UTRECHT STUDIES IN EARTH SCIENCES

Mededeling van de
Faculteit Geowetenschappen
Universiteit Utrecht
No. 030

Imaging the Earth's small-scale structure using full-waveform inversion: Theory and application to the imaging of mantle plumes

Florian Rickers

Utrecht 2013

Members of the dissertation committee:

Prof. dr. J. Ritsema
Department of Earth and Environmental Sciences
University of Michigan, USA

Prof. dr. B. Romanowicz
Department of Earth and Planetary Science
University of California, Berkeley, USA

Prof. dr. C. Thomas
Department of Geophysics
University of Münster, Germany

Prof. dr. M.J.R. Wortel
Faculty of Geosciences
Utrecht University, the Netherlands

Dr. S. Lebedev
School of Cosmic Physics, Geophysics Section
Dublin Institute for Advanced Studies, Ireland

Printed in the Netherlands by Wöhrmann Print Service, Zutphen.

ISBN: 978-90-6266-322-4

Imaging the Earth's small-scale structure using full-waveform inversion: Theory and application to the imaging of mantle plumes

Beeldvorming van de kleinschalige structuur van de Aarde
door inversie van volledige golfvormen: Theorie en toepassing
op de beeldvorming van mantelpluimen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op vrijdag 8 maart 2012 des ochtends te 10.30 uur

door

Florian Rickers
geboren op 27 juni 1981, te Münster, Duitsland

Promotor: Prof. dr. J. Trampert
Co-promotor: Dr. A. Fichtner

This study was funded by the Netherlands Research Centre for Integrated Solid Earth Science (ISES).

Contents

Chapter 1

Introduction

Seismic tomography has developed into a powerful tool to constrain lateral variations of the Earth's structure since the pioneering studies of the late 1970's. The inversion of body-wave traveltime residuals, surface-wave phase dispersion and free-oscillation splitting data resulted in increasingly detailed models, giving insight into details of mantle convection and tectonics. Significant further refinement of tomographic models requires either improvements in global data coverage, particularly in oceanic regions, or a better exploitation of the available data. Approximate inversion techniques, which the majority of tomographic models are based on, pose a limit to the achievable resolution and leave part of the available information unused. The resolution of current global- and continental-scale tomographic models is on the order of several hundreds kilometers in the upper mantle to a thousand kilometres in the lower mantle, with regional variations.

Over the last decade, increasing computational power and theoretical advances have initiated the transition from approximate tomographic methods to iterative full-waveform techniques. These rely on numerical methods to accurately compute the full wavefield in realistic media, potentially allowing the exploitation of the complete seismogram. Full-waveform methods have the potential to further refine tomographic models and to constrain structures on a scale of a few hundred kilometres and below. Such small-scale structures give rise to strong diffraction effects of seismic wave propagation, which are difficult to account for with approximate methods.

In recent years, the first regional- and continental-scale full-waveform models have been published (Tape et al., 2007, 2009; Fichtner et al., 2009; Tape et al., 2010; Fichtner and Trampert, 2011b; Zhu et al., 2012). A num-

ber of subjective choices are involved in the construction of such models, and the complexity of the full seismic wavefield together with the non-linear character of full-waveform inversions make it a challenging task to determine optimum inversion strategies. Such strategies are necessarily problem-dependent, and this thesis investigates how full-waveform methods can be used efficiently to constrain small-scale structures in the Earth's mantle.

1.1 Seismic tomography based on approximate methods

While seismic observations have been used to make inferences on the Earth's interior as early as the beginning of the 20th century (Oldham, 1906; Lehmann, 1936), the beginning of seismic tomography as such dates back to the 1970's, when first studies constrained lateral variations of seismic wavespeeds inside the Earth (Aki et al., 1977; Dziewonski et al., 1977).

These pioneering tomographic studies were based on body-wave traveltime residuals with respect to 1-D velocity models. Using perturbation theory and the ray-theoretical approximation, which states that the traveltime of a wave is only determined by structures along an infinitesimally thin ray path between source and receiver, the traveltime residuals were mapped into three-dimensional variations of seismic wavespeeds. Traveltime tomography has remained a popular technique until today, forming the base of many global models (e.g. Dziewonski, 1984; Zhou, 1996; van der Hilst et al., 1997; Bijwaard et al., 1998; Kárason and van der Hilst, 2001; Grand, 2002; Amaru, 2007; Li et al., 2008; Simmons et al., 2012).

Besides body-wave traveltimes, other types of data were used to construct tomographic models. Among these, the frequency-dependent dispersion of surface-waves and free-oscillation splitting measurements are the most frequently used. The dispersion of surface-waves, mainly sensitive to upper-mantle structure, provides better constraints in regions where seismic stations are sparse, in particular the oceans. A considerable number of models are exclusively based on surface-wave data (e.g. Zhang and Tanimoto, 1993; Trampert and Woodhouse, 1995; Ekström et al., 1997; Shapiro and Ritzwoller, 2002). Measuring the splitting of free oscillations of the Earth, which are excited by large earthquakes, allows us to constrain the long-wavelength structure of the whole mantle and the core (e.g. Giardini et al., 1987; Ritzwoller et al., 1988; Li et al., 1991; He and Tromp, 1996;

Resovsky and Ritzwoller, 1998; Deuss et al., 2010).

The different types of data each have their advantages and shortcomings, which inspired researchers to combine different datasets in joint inversions. Surface-waves have been used in combination with body-waves (e.g. Li and Romanowicz, 1996; Mégnin and Romanowicz, 2000; Lebedev and Van Der Hilst, 2008) and with free-oscillation splitting measurements(e.g. Masters et al., 1982; Woodhouse and Dziewonski, 1984). Other models combine body-waves, surface-waves and free-oscillation splitting measurements (e.g. Masters et al., 1996; Ritsema et al., 1999, 2011).

Various features of mantle convection and tectonics have been revealed by these models. These include subducted oceanic lithosphere penetrating into the lower mantle (Fukao et al., 1992, 2001; Grand, 1994; Kárason and van der Hilst, 2000; Grand, 2002), large low-wavespeed anomalies in the lower mantle beneath the Pacific Ocean and beneath Africa (e.g. Su et al., 1994; Ritsema et al., 1999; Romanowicz and Gung, 2002). Other features clearly imaged in numerous models include low wavepeeds beneath Mid-Ocean ridges and high wavespeeds associated with continental cratons (e.g. Zhang and Tanimoto, 1993; Gung et al., 2003; Ritsema et al., 2004; Lebedev and Van Der Hilst, 2008).

However, all the models listed above rely on approximations in the sense that the true sensitivity of seismic waves is not or only partly honoured. At present, it is becoming evident that these approximations pose a limit to the further refinement of tomographic models. The ray-theoretical infinite-frequency approximation, which most techniques are based on, does not account for the sensitivity of seismic waves to structure off the infinitesimally thin ray path. Such an approximation is appropriate when the length-scale of an anomaly exceeds the Fresnel zone width of seismic waves, but it breaks down for smaller anomalies. The reason is that the complexity of the wavefield increases due to finite-frequency wave propagation effects for anomalies smaller than the width of the first Fresnel zone. In that case, information tends to be contained in the diffracted rather than in the transmitted wavefield. The loss of information in the transmitted wavefield, resulting from diffraction effects, is referred to as wavefront healing (e.g. Wielandt, 1987; Nolet and Dahlen, 2000; Hung et al., 2001; Dahlen, 2004; Malcolm and Trampert, 2011).

Several techniques have been developed with the aim of incorporating finite-frequency wave propagation effects into tomographic models, while avoiding the computationally costly numerical simulation of the full seismic

wavefield. Based on normal mode coupling theory, NACT-kernels take off-raypath sensitivity within the vertical plane into account (Li and Romanowicz, 1996; Mégnin and Romanowicz, 2000; Panning and Romanowicz, 2006). A different technique, commonly referred to as linearised finite-frequency tomography, replaces ray theory by the single-scattering approximation of Born theory (Hudson, 1977; Wu and Aki, 1985). In the forward problem, this results in a better estimation of traveltime residuals for anomalies smaller than the width of the Fresnel zone (Baig et al., 2003; Dahlen, 2004). The technique was initially established for surface-waves (Woodhouse and Girnius, 1982; Snieder, 1993) and later modified to include body-waves (Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2001). Several global models were constructed using finite-frequency theory (e.g. Montelli et al., 2004; Zhou et al., 2005; Montelli et al., 2006b). It is a matter of ongoing debate whether inversions using linearised finite-frequency theory lead to better-resolved tomographic images compared to ray theory (de Hoop and van der Hilst, 2005; van der Hilst and de Hoop, 2005, 2006; Montelli et al., 2006a; Trampert and Spetzler, 2006; Boschi et al., 2006).

1.2 Full-waveform tomography

Today, available computational resources along with advances in numerical techniques make it possible to compute the full, albeit band-limited, seismic wavefield in heterogeneous 3-D media using, for instance, finite-difference (e.g. Alterman and Karal, 1968; Boore, 1972; Alford and Kelly, 1974; Virieux, 1984) or spectral-element methods (e.g. Komatitsch and Tromp, 2002; Komatitsch et al., 2005; Chaljub et al., 2007; Fichtner and Igel, 2008). Because the computational costs increase rapidly with the minimum frequency of the simulated wavefield, global-scale full-wavefield simulations are currently only feasible at periods of at least several seconds. The availability of the full wavefield potentially allows us to evaluate the misfit between synthetic and observed waveforms, not only for the major phases, but for the whole seismogram, including waves diffracting at small-scale anomalies. The misfit is iteratively minimised using a gradient-based optimisation scheme, such as the conjugate gradient method (Fletcher and Reeves, 1964). An intermediate step at every iteration is the computation of the gradient, which is efficiently obtained with the adjoint method at the same computational cost as a forward simulation (e.g. Tarantola, 1984; Tromp et al., 2005; Fichtner et al., 2006; Tape et al., 2007). This

tomographic technique, usually referred to as either adjoint tomography or full-waveform tomography, is accurate in the sense that it does not rely on approximations in the computation of the forward and the adjoint wavefield. The term full-waveform inversion is not always unambiguously understood, however. To prevent confusion, we explicitly mean adjoint-based iterative full-waveform tomography when referring to full-waveform tomography throughout this thesis.

Such full-waveform techniques have been successfully used to construct regional and continental-scale tomographic models (Tape et al., 2007; Fichtner et al., 2009; Tape et al., 2010; Fichtner and Trampert, 2011b; Zhu et al., 2012). While not falling into our strict definition of adjoint-based full-waveform tomography, Lekić and Romanowicz (2011) used a computationally less demanding hybrid approach, combining spectral-element forward simulations with 2-D NACT kernels instead of 3-D adjoint gradients, to construct a global upper-mantle model.

Full-waveform techniques open new possibilities for seismic tomography, but also a new level of complexity and a number of subjective choices, which affect the tomographic model.

The misfit, used to quantify the difference between synthetic and observed seismograms, needs to be chosen carefully, because it determines the information which constrains the tomographic model. The options are countless, and the only requirement is that the measurement is physically meaningful with respect to the parameters to be resolved. Possible choices include cross-correlation traveltime measurements either in a single frequency band (Luo and Schuster, 1991; Tape et al., 2007) or multiple frequency bands (Sigloch and Nolet, 2006; Sigloch et al., 2008), multi-taper measurements (Tape et al., 2009, 2010; Zhu et al., 2012), L_2-norm waveform difference (Tarantola, 1984; Tromp et al., 2005), time-frequency phase misfits (Fichtner et al., 2009), instantaneous phase (Taner et al., 1979; Bozdağ et al., 2011) or instantaneous amplitude misfits (Bozdağ et al., 2011).

While full-waveform tomography is a non-linear problem, the convergence of the conjugate gradient algorithm depends on the degree of non-linearity of the misfit function. For instance, the L_2-norm waveform difference combines amplitude and phase information, resulting in a strongly non-linear misfit function. This slows down the convergence towards the optimum model, and the iterative algorithm may even converge towards a local minimum of the misfit function, instead of the global minimum (Luo and Schuster, 1991). Pure phase measurements are a better choice because

their relation to the velocity structure is less non-linear (Fichtner et al., 2009).

The convergence can be improved through smoothing and pre-conditioning of the gradient, reflecting the estimated resolution of the model. Multi-scale inversion strategies, in which the frequency content of the seismograms is increased over the course of the iterations, are useful to support the convergence towards the global minimum of the misfit function (Bunks et al., 1995; Sirgue and Pratt, 2004).

The complexity of the full wavefield and the number of subjective choices in full-waveform inversion require thorough investigation. With better understanding of the implication of the various choices, the potential of full-waveform techniques can be used more efficiently.

1.3 Motivation and outline of the thesis

This thesis explores the potential of adjoint-based full-waveform inversion to image small-scale structures within the Earth. While the technical and theoretical tools have been developed and are now computationally feasible, questions concerning optimum inversion strategies and the quantification of waveform differences remain.

In this thesis, we first use synthetic experiments to gain insight into the resolution capabilities of full-waveform methods, which are then applied to construct a model of the North Atlantic region. While this study explores the tomographic imaging of smaller-scale (less than several hundred kilometers) velocity anomalies in deeper regions of the mantle, the focus and the experimental setup is on the tomographic detection of possible mantle plumes (Morgan, 1971). Whether they exist as features of mantle convection next to the well-documented plate mode (Kárason and van der Hilst, 2000; Fukao et al., 2001; Grand, 2002) is one of the much-debated issues of seismic tomography and Earth sciences in general. Their small lateral extent and their possible lower-mantle source region pushes tomographic methods to their limits. Furthermore, most hotspot regions are only sparsely sampled by seismic waves. The recent study by Hwang et al. (2011) concludes that mantle plumes are not detectable using cross-correlation traveltime measurements. These properties make mantle plumes a challenging target for exploring the resolution capabilities of full-waveform methods. The structure of this thesis is based on the following outline:

Chapter 2 gives an overview of the technical and theoretical aspects of full-waveform inversion. This includes a description of the spectral-element method, the adjoint method and the iterative conjugate gradient scheme, strategies to ensure a fast convergence towards an optimum model and techniques to assess the resolution of full-waveform models.

Chapter 3 discusses the characteristics and implications of finite-frequency wave propagation effects, i.e. diffraction and the resulting wavefront healing. Full-waveform tomographic methods offer the possibility to accurately model diffracted waves, which is one of their major advantages over classical methods. The information of small-scale anomalies is imprinted on the diffracted rather than the transmitted wavefield, therefore the correct treatment of diffracted waves is essential to resolve small-scale structure in an inversion.

Chapter 4 explores the possibility of imaging a narrow mantle plume in synthetic tomographic experiments. Two misfit functions are compared: the simple cross-correlation traveltime misfit and the time-dependent instantaneous phase misfit. To facilitate the comparison, we restrict the study to the P-wave and its coda. We find that narrow plumes can only be imaged in the lower mantle when diffracted waves are included and the misfit accounts for the time- and/or frequency-dependence of the time series. The instantaneous phase misfit has ideal properties for this purpose. We further confirm that the simple cross-correlation traveltime misfit does not extract the information required to image lower-mantle plumes.

Chapter 5 extends the study of the previous chapter to longer measurement windows including S-waves and surface-waves. We further use three-component seismograms and consider radial anisotropy by inverting for the vertically and horizontally polarised S-velocities β_{sv} and β_{sh} in addition to the P-velocity α, investigating whether plume information is preferentially contained in one of the polarisation directions.

In **Chapter 6**, the full-waveform instantaneous phase misfit is applied to a set of real seismic data covering the North Atlantic region. The resulting S-velocity model reveals features in unprecedented detail. Highlights are clearly separated hotspots associated with Iceland and Jan Mayen, a low-velocity layer beneath much of the oceanic and part of the continental lithosphere and separate plume conduits beneath Iceland and Jan Mayen, reaching into the lower mantle. Resolution tests using Point Spread Functions (Fichtner and Trampert, 2011b) indicate that these structures are well-resolved.

In **Chapter 7**, a concluding discussion is given.

Chapter 2

Full-waveform inversion using the adjoint method

In this chapter, we describe the components and the general scheme of a three-dimensional (3-D) iterative full-waveform inversion, based on the adjoint method. The aim is to minimise a misfit function $\chi(\mathbf{m})$, which quantifies the similarity between observed seismograms and synthetic seismograms computed for an Earth model $\mathbf{m}$. The misfit function can be based on an arbitrary measure of misfit. Examples are cross-correlation traveltime misfits of specific phases (Luo and Schuster, 1991), time-frequency misfits (Fichtner et al., 2009) or the instantaneous phase difference (Bozdağ et al., 2011). While the choice of misfit function determines the information which enters the inversion and hence constrains the resulting model, the general iterative inversion scheme is applicable to any physically meaningful and differentiable misfit. We therefore do not discuss a particular choice of misfit in this chapter. The three components of a full-waveform inversion are (1) the computation of the synthetic forward wavefield in the current model $\mathbf{m_n}$, (2) the construction of adjoint sources and the computation of the gradient (3) the update of model $\mathbf{m_n}$ based on the gradient, yielding a model $\mathbf{m_{n+1}}$ with a reduced misfit $\chi(\mathbf{m_{n+1}}) < \chi(\mathbf{m_n})$. These steps are repeated until a model with a sufficiently small misfit is obtained. In the last two sections of this chapter, we discuss strategies to improve the convergence of iterative inversion schemes and the resolution analysis in full-waveform inversion.

2.1 Computation of the forward wavefield - the spectral-element method

With today's computational resources, it has become feasible to numerically compute the full seismic wavefield in heterogeneous 3-D media at periods down to a few seconds on regional or even global scales. This has opened the possibility of exploiting the full seismic waveform information to constrain tomographic models in seismology. Due to its favourable properties which make it well-suited for problems of numerical wave propagation, the spectral-element method (SEM) (e.g. Komatitsch and Tromp, 2002; Komatitsch et al., 2005; Chaljub et al., 2007; Fichtner and Igel, 2008) has become increasingly popular within the last decade. SEM solutions are of high accuracy, and SEMs are computationally efficient compared to other similarly accurate methods. With respect to finite-difference (FD) schemes (e.g. Alterman and Karal, 1968; Boore, 1972; Alford and Kelly, 1974; Virieux, 1984), SEMs have the advantage of reduced numerical dispersion and an accurate treatment of the free surface, thereby enabling the efficient simulation of surface-waves. This is because the weak formulation of the elastic wave equation, which SEMs are based on, implicitly accounts for the free-surface boundary condition. A disadvantage are the significantly higher computational demands of SEMs with respect to FD methods. Compared to numerical methods of similar accuracy, for instance the Discontinuous-Galerkin (DG) method (e.g. Käser and Dumbser, 2006; de la Puente et al., 2008), SEMs are computationally more efficient. However, SEMs are less flexible when large variations in grid cell size are required, as for instance is the case in finite-source inversions. The DG method can therefore be more efficient in complex geometries. For model geometries which can be represented by a relatively regular grid, as is the case of most regional and global Earth models, the spectral-element method is the preferred choice.

The computational mesh for SEMs is composed of hexahedral elements, which may be deformed to honour discontinuities or major anomalies within the Earth. Within each element, the displacement vector is expressed in terms of Lagrange polynomials (polynomials of degree 4 are used in our case). Gauss-Lobatto-Legendre (GLL) points are used as collocation points for the Lagrange polynomials, providing a finer discretisation within each element. The use of GLL points for the numerical integration results in a diagonal mass matrix, which is one of the factors making SEMs computationally efficient. For further details about the technical aspects of SEMs,

the reader is referred to descriptions of the method in Komatitsch et al. (2005) and Chaljub et al. (2007).

