency.

In this thesis, the applicability of the first order method to lower mantle structures has been investigated. On a smaller scale, regional seismology could also benefit from the computational speed-up offered by the first order method. In this context, it is important to consider the source and receiver effects resulting in a complex kernel with large magnitude in the region of the source and receiver. It is expected that higher frequency kernels require a smaller volume of kernel to be masked, and that the method could be scaled down to include a finer, regional mesh.

6.1.3 Advantages and applications

The significant advantage of the first order method described in this thesis is the very low computational cost of computing seismograms for additional models once the waveform kernel has been calculated. For the example in Chapter 5 of applying this method to an ALLVP-like structure, the additional cost per model is $\approx 7\ s$. We tested 500 models in a 4-D parameter space to create a database of cross-correlation travel time measurements at four stations. Such a database can be probed to determine the most likely set of models that meet certain criteria that are based on observations, in a probabilistic approach to forward waveform modelling.

6.2 Further work

6.2.1 First order method

This work leaves the potential for more tests to be done on the first order approximations and their validity. Although we identify where the approximations are valid and where they are not within our set of models, we aim towards a definitive rule as

to where each approximation can be applied. This could involve tests on variants of tomography models arrived at by changing the amplitude or the typical lengthscale of velocity variations within the mantle in a similar manner to *Parisi et al.* (2015). The set of bubble models could be expanded, and the effect of off-path structure could be examined.

In this thesis we have calculated Marks seismograms, but tended to focus on the two end members of the Born and Rytov approximations. Further work could investigate more thoroughly the effect of changing n and perhaps determining which value is most appropriate depending on the set-up.

The work in this thesis on generating synthetic waveforms using first order approximations has shown that the Rytov approximation performs better than the Born approximation, particularly regarding phase. Since finite frequency tomography is based on phase measurements from cross-correlation, this makes a strong case for a Rytov-based tomography. The Rytov approximation could be used to compute measurement kernels as well as to generate synthetic seismograms in full waveform inversion, since it seems to match the reference solutions in the case of S40RTS as shown in Chapter 5.

6.2.2 Applications

In Chapter 5, we gave a simple example of how our first order method can be applied to a simplified ALLVP-like structure in the lower mantle. More realistic models could include more complex LLVP shapes and those with internal structure, which may be based on tomography models. It would also be interesting to include the effects of ultra-low velocity zones, which despite their significant velocity reductions are sufficiently small in size for first order approximations to be applicable. We can also extend to more parameters and many more models. This method raises the possibility of probabilistic, statistical probing of many Earth models. It has the potential to be embedded into a full MCMC/Bayesian context of lower mantle inference.

An obvious and important next step is to compare our first order synthetic seismo-

grams to observed data. There is observed data from temporary seismic arrays located in Africa which has not been studied in the context of the lower mantle, for example from the NARS array in Botswana (*Fadel et al.*, 2018). This will provide constraints on which to probe a synthetic seismogram database, in order to find a set of models which could realistically describe the data. Some of the details that we observed in the synthetic data.

In this thesis we have not examined in detail the finite frequency nature of the first order method, but it is likely to be a valuable tool in analysing observed data. With the spatial dependence of sensitivity kernels, we can quantify exactly where the structure is that affects the waveform. Since the shape of the kernels changes with frequency, examining the seismogram in different frequency bands can illuminate lower mantle structures to a greater extent.

Bibliography

Ammon, C. J. (1991), The isolation of receiver effects from teleseismic P waveforms, *Bulletin of the Seismological Society of America*, *81*(6), 2504–2510.

Beydoun, W. B., and A. Tarantola (1988), First Born and Rytov approximations: Modeling and inversion conditions in a canonical example, *The Journal of the Acoustical Society of America*, *83*(3), 1045.

Brandenburg, J. P., and P. E. van Keken (2007), Deep storage of oceanic crust in a vigorously convecting mantle, *Journal of Geophysical Research: Solid Earth*, *112*(6), 1–15.

Bull, A. L., A. K. McNamara, and J. Ritsema (2009), Synthetic tomography of plume clusters and thermochemical piles, *Earth and Planetary Science Letters*, *278*(3-4), 152–162.

Capdeville, Y. (2005), An efficient Born normal mode method to compute sensitivity kernels and synthetic seismograms in the Earth, *Geophysical Journal International*, *163*(2), 639–646.

Capdeville, Y., E. Stutzmann, and J. P. Montagner (2000), Effect of a plume on long period surface waves computed with normal modes coupling, *Physics of the Earth and Planetary Interiors*.

Capdeville, Y., C. Larmat, J.-P. Vilotte, and J.-P. Montagner (2002), A new coupled spectral element and modal solution method for global seismology: A first application to the scattering induced by a plume-like anomaly.

Červený, V., M. M. Popov, and I. Pšenčík (1982), Computation of wave fields in inhomogeneous media Gaussian beam approach, *Geophysical Journal of the Royal Astronomical Society*, *70*(1), 109–128.

Chaljub, E., and A. Tarantola (1997), Sensitivity of SS precursors to topography on the upper-mantle 660-km discontinuity, *24*(21), 2613–2616.

Chaljub, E., D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette, and G. Festa (2007), Spectral-element analysis in seismology, *Advances in Geophysics*, *48*, 365–419.

Chapman, C. H. (1978), A new method for computing synthetic seismograms, *Geophysical Journal of the Royal Astronomical Society*, *54*(3), 481–518.

Chapman, C. H., and R. Drummond (1982), Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory, *Bulletin of the Seismological Society of America*, *72*(6), S277–S317.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities, *Seismological Research Letters*, *70*(2), 154–160.

Dahlen, F. A. (1987), Multiplet coupling and the calculation of synthetic longperiod seismograms, *Geophysical Journal of the Royal Astronomical Society*, *91*(1), 241–254.

Dahlen, F. A., and J. Tromp (1998), *Theoretical Global Seismology*, 944 pp., Princeton University.

Dahlen, F. a., S. H. Hung, and G. Nolet (2000), Fréchet kernels for finite-frequency traveltimesI. Theory, *Geophysical Journal International*, *141*(1), 157–174.

Diego Mercerat, E., and G. Nolet (2013), On the linearity of cross-correlation delay times in finite-frequency tomography, *Geophysical Journal International*, *192*(2), 681–687.

Dubrovinsky, L., N. Dubrovinskaia, F. Langenhorst, D. Dobson, D. Rubie, C. Geßmann, I. A. Abrikosov, B. Johansson, V. I. Baykov, L. Vitos, T. L. Bihan, W. A. Crichton, and H.-P. Dmitriev, V.Weber (2003), Iron silica interaction at extreme conditions and the electrically conducting layer at the base of Earth's mantle, *Nature*, *422*, 58–61.

Dumbser, M., and M. Käser (2006), An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - II. The three-dimensional isotropic case, *Geophysical Journal International*, *167*(1), 319–336.

Dziewonski, A. M. (1984), Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6., *Journal of Geophysical Research*, *89*(B7), 5929–5952.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356.

Fadel, I., M. van der Meijde, and H. Paulssen (2018), Crustal Structure and Dynamics of Botswana, *Journal of Geophysical Research: Solid Earth*, *123*(12), 10,659–10,671.

Fichtner, A. (2011), *Full Seismic Waveform Modelling and Inversion*, Springer-Verlag Berlin Heidelberg.

Fichtner, A., B. L. Kennett, H. Igel, and H. P. Bunge (2009), Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophysical Journal International*, *179*(3), 1703–1725.

Ford, H. A., M. D. Long, X. He, and C. Lynner (2015), Lowermost mantle flow at the eastern edge of the African Large Low Shear Velocity Province, *Earth and Planetary Science Letters*, *420*, 12–22.

Frost, D. a., and S. Rost (2014), The P-wave boundary of the Large-Low Shear Velocity Province beneath the Pacific, *Earth and Planetary Science Letters*, *403*, 380–392.

Fuchs, K., and G. Müller (1971), Computation of Synthetic Seismograms with the Reflectivity Method and Comparison with Observations, *Geophysical Journal of the Royal Astronomical Society*, *23*(4), 417–433.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geoscience*, *9*(June), 481.

Grand, S. P., R. D. Van Der Hilst, and S. Widiyantoro (1997), Global seismic tomography: A snapshot of convection in the earth, *GSA Today*, *7*(4), 1–7.

He, Y., and L. Wen (2009), Structural features and shear-velocity structure of the "Pacific Anomaly", *Journal of Geophysical Research: Solid Earth*, *114*(2), 1–17.

He, Y., and L. Wen (2012), Geographic boundary of the "Pacific Anomaly" and its geometry and transitional structure in the north, *Journal of Geophysical Research: Solid Earth*, *117*(9), 1–16.