2.2 Computation of the gradient - the adjoint method

A synthetic seismogram $\mathbf{u}_{synt}(t, \mathbf{m})$, computed for model $\mathbf{m}$, and the corresponding observed seismogram $\mathbf{u}_{obs}(t)$ are usually different. Besides noise in the data and inaccuracies in the source parameters, the largest part of the difference is due to the imperfect model $\mathbf{m}$, which does not represent the true Earth accurately in the considered frequency range. To construct a more accurate model, the quantification of the differences between synthetics and observations in a physically meaningful way is the first step. Such a measure of waveform difference is referred to as the misfit, and the total misfit of all seismograms in an inversion is expressed by the misfit function

$$\chi(\mathbf{m}) = \sum_{r=1}^{N} \int_0^T \mathbf{h}(\mathbf{x}_r, t, \mathbf{m}) dt, \tag{2.1}$$

where $\mathbf{h}(\mathbf{x}_r, t, \mathbf{m})$ is an arbitrary measure of seismogram differences, N the total number of seismograms in an inversion, $\mathbf{x}_r$ the position of the receiver r in the computational grid $\mathbf{x}$, t the time and T the end time of the forward simulation.

In the following, we briefly summarise the main aspects involved in the computation of the gradient $\mathbf{g}(\mathbf{m})$ of a misfit function $\chi(\mathbf{m})$ with respect to a model $\mathbf{m}$. This is done efficiently via the adjoint method (Tarantola, 1984; Tromp et al., 2005; Fichtner et al., 2006).

The gradient $\mathbf{g}(\mathbf{m})$ of a misfit function $\chi(\mathbf{m})$ is obtained from the Fréchet derivative $\delta\chi(\mathbf{m})$:

$$\begin{aligned}
\delta\chi(\mathbf{m}) &= \mathbf{g}(\mathbf{m})\delta\mathbf{m} \\
&= \sum_{r=1}^{N} \int_0^T \left[\frac{\partial \mathbf{h}(\mathbf{x}_r, t, \mathbf{m})}{\partial \mathbf{m}} \delta\mathbf{m} \right] dt \\
&= \sum_{r=1}^{N} \int_0^T \left[\frac{\partial \mathbf{h}(\mathbf{x}_r, t, \mathbf{m})}{\partial \mathbf{u}(\mathbf{x}_r, t, \mathbf{m})} \delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m}) \right] dt,
\end{aligned} \tag{2.2}$$

where $\delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m})$ describes the variation in displacement $\mathbf{u}(\mathbf{x}_r, t, \mathbf{m})$ for receiver r at position $\mathbf{x}_r$, resulting from perturbations in the model parameters $\mathbf{m}$. It is not feasible from a computational point of view to compute $\delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m})$ explicitly, as this would require a forward simulation of the full wavefield for a change in each element of the model vector $\mathbf{m}$.

The adjoint method circumvents the need for an explicit computation of $\delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m})$. Instead, the Fréchet derivative is transformed into a volumetric integral by expressing $\delta\mathbf{u}(\mathbf{x}_r, t, \mathbf{m})$ in terms of the single-scattering Born approximation and the so-called adjoint wavefield $\mathbf{u}^\dagger(\mathbf{x}, T-t, \mathbf{m})$. Detailed derivations are given in Tromp et al. (2005) and Fichtner et al. (2006). In the interest of brevity, the dependence on the model $\mathbf{m}$ is omitted in the following equations. For an isotropic medium, the Fréchet derivative $\delta\chi$ can be expressed as

$$\delta\chi = \int_V \left[K_\alpha(\mathbf{x})\delta\ln\alpha(\mathbf{x}) + K_\beta(\mathbf{x})\delta\ln\beta(\mathbf{x}) + K_\rho(\mathbf{x})\delta\ln\rho(\mathbf{x})\right] d^3\mathbf{x} \tag{2.3}$$

with

$$K_\alpha(\mathbf{x}) = -2\left[\kappa(\mathbf{x}) + \frac{4}{3}\mu(\mathbf{x})\right]\int_0^T \Big[\nabla\cdot\mathbf{u}^\dagger(\mathbf{x}, T-t)\Big]\left[\nabla\cdot\mathbf{u}(\mathbf{x}, t)\right]dt, \tag{2.4}$$

$$\begin{aligned} K_\beta(\mathbf{x}) = &-4\mu(\mathbf{x})\int_0^T \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)dt \\ &+ \frac{8}{3}\mu(\mathbf{x})\int_0^T \Big[\nabla\cdot\mathbf{u}^\dagger(\mathbf{x}, T-t)\Big]\left[\nabla\cdot\mathbf{u}(\mathbf{x}, t)\right]dt, \end{aligned} \tag{2.5}$$

$$\begin{aligned} K_\rho(\mathbf{x}) = &-\kappa(\mathbf{x})\int_0^T \Big[\nabla\cdot\mathbf{u}^\dagger(\mathbf{x}, T-t)\Big]\left[\nabla\cdot\mathbf{u}(\mathbf{x}, t)\right]dt \\ &-\rho(\mathbf{x})\int_0^T \mathbf{u}^\dagger(\mathbf{x}, T-t)\cdot\frac{\partial^2\mathbf{u}(\mathbf{x}, t)}{\partial \mathrm{t}^2}dt \\ &-2\mu(\mathbf{x})\int_0^T \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)dt, \end{aligned} \tag{2.6}$$

where $K_\alpha(\mathbf{x})$, $K_\beta(\mathbf{x})$ and $K_\rho(\mathbf{x})$ are referred to as Fréchet kernels of the P-velocity α, the S-velocity β and the density ρ, respectively. The Fréchet kernels are defined in terms of the bulk modulus $\kappa(\mathbf{x})$, the shear modulus $\mu(\mathbf{x})$, the density $\rho(\mathbf{x})$, the forward and adjoint wavefields $\mathbf{u}(\mathbf{x}, t)$ and

$\mathbf{u}^\dagger(\mathbf{x}, T-t)$ and the corresponding forward and adjoint traceless strain deviators $\mathbf{D}(\mathbf{x}, t)$ and $\mathbf{D}^\dagger(\mathbf{x}, T-t)$.

The adjoint wavefield $\mathbf{u}^\dagger(\mathbf{x}, T-t)$ is obtained by time-reversed propagation of an adjoint source $f^\dagger(\mathbf{x}, t, \mathbf{m})$ from the receiver locations during an adjoint simulation, which comes at the same computational cost as a forward simulation of the wavefield. The adjoint source can be computed and simultaneously propagated for all receivers associated to the same event. The adjoint source is based on the difference between observed and synthetic data and depends on the chosen misfit function. For the generalised misfit function in eq. 2.1, the adjoint source is

$$f^\dagger(\mathbf{x}, t, \mathbf{m}) = \sum_{r=1}^{N} \frac{\partial \mathbf{h}(\mathbf{x}_r, T-t, \mathbf{m})}{\partial \mathbf{u}(\mathbf{x}_r, T-t, \mathbf{m})} \delta(\mathbf{x}-\mathbf{x}_r). \tag{2.7}$$

The Fréchet kernels are constructed during the adjoint simulation through the interaction of the adjoint wavefield $\mathbf{u}^\dagger(\mathbf{x}, T-t)$ with the stored forward wavefield $\mathbf{u}(\mathbf{x}, t)$ according to eq. (2.4)-(2.6), and are thus obtained at the computational cost of one forward and one adjoint simulation per event. The temporal and spatial derivatives involved in the construction of the Fréchet kernels can be evaluated, for instance, with a finite-difference scheme.

A Fréchet kernel based on a body-wave cross-correlation traveltime measurement at a single receiver in a 1-D model has the typical 'banana-doughnut' shape (Tromp et al., 2005), similar to analytically derived finite-frequency kernels (Marquering et al., 1999; Dahlen et al., 2000, e.g.). Fréchet kernels based on more sophisticated misfits, such as time- or frequency-dependent measurements, can have a more complex shape. The kernels are strongly dependent on the choice of misfit and the part of the seismogram which is measured.

If adjoint sources are back-propagated from all receivers associated with the same event, the resulting superposition of individual kernels is referred to as the 'event kernel' (Tape et al., 2007). Similarly, the superposition of the event kernels for all events used in an inversion is referred to as the 'misfit kernel'.

Fréchet kernels are computed on the discrete numerical grid $\mathbf{x}$. The model $\mathbf{m}$ can be defined on the same grid or, as in our case, in terms of basis functions. The gradient $\mathbf{g}(\mathbf{m})$, which is used in the iterative optimisation scheme in the following section, is in the latter case obtained by projecting the Fréchet kernels from the numerical grid onto the basis functions of the

model vector $\mathbf{m}$.

2.3 Iterative optimisation - the conjugate gradient algorithm

The gradient $\mathbf{g}(\mathbf{m})$ represents the first derivative of the misfit function $\chi(\mathbf{m})$ with respect to the model parameters $\mathbf{m}$, containing information on the direction in which the model parameters $\mathbf{m}$ need to be perturbed to reduce the misfit $\chi(\mathbf{m})$. Knowledge of the Hessian $\mathbf{H}(\mathbf{m})$, containing second-derivative information, would provide information on how far each individual element of the model vector $\mathbf{m}$ has to be perturbed to achieve an optimum decrease of the misfit function $\chi(\mathbf{m})$.

Because the computation of the Hessian $\mathbf{H}(\mathbf{m})$ is currently not routinely possible for 3-D full-waveform tomography, second-derivative information is not available and the inversion has to be based on the gradient $\mathbf{g}(\mathbf{m})$ alone. In this case, iterative optimisation methods are used. An efficient iterative scheme is the conjugate gradient method (Fletcher and Reeves, 1964). It relies on successive gradient-based updates of the model $\mathbf{m}$, leading to the convergence towards the minimum of $\chi(\mathbf{m})$. The gradient is recalculated in the updated model at each iteration step.

Different from the method of steepest descent, where the model update is only based on the current gradient, the descent direction of the conjugate gradient method is built from the current gradient and a recursive term. For $\chi(\mathbf{m})$ quadratic in $\mathbf{m}$, this term ensures that the current descent direction is conjugate to all previous descent directions. Although $\chi(\mathbf{m})$ is usually not exactly quadratic in full-waveform inversion and the descent directions therefore are not exactly conjugate, the convergence towards the global minimum is accelerated compared to the method of steepest descent. The conjugate gradient method has successfully been used in many full-waveform tomographic studies (Tape et al., 2007, 2009; Fichtner et al., 2009; Fichtner and Trampert, 2011b; Zhu et al., 2012).

The following steps outline the conjugate gradient algorithm (Fletcher and Reeves, 1964), starting at iteration $n = 0$ with an initial model $\mathbf{m}_0$.

1. Compute the gradient $\mathbf{g}(\mathbf{m}_n)$ for the current model $\mathbf{m}_n$.
   This requires one forward and one adjoint simulation per event.

2. Set the descent direction $\mathbf{h}_n$ to

$$\mathbf{h}_n = -\mathbf{g}(\mathbf{m}_n) + \beta_n \mathbf{h}_{n-1}. \tag{2.8}$$

In the first iteration ($n = 0$), $\beta_0 = 0$. In subsequent iterations, β_n is obtained by

$$\beta_n = \frac{\|\mathbf{g}(\mathbf{m}_n)\|}{\|\mathbf{g}(\mathbf{m}_{n-1})\|}. \tag{2.9}$$

3. Add a multiple of $\mathbf{h}_n$ to the current model $\mathbf{m}_n$. This yields the updated model $\mathbf{m}_{n+1}$:

   $$\mathbf{m}_{n+1} = \mathbf{m}_n + \gamma_n \mathbf{h}_n. \tag{2.10}$$

   The optimum step length γ_n is determined with a quadratic line search, involving two forward simulations per event (see below).

4. Case 1: If $\chi(\mathbf{m}_n)$ is sufficiently small → The inversion is finished, $\mathbf{m}_n$ is the final model.
   Case 2: If $\chi(\mathbf{m}_n)$ needs to be reduced further → go to (1.) and repeat.

In total, four numerical wavefield calculations per event are required to perform one iteration to update the model from $\mathbf{m}_n$ to $\mathbf{m}_{n+1}$. Two calculations are needed to obtain the gradient $\mathbf{g}(\mathbf{m}_n)$, and two more to determine the optimum step length γ_n.

To estimate the optimum step length γ_n, which leads to a maximum misfit reduction in the current descent direction $\mathbf{h}_n$, a quadratic line search is used. This involves the assumption that the misfit function behaves approximately quadratic in the vicinity of the current model $\mathbf{m}_n$ with respect to perturbations in the current descent direction $\mathbf{h}_n$. Two test step lengths γ_{t1} and γ_{t2} are chosen, yielding two test models $\mathbf{m}_{t1}$ and $\mathbf{m}_{t1}$

$$\mathbf{m}_{t1} = \mathbf{m}_n + \gamma_{t1} \mathbf{h}_n, \tag{2.11}$$

$$\mathbf{m}_{t2} = \mathbf{m}_n + \gamma_{t2} \mathbf{h}_n. \tag{2.12}$$

The forward wavefield is computed for each of the test models, and the corresponding values of the misfit function $\chi(\mathbf{m}_{t1})$ and $\chi(\mathbf{m}_{t2})$ are evaluated. Together with the misfit of the unperturbed model $\chi(\mathbf{m}_n)$, three data pairs of the form (step length / misfit) are then available. A quadratic polynomial can be fitted through these three data points, and the step length γ at its minimum represents the optimum step length γ_n.

The test step lengths γ_{t1} and γ_{t2} need to be chosen carefully. The quadratic approximation usually only holds for small perturbations, therefore the test step lengths should not be too far from the actual minimum.

We noticed for our inversions that the optimum step length in most cases does not change dramatically between two iterations. We obtained a stable scheme by relating the test step lengths of the current iteration to the optimum step length γ_{n-1} of the previous iteration through

$$\gamma_{t1} = \frac{\gamma_{n-1}}{2}, \tag{2.13}$$

$$\gamma_{t2} = \gamma_{n-1}. \tag{2.14}$$

The conjugate gradient algorithm is straightforward in theory. However, the misfit function $\chi(\mathbf{m})$ is usually non-linear and may display multiple local minima besides the global minimum. Because of the quadratic approximation involved in determining the optimum step length γ_n at each iteration, the algorithm tends to converge towards the minimum closest to the current misfit value $\chi(\mathbf{m}_n)$. Strategies which guide the algorithm towards the desired global minimum are therefore essential and are described in the following section.

2.4 Strategies to improve the convergence in iterative inversions

As mentioned in the previous section, the non-linearity of the misfit function may cause the iterative algorithm to converge towards a local minimum instead of the global one. Furthermore, the convergence rate of the algorithm usually decreases with the degree of non-linearity, making more iterations necessary which increases the computation demands. Strategies to guide the algorithm efficiently towards the optimum model are aimed at keeping the degree of non-linearity low and at starting the inversion close enough to the global minimum to avoid possible local minima. Important strategies include:

- Choosing a 3-D starting model which explains the data relatively well at the beginning of the inversion.
- Choosing a good misfit function. The shape of the misfit function is determined by the type of measurement used to quantify the remaining difference between synthetic and observed seismograms. Ideally, the measurement should be as linearly related to the model parameters as possible. At the same time, it is desirable that the misfit function

extracts the maximum amount of information from the seismograms, thereby increasing the constraints on the model. These two properties usually trade-off to some degree, but certain misfits combine both properties better than others.

- The non-linearity of the misfit function is usually stronger for seismograms with a higher frequency content, due to more complex wave propagation effects. A so-called multi-scale approach helps to guide the inversion towards the global minimum by successively increasing the frequency content during the iterative inversion (Bunks et al., 1995; Sirgue and Pratt, 2004). In this way, the long-wavelength structure is constrained early in the inversion and the misfit function is closer to the global minimum when higher frequencies are admitted at a later point in the inversion, reducing the likelihood of convergence towards a local minimum.

- Smoothing of the gradient is essential to prevent the propagation of noise into the model and the manifestation of unresolved small-scale structure. Noise may be present in the gradient for various reasons, but mainly because of noisy data or measurement errors such as phase jumps (see Section 4.5.3). Furthermore, depending on the misfit and the specific seismogram, the adjoint source can contain wavelengths shorter than the minimum wavelength resolved by the numerical wavefield simulation, leading to numerical noise. A sensible approach is the smoothing of the gradient over a length-scale approximately corresponding to the shortest wavelength resolved by the numerical scheme (Tape et al., 2007). This may be done, for instance, by convolving the gradient with a 3-D Gaussian kernel. Smoothing reduces the non-linearity of the misfit function by enforcing a minimum length-scale of the resolved structure.

- Pre-conditioning of the gradient can be used to correct for differences in the sampling density or to balance contributions from different datasets with a differing sensitivity. The term pre-conditioning refers to modifications of the gradient aiming to improve the convergence. Smoothing is, strictly speaking, also a form of pre-conditioning. For instance, the sensitivity of the gradient may be higher in the uppermost mantle if different wave types are combined in the same inversion. Surface-waves and (multiple) surface reflections of body-waves provide a denser sampling in these regions compared to the deeper regions of the model which are sampled by direct body-waves only. To accelerate the convergence,

it may be useful in such a case to apply pre-conditioning by damping the gradient in the upper mantle to a certain degree. We use a pre-conditioned gradient $\mathbf{g}^{pre}(\mathbf{m})$, computed as the square root of the absolute value of the original gradient, multiplied by the original sign. For each element $g_i(\mathbf{m})$ of the gradient vector $\mathbf{g}(\mathbf{m})$, it is derived by

$$g_i^{pre}(\mathbf{m}) = \text{sign}\left[g_i(\mathbf{m})\right] \sqrt{|g_i(\mathbf{m})|}. \tag{2.15}$$

While pre-conditioning of the gradient may appear to be a rather subjective process, it simply aims at accelerating the convergence of the iterative scheme. The Hessian $\mathbf{H}(\mathbf{m})$, which is not available in full-waveform tomography, can be viewed as an ideal pre-conditioner. If the gradient was scaled by the inverse Hessian, a method referred to as Newton method, convergence towards the minimum of $\chi(\mathbf{m})$ would be quadratic (e.g. Pratt et al., 2002; Fichtner and Trampert, 2011a). The subjective pre-conditioning of the gradient attempts to mimic the missing Hessian to a certain degree. As long as it is applied carefully and the sign of the original gradient is not changed, it can help to accelerate the convergence towards the optimum model.

- Weighting of different wave types or individual sensitivity kernels can, similar to pre-conditioning, help to obtain a more even sensitivity of the gradient in different model regions. Because the misfit kernel is a superposition of the individual sensitivity kernels between sources and receivers, a higher station or event density in a particular region can result in a locally increased sensitivity in the misfit kernel. Similarly, regions of sparser ray density may be under-represented in the misfit kernel. This bias can be reduced by appropriate scaling of the adjoint sources based on the local station or receiver density. Weighting may also be considered to balance the sensitivities of different wave types when they are combined in the same inversion.

2.5 Local resolution analysis using Point Spread Functions

The non-linearity of full-waveform inversion schemes results in a complex relation between the data and the tomographic model. A full quantitative resolution analysis of realistic full-waveform models is currently not feasible, because the computational cost to derive the full Hessian $\mathbf{H}(\mathbf{m}) = \frac{\partial^2 \chi(\mathbf{m})}{\partial \mathbf{m} \partial \mathbf{m}}$

is prohibitively high. The Hessian contains the second-order derivatives of the misfit function $\chi(\mathbf{m})$, describing its local curvature. Computed for the optimum model $\mathbf{m}_{opt}$, the Hessian provides information on the resolution of the individual elements of the model vector $\mathbf{m}$, and their mutual trade-offs (e.g. Fichtner and Trampert, 2011b).

Although the full Hessian is not available for realistic full-waveform models, Fichtner and Trampert (2011b) describe a computationally affordable technique to evaluate the Hessian, applied to a localised model perturbation $\delta\mathbf{m}$, for the optimum model $\mathbf{m}_{opt}$. The resulting response is referred to as Point Spread Function (PSF), and allows an assessment of the resolution of localised features in a full-waveform model.

To compute a PSF, a smooth, point-localised Gaussian perturbation $\delta\mathbf{m}$, centred on the localised feature to be assessed, is superimposed onto $\mathbf{m}_{opt}$:

$$\mathbf{m}_{pert} = \mathbf{m}_{opt} + \delta\mathbf{m}. \tag{2.16}$$

A 'perturbed' gradient $\mathbf{g}(\mathbf{m}_{pert})$ is then obtained with one forward simulation for the perturbed model and one adjoint simulation for the optimum model, per event. The adjoint source is hereby based on the misfit between perturbed synthetics and data. Similarly, an 'unperturbed' gradient $\mathbf{g}(\mathbf{m}_{opt})$ is obtained with forward and adjoint simulations both for the unperturbed optimum model. The adjoint source is based on the misfit between unperturbed synthetics and data.

The PSF, representing the linearised response of the model to the perturbation $\delta\mathbf{m}$, is obtained by subtracting the 'unperturbed' gradient from the 'perturbed' gradient:

$$\mathbf{H}(\mathbf{m})\delta\mathbf{m} = \mathbf{g}(\mathbf{m}_{pert}) - \mathbf{g}(\mathbf{m}_{opt}). \tag{2.17}$$

The PSF can be understood as the first iteration of a fully non-linear 'spike test', representing a conservative first estimate of the local model resolution and possible trade-offs. Multiple iterations can be performed using a conjugate gradient scheme, resulting in an improved recovery of the spike.

Using PSFs, samples of the local resolution of a full-waveform model can be computed for various model locations. The first sample requires two forward and two adjoint simulations per event to construct the perturbed and unperturbed gradients. Because the unperturbed gradient can be re-used, each additional sample requires one forward and one adjoint simulation per event to construct the perturbed gradient.

Chapter 3

Finite-frequency effects in seismic wave propagation

The complexity of the seismic wavefield increases considerably when the scale-length of the heterogeneities decreases below the width of the first Fresnel zone. While large-scale lateral velocity anomalies mainly result in an advance or delay of a transmitted waveform with respect to that in a reference model, small-scale heterogeneities give rise to a strongly diffracted wavefield. The diffracted waves interfere with the transmitted waves, increasingly concealing the acquired advance or delay with growing propagation distance from the anomaly. Ray theory ignores diffraction effects and implicitly makes the assumption that the full information is contained in the transmitted wavefield. The discrepancy between predicted ray-theoretical traveltimes and observed traveltimes is referred to as wavefront healing, because the perturbed wavefront 'heals', or rather is increasingly concealed, by interfering diffracted waves. In the last section of this chapter, we investigate the perturbation of the seismic wavefield by a cylindrical velocity anomaly, reminiscent of a mantle plume, in a synthetic 3-D simulation. The full wavefield is computed, and both, a fast and a slow anomaly, are considered.

3.1 Diffraction

Diffraction has been studied in all disciplines of science dealing with wave propagation. Different types of waves, for instance electromagnetic waves,

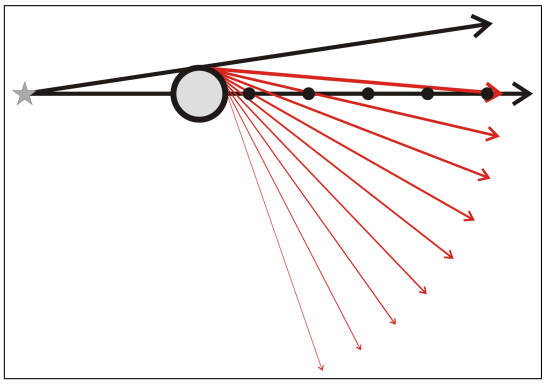


Figure 3.1: Scheme depicting diffraction from a circular anomaly, according to the theory of Keller (1962). The grey star denotes the source, the grey circle a velocity anomaly with respect to the surrounding medium, and the black dots receivers at different distance behind the anomaly. The black lines indicate direct rays, one travelling through the center of the anomaly, the other with a tangential incident on the anomaly boundary. The red lines indicate diffracted rays which are emitted tangential to the anomaly. The varying thickness of the red lines indicates the (exponential) amplitude decay of the diffracted rays with increasing propagation distance along the boundary of the anomaly.

water waves, sound waves, particle waves and of course seismic waves, are all governed by the same diffraction principles. An intuitive way to understand the occurrence of diffraction is the Huygens-Fresnel principle. It states that every point along a propagating wavefront is the source of a spherical secondary wavefront. For an infinitely long or unbounded wavefront of constant amplitude and constant propagation speed, all the components of the secondary waves that do not propagate normal to the wavefront interfere destructively. As a result, the wavefront advances unperturbed.

If a fraction of the wavefront passes through a velocity anomaly, the wavefront is delayed or advanced locally, depending on the sign of the anomaly. The secondary spherical waves emitted from the unperturbed wavefront do not fully cancel out any more and generate a diffracted wavefront which 'bends' around the anomaly, thereby emitting waves in directions tangential to the anomaly.

In his 'Geometrical theory of diffraction', Keller (1962) presented a ray-based geometrical theory which introduced diffracted rays in addition to direct or transmitted rays. This theory is strongly simplifying, but it ap-

proximately describes the properties of the diffracted wavefield an an intuitive way. The main properties of the diffracted rays according to Keller's theory were summarised by Malcolm and Trampert (2011). Three main properties were identified, these are: (1) The source of the diffracted rays are those rays with a tangential incident on the boundary of an anomaly. The diffracted rays also leave the anomaly in a direction tangential to the boundary. (2) The amplitude of diffracted rays decays exponentially with increasing distance travelled along the anomaly boundary. Since diffracted rays are constantly emitted during the propagation along the boundary, the amplitude of the leaving ray must be proportional to the amplitude of the ray propagating along the boundary. This is expressed as

$$\frac{da}{ds} \propto -a \Rightarrow a \propto e^{-s}, \tag{3.1}$$

with a denoting the amplitude of the diffracted ray and s the arc length along the anomaly. (3) The strength of the diffracted ray is proportional to $\sqrt{\lambda}$, where λ is the wavelength. Diffracted rays are therefore stronger for lower frequencies. In the high-frequency limited, the diffracted rays vanish.

In Fig. 3.1, we illustrate the generation of diffracted rays by a circular wavespeed anomaly according to Keller's theory. The anomaly represents, for instance, the low-velocity conduit of an idealised mantle plume. Rays are emitted from a point source, and the transmitted rays traversing the anomaly acquire either a traveltime advance or delay, depending on the sign of the velocity perturbation within the anomaly. Those rays with a tangential incidence angle onto the anomaly boundary give rise to diffracted rays, which are continuously emitted in directions tangential to the boundary during propagation of the ray along the boundary. The amplitude of the emitted diffracted rays decreases exponentially with the propagation distance along the boundary, the diffracted rays emitted at a larger angle with respect to the original ray are therefore weaker. This simple scheme indicates that a simple anomaly gives rise to a rather complex field of diffracted rays. The diffracted rays interfere with the transmitted rays, partly concealing time delays or advances carried by the transmitted rays. In the following sections, we refer to Fig. 3.1 to further explain the characteristics of finite-frequency wave propagation effects.

Keller's theory explains the basic principle of diffraction, which is sufficient for understanding the mechanism of the finite-frequency effects discussed in this chapter. However, the theory is simple and describes real observations only approximately. Through synthetic experiments, Malcolm

and Trampert (2011) confirmed that the theory predicts the diffraction of seismic waves reasonably well. Remaining differences are likely to be caused by the sensitivity of seismic waves to a finite volume surrounding the ray path, the diffracted waves therefore sample the interior of the anomaly to some extent. This is not accounted for in Keller's ray-based theory, which assumes the diffracted rays to be independent from the sign of the anomaly.

3.2 Wavefront healing

With the evolution of tomographic techniques and models, it became evident that ray theory does not explain the observed decay of traveltime residuals with increasing propagation distance behind small-scale anomalies. This was attributed to interfering diffracted waves by Wielandt (1987). Later studies investigated this effect, which is referred to as wavefront healing, in more detail (e.g. Nolet and Dahlen, 2000; Dahlen et al., 2000; Hung et al., 2001; Malcolm and Trampert, 2011).

The mechanism of wavefront healing can be inferred from Fig. 3.1. From a ray-theoretical perspective, a wave travelling through a velocity anomaly acquires a traveltime advance or delay. The direct ray through the center of the anomaly in Fig. 3.1 represents such a transmitted wave. The information loss of the transmitted wave is a result of diffracted waves interfering with the transmitted waves, concealing their traveltime shifts. The pattern of this interference depends on the relative timing and amplitude of the transmitted and diffracted waves.

For receivers behind the plume, the decrease of traveltime residuals with increasing distance from the anomaly results from two effects: (1) The amplitude of diffracted waves arriving at a receiver further from the anomaly is larger, because their propagation distance along the anomaly boundary is shorter. (2) The relative difference in the source-receiver path length between transmitted and diffracted waves decreases with increasing distance from the anomaly. In case of a low-velocity anomaly, this leads to an earlier arrival of the diffracted wave relative to the transmitted wave at a certain distance beyond the anomaly. The reason is that the diffracted wave does not traverse the anomaly and therefore is not delayed on its path. In case of a high-velocity anomaly, the transmitted wave acquires a phase advance. The transmitted wave now arrives earlier than the diffracted wave at all receiver distances. However, the advanced wavefront quickly loses energy during propagation because waves are emitted sideways. At the same

time, the interfering diffracted waves gain in amplitude. Beyond a certain distance from the anomaly, the resulting waveform is dominated by the diffracted waves.

For a given epicentral distance, the strength of wavefront healing depends on the size of the anomaly and the seismic wavelength. A larger anomaly implies a greater path-length difference between diffracted and transmitted waves, the wavefront heals more slowly. A larger anomaly furthermore prolongs the distance travelled along the boundary of the anomaly, which weakens the diffracted waves. The wavelength λ affects the amplitude of the diffracted waves. According to Keller's theory, their amplitude is proportional to $\sqrt{\lambda}$.

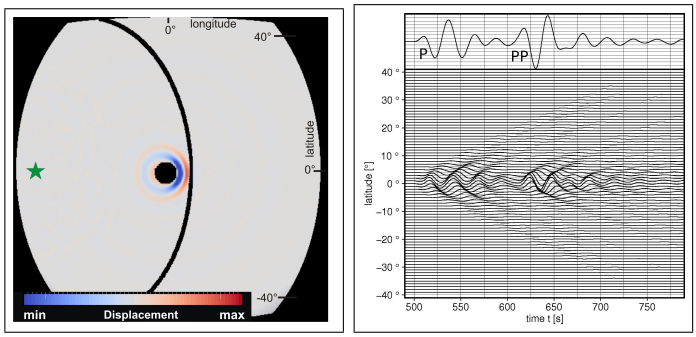
This dependence of wavefront healing on both anomaly size and wavelength can be described in terms of the first Fresnel zone between source and a receiver. Diffracted waves only interfere constructively with the transmitted wave if their additional traveltime is within half a wave period of the transmitted wave. The first Fresnel zone describes all the possible ray paths that fulfil this condition. It is of ellipsoidal shape and its maximum width is $d_F = \sqrt{\lambda L}$, where L denotes the path-length between source and receiver.

Based on this description, it is evident that cross-correlation traveltime shifts (Luo and Schuster, 1991, see also Section 4.4), induced by an anomaly that is large with respect to the first Fresnel zone, are well-described by ray theory. Diffracted waves do not, or only to a small extent, interfere with the transmitted wave in this case. The traveltime shifts induced by smaller-scale anomalies, however, are increasingly influenced by diffracted waves which conceal the traveltime shift of the transmitted waves, resulting in wavefront healing.

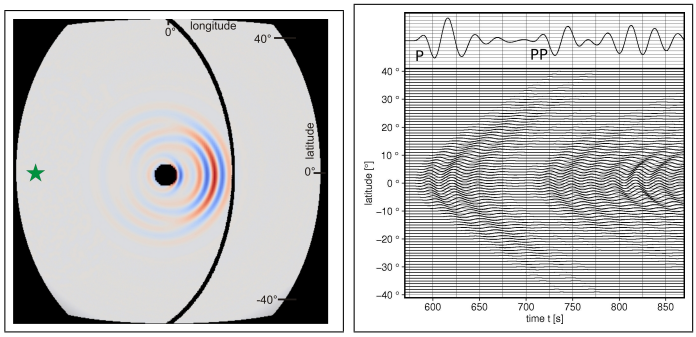
In fact, the term 'wavefront healing' is somewhat misleading. The wavefront does not fully heal and differences are still observed at large distances. It is rather the limitation of the measurement, such as the cross-correlation traveltime or the first-arrival picking, that does not permit the extraction of the information.

3.3 The perturbation of the wavefield by a cylindrical anomaly

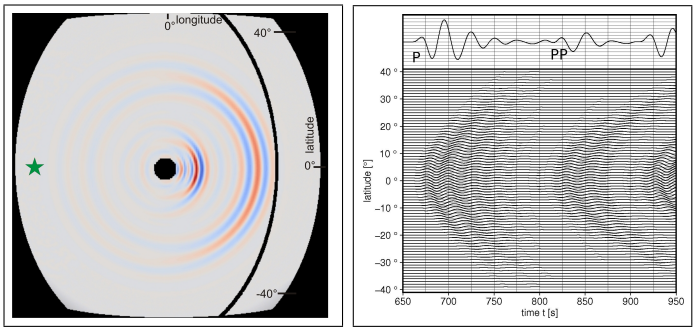
In this section, the perturbation of the seismic wavefield caused by a narrow cylindrical anomaly is investigated, for a slow and a fast anomaly. The full



(a) t_s=520 s

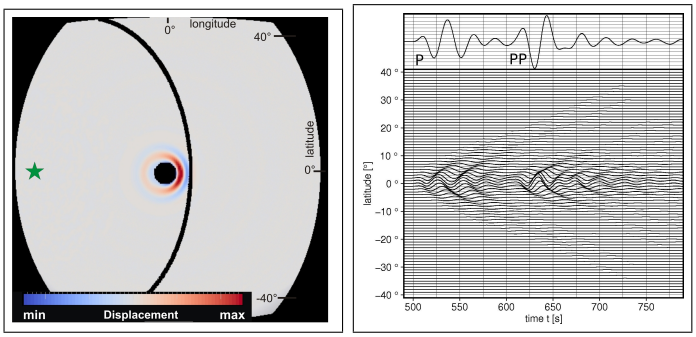


(b) t_s=600 s

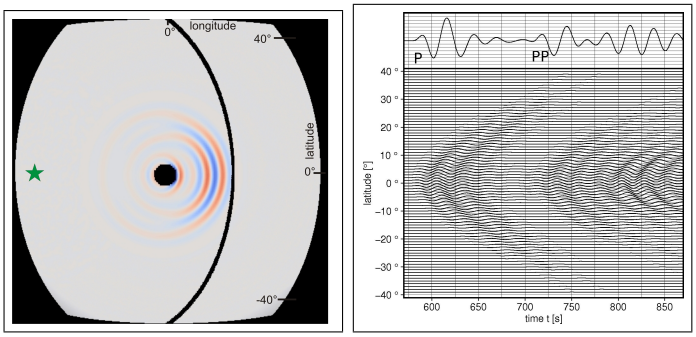


(c) t_s=680 s

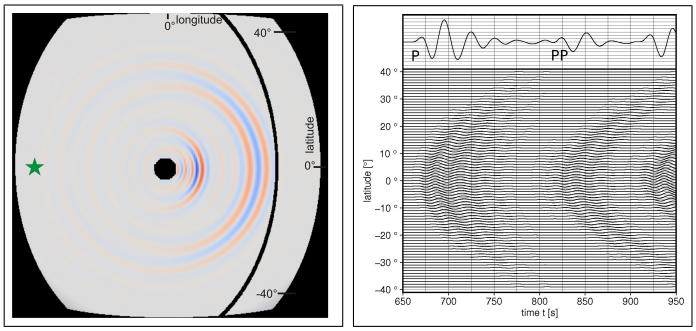
Figure 3.2: Diffraction due to a low-velocity anomaly. Left-hand panels (a)-(c): Spatial wavefield difference $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ at the surface, for snapshots at three different times t_s. The difference results from a cylindrical P-velocity perturbation of -5% with respect to PREM, the diameter of the anomaly is 300 km. The dominant period is 25 s. The curved black line indicates the location of the unperturbed P-wavefront. The same colour scale is used for all snapshots. Right-hand panels (a)-(c): Temporal wavefield difference $u_{dif}(\mathbf{x}_r, t)$ at the seismogram locations $\mathbf{x}_r$. The seismograms are placed at every latitudinal degree along the curved black lines in the left-hand panel, corresponding to the unperturbed P-wavefront. The seismogram at the top is the unperturbed seismogram $u_{ref}(\mathbf{x}_r, t)$, which is identical for all receivers along the unperturbed P-wavefront. The same scale is used to plot $u_{ref}(\mathbf{x}_r, t)$ and $u_{dif}(\mathbf{x}_r, t)$, permitting a comparison of their relative amplitudes.



(a) t_s=520 s



(b) t_s=600 s



(c) t_s=680 s

Figure 3.3: Diffraction due to a high-velocity anomaly. Left-hand panels (a)-(c): Spatial wavefield difference $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ at the surface, for snapshots at three different times t_s. The difference results from a cylindrical P-velocity perturbation of +5% with respect to PREM, the diameter of the anomaly is 300 km. The dominant period is 25 s. The curved black line indicates the location of the unperturbed P-wavefront. The same colour scale is used for all snapshots. Right-hand panels (a)-(c): Temporal wavefield difference $u_{dif}(\mathbf{x}_r, t)$ at the seismogram locations $\mathbf{x}_r$. The seismograms are placed at every latitudinal degree along the curved black lines in the left-hand panel, corresponding to the unperturbed P-wavefront. The seismogram at the top is the unperturbed seismogram $u_{ref}(\mathbf{x}_r, t)$, which is identical for all receivers along the unperturbed P-wavefront. The same scale is used to plot $u_{ref}(\mathbf{x}_r, t)$ and $u_{dif}(\mathbf{x}_r, t)$, permitting a comparison of their relative amplitudes.

seismic wavefield is computed with the spectral-element method.

3.3.1 Method

The wavefield is calculated numerically with the spectral-element code SES3D (Fichtner and Igel, 2008), which accounts for the full complexity of wave propagation in realistic Earth models. The model extends 90° in both lateral directions and 2500 km in depth. The numerical grid consists of hexahedral elements with dimensions of $1^\circ \times 1^\circ \times 50$km, which are deformed to honour the spherical geometry of the Earth. The dominant period of the simulated wavefield is 25 s, which is well-resolved by the numerical grid. The wavefield is calculated for three different models. The first is the 1-D model PREM (Dziewonski and Anderson, 1981). The resulting wavefield $\mathbf{u}_{ref}$ serves as reference for the following perturbations. In the second simulation, a vertically oriented, cylindrical low-velocity perturbation with a diameter of 300 km is placed in the centre of the model. The P-velocity is perturbed by -5% with respect to PREM within the cylinder. Only the P-velocity is perturbed, the other parameters keep their PREM value. The third simulation concerns a fast anomaly, where the P-velocity is perturbed by $+5\%$. The perturbed wavefields are denoted by $\mathbf{u}_{cyl}$.

The source is placed at 0° longitude and -40° latitude. The distance to the anomaly, which sits at 0° latitude and 0° longitude, is 40°. Receivers are placed at the surface along lines of constant epicentral distance. They are spaced to coincide with each latitudinal degree between -40° and 40°, resulting in 81 receivers per line.