Hirose, K., Y. Fei, Y. Ma, and H. K. Mao (1999), The fate of subducted basaltic crust in the Earth's lower mantle, *Nature*, *397*(6714), 53–56.

Hosseini, K., K. J. Matthews, K. Sigloch, G. E. Shephard, M. Domeier, and M. Tsekhmistrenko (2018), SubMachine: Web-Based Tools for Exploring Seismic Tomography and Other Models of Earth's Deep Interior, *Geochemistry, Geophysics, Geosystems*, *19*(5), 1464–1483.

Hosseini, K., K. Sigloch, M. Tsekhmistrenko, A. Zaheri, T. Nissen-Meyer, and H. Igel (2019), Global mantle structure from multi-frequency tomography using P, PP and P-diffracted waves, *Geophysical Journal International*, pp. 96–141.

Igel, H., T. Nissen-Meyer, and G. Jahnke (2002), Wave propagation in 3D spherical sections: Effects of subduction zones, *Physics of the Earth and Planetary Interiors*, *132*(1-3), 219–234.

Jeffreys, H., and K. E. Bullen (1940), *Seismological Tables*, British Association for the Advancement of Science, London.

Kästle, E. D., M. Weber, and F. Krüger (2017), Complex Deep Structure of the African Low-Velocity Zone, *Bulletin of the Seismological Society of America*, *107*(4), 1688–1703.

Kennett, B. L. (1983), *Seismic wave propagation in Stratified Media*, Cambridge University Press.

Kennett, B. L. (2001), *The seismic wavefield vols. I and II*, Cambridge University Press.

Kennett, B. L., and E. R. Engdahl (1991), Traveltimes for global earthquake location and phase identification, *Geophysical Journal International*, *105*(2), 429–465.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophysical Journal International*, *122*(1), 108–124.

Kind, R. (1985), The reflectivity method for different source and receiver structures and comparison with GRF data., *Journal of Geophysics - Zeitschrift fur Geophysik*, *58*(1-3), 146–152.

Knittle, E., and R. Jeanloz (1989), Simulating the coremantle boundary: An experimental study of highpressure reactions between silicates and liquid iron, *Geophysical Research Letters*, *16*(7), 609–612.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. van Heijst (2016), SP12RTS: A degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039.

Koelemeijer, P., A. Deuss, and J. Ritsema (2017), Density structure of Earth's lowermost mantle from Stoneley mode splitting observations, *Nature Communications*, *8*(May), 1–10.

Komatitsch, D., and J. Tromp (2002a), Spectral-element simulations of global seismic wave propagation - I. Validation, *Geophysical Journal International*, *149*(2), 390–412.

Komatitsch, D., and J. Tromp (2002b), Spectral-element simulations of global seismic wave propagation - II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophysical Journal International*, *150*, 303–318.

Kristeková, M., J. Kristek, P. Moczo, and S. M. Day (2006), Misfit criteria for quantitative comparison of seismograms, *Bulletin of the Seismological Society of America*, *96*(5), 1836–1850.

Kristeková, M., J. Kristek, and P. Moczo (2009), Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophysical Journal International*, *178*, 813–825.

Lau, H. C., J. X. Mitrovica, J. L. Davis, J. Tromp, H. Y. Yang, and D. Al-Attar (2017), Tidal tomography constrains Earth's deep-mantle buoyancy, *Nature*, *551*(7680), 321–326.

Lay, T., J. Hernlund, E. J. Garnero, and M. S. Thorne (2006), A post-perovskite lens and D" heat flux beneath the central Pacific., *Science (New York, N.Y.)*, *314*(5803), 1272–6.

Leng, K., T. Nissen-Meyer, and M. van Driel (2016), Efficient global wave propagation adapted to 3-D structural complexity: A pseudospectral/spectral-element approach, *Geophysical Journal International*.

Leng, K., T. Nissen-Meyer, M. Van Driel, K. Hosseini, and D. Al-Attar (2019), AxiSEM3D: Broad-band seismic wavefields in 3-D global earth models with undulating discontinuities, *Geophysical Journal International*, *217*(3), 2125–2146.

Li, X. D., and B. Romanowicz (1996), Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *Journal of Geophysical Research: Solid Earth*, *101*(B10), 22,245–22,272.

Li, X. ., and T. Tanimoto (1993), Waveforms of LongPeriod Body Waves In A Slightly Aspherical Earth Model: Waveforms of longperiod body waves in a slightly aspherical Earth model, *Geophysical Journal International*, *112*(1), 92–102.

Liu, Q., and J. Tromp (2008), Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophysical Journal International*, *174*(1), 265–286.

Lynner, C., and M. D. Long (2014), Lowermost mantle anisotropy and deformation along the boundary of the African LLSVP, (41), 3447–3454.

Maneewongvatana, S., and D. M. Mount (1999), It's okay to be skinny , if your friends are fat, *ReCALL*, pp. 1–8.

Marks, D. L. (2006), A family of approximations spanning the Born and Rytov scattering series, *Optics Express*, *14*(19), 8837.

Marquering, H., F. A. Dahlen, and G. Nolet (1999), Three-dimensional sensitivity kernels for finite-frequency traveltimes : the banana - doughnut paradox, *Geophysical Journal International*, *137*, 805–815.

McNamara, A. K., and S. Zhong (2005), Thermochemical structures beneath Africa and the Pacific Ocean, *Nature*, *437*(7062), 1136–1139.

Mohorovicic, A. (1909), Earthquake of 8 October 1909 (Potres od 8. X 1909; Das Beben vom 8. X. 1909), *Yearly Report of the Zagreb Meteorological Observatory for the Year, 1909*.

Montelli, R., G. Nolet, F. a. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung (2004), Finite-frequency tomography reveals a variety of plumes in the mantle., *Science (New York, N.Y.)*, *303*(5656), 338–343.

Mulyukova, E., B. Steinberger, M. Dabrowski, and S. V. Sobolev (2015), Survival of LLSVPs for billions of years in a vigorously convecting mantle: Replenishment and destruction of chemical anomaly, *Journal of Geophysical Research: Solid Earth*, *120*(5), 3824–3847.

Nakagawa, T., P. J. Tackley, F. Deschamps, and J. A. Connolly (2009), Incorporating self-consistently calculated mineral physics into thermochemical mantle convection simulations in a 3-D spherical shell and its influence on seismic anomalies in Earth's mantle, *Geochemistry, Geophysics, Geosystems*, *10*(3), 1–17.

Ni, S., and D. V. Helmberger (2003a), Seismological constraints on the South African superplume: Could be the oldest distinct structure on earth, *Earth and Planetary Science Letters*, *206*(1-2), 119–131.

Ni, S., and D. V. Helmberger (2003b), Further constraints on the African superplume structure, *Physics of the Earth and Planetary Interiors*, *140*(1-3), 243–251.

Ni, S., and D. V. Helmberger (2003c), Ridge-like lower mantle structure beneath South Africa, *Journal of Geophysical Research: Solid Earth*, *108*(B2), 2094.

Ni, S., E. Tan, M. Gurnis, and D. Helmberger (2002), Sharp Sides to the African Superplume, *Science*, *296*(5574), 1850–1852.

Nissen-meyer, T., A. Fournier, and F. A. Dahlen (2007), A two-dimensional spectral-element method for computing spherical-earth seismograms - I. Moment-tensor source, *Geophysical Journal International*, *168*(3), 1067–1092.

Nissen-Meyer, T., F. A. Dahlen, and A. Fournier (2007), Spherical-earth Fréchet sensitivity kernels, *Geophysical Journal International*, *168*(3), 1051–1066.

Nissen-Meyer, T., M. Van Driel, S. C. Staehler, K. Hosseini, S. Hempel, L. Auer, A. Colombi, and A. Fournier (2014), AxiSEM: Broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth*, *5*(1), 425–445.

Nolet, G. (2008), *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*, Cambridge University Press.

Oldham, R. (1906), he constitution of the interior of the Earth, as revealed by earthquakes, *Quarterly Journal of the Geological Society*, *62*, 456–475.

Panning, M. P., Y. Capdeville, and B. A. Romanowicz (2009), Seismic waveform modelling in a 3-D Earth using the Born approximation: Potential shortcomings and a remedy, *Geophysical Journal International*, *177*(1), 161–178.

Parisi, L., A. M. Ferreira, and Y. Capdeville (2015), Validity domain of the Born approximation for seismic waveform modelling in realistic 3-D Earth structure, *Geophysical Journal International*, *200*(2), 908–914.

Persson, P. O., and G. Strang (2004), A simple mesh generator in Matlab, *SIAM Review*, *46*(2), 329–345.