Snapshots of the spatial wavefield difference $\mathbf{u}_{dif}(\mathbf{x}, t_s) = \mathbf{u}_{cyl}(\mathbf{x}, t_s) - \mathbf{u}_{ref}(\mathbf{x}, t_s)$ are computed at specific times t_s on the spatial numerical grid $\mathbf{x}$. The perturbed wavefield is denoted by $\mathbf{u}_{cyl}(\mathbf{x}, t_s)$ and the reference wavefield by $\mathbf{u}_{ref}(\mathbf{x}, t_s)$. The temporal wavefield difference, which is the difference of seismograms at specific receiver locations $\mathbf{x}_r$, is obtained as $\mathbf{u}_{dif}(\mathbf{x}_r, t) = \mathbf{u}_{cyl}(\mathbf{x}_r, t) - \mathbf{u}_{ref}(\mathbf{x}_r, t)$.

3.3.2 Results

The results are first presented for the low-velocity anomaly. The left-hand panels in Fig. 3.2 (a)-(c) show snapshots of the spatial wavefield difference $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ at the surface of the model, at times $t_s = 520$ s, $t_s = 600$ s and $t_s = 680$ s. The curved black line denotes the position of the unperturbed P-wavefront at time t_s. The snapshots indicate that the anomaly acts as

a source of a diffracted wavefield. The wavefield difference spreads radially from the anomaly, while the unperturbed wavefront spreads radially from the source. The wavefield difference is distributed over a wide azimuthal range. While the largest difference is observed in the forward direction, weak diffractions are observed at all angles, and even some backscattering occurs. The non-zero values of $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ above a certain receiver azimuth must result from diffracted waves, because the anomaly is sampled by direct waves only within a narrow range of receiver azimuths. Within the azimuthal range that is covered by transmitted waves, the wavefield difference results from the interference of both transmitted and diffracted waves. The maximum of $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ is observed, as expected, behind the anomaly along the central ray path at 0° latitude. It is largest after the wavefront has just passed the anomaly at $t_s = 520$ s. In the later snapshots at $t_s = 600$ s and $t_s = 680$ s, the difference is spread more evenly over a wider azimuthal range.

In the right-hand panels of Fig. 3.2 (a)-(c), the temporal wavefield difference $\mathbf{u}_{dif}(\mathbf{x}_r, t)$ is presented. This is the difference of the seismograms, recorded at receivers along the unperturbed P-wavefront. The P-wavefront is indicated by the curved black line in the corresponding left-hand panels of Fig. 3.2 (a)-(c). The receivers, 81 per line, are placed along the line at every latitudinal degree between $-40°$ and 40°. For reference, an unperturbed seismogram is shown at the top of each panel. The latter is identical for all receivers, since they are located at the same distance from the source. The reference seismogram and the seismogram difference $u_{dif}(\mathbf{x}_r, t)$ are plotted at the same scale, allowing the comparison of their relative amplitudes.

The information of the temporal wavefield difference $u_{dif}(\mathbf{x}_r, t)$ is qualitatively similar to the spatial wavefield difference $\mathbf{u}_{dif}(\mathbf{x}, t_s)$. However, it gives a better picture of the amplitudes and the arrival times of diffracted waves with varying receiver azimuth. A short distance behind the anomaly (Fig. 3.2(a)), the difference in the seismograms is greatest within a narrow azimuthal range behind the anomaly, which is likely due to the travel-time shift of the transmitted waves. The diffractions observed at larger azimuths are comparatively weak. This is consistent with Keller's theory of weak diffracted waves for a large diffraction angle. The late arrival of the diffracted waves compared to the direct P-wave reflects their increasing path-length with growing receiver azimuth. At a larger distance from the anomaly (Fig. 3.2(b)), the seismogram difference is distributed more evenly over a wider azimuthal range. The highest amplitude is recorded

at the receivers in line with source and anomaly, but the transition to the diffracted wavefield at larger receiver azimuths is smooth. At 0° latitude, the amplitude of $u_{dif}(\mathbf{x}_r, t)$ reaches $\sim 15\%$ with respect to the P-wave amplitude. It decays to $\sim 8\%$ at a latitude of 15°. The longer path-length of the diffracted waves is again reflected in their later arrival time relative to the direct P-wave with increasing azimuth. At the furthest receiver distance considered in this experiment (Fig. 3.2(c)), the wavefield difference is distributed relatively evenly over a wide azimuthal range. At this distance, the amplitude of $u_{dif}(\mathbf{x}_r, t)$ relative to the amplitude of the P-wave is $\sim 12\%$ at a latitude of 0° and $\sim 7\%$ at a latitude of 15°.

The same experiment for a high-velocity anomaly is presented in Fig. 3.3. The resulting pattern of $\mathbf{u}_{dif}(\mathbf{x}, t_s)$ and $u_{dif}(\mathbf{x}_r, t)$ are nearly the same, but of the opposite sign. The only notable difference is a slightly lower amplitude of the wavefield difference for the fast anomaly compared to the slow anomaly within a narrow azimuthal range behind the anomaly. This is likely a result of focusing or defocusing of the transmitted waves behind a slow and fast anomaly, respectively.

3.3.3 Discussion and Conclusions

The cylindrical anomaly of diameter 300 km used in this experiment is narrow compared to the width of the first Fresnel zone. At an epicentral distance of 80° and a period of 25 s, the first Fresnel zone of a P-wave has a maximum width of ~ 1400 km. Our results indicate that for such a narrow anomaly, diffraction effects dominate the wavefield difference already at a short distance behind the anomaly. The initial waveform difference, just after the wavefront has travelled through the anomaly, is dominated by the traveltime delay of the perturbed wavefront. With further propagation, it rapidly transforms into a diffraction-dominated pattern which is relatively evenly distributed over a wide azimuthal range. At increasing receiver azimuth, the diffracted waves interfere at a later time with the direct waves. This is a result of the longer path-length travelled by the diffracted waves compared to the direct waves.

Our simulations indicate that diffracted waves are sensitive to the sign of the perturbation, leading to an opposite sign of the diffracted waves for slow and fast anomalies. This is not considered in Keller's theory, where the diffracted waves are independent from the anomaly's interior properties. Similar effects were noticed by Wielandt (1987) and Malcolm and Trampert (2011). They suggest that, due to their finite-frequency nature, real seismic

waves sample the interior of the anomaly during the propagation along its boundary. In a forward-modelling study of CMB-diffracted waves, To and Romanowicz (2009) also observe that diffracted waves are affected by surrounding structures.

For a fast anomaly, the wavefield difference is generally of opposite sign compared to a slow anomaly. The amplitude of the difference is slightly larger behind a slow anomaly. This is likely a result of focusing or defocusing of the transmitted wavefield behind a fast or slow anomaly, respectively (e.g. Hung et al., 2001).

The simple measure of wavefield difference used in this experiment does not permit a clear separation of the waveform difference into acquired phase shifts of transmitted waves and diffraction effects. However, the consistency of the observed pattern with the expected diffraction pattern from theoretical considerations indicates that the effect of the perturbation is mainly observed in the diffracted wavefield.

In the following Chapter 4, finite-frequency effects and their implications for tomographic inversions are investigated in greater detail with a similar experimental setup. We first study the effect of the diffracted wavefield on the measurement and inversion of cross-correlation traveltime residuals. Following that, we demonstrate that the information contained in the diffracted wavefield can be extracted and used in tomographic inversions to constrain anomalies narrower than the width of the first Fresnel zone.

Chapter 4

Imaging mantle plumes with instantaneous phase measurements of diffracted waves: P-waves

4.1 Summary

In a synthetic tomographic experiment, we succeeded to recover an idealised narrow mantle plume reaching deep into the lower mantle by using a misfit based on the instantaneous phase difference. A misfit based on simple cross-correlation traveltime shifts leaves the lower mantle part of the plume largely unresolved, despite the use of finite-frequency sensitivity kernels. The time-continuous and amplitude-independent instantaneous phase misfit allows us to measure the interaction between direct and diffracted waves as a function of time, which is difficult to capture by simple cross-correlation traveltime measurements. The diffracted waves arriving later than the main phase are essential to improve the tomographic result. The measurement of diffracted waves yields the necessary information to recover the plume correctly even in the lower mantle. The instantaneous phase measurement is ideal to capture this interaction, but other time- or

The content of this chapter was published as: Rickers, F., Fichtner, A. and Trampert, J., 2012. Imaging mantle plumes with instantaneous phase measurements of diffracted waves. Geophys. J. Int. 190, 650–664. doi: 10.1111/j.1365-246X.2012.05515.x

frequency-dependent measurement may give similar results. We also investigated the effect of wavefront healing on cross-correlation traveltime shifts for a range of differently sized idealised mantle plumes. We confirm that wavefront healing severely reduces traveltime shifts when the plume conduit is considerably thinner than the width of the first Fresnel zone. For plume conduits with a diameter on the order of 100 km, even traveltime shifts measured at periods as short as $T = 1$ s are affected.

4.2 Introduction

Morgan (1971) suggested thin, hot plumes as a second mode of convection of the Earth's mantle besides the large-scale convection related to plate tectonics. Assumed to develop from instabilities in the core-mantle thermal boundary layer and rising to the surface, mantle plumes are a possible explanation for intra-plate volcanism observed at locations such as Hawaii and for locally increased basalt production along ridges as in Iceland. Plumes are postulated to consist of a narrow plume conduit connecting the plume head to the source region of the hot material. Indirect evidence in favor of the plume theory comes from geology, geochemistry as well as from numerical and laboratory fluid dynamics. However, alternative explanations for the surface observations cannot be ruled out (e.g. Foulger et al., 2005).

While the plate mode has been well-documented by seismic tomography (e.g. Fukao et al., 2001), results concerning mantle plumes are inconclusive. In the upper mantle, low seismic wavespeed regions have been observed below many of the known hotspots in global and regional tomographic models (e.g. Wolfe et al., 1997; Allen et al., 2002; Christiansen et al., 2002; Ritsema and Allen, 2003). A small number of tomographic studies show the extension of upper-mantle low-velocity anomalies into the lower mantle using either dense regional networks (Bijwaard and Spakman, 1999; Wolfe et al., 2009) or finite-frequency sensitivity kernels (Montelli et al., 2004, 2006b). It is not established yet whether these low-velocity observations can be unambiguously linked to mantle plumes.

The possibly small horizontal extent of plume conduits poses a challenge to tomographic methods especially in the lower mantle. Whenever a low-velocity anomaly is not fully contained within the first Fresnel zone, seismic waves can diffract around it and interfere constructively with the directly travelling wave. The waveform of the directly travelling wave becomes distorted, which can lead to a decrease in the measurable time shift

(Wielandt, 1987; Nolet and Dahlen, 2000; Malcolm and Trampert, 2011). This effect, referred to as wavefront healing, generally tends to become stronger with decreasing ratio of anomaly to Fresnel zone size and with increasing epicentral distance. If lower-mantle plumes are on the order of 100 km in diameter, even high-frequency waves are affected by diffraction effects - the first Fresnel zone width of a $T = 1$ s P-wave exceeds 300 km in the lower mantle.

Hwang et al. (2011) studied S-wave traveltime shifts for a range of possible plume geometries derived from geodynamic models. They conclude that wavefront healing renders plumes seismically invisible in the lower mantle, even at periods as short as $T = 5$ s. We agree with their analysis when simple traveltime measurements are used. They make it difficult to capture the details of the interference between direct and diffracted waves and therefore leave part of the available information unused.

In an idealised synthetic experiment, we confirm that traveltime shifts of deep mantle plumes are strongly reduced by wavefront healing when the width of the first Fresnel zone is considerably larger than the plume diameter. We perform an iterative adjoint inversion based on the measured traveltime shifts of such a narrow plume and do not succeed to recover the lower-mantle part, despite the use of finite-frequency sensitivity kernels and a non-linear optimisation scheme. A thicker plume with a diameter comparable to the width of the first Fresnel zone is well-recovered.

A better choice of a misfit function is the instantaneous phase difference (Bozdağ et al., 2011), a time-continuous, amplitude-independent measure of phase differences between waveforms that can capture the details of the interference between diffracted and direct waves. Using this misfit and a sufficiently large measurement window to capture the diffracted waves, we show that an idealised mantle plume with a diameter significantly smaller than the width of the first Fresnel zone can be imaged correctly, in shape as well as in amplitude, even at depth.

4.3 Method

4.3.1 Numerical method

We simulated the forward and adjoint wavefields using the spectral-element code SES3D (Fichtner and Igel, 2008). It solves the weak form of the elastic wave equation on a hexahedral grid in spherical coordinates. SES3D is

designed for regional- and continental-scale simulations, and easily adaptable to our purpose. At the model boundaries, waves are absorbed by a perfectly matched layer (PML) to suppress reflections.

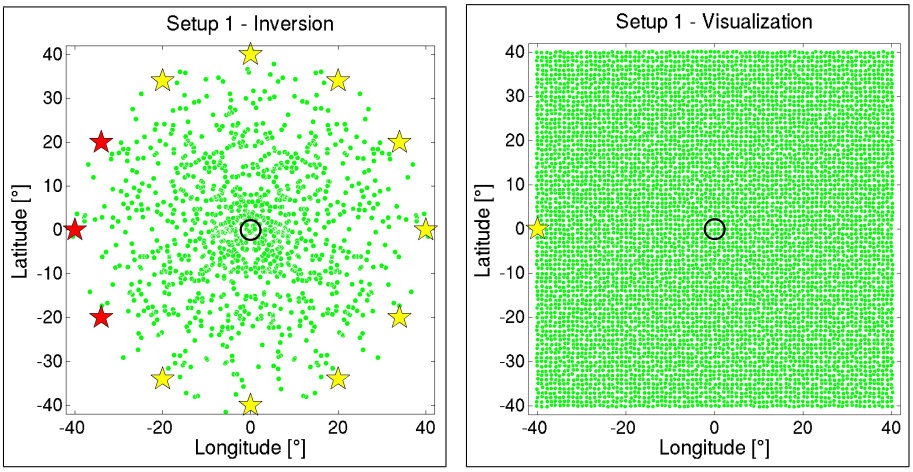
4.3.2 Setup of the synthetic experiments

Model

We aim to study the lower mantle, which imposes certain requirements on the horizontal and vertical extent of the model volume. We defined it to extend to 2500 km in the vertical and 90° in the latitudinal as well as in the longitudinal direction. The computational grid is divided into $108 \times 108 \times 36$ elements in the latitudinal, longitudinal and vertical directions. This results in a maximum element size of approximately $90 \times 90 \times 70$ km at the surface close to the equator. Moving away from the equator and with increasing depth, the element size decreases due to the spherical coordinate system. Each element contains 125 grid points on which the wave equation is discretised. To ensure a reliable solution, one wavelength should be discretised over at least two elements, which limits the minimum wavelength that can be accurately modelled to about $\lambda = 180$ km. In the case of P-waves, which we focus on in this study, an accurate simulation is ensured by using a minimum wave period of $T = 25$ s. This requirement is fulfilled by low-pass filtering the source-time function accordingly. The model parameters themselves are defined on blocks with a side length of 45 km, the model parameters are constant within each block. For the maximum epicentral distance of 80° that we use in this study, rays reach a maximum depth of around 2300 km according to ray theory. As a starting model for the tomographic inversion we chose PREM (Dziewonski and Anderson, 1981) with a uniform, 24 km thick crustal layer and no ocean. The minimum P-velocity is 6.8 km/s.

Implementation of the synthetic plume

The target model in which the data are calculated consists of PREM with a superimposed, centrally placed vertical cylinder of reduced P-velocity α, which we refer to as 'plume' throughout this study. We are mainly interested in studying the lower part of plumes and consider this a sufficient approximation to a plume conduit, without worrying about the possible geometry of a plume head. Within the plume, α is reduced by 5% with respect to PREM. The boundaries of the cylinder are slightly smoothed by



(a) Inversion setup (b) Visualization setup

Figure 4.1: Location of events (red and yellow stars) and receivers (green dots) on the model surface. The red events denote actually computed events, the yellow events are rotated versions of the red events. The black circle indicates the location of the plume. (a) Configuration used for the tomographic inversions. (b) Configuration used for the visualisation of the spatial misfit distribution.

averaging over neighbouring model blocks to prevent strong velocity contrasts between discretisation points. As long as the smoothing is moderate it does not affect the computed seismograms visibly. We confirmed this by testing various degrees of smoothing and inspecting the resulting seismograms. We studied plumes of different diameters, from 220 to 1000 km. The tomographic inversions are performed for plumes of diameter $d = 300$ km and $d = 1000$ km, using a dominant P-wave period of $T = 25$ s.

Source and receiver distribution

The source is modelled as an explosive point source which leads to a simple radiation pattern. We use two different source-receiver configurations throughout this paper (Fig. 4.1):

1. For the tomographic inversions, we want to ensure a good illumination of the plume at all depths from various angles. This is achieved by placing twelve seismic sources evenly around the plume, at distances of 40° from the plume axis and at a depth of 10 km. Receivers are distributed in a

random manner around the plume, with a maximum epicentral distance of 80°. Because we want to test the limits of the resolution capabilities of different methods, we use 620 receivers which ensures a better plume illumination than would usually be achievable in a realistic tomographic inversion. We are here only imaging the P-velocity α, and therefore all measurements are made on the vertical component of the seismograms.

To save computational resources, and because we are dealing with a problem that is approximately symmetric with respect to the plume axis, we actually compute the event kernels - volumetric sensitivities combining all sensitivity kernels for a single event - for three of the sources only, marked red in Fig. 4.1(a). Each of these three event kernels is then rotated with respect to the vertical axis by 90°, 180° and 270° and the rotated and original kernels are superimposed, resulting in a misfit kernel consisting of 12 sources.

2. For the purpose of visualising the spatial distribution of measurable misfits, we place one source at 40° distance from the plume, and cover the whole model surface by a dense receiver grid with interstation-distances of about 1°. By evaluating a chosen misfit measure at all receivers, we can visualise the spatial misfit distribution at the surface.

4.3.3 Inversion technique

We used a fully three-dimensional iterative algorithm to perform the nonlinear tomographic inversion. The aim is to minimise a misfit function $\chi(\mathbf{m})$, which is defined in terms of a chosen measure of misfit between observed and synthetic seismograms, $\mathbf{m}$ is the vector of the current model parameters, in our case the P-velocity α. We considered two misfit functions in this study, based on cross-correlation traveltime shifts and on the instantaneous phase difference, respectively.

We used a conjugate gradient algorithm (Fletcher and Reeves, 1964) to minimise the misfit function, which has been successfully used in various tomographic studies (e.g. Fichtner et al., 2009; Tape et al., 2007). The conjugate gradient algorithm iteratively updates the starting model (PREM in our case) towards a model with a smaller misfit $\chi(\mathbf{m})$, using the information provided by the gradient $\mathbf{g}(\mathbf{m})$ of the misfit function. The gradient is calculated at every iteration step via the adjoint method (Tarantola, 1984; Tromp et al., 2005; Fichtner et al., 2006). We ended the iterative procedure when the value of the misfit function had decreased by 95% with respect

to the initial misfit.

We applied pre-conditioning to the gradient at each iteration step to improve the convergence of the conjugate gradient algorithm. By reducing the value of the gradient where it is extremely large, based on a threshold value, we accounted for the often high sensitivities in the immediate vicinity of sources and receivers, which are physically not justified and result from singularities. We furthermore applied Gaussian smoothing to the gradient at each iteration step, over a length-scale of 200 km, approximately corresponding to the dominant wavelength. This is intended to prevent unresolved small-scale artefacts in the tomographic model.

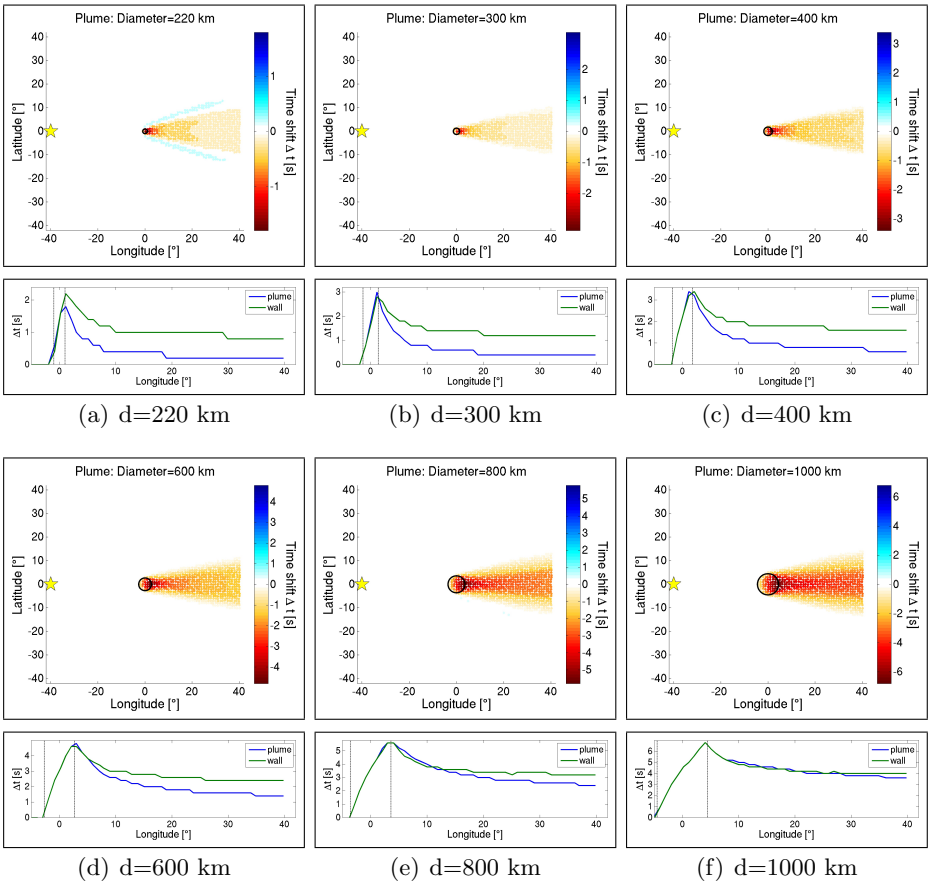
4.4 Misfit 1: Cross-correlation traveltime shifts

4.4.1 Measurement of cross-correlation traveltime shifts for plumes of different diameter

We measured traveltime shifts Δt by cross-correlating the observed (PREM + plume) and synthetic (PREM alone) waveforms, which have been windowed using a cosine taper to isolate the P-wave. The length of the tapering window is 25 s, which corresponds to the dominant P-wave period of the forward wavefield. Because we are analysing P-waves, we performed the measurement on the vertical component only. Prior to an inversion, we computed and compared traveltime shifts for a range of differently sized mantle plumes. For all plume diameters, we used a dominant P-wave period of $T = 25$ s, corresponding to a wavelength in the lower mantle of approximately $\lambda = 300$ km. The diameter of the first Fresnel zone at the maximum epicentral distance of 80° reaches approximately 1400 km.

We varied the plume diameter between 220 and 1000 km, covering a range expected to show different susceptibilities to wavefront healing. We performed a simulation of the forward wavefield for each plume model, using the dense receiver grid in Fig. 4.1(b). Computing Δt at each receiver, we obtained a visual impression of the spatial distribution of traveltime shifts for the differently sized plumes. The upper panel of each subpart in Fig. 4.2 shows the misfit distribution.

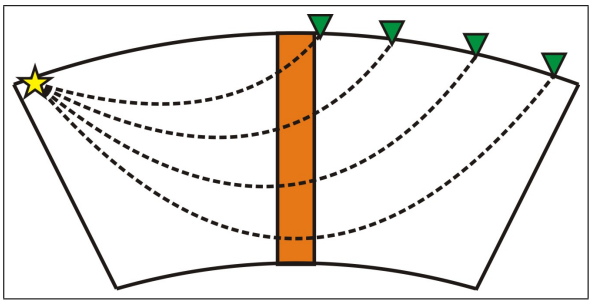
The lower panel of each subpart in Fig. 4.2 shows the traveltime shifts at receivers behind the plume along the equator, as a function of epicentral distance. A decrease of Δt with increasing distance from the plume is observed for all plumes. This decrease is partly due to a decrease in the path



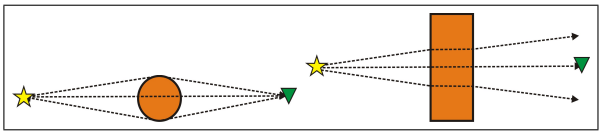
(a) d=220 km (b) d=300 km (c) d=400 km

(d) d=600 km (e) d=800 km (f) d=1000 km

Figure 4.2: Upper panel of subparts (a) to (f): P-wave cross-correlation travel time shifts for plumes of different diameter, measured in the dense station network shown in Fig. 4.1(b). The dominant P-wave period is $T = 25$ s, the source depth is 10 km and the perturbation of the P-wave speed within the plume is -5% with respect to PREM. The yellow star indicates the location of the source, the black circle the location and thickness of the plume. The measured time shifts at each station are indicated by the colour. Lower panel of subparts (a) to (f): Absolute value of the travel time shifts for receivers located on the direct source-plume path along the equator (blue). For comparison, approximate travel time shifts without the influence of wavefront healing are given (green). For the computation of the latter, the plume has been replaced by a vertical wall of the same thickness. The black dotted lines indicate the extent of the plume.



(a) Path length within the plume (cross-section through equatorial plane)



(b) Plume vs. wall (cross-section through horizontal plane)

Figure 4.3: (a) The path length travelled by the wave within the plume decreases with increasing epicentral distance. (b) Left: For a plume of diameter smaller than the width of the first Fresnel zone, the diffracted waves interfere constructively with the direct wave at the receiver which can lead to wavefront healing. Right: If the plume is replaced by a vertical wall of the same thickness, no diffraction occurs and the direct wave, uninfluenced by wavefront healing, is measured.

length that the wave travels within the plume for increasing epicentral distance (see Fig. 4.3(a) for a cartoon), and partly due to wavefront healing. To estimate the influence of wavefront healing alone, we show for comparison the corresponding traveltime shifts without the influence of wavefront healing. The latter are estimated by replacing the plume with a vertical wall in latitudinal direction, having the same thickness as the plume. Waves cannot diffract around this wall, and the measured traveltime shifts are due to path differences alone. The cartoon in Fig. 4.3(b) shows how wavefront healing affects the traveltime for a plume, but not for a vertical wall of the same thickness as the plume. The strength of wavefront healing can then be estimated from the difference between the measured time shifts for the plume and the wall.

As documented by previous studies (e.g. Hung et al., 2001; Malcolm and Trampert, 2011; Hwang et al., 2011), the influence of wavefront healing on

traveltime shifts increases with increasing epicentral distance as well as with decreasing plume diameter. For a plume of diameter $d = 300$ km, the traveltime shift has decreased by two third due to wavefront healing already at 20° behind the plume (Fig. 4.2(b)). With increasing plume diameter, the strength of wavefront healing decreases. However, even for a plume of diameter $d = 600$ km, Δt is reduced by 50% at the largest epicentral distance of 80°(Fig. 4.2(d)). A plume of diameter $d = 1000$ km is only weakly affected by wavefront healing for the range of epicentral distances considered in our experiment (Fig. 4.2(f)). Although the maximum width of the first Fresnel zone still exceeds the plume diameter in this case, the diffracted waves do not severely distort the waveforms any more. For a plume of diameter $d = 220$ km, which is below the dominant wavelength in the lower mantle, wavefront healing is very strong. We even observe lines of weakly positive traveltime shifts surrounding the area of negative traveltime shifts, indicating strong diffraction effects (Fig. 4.2(a)).

4.4.2 Inversion of the traveltime shifts

The misfit function $\chi(\mathbf{m})$ for cross-correlation traveltime shifts Δt is defined as

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}[\Delta t_r(\mathbf{m})]^2, \tag{4.1}$$

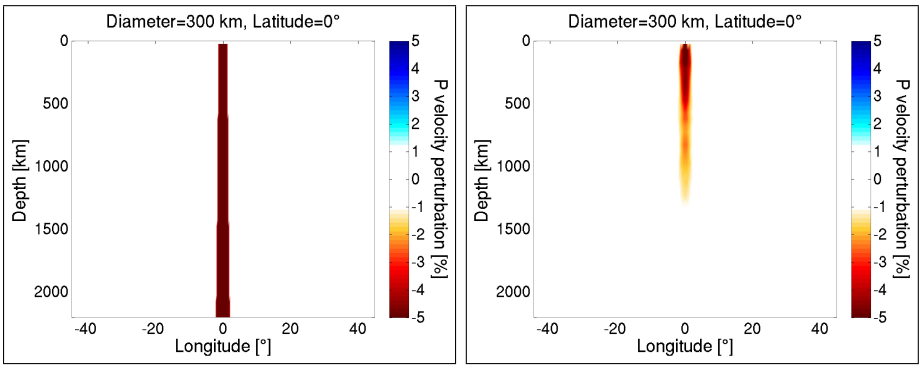
where r denotes the receiver index and N the total number of receivers.

The corresponding adjoint source $f^\dagger(\mathbf{x}, t, \mathbf{m})$ is computed as (Luo and Schuster, 1991; Tromp et al., 2005)

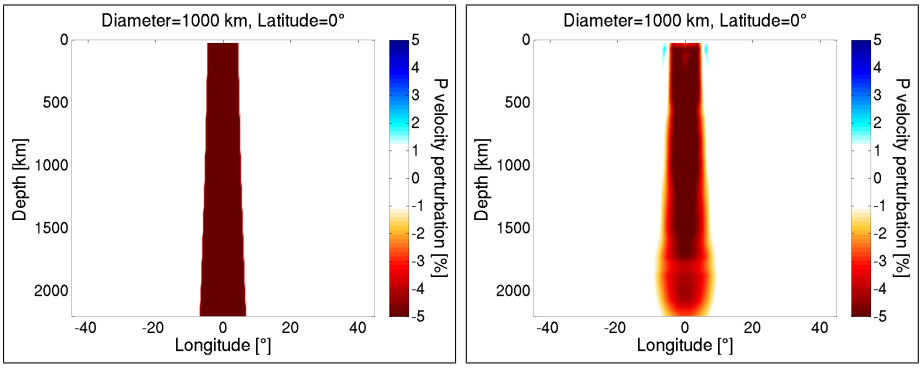
$$f^\dagger(\mathbf{x}, t, \mathbf{m}) = -\sum_{r=1}^{N}\Delta t_r(\mathbf{m})\frac{\dot{u}_{ref}(\mathbf{x}_r, T-t, \mathbf{m})}{\|\dot{u}_{ref}(\mathbf{x}_r, T-t, \mathbf{m})\|^2}\delta(\mathbf{x}-\mathbf{x}_r), \tag{4.2}$$

where T is the end time of the forward simulation.

We used a conjugate gradient algorithm to minimise the cross-correlation misfit function in eq. (4.1). We performed the inversion for a plume of diameter $d = 300$ km, using the source-receiver configuration in Fig. 4.1(a) which ensures a good plume illumination at all depths. The plume diameter is significantly smaller than the maximum width of the first Fresnel zone at the dominant period of $T = 25$ s. The latter has a width of approximately 1400 km in the lower mantle. For comparison, we perform the



(a) d=300 km



(b) d=1000 km

Figure 4.4: (a) Left: Original plume with diameter $d = 300$ km. Right: Recovered plume using cross-correlation travel time shifts. (b) Left: Original plume with diameter $d = 1000$ km. Right: Recovered plume using cross-correlation travel time shifts.

same inversion for the larger plume of diameter $d = 1000$ km as well, which is barely affected by wavefront healing.

If wavefront healing can be accounted for by the use of finite-frequency kernels, we expect both plumes to be recovered reasonably well in the tomographic inversion. We stopped the inversions when $\chi(\mathbf{m})$ had been reduced by 95% with respect to the initial misfit $\chi(\mathbf{m_0})$. This was the case after 15 iterations for the $d = 300$ km plume and after 10 iterations for the $d = 1000$ km plume.

On the left hand side of Fig. 4.4, we show the original plumes which

have been used to generate the observed data. On the right hand side, the corresponding final tomographic models are presented. The figure displays the relative P-velocity perturbation α with respect to PREM in a cross-section through the equatorial plane. The colour scale is adjusted to show only those parts of the plume which are imaged with at least 20% of the original velocity perturbation. The apparent thickening of the plumes towards the bottom results from the cartesian coordinate system that is used for plotting, while the actual model is defined in a spherical coordinate system.

It is clearly visible that the loss of information due to wavefront healing is reflected in the tomographic model of the thinner plume. It can only be considered well-resolved within the upper 500 km of the mantle. Down to 1200 km, part of the plume is recovered with at least 20% of its original amplitude. Below that depth, the plume could not be recovered. The thicker plume is recovered well throughout the whole depth of the model. In the lowermost 500 km, the plume is slightly broadened and weakened compared to the original plume.

4.4.3 Discussion of traveltime results

As the tomographic model in Fig. 4.4(a) shows, we cannot image mantle plumes of a diameter much smaller than the size of the first Fresnel zone with simple travel time shifts. Despite the use of fully three-dimensional wave propagation, an idealised source-receiver configuration and finite-frequency sensitivity kernels, wavefront healing is not properly accounted for.

Simple cross-correlation measurements give only one value to describe the difference between the waveforms over the length of the measurement window. In fact there are different waves arriving within the P-wave window (the diffracted and the direct waves), but we are unable to distinguish their individual arrivals with a single measurement. The cross-correlation measurement implicitly makes the assumption of a constant time shift over the whole length of the P-wave window. As a result, the inversion algorithm converges towards an anomaly which is large enough to be consistent with such a constant time shift, leading to a recovered anomaly determined by the size of the first Fresnel zone. Due to an increase of the volume of the Fresnel zone and a decrease of time shifts with increasing epicentral distance, the deeper part of the $d = 300$ km plume is strongly broadened and weakened with respect to the original plume. The $d = 1000$ km plume

(Fig. 4.4(b)) is well-recovered, the slight broadening and weakening in the lowermost part can be explained by the slightly smaller plume diameter compared to the first Fresnel zone width at this depth.

We can scale these results, which are derived at rather long periods and for a thick plume, to shorter periods and a thinner plume. The influence of diffraction effects on cross-correlation traveltimes is dependent on the ratio of anomaly width d to Fresnel zone width d_F (Hung et al., 2001; Baig et al., 2003). The Fresnel zone width d_F for a given epicentral distance is proportional to the square root of the wavelength: $d_F \propto \sqrt{\lambda}$. Our tomographic result for a plume of diameter $d = 300$ km and a dominant P-wave period of $T = 25$ s is therefore similar to a $d = 100$ km plume and a dominant P-wave period of $T = 3$ s, as their respective ratios d/d_F are comparable.

4.5 Misfit 2: Instantaneous phase difference

4.5.1 Motivation for choosing the instantaneous phase difference

The results from the traveltime based inversion indicate that it is not sufficient to characterise the misfit between waveforms in the presence of diffraction effects by only one single traveltime measurement. To capture the details of the interaction between diffracted and direct waves, a measurement is required that can account for complexities within the measurement window. Such a measurement also enables us to take those diffracted waves into account which arrive later than the main phase, simply by extending the measurement window. Extending the measurement window for a cross-correlation measurement will not result in improvements, because the measured traveltime shift will still be dominated by the high-amplitude parts of the signal.

An ideal misfit for our purpose in this synthetic experiment is the instantaneous phase misfit (Bozdağ et al., 2011). It is derived in the time domain and provides a time-dependent measure of the phase difference, independent of amplitude. The last property is of great importance, because it enables us to consider the phase information contained in lower-amplitude parts of the signal in the tomographic inversion with an equal weight compared to main phases. A further advantage of a pure phase measurement is that it is less non-linear with respect to the model parameters that we

invert for, compared to misfits that mix phase and amplitude information. This improves the convergence of the misfit function towards its (global) minimum during the iterative inversion. The instantaneous phase misfit is not the only possible choice, other misfits designed to extract more information from waveforms include:

- Time-frequency misfits which are used by Fichtner et al. (2009). They are similar to the instantaneous phase misfit, with the difference that the latter is derived purely in the time-domain. We therefore avoid a transformation to the time-frequency domain which involves an averaging process imposed by the width of the sliding window.

- Multitaper measurements which are used by Tape et al. (2009, 2010) for time windows automatically selected by FLEXWIN (Maggi et al., 2009). The measurement is aimed at obtaining the frequency-dependent transfer function which transforms the synthetic waveform within each selected time window into the observed waveform.

- Cross-correlation measurements at multiple frequency bands (Sigloch and Nolet, 2006; Sigloch et al., 2008).

- Waveform difference misfits (e.g. Tarantola, 1984; Tromp et al., 2005), which are the most straightforward measurement by subtracting one seismogram from the other and inverting for the L_2-norm of the time-dependent difference. It has not proven suitable for our purposes, mainly because it mixes amplitude and phase information which results in strongly non-linear behaviour of the measurement with respect to the model parameters (Luo and Schuster, 1991). It is furthermore dominated by the high amplitudes of the signal, favouring the direct rather than the diffracted wave.

We chose the instantaneous phase misfit because its implicit amplitude-independence and high time resolution are not easily achievable by other misfits. The instantaneous phase misfit has, to our knowledge, not been applied to actual tomographic inversions previously. Although the concept of the instantaneous phase has been known for decades, it has mainly been used for visualisation purposes in exploration seismics (Taner et al., 1979; Barnes, 2007). The theoretical framework necessary to use the instantaneous phase difference in adjoint tomography has been derived by Bozdağ et al. (2011). They present Fréchet kernels based on instantaneous phase misfits in their paper.

4.5.2 Theory: Mathematical derivation of the instantaneous phase difference

Because a real-valued seismic time series $u(t)$ cannot simply be assigned a phase at an instant in time, we compute the complex analytical signal $a(t)$, which has the original signal $u(t)$ as real part and the negative of its Hilbert transform $H\left[u(t)\right]$ as imaginary part

$$a(t) = u(t) - iH\left[u(t)\right]. \tag{4.3}$$

The Hilbert transform $H\left[u(t)\right]$ is a phase shifted version of $u(t)$, with each frequency shifted by $\frac{\pi}{2}$.

Being a complex signal, $a(t)$ may be written in exponential form

$$a(t) = A(t)e^{i\phi(t)}, \tag{4.4}$$

which naturally defines a time-dependent phase $\phi(t)$ and amplitude $A(t)$

$$\phi(t) = \tan^{-1}\frac{\Im[a(t)]}{\Re[a(t)]} \tag{4.5}$$

$$A(t) = \sqrt{\Re[a(t)]^2 + \Im[a(t)]^2} \tag{4.6}$$

The time-dependent phase $\phi(t)$ is referred to as the instantaneous phase and $A(t)$ as the instantaneous amplitude or envelope of $u(t)$.

To define a misfit function, we need to consider the instantaneous phase difference between two signals $u_1(t)$ and $u_2(t)$. We multiply the complex conjugate of the first analytical signal $a_1(t)$ with the second analytical signal $a_2(t)$

$$a_1^*(t)a_2(t) = A_1(t)A_2(t)e^{i[\phi_2(t)-\phi_1(t)]}. \tag{4.7}$$

Dividing (4.7) by the product of the envelopes $A_1(t)A_2(t)$ results in a complex signal $a_p(t)$, containing the amplitude-independent instantaneous phase difference $\Delta\phi(t) = \phi_2(t) - \phi_1(t)$ of the two signals

$$a_p(t) = \frac{a_1^*(t)a_2(t)}{A_1(t)A_2(t)} = e^{i[\Delta\phi(t)]}. \tag{4.8}$$

In principle, we can now extract $\Delta\phi(t)$ by

$$\Delta\phi(t) = \tan^{-1}\frac{\Im[a_p(t)]}{\Re[a_p(t)]}. \tag{4.9}$$

Based on $\Delta\phi(t)$, we can define a misfit function $\chi(\mathbf{m})$ similar to the traveltime misfit function, which contains the squared L_2-norm of $\Delta\phi(t)$ at all receivers r, time-integrated over the duration of the seismograms T

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \|\Delta\phi_r(t,\mathbf{m})\|^2 dt. \tag{4.10}$$

The adjoint source $f_r^\dagger(\mathbf{x}_r, t, \mathbf{m})$ at one receiver with index r and position $\mathbf{x}_r$ for this misfit function can be expressed as (Bozdağ et al., 2011)

$$\begin{aligned} f_r^\dagger(\mathbf{x}_r, t, \mathbf{m}) = &-\Delta\phi(\mathbf{x}_r, T-t, \mathbf{m})\frac{\mathrm{H}\left[u_{ref}(\mathbf{x}_r, T-t, \mathbf{m})\right]}{A(\mathbf{x}_r, T-t, \mathbf{m})^2} \\ &+ \mathrm{H}\left[\Delta\phi(\mathbf{x}_r, T-t, \mathbf{m})\frac{u_{ref}(\mathbf{x}_r, T-t, \mathbf{m})}{A(\mathbf{x}_r, T-t, \mathbf{m})^2}\right], \end{aligned} \tag{4.11}$$

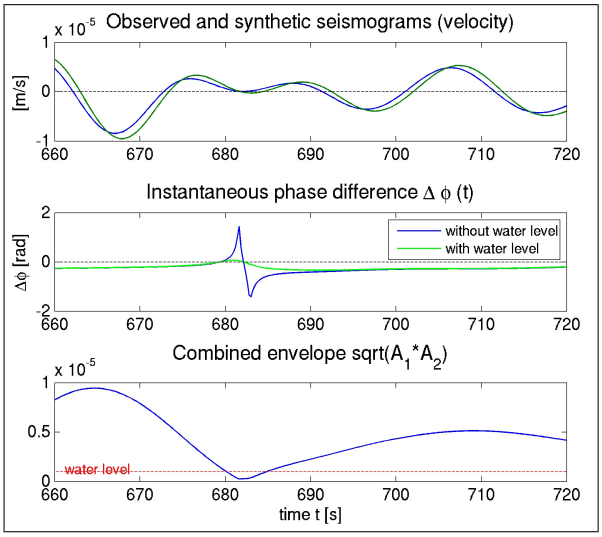
and the combined adjoint source for all N receivers results as

$$f^\dagger(\mathbf{x}, t, \mathbf{m}) = \sum_{r=1}^{N} f_r^\dagger(\mathbf{x}_r, t, \mathbf{m})\delta(\mathbf{x}-\mathbf{x}_r). \tag{4.12}$$

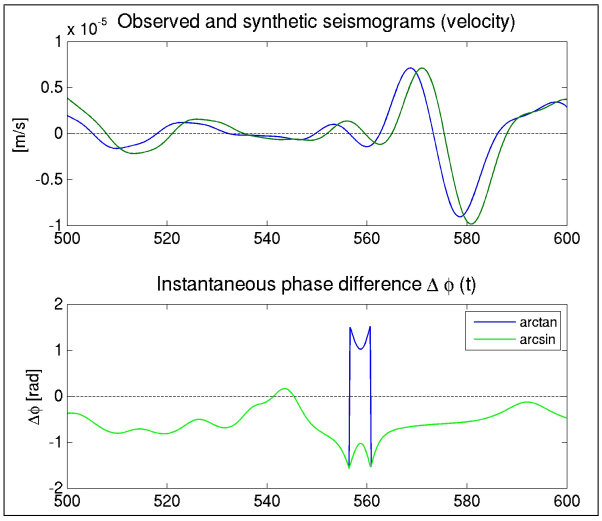
The adjoint source implicitly accounts for amplitude variations in the seismograms through division by the squared envelope $A(\mathbf{x}_r, T-t, \mathbf{m})^2$. For the same phase difference, low amplitude phases obtain a larger weight compared to high amplitude phases in the adjoint source. This results in amplitude-independent Fréchet kernels when the adjoint wavefield interacts with the forward wavefield. In practice however, we have to introduce an amplitude-dependent water level to prevent the magnification of noise and to suppress discontinuities in $\Delta\phi(t)$, which would otherwise dominate the Fréchet kernels.

4.5.3 Practice: Regularisation of the instantaneous phase difference

The computation of the instantaneous phase difference $\Delta\phi(t)$ between two signals $u_1(t)$ and $u_2(t)$ is straightforward in theory, however in practice care has to be taken. The instantaneous phase difference can be discontinuous due to phase jumps. The arctangent function is uniquely defined only within the interval $[-\frac{\pi}{2}; \frac{\pi}{2}]$, which imposes the requirement that the two signals $u_1(t)$ and $u_2(t)$ have to be close enough, so that $\Delta\phi(t)$ is confined to this interval throughout the measurement window. Because we are



(a) Phase jump (missed zero crossing)



(b) Phase jump ($\Delta\phi(t)$ locally too large)

Figure 4.5: Two reasons for the appearance of discontinuities in the instantaneous phase difference $\Delta\phi(t)$ and their remedies. (a) A phase jump occurs when one of the two signals misses a zero crossing, which determines the beginning of a new cycle. This occurs mainly in those parts of the signal where the envelope approaches zero. An envelope-dependent water level smoothly damps $\Delta\phi(t)$ in the critical parts. (b) A phase jump of π occurs when $\Delta\phi(t)$ locally exceeds the interval $\pm\frac{\pi}{2}$. Using the arcsine instead of the arctangent when extracting the phase from equation (4.8) reduces the error mapped into the kernel.

dealing with broadband signals, it can occur that $\Delta\phi(t)$ exceeds this interval, even when the two signals appear similar visually. We identified two main scenarios that cause discontinuities in $\Delta\phi(t)$, and present solutions to suppress them:

1. One reason for the occurrence of phase jumps is that the beginning of each new instantaneous phase cycle is determined by a zero crossing of the signal $u(t)$, from positive to negative. For the larger amplitude parts of two sufficiently similar signals $u_1(t)$ and $u_2(t)$, these zero crossings of both signals usually occur close enough to ensure that both signals enter a new cycle before their instantaneous phase difference exceeds $\pm\frac{\pi}{2}$. In very low amplitude parts of the signals, however, noise or interference effects can locally distort one of the signals sufficiently to prevent the zero crossing and the beginning of a new cycle, while the other signal enters a new cycle. This causes $\Delta\phi(t)$ to exceed $\pm\frac{\pi}{2}$ and results in a discontinuity which introduces a large error in the measurement.

   The example in Fig. 4.5(a) shows such a scenario. Although both waveforms appear very similar, a phase jump occurs. The resulting instantaneous phase difference $\Delta\phi(t)$ contains a large discontinuity, which renders it unusable for the inversion. In the lowermost plot of the same figure, we see that the envelope of the signal is very small at the location of the phase jump.

   We found that as long as the two signals are similar in general, phase jumps of this type are confined to the parts of the signal where the envelope almost reaches zero. We can resolve this problem by defining an envelope water level b, the square of which we add to the product of the envelopes in the denominator of eq. (4.8)

$$a_p = \frac{a_1^*(t)a_2(t)}{A_1(t)A_2(t) + b^2} = e^{i(\phi_2(t)-\phi_1(t))}. \tag{4.13}$$

   As a result, $\Delta\phi(t)$ is damped smoothly in the critical low-envelope regions, as is shown in the example in Fig. 4.5(a), and the signal remains usable for the inversion. It remains to choose a suitable value for b which provides enough damping to suppress the phase jumps, but leaves as much as possible of the remaining signal unchanged. In our case of noise-free, synthetic data, we chose $b = 5\%$ of the maximum P-wave amplitude of the respective seismogram. In the case of noise-affected real data, b can be increased to damp the influence of noise on the measurement.

2. A different type of phase jump is shown in Fig. 4.5(b). Although most of the signal is confined to the $\pm\frac{\pi}{2}$-interval, a small perturbation in the signal which already has a considerable phase difference can cause $\Delta\phi(t)$ to grow out of the interval limits. This case is different from the previous one, here both signals are within the same instantaneous phase cycle, but the phase difference locally grows too large. A solution which still allows us to work with this signal is the use of the arcsine instead of the arctangent when $\Delta\phi(t)$ is extracted from eq. (4.8)

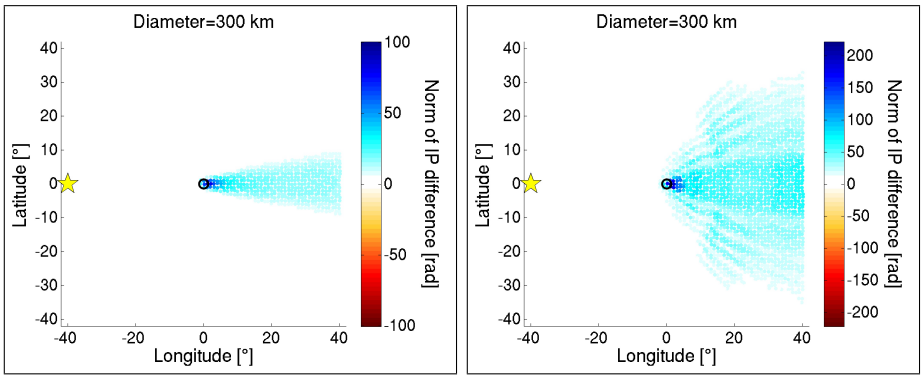
$$\Delta\phi(t) = \sin^{-1}(\Im[a_p]). \tag{4.14}$$

The difference is that the arcsine, although like the arctangent only uniquely defined within the $\pm\frac{\pi}{2}$-interval, does not switch sign as soon as the value grows out of the interval limits. Instead, the phase difference is symmetric around the values $\pm\frac{\pi}{2}$. This will still result in a small error in the measured instantaneous phase difference, but it does not lead to a phase jump of π as soon as $\Delta\phi(t)$ grows too large. Obviously this solution is not ideal when too many phase jumps occur in a signal. In that case a better solution could be to unwrap the phase difference through shifting $\Delta\phi(t)$ by $\pm\pi$ whenever a phase jump of this type is detected. To automate this process, however, care has to be taken that all phase jumps of the first type (due to missed zero crossings) are properly removed before unwrapping, otherwise even larger errors can be introduced.

Although these two techniques of avoiding discontinuities in $\Delta\phi(t)$ do leave small errors in the measured instantaneous phase difference, they keep it sufficiently well-behaved. That makes it possible to use the instantaneous phase difference as a misfit function in the following inversions, without having to interfere manually during the iterative procedure. The convergence rate of the misfit function is comparable to the cross-correlation case, indicating that no substantial errors are introduced into the kernels.

4.5.4 Spatial distribution of the instantaneous phase misfit depending on the measurement window

The instantaneous phase difference allows a measurement for arbitrary parts of the seismograms, provided the signals resemble each other sufficiently to confine $\Delta\phi(t)$ to the $\pm\frac{\pi}{2}$ interval limits.



(a) Window: P-wave (b) Window: extended P-wave

Figure 4.6: Spatial distribution of the L_2-norm of the instantaneous phase difference $||\Delta\phi(t)||$, measured using the dense station network in Figure 4.1(b). (a) For the same P-wave window as in the cross-correlation case. (b) For an extended P-wave window, reaching up to the onset of the PP-wave.

We can, similar to the traveltime case in Section 4.4.1, visualise the spatial distribution of the instantaneous phase misfit by performing a forward simulation using the dense receiver grid in Fig. 4.1(b).

We computed the misfit for a $d = 300$ km plume, comparing the spatial misfit distribution for two different measurement windows. In the first case, we used the same window (the 25 second P-wave window) as we did in the inversion based on cross-correlation traveltime shifts. In the second case, we extended the measurement window to not only cover the P-wave itself, but part of the later-arriving signal as well. This window extends until the arrival of the surface-reflected PP-wave (which is excluded). It covers a large part of the plume-diffracted P-waves, which arrive later than the main phase at receivers with larger azimuth. We excluded the surface-reflected PP-wave and the later-arriving phases to keep the experiment simple, and because we are specifically aiming at improving the recovery of the deeper part of the plume, which the surface-reflected waves do not traverse.

Fig. 4.6(a) shows the L_2-norm $||\Delta\phi(t)||$ of the instantaneous phase difference at each receiver for the same P-wave window as in the cross-correlation case. The spatial distribution of receivers where the misfit is non-zero is similar as in the cross-correlation case. However, the norm of the instantaneous phase difference decreases more slowly with increasing

epicentral distance compared to the cross-correlation traveltime shifts for a plume diameter of 300 km.

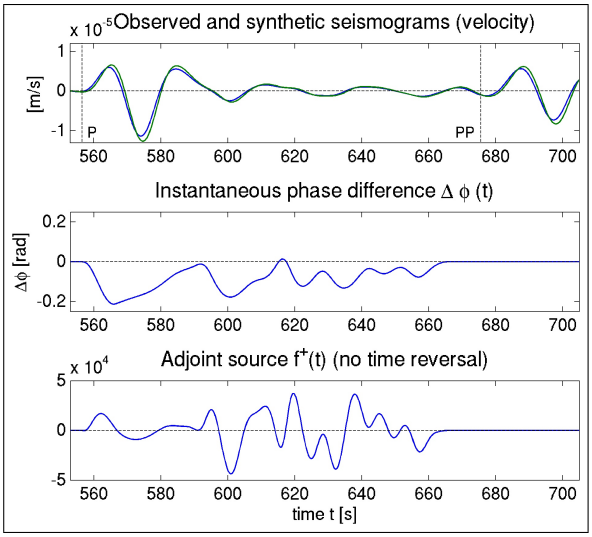
Fig. 4.6(b) shows the misfit for the extended window, using the signal up to the onset of the PP-wave. The number of receivers with a non-zero misfit has increased, now including receivers at larger azimuth. Above a certain receiver azimuth, the diffracted waves arrive later than the direct P-wave as a result of their longer ray path. These waves are covered by the extended window. There is a distinct periodic pattern noticeable in the spatial misfit distribution, reflecting the periodic sensitivity pattern of the higher Fresnel zones of the Fréchet kernels.

4.5.5 Two examples of instantaneous phase measurements

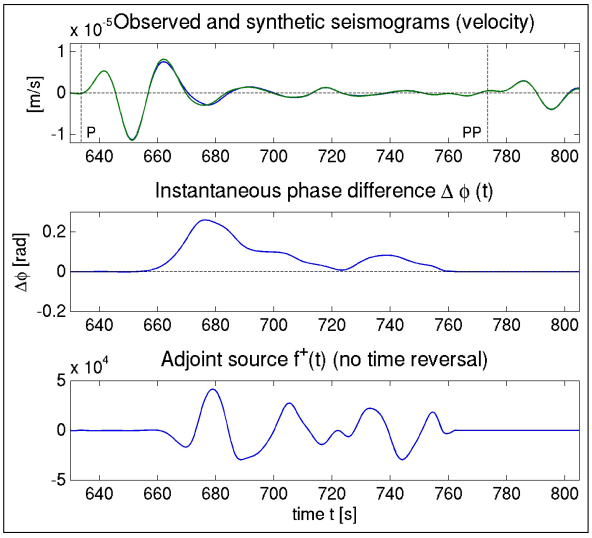
Fig. 4.7 presents two examples for the computation of instantaneous phase difference and adjoint source from the observed and synthetic signals, using the extended measurement window up to the onset of the PP-wave. The first example in Fig. 4.7(a) is for a receiver located behind the plume, in line with the source and the plume axis. The two seismograms are slightly time shifted with respect to each other over the whole signal length. The instantaneous phase measurement reveals the time-dependence of the phase shift within the measurement window. Even over the length of the main P-wave, the phase difference varies. It is larger at the beginning and decreases towards the end. The highest amplitudes of the adjoint source coincide with the low-amplitude parts of the seismograms.

The example in Fig. 4.7(b) is for a receiver with a larger azimuth. Visually, the largest difference between both seismograms extends from the end of the main P-wave into the lower-amplitude part of the signal and is caused by an interfering diffracted wave perturbing the waveform. The instantaneous phase difference extracts this information accurately as a function of time. A cross-correlation measurement however, even if the window covers the diffracted waves, would measure a zero time shift. The highest-amplitude part of the P-wave, which is not time-shifted, will dominate the measurement.

Inspecting the adjoint sources in Fig. 4.7, we notice that they contain higher frequencies than the original seismograms do. This is because the time-dependent instantaneous phase difference can vary at a higher frequency than the dominant frequency of the signals. These faster oscillations are mapped into the adjoint source. Because our wavefield simulations can be considered accurate only for periods above $T = 25$ s, the

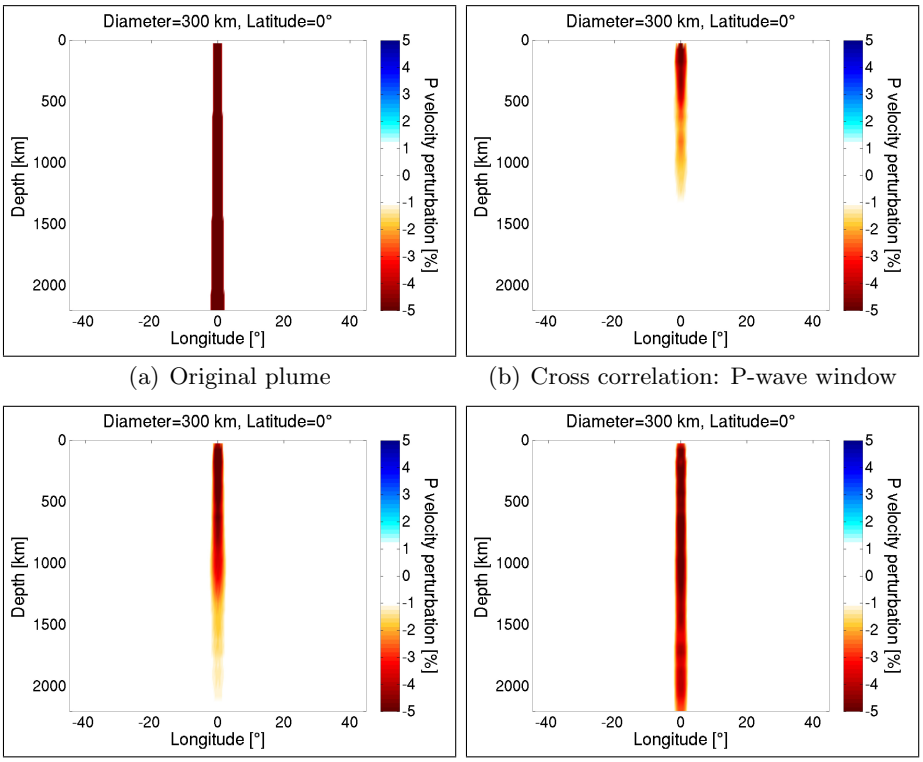


(a) A receiver located behind the plume



(b) A receiver with larger azimuth

Figure 4.7: Seismograms $u(t)$, instantaneous phase difference $\Delta\phi(t)$ and adjoint source $f^{\dagger}(t)$ for the extended P-wave window at two different receivers. (a) A receiver behind the plume on the direct wave path along the equator. (b) A receiver off the direct wave path, the difference between the seismograms is due to diffracted waves.



(a) Original plume

(b) Cross correlation: P-wave window

(c) Instantaneous phase: P-wave window

(d) Instantaneous phase: extended P-wave window

Figure 4.8: (a) The original plume of diameter 300 km. (b) Final model using cross-correlation travel times. (c) Final model using the instantaneous phase difference of the same P-wave window as in (b). (d) Final model using the instantaneous phase difference of the extended P-wave window which reaches up to the onset of the PP-wave.

higher-frequency adjoint source introduces a certain amount of noise into the Fréchet kernels. To avoid the possible manifestation of small-scale artefacts during the inversion, we applied Gaussian smoothing to the gradient over a length-scale corresponding to the dominant wavelength at each iteration step.

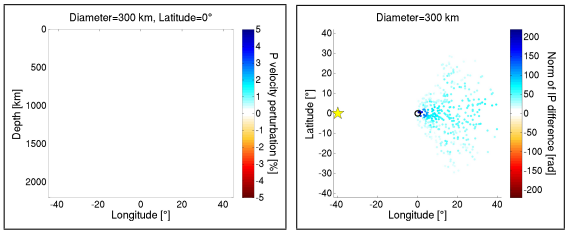
4.5.6 Inversion using the instantaneous phase difference

We applied the same inversion procedure as in the cross-correlation case, except that now we minimised the instantaneous phase misfit function in eq. (4.10), using the corresponding adjoint source in eq. (4.12). We performed the inversion for two different measurement windows of different length. First the same P-wave window as in the cross-correlation case, and then the extended window up to the onset of the PP-wave. We again ended the iterative algorithm when the value of $\chi(\mathbf{m})$ had decreased by 95% compared to the initial value.

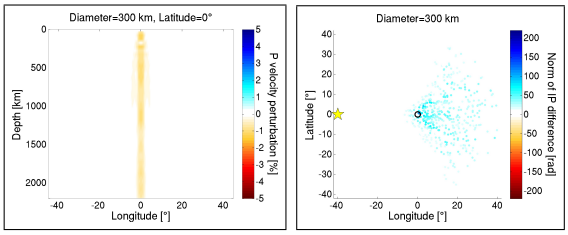
Fig. 4.8(a) shows the original plume which we used to generate the data and, for comparison, Fig. 4.8(b) again shows the final tomographic model resulting from the cross-correlation based inversion. In Fig. 4.8(c), we present the resulting tomographic model using instantaneous phases measurements of the 'short' 25 second P-wave window. We notice a significant improvement in the recovery of the plume compared to the traveltime based model. Down to a depth of 1000 km the plume is resolved with more than 50% of its original velocity perturbation. Below 1200 km however, the plume is still recovered much weaker compared to the original plume. Fig. 4.8(d) shows the final tomographic model using the extended P-wave window, taking into account more of the diffracted wave field. Although slightly weaker in amplitude and thinner in diameter in the lower mantle compared to the original plume, the plume is well-constrained throughout the whole depth of the model.

Fig. 4.9 shows the evolution of the tomographic model and the corresponding spatial misfit distribution at different iteration steps.

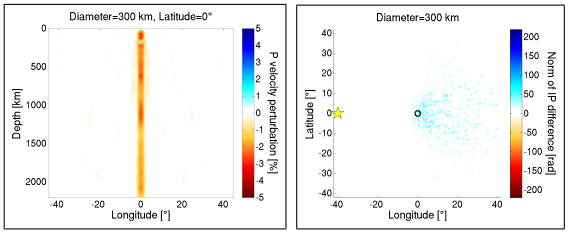
Regarding the convergence of the misfit function, we note that the instantaneous phase misfit for the extended window converges slightly slower compared to the inversions based on just the P-wave window. 15 iterations had to be performed for the P-wave window and 17 for the extended window to reduce the misfit by 95% with respect to the initial value. Considering the larger amount of information extracted from the seismograms, the instantaneous phase misfit does not appear to behave more non-linear with respect to the model parameters than the cross-correlation misfit function does.



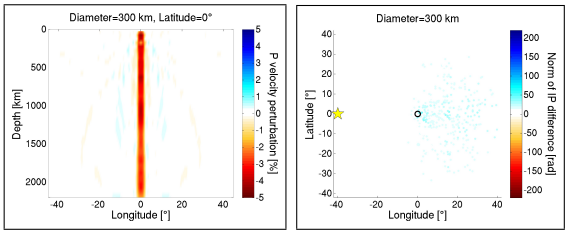
(a) Model $\mathbf{m_0}$



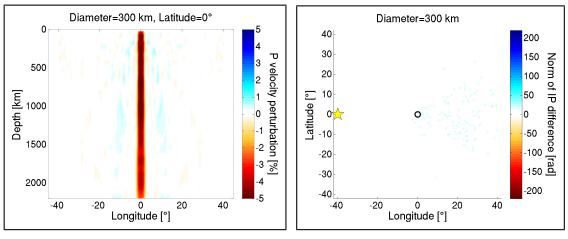
(b) Model $\mathbf{m_2}$



(c) Model $\mathbf{m_5}$



(d) Model $\mathbf{m_{11}}$



(e) Model $\mathbf{m_{17}}$

Figure 4.9: Evolution of the tomographic model at different iteration steps (left) and the corresponding spatial misfit distribution for one event (right), using the instantaneous phase misfit for the extended P-wave window. (a) Starting model $\mathbf{m_0}$ (PREM), (b) Model $\mathbf{m_2}$, (c) Model $\mathbf{m_5}$, (d) Model $\mathbf{m_{11}}$, (e) Final model $\mathbf{m_{17}}$.

4.5.7 Discussion of the instantaneous phase results

The instantaneous phase measurement captures finite-frequency effects significantly better than simple cross-correlation measurements can. The consideration of later-arriving diffracted waves makes it possible to recover the idealised mantle plume, although its diameter is only a fraction of the width of the first Fresnel zone.

The direct comparison of the two misfit functions using the same 'short' P-wave window (Figs 4.8(b) and 4.8(c)) reveals that simple traveltime measurements leave part of the available information unused, due to their inability to measure phase differences as a function of time. The instantaneous phase difference is able to capture these time-dependent phase variations, which result from the interference of direct and diffracted waves. This information is mapped into the Fréchet kernel and helps to, at least partly, account for these interference effects in the inversion. Information on the deep part of the plume appears to be found later in the P-wave coda.

An advantage of the instantaneous phase misfit is that the measurement window can be extended arbitrarily. Later-arriving diffracted waves can be considered in the measurement. Due to the amplitude-independence of the instantaneous phase, they contribute with a similar weight to the total value of the misfit function as the main phases do. The measurement of the later-arriving diffracted waves imposes strong constraints on the plume diameter even at depth. A mis-placed and mis-sized plume even in the deep mantle causes a large misfit at receivers where the later-arriving, diffracted waves are recorded. This misfit can only be accounted for by a correctly constrained plume.

As we can infer from the left-hand panel of Fig. 4.9, the shape of the plume is well-constrained from the beginning of the iterative procedure by the diffracted waves. The consideration of the diffracted waves forces the iterative algorithm to keep the plume narrow at all depths and to steadily increase its amplitude over the course of the iterations. Both the correct amplitude and shape are required to reduce the misfit of the direct as well as the diffracted waves. The spatial misfit distribution in the right-hand panel of Fig. 4.9 shows how the misfit is reduced at all receivers relatively evenly while the amplitude of the velocity perturbation increases.

To intuitively understand the ability of diffracted waves to constrain the plume at depth, it is helpful to analyse the structure of the Fréchet kernels. A classic traveltime Fréchet kernel for a body-wave consists of a central first Fresnel zone surrounded by multiple higher Fresnel zones, which de-

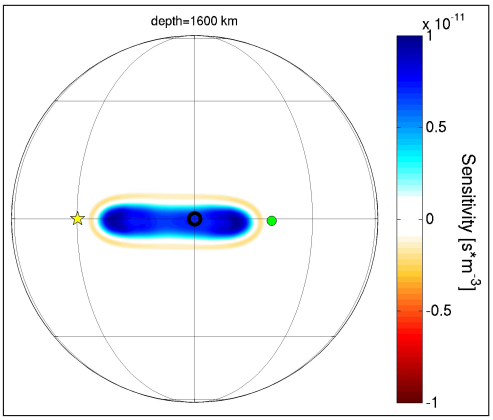


Figure 4.10: Horizontal slice at a depth of 1600 km through the Fréchet kernel for a cross-correlation P-wave measurement, measured at a receiver located behind the plume. The diffracted wave arrives within the P-wave window, the plume (black circle) is contained within the first Fresnel zone of the kernel. The source is denoted by the yellow star, the receiver by the green dot.

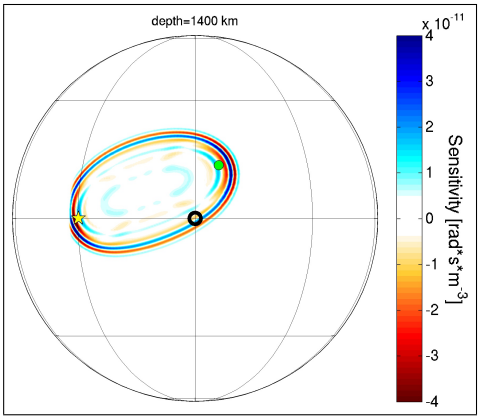


Figure 4.11: Horizontal slice at a depth of 1400 km through the Fréchet kernel for an instantaneous phase measurement, measured at a receiver with larger azimuth. The diffracted wave arrives later than the P-wave does, the plume (black circle) is located within higher order Fresnel zones of the kernel. The source is denoted by the yellow star, the receiver by the green dot.

crease in thickness with increasing Fresnel zone order. Each Fresnel zone represents a time window of half a wave period length, in which the adjoint wavefield and the forward wavefield alternately interfere either constructively or destructively. In the case of simple cross-correlation traveltime measurements, the largest amplitudes of the forward and adjoint wavefield interact within the first Fresnel zone, which consequently has the highest sensitivity. The reason is that cross-correlation measurements are dominated by high amplitudes, and as a result the seismogram and the (not time-reversed) adjoint source have their respective maximum amplitudes within less than half a wave period difference. An example is given in Fig. 4.10 for a sensitivity kernel based on a P-wave cross-correlation traveltime measurement.

The Fréchet kernel for the more complex measurement of the instantaneous phase difference can have a different appearance, in particular can it have the highest sensitivity in higher-order Fresnel zones. When we perform an instantaneous phase measurement of a later-arriving, diffracted wave as in Fig. 4.7(b), the maximum amplitude of the seismogram (which is the P-wave) and the maximum amplitude of the adjoint source (which is the diffracted wave) are time-shifted with respect to each other by more than half a wave period, which is due to the later-arriving diffracted wave and the amplitude-independence of the instantaneous phase measurement. During the adjoint simulation, the largest amplitudes of the forward and the adjoint wavefield interact with each other in higher order Fresnel zones. The later the diffracted wave arrives relative to the main P-wave, the higher is the order of the Fresnel zone which is most sensitive to the plume. In Fig. 4.11, we present such a Fréchet kernel. The highest sensitivity of the Fréchet kernel coincides with the Fresnel zone(s) containing the plume. The finer sensitivity pattern of the higher-order Fresnel zones is the reason for their ability to spatially constrain an anomaly which is below the width of the first Fresnel zone.

Although this explanation might be somewhat over-simplified due to the complex interaction process of forward and adjoint wavefield, it is clear that the misfit resulting from the diffracted waves requires a correctly-located and -sized plume to be accounted for, while the direct waves allow a larger tolerance in the position and dimension of the plume within the first Fresnel zone without a large influence on the misfit.

As a side note, we also investigated the case of a plume with a fast instead of a slow velocity perturbation and otherwise identical properties.

We did not notice qualitative differences compared to a slow plume.

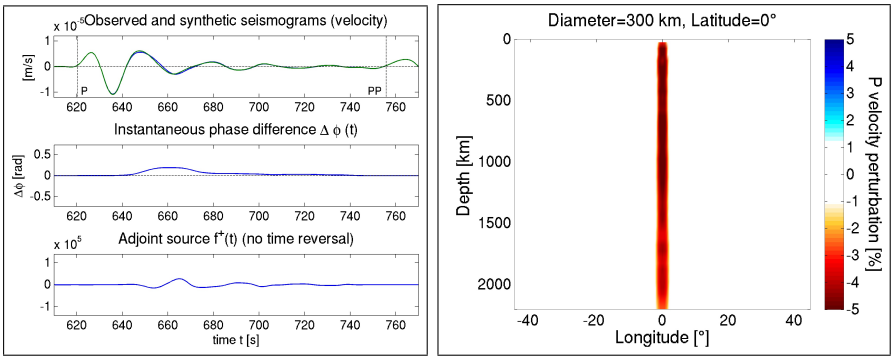
4.6 Towards the use of the instantaneous phase misfit with real seismic data

4.6.1 The effect of noise on the instantaneous phase misfit

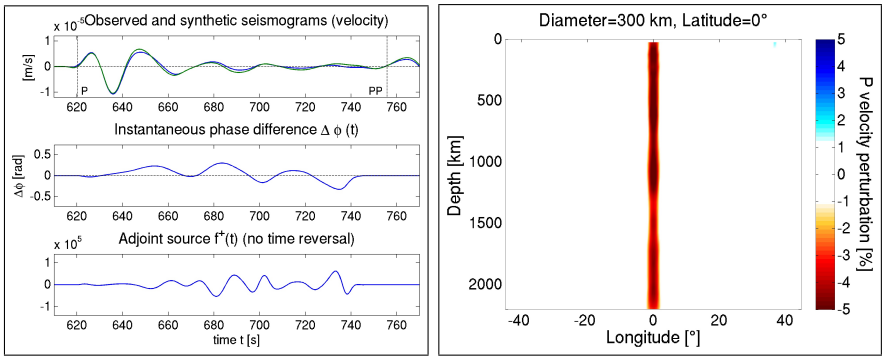
While the synthetic experiments in the preceding sections have been performed on noise-free seismograms with the aim to assess and compare the different misfits, in this section we present results which demonstrate that the better resolution capabilities of the instantaneous phase misfit still persist in the presence of noise.

We first investigated the effect of random noise on the instantaneous phase measurement and on the resulting tomographic model. We generated a trace of Gaussian-distributed white noise for each observed seismogram, which was low-pass filtered at a cut-off period of $T = 25$ s. We scaled the amplitude of the noise traces to 10% of the P-wave amplitude at an epicentral distance of 40° and superimposed them on the observed seismograms. Depending on the receiver location, the amplitude of the random noise is comparable or larger than the amplitude of the low-amplitude diffracted waves whose accurate measurement we assume to be important for the better resolution capabilities of the instantaneous phase misfit. The upper panels of Fig. 4.12(a) and Fig. 4.12(b) show examples of seismograms recorded at a receiver where a diffracted wave arrives in the later part of the P-wave, before and after the random noise trace was added. With the added noise, the diffracted wave is not clearly identifiable any more, and the measured instantaneous phase difference and resulting adjoint source differ considerably from the noise-free case.

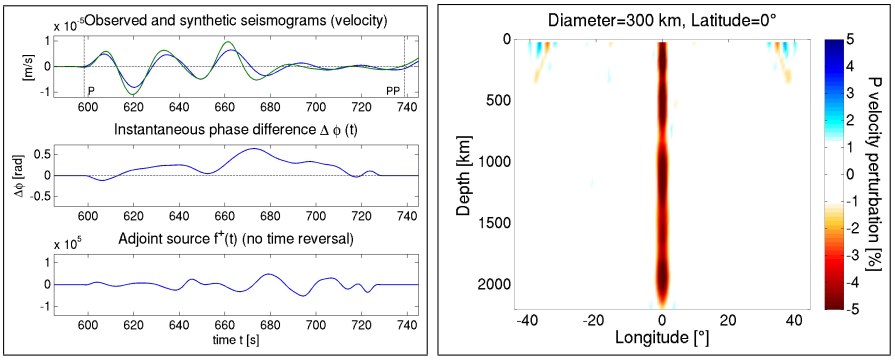
Inverting the instantaneous phase difference of the extended P-wave window for these noisy observed seismograms, we obtained a tomographic model which resolves the plume nearly as well as when noise-free seismograms were used (see lower panel in Fig. 4.12(b)). While this seems surprising at first, the explanation is simple: The noise propagates into the sensitivity kernels, but due to its random nature the noise of the individual kernels does not interfere constructively when they are combined to form the misfit kernel. The contributions related to random noise cancel out to a large extent, while the contributions related to real structure interfere constructively. While the number of iterations required to recover the plume



(a) No noise



(b) Random noise



(c) Random + correlated noise

Figure 4.12: Left-hand panels of (a) to (c): Seismograms $u(t)$, instantaneous phase difference $\Delta\phi(t)$ and adjoint source $f^{\dagger}(t)$ at the same receiver which records a diffracted wave in the coda of the P-wave. (a) for noise-free seismograms (b) for random noise added to the observed seismograms (10% of the P-wave amplitude) (c) for both random and correlated noise due to a missing crustal layer in the initial model. The source depth is 200 km in this experiment. Right-hand panels of (a) to (c): The corresponding tomographic results using the instantaneous phase misfit of the extended P-wave window reaching up to the onset of the PP-wave.

was comparable to the noise-free case, we noticed that the value of the misfit function $\chi(\mathbf{m})$ converged towards a value representing the random noise part of the measured misfit. This is desirable, indicating that the remaining misfit did not propagate into the tomographic model as mis-placed structure. In our experiment, we could reduce $\chi(\mathbf{m})$ to about 50% of its initial value.

In the next step, we investigated the influence of correlated noise. We again used PREM as initial model in the inversion, but this time without the crustal layer which has so far been present in the target model as well as in the initial model. In this experiment, we furthermore changed the source depth to 200 km to introduce the effect of the missing crust only at the location of the receivers and not at location of the source. The crustal layer is a uniform, 24 km thick layer of lower wavespeed with respect to the underlying mantle. We replace it by extending the underlying mantle's properties to the surface. The absence of the crustal layer in the initial model leads to larger differences in phase as well as in amplitude between observed and reference seismograms. The upper panel of Fig. 4.12(c) shows that the instantaneous phase difference is now dominated by effects caused by the absence of the crustal layer. In addition, the random noise trace of the previous experiment was added to the observed data.

The crustal layer is too thin to be recovered by the iterative inversion procedure, due to the Gaussian smoothing over a length-scale of approximately 200 km which is applied to the gradient. The difference in the seismograms can therefore be considered as correlated noise caused by unknown crustal structure. Such noise is expected to be more problematic than random noise, because the contributions of the correlated noise in the individual sensitivity kernels do not necessarily cancel out when they are combined to form the misfit kernel but instead, due to being correlated, might interfere constructively and introduce structure in wrong areas of the model.

Inverting the instantaneous phase difference affected by both random and correlated noise, we still succeeded to recover the plume to full depth (lower panel in Fig. 4.12(c)). Some additional structure was introduced mainly in the upper mantle, at the edges of the model. Some weaker misplaced structure appeared in the tomographic model, but is hardly visible. What is important to note is that despite being concealed by noise in the seismograms, the information about the plume is still extracted correctly by the instantaneous phase-based sensitivity kernels. This is of great practical

relevance, as the crustal model is often not well known in tomographic inversions, giving raise to correlated noise.

4.6.2 Other factors to consider

While the presence of random as well as correlated noise in the data is probably the largest concern when it comes to transitioning the synthetic experiment to an experiment based on real data, there are other issues which need to be considered. The most severe concern might be that the distribution of seismic events and receivers in most plume-suspected regions is far from perfect. In most cases, full-waveform tomographic methods rely on a smaller number of seismic records than traditional methods. However, large gaps in the receiver distribution or a biased distribution of seismic events might lead to artefacts in the tomographic model which could be misinterpreted as imaged structure. A careful selection of stations and events is therefore essential. The chosen data should have a high signal-to-noise ratio and sample the area as evenly as possible.

While in the synthetic experiments the mantle structure of the initial model is, apart from the idealised plume, identical to the target model, it will usually not be that well-known in a real experiment. We don't expect this to be a problem to the applicability of the instantaneous phase measurements, however we need to make sure that the waveforms do not differ by more than a quarter of a wavelength for a meaningful measurement. A multi-scale approach might therefore prove useful, starting the inversion at longer periods and increase the frequency at later iterations, when the long-wavelength structure has been adjusted.

Possible sharp amplitude contrasts between plume and surrounding mantle are likely to be not perfectly recovered due to the smoothing that is applied to the gradient of the misfit function, however the perturbation magnitude of a plume itself should be recoverable even deep in the mantle. The strong constraints on the shape of the plume require a correct magnitude to explain the amplitude of the direct as well as the plume-diffracted waves.

Eventually, to assess whether imaged structure in an inversion using adjoint techniques is well-resolved, a resolution analysis based on Hessian kernels can be performed (Fichtner and Trampert, 2011b).

4.7 Discussion and conclusions

We confirm that cross-correlation traveltime shifts for an idealised mantle plume conduit are affected by wavefront healing when the plume conduit is significantly narrower than the width of the first Fresnel zone. As the plume diameter increases to the width of the first Fresnel zone, wavefront healing becomes negligible.

We have demonstrated that an idealised plume conduit of diameter $d = 300$ km cannot be imaged tomographically when simple P-wave traveltime shifts, measured at a dominant period of $T = 25$ s, are used. Despite an idealised source-receiver distribution and the use of iterative fully three-dimensional inversion techniques, the lower mantle part of the plume remains largely unresolved. This result is scalable to the case of shorter wavelengths and a thinner plumes, because the influence of wavefront healing on cross-correlation traveltimes depends on the ratio of plume diameter d to the width of the first Fresnel zone d_F (Hung et al., 2001; Baig et al., 2003). The tomographic result using $T = 25$ s waves to image a $d = 300$ km plume is therefore considered similar to the use of $T = 3$ s waves to image a $d = 100$ km plume, as the respective ratios d/d_F are comparable. For teleseismic P-waves, traveltime measurements on real data can be realistically made down to periods of $T = 1$ s. This translates into a width of the first Fresnel zone in the deep lower mantle of approximately 300 km and implies that traveltime measurements of plumes considerably smaller than 300 km will be strongly affected by wavefront healing. In the case of teleseismic S-waves, real-data measurements can be made at periods down to approximately $T = 10$ s, resulting in a maximum Fresnel zone width of about 700 km.

Under the assumption that our idealised plume conduit is a reasonable model for a real plume conduit, we conclude that it is not possible to obtain well-resolved tomographic images of narrow plume conduits in the lower mantle using simple cross-correlation traveltime measurements of main body-wave phases, even when finite-frequency kernels are used. Possibly existing lower-mantle plume conduits will have to extend several hundreds of kilometres in diameter to not be affected by wavefront healing. This result agrees with Hwang et al. (2011), who conclude that deep plumes are seismically invisible based on traveltime measurements of S-waves.

We explain the limited resolution of simple cross-correlation traveltime shifts, even when finite-frequency kernels are used, with the fact that a

single cross-correlation traveltime measurement cannot capture the details of the interaction between diffracted and direct waves. The measured time shift is implicitly assumed to be constant over the whole measurement window. Consequently, the iterative procedure converges towards an anomaly whose dimensions are determined by the size of the first Fresnel zone, and smaller structures cannot be constrained.

We demonstrated that the tomographic result can be improved by using a misfit based on the instantaneous phase difference, which is a time-dependent, amplitude-independent measure of phase differences between two signals. By capturing the time-dependent interference between direct and diffracted waves, it provides more information compared to simple traveltime shifts. It can be applied to arbitrary measurement windows, and the amplitude-independence ensures that small-amplitude parts of the signal contribute with an equal weight as the main phase to the Fréchet kernel. Inverting the instantaneous phase difference instead of cross-correlation traveltime shifts for the P-wave window only already results in a significantly improved tomographic image of the mantle plume. Local variations of the phase shift within the P-wave window are extracted and mapped into the Fréchet kernel, constraining the plume tighter. Still we cannot completely separate direct and diffracted waves, and the lowermost part of the plume is not well-constrained.

The plume can be recovered almost completely in shape as well as in amplitude when we extend the measurement window to include later-arriving diffracted waves, which are difficult to capture with simple cross-correlation measurements. With the instantaneous phase, their information can be extracted despite their small amplitude, and they contribute strongly to the total value of the misfit function. These diffracted waves constrain the shape of the plume well even at depth. The Fréchet kernel for a source-receiver pair recording a later-arriving diffracted wave has the highest sensitivity in higher-order Fresnel zones, allowing for better spatial constraints on the plume due to their finer structure compared to the first Fresnel zone. The instantaneous phase misfit together with a measurement window sufficiently long to capture the diffracted waves thus allows the imaging of deep, narrow plume conduits, at least within the idealized setup of this synthetic study.

We consider the instantaneous phase misfit a well-suited choice for our purpose of considering diffracted waves in an inversion, as it naturally combines a high resolution in time with amplitude-independence. Being a

pure phase misfit, the measurement is reasonably linear with respect to the model parameters of interest. Other time- and / or frequency-dependent misfits may lead to similar results if they can capture the details of the interference between direct and diffracted waves and assign the low-amplitude diffracted waves a sufficiently high weight in the misfit function.

We have successfully applied the instantaneous phase misfit within the framework of this idealised synthetic experiment. We have also shown that the method is still applicable in the presence of random and correlated noise due to, for instance, unknown crustal structure, even when the noise completely conceals the small-amplitude diffracted waves. However, an application of the method in a real seismic experiment is likely to pose additional challenges. What might be the most limiting factor is the distribution of events and receivers in many plume-suspected regions. Care should be taken to ensure a ray coverage which is as dense and unbiased as possible to prevent the appearance of artefacts in the tomographic model. To deal with a not well-constrained mantle model (apart from a possible plume), it might be helpful to use a multi-scale approach and begin the iterative procedure at lower frequencies. Once the long-wavelength structure of the mantle has been adjusted, the frequency content can be increased in later iterations. To assess the resolution of specific features in a tomographic model, a resolution analysis based on Hessian kernels can be performed (Fichtner and Trampert, 2011b).

Acknowledgements

We thank Carl Tape and an anonymous reviewer for their detailed and constructive comments which helped to improve the manuscript. We furthermore thank Ebru Bozdağ for insightful discussions regarding the instantaneous phase and Theo van Zessen for maintaining the STIG cluster. This work was financed by the Netherlands Research Centre for Integrated Solid Earth Sciences under grants ISES-NorMar-2.6 and ISES-UU-PC-cluster.

Chapter 5

Imaging mantle plumes with instantaneous phase measurements of diffracted waves: S-waves and radial anisotropy

In this chapter, the experimental setup of the previous chapter is extended to the case where an idealised plume is perturbed in both P- and S-velocity. The model is parametrised in terms of the P-velocity α and both the horizontally and vertically polarised S-velocities β_{sh} and β_{sv}. The instantaneous phase misfit is measured on three-component seismograms and the measurement window is extended to include the S-wave. We investigate the applicability of the instantaneous phase misfit to longer time series and whether the polarisation of the waves influences their sensitivity to the plume.

5.1 Introduction

In the previous chapter, we demonstrated that a deep-reaching, narrow plume conduit, represented by a vertical cylinder of reduced P-velocity, can, under idealised conditions, be fully imaged with a misfit based on the instantaneous phase difference. The time-dependent measurement of

diffracted waves in the P-wave coda plays a crucial role for the successful plume recovery. To keep the experiment simple and to understand the basic principles, the previous study was restricted to the vertical-component seismograms, and the measurement was limited to the P-wave and its coda.

In principle, the instantaneous phase misfit can be applied to full seismograms. The inclusion of S-waves and surface-waves is easily achieved by extending the measurement window, and the expectation is that a S-wave anomaly can be recovered equally well as a P-wave anomaly.

Tomographic models often include radial S-wave anisotropy (e.g. Nataf et al., 1984; Ekström and Dziewonski, 1998; Pilidou et al., 2005; Panning and Romanowicz, 2006). This means that the horizontally and vertically polarised S-velocities β_{sh} and β_{sv} are independent parameters. In inversions of surface-wave data, radial anisotropy within the uppermost mantle is often required to explain the dispersion of both Rayleigh and Love waves simultaneously. The constrains of both cannot be combined in an isotropic model, a phenomenon referred to as Love-Rayleigh discrepancy (Anderson, 1961; Aki and Kaminuma, 1963; McEvilly, 1964). It has been interpreted to result from the flow-induced alignment of the crystal axis of anisotropic mantle minerals (Birch, 1960), generating large-scale anisotropic regions which affect the wavespeed differently depending on the polarisation of the waves. Radial anisotropy may also be caused by aligned small-scale heterogeneities such as thin layers (Backus, 1962) or melt pockets (Mainprice, 1997). While anisotropy in most models is strongest in the uppermost mantle, there is evidence for the presence of anisotropy in deeper mantle regions (Trampert and van Heijst, 2002; Panning and Romanowicz, 2006; Visser et al., 2008). The reason that deeper anisotropy is less apparent in most tomographic models may be a result of the decreasing resolution with increasing depth.

However, the appearance of radial anisotropy may also result from differences in the resolution of the β_{sv}- and β_{sh}-components of tomographic models. This could, for instance, be caused by a different sampling of the medium or a different sensitivity of horizontally (SH) and vertically (SV) polarized waves to the heterogeneities. Tilmann et al. (1998) found an asymmetry in the coupling between SH- and SV-waves when scattered at a cylindrical vertical anomaly. The excitation of the vertical component by scattered SH-waves was thereby stronger than vice versa.

In this study, we investigate whether the recovery of the plume depends on the polarisation of the waves. At the same time, we demonstrate the

applicability of the instantaneous phase misfit to longer time series and three-component seismograms.

5.2 Method

The method and experimental setup are similar to the P-wave study in Chapter 4. In the following, only differences to the previous setup are described, the reader is referred to Section 4.3 for a detailed description of the method. Different from the previous setup, the idealised plume is also perturbed in S-velocity, in addition to the P-velocity perturbation of -5%. The amplitude of the perturbation in β_{sv} as well as β_{sh} is -8% with respect to PREM. To generate S-waves, the source type is changed from an explosive point source to a double-couple point source.

The instantaneous phase misfit is measured in the same way as in the P-wave case, but now the measurement window extends from the P-wave to the onset of the surface-reflected SS-wave. All three seismogram components are used in this inversion.

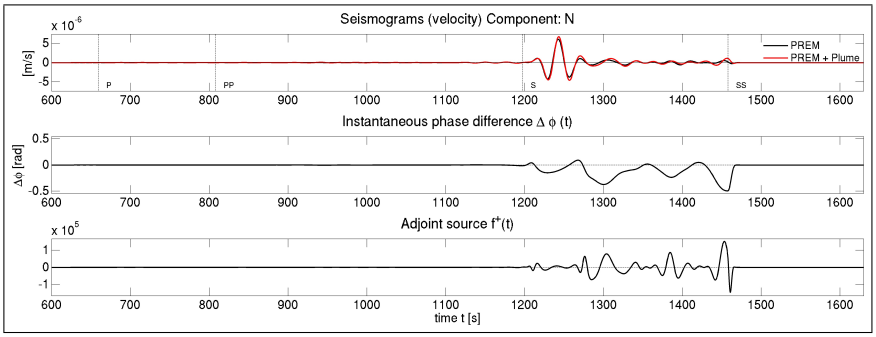
The isotropic S-velocity β can be calculated from the horizontally and vertically polarised components as the Voigt average (Babuska and Cara, 1991), defined as

$$\beta_s = \sqrt{\frac{2\beta_{sv}^2 + \beta_{sh}^2}{3}}. \tag{5.1}$$

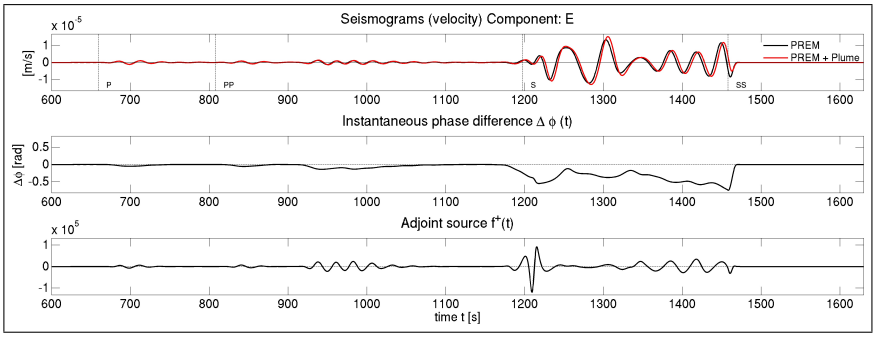
5.3 Inversion and results

The instantaneous phase misfit is measured within a window extending from the onset of the P-wave to the onset of the SS-wave, on all three components of the seismograms. We do not include surface-waves because our aim is to investigate the recovery of the deeper part of the idealised mantle plume, but an extension of the measurement window to include surface-waves would be straightforward. In this inversion, the damping parameter b (see Section 4.5.3) is chosen to be 5% of the maximum amplitude within the measurement window, which includes P- as well as S-waves. This results in the damping of part of the diffracted P-wave information, which is discussed further below.

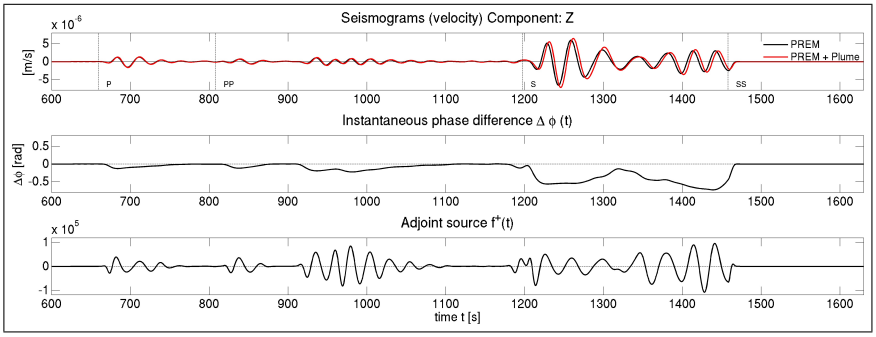
In Fig. 5.1, the measured instantaneous phase and the corresponding adjoint sources are presented for a three-component seismogram at a re-



(a) N-component



(b) E-component



(c) Z-component

Figure 5.1: Windowed seismograms, instantaneous phase difference and adjoint sources for a three-component seismogram, recorded at a distance of 28° behind the plume, in line with the source and the plume axis. The seismograms are windowed between the onset of the P-wave and the onset of the SS-wave, using a cosine-taper. The predicted arrival times of the major phases, based on PREM, are indicated by the vertical black lines in the seismograms.

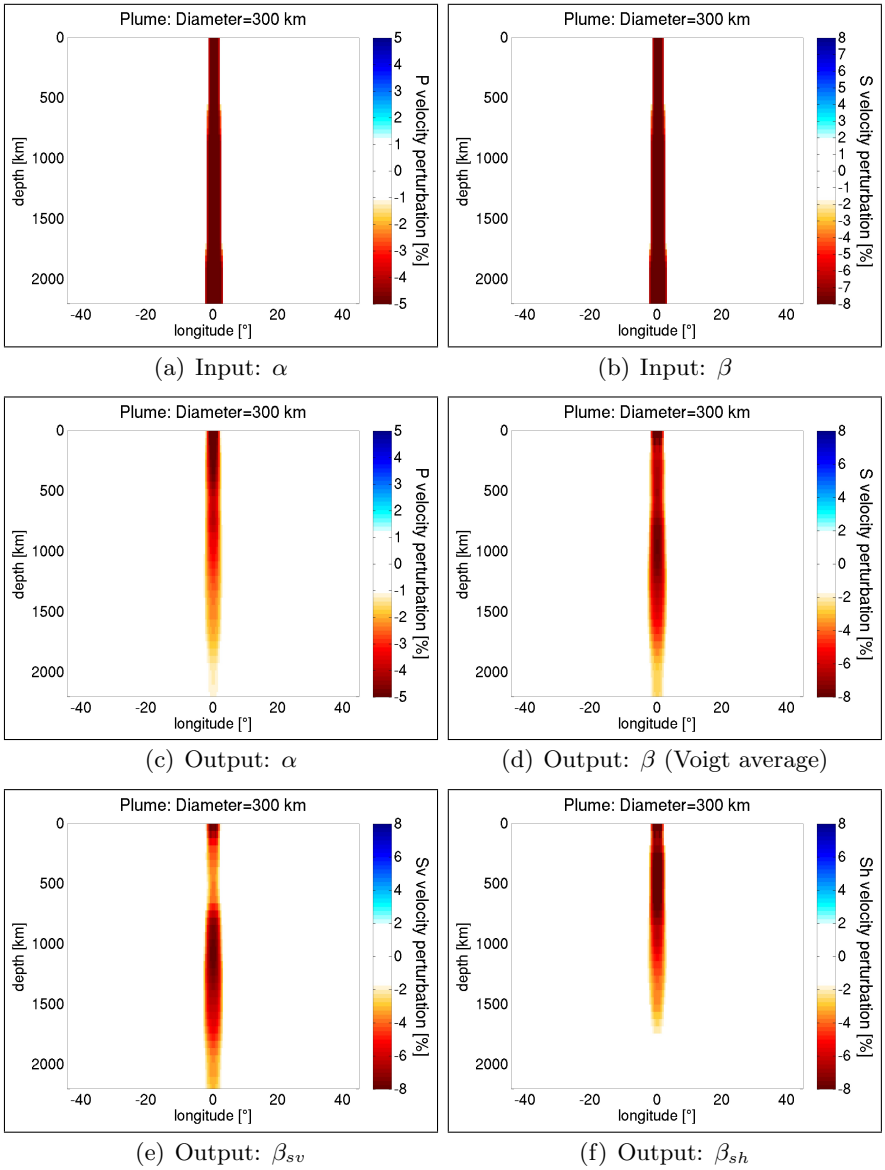
ceiver located 28° behind the plume, in line with source and plume axis. The measured instantaneous phase difference is larger for the S-wave than for the P-wave, which is partly due to the stronger S-velocity perturbation of the plume. The shorter wavelength of S-waves compared to P-waves at the same dominant period of 25 s also plays a role. Since the strength of the diffracted wavefield scales approximately as $\sqrt{\lambda}$ (Keller, 1962), the diffracted wavefield is weaker for waves of shorter wavelength.

We performed 10 iterations, in the course of which the instantaneous phase misfit function $\chi(\mathbf{m})$ (see Section 4.5) was reduced by 87%. The misfit could have been reduced further, but we chose to end the inversion after 10 iterations since the general pattern had emerged. The original and recovered plumes for the model parameters α, β_{sv} and β_{sh}, as well as for the isotropic S-velocity β, are presented in Fig. 5.2. Compared to the previous study in Chapter 4, the plume recovery is less complete for the P-velocity α (Fig. 5.2(c)). The damping parameter b, chosen to damp the instantaneous phase measurements in those parts where the seismogram amplitude is below 5% of the S-wave amplitude, probably results in over-damping of the low-amplitude P-wave information. Diffracted P-waves therefore