Ritsema, J., S. Ni, D. V. Helmberger, and H. P. Crotwell (1998), Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa, *Geophysical Research Letters*, *25*(23), 4245–4248.

Ritsema, J., a. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236.

Romanowicz, B. A., M. P. Panning, Y. Gung, and Y. Capdeville (2008), On the computation of long period seismograms in a 3-D earth using normal mode based approximations, *Geophysical Journal International*, *175*, 520–536.

Schubert, G., G. Masters, P. Olson, and P. Tackley (2004), Superplumes or plume clusters?, *Physics of the Earth and Planetary Interiors*, *146*(1-2), 147–162.

Schumacher, L., C. Thomas, and R. Abreu (2018), Out-of-Plane Seismic Reflections Beneath the Pacific and Their Geophysical Implications, *J. Geophys. Res*, *123*, 1–17.

Shearer, P. (2009), *Introduction to seismology*, Cambridge University Press.

Spetzler, J., and R. Snieder (2001), The effect of small scale heterogeneity on the arrival time of waves, *Geophysical Journal International*, *145*, 786–796.

Sun, D., and M. S. Miller (2013), Study of the western edge of the African large low shear velocity province, *Geochemistry, Geophysics, Geosystems*, *14*(8), 3109–3125.

Tao, K., S. P. Grand, and F. Niu (2018), Seismic Structure of the Upper Mantle Beneath Eastern Asia From Full Waveform Seismic Tomography, *Geochemistry, Geophysics, Geosystems*, *19*(8), 2732–2763.

Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009), Adjoint Tomography of the Southern California Crust, *Science*, *325*(August), 988–993.

Tarantola, A., and B. Valette (1982), Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion, *Reviews of Geophysics and Space Physics*, *20*(2), 219–232.

Tesoniero, A., L. Auer, L. Boschi, and F. Cammarano (2015), Journal of Geophysical Research : Solid Earth Hydration of marginal basins and compositional variations within the continental lithospheric mantle inferred from a new global model of shear and compressional velocity, *J. Geophys. Res.*, *120*, 7789–7813.

Tian, Y., R. Montelli, G. Nolet, and F. A. Dahlen (2007), Computing traveltime and amplitude sensitivity kernels in finite-frequency tomography, *Journal of Computational Physics*, *226*(2), 2271–2288.

Torsvik, T. H., K. Burke, B. Steinberger, S. J. Webb, and L. D. Ashwal (2010), Diamonds sampled by plumes from the core-mantle boundary., *Nature*, *466*(7304), 352–355.

Trampert, J., F. Deschamps, J. Resovsky, and D. Yuen (2004), Probabilistic tomography maps chemical heterogeneities throughout the lower mantle., *Science (New York, N.Y.)*, *306*(5697), 853–856.

Tromp, J., C. Tape, and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, *160*(1), 195–216.

Van Driel, M. (2015), Efficient Methods in Global Seismic Wave Propagation, *PhD thesis*.

Wang, Y., and L. Wen (2004), Mapping the geometry and geographic distribution of a very low velocity province at the base of the Earth's mantle, *Journal of Geophysical Research B: Solid Earth*, *109*(10), 1–18.

Wang, Y., and L. Wen (2007a), Geometry and P and S velocity structure of the "African Anomaly", *Journal of Geophysical Research: Solid Earth*, *112*(5), 1–26.

Wang, Y., and L. Wen (2007b), Complex seismic anisotropy at the border of a very low velocity province at the base of the Earth's mantle, *Journal of Geophysical Research*, *112*, 1–11.

Wen, L., P. Silver, D. James, and R. Kuehnel (2001), Seismic evidence for a thermochemical boundary at the base of the Earth's mantle, *Earth and Planetary Science Letters*, *189*(3-4), 141–153.

Zaroli, C., S. Lambotte, and J. J. Lévêque (2015), Joint inversion of normal-mode and finite-frequency S-wave data using an irregular tomographic grid, *Geophysical Journal International*, *203*(3), 1665–1681.

Zhao, C., E. J. Garnero, A. K. McNamara, N. Schmerr, and R. W. Carlson (2015), Seismic evidence for a chemically distinct thermochemical reservoir in Earth's deep mantle beneath Hawaii, *Earth and Planetary Science Letters*, *426*, 143–153.

Zhu, H., E. Bozdg, and J. Tromp (2015), Seismic structure of the European upper mantle based on adjoint tomography, *Geophysical Journal International*, *201*(1), 18–52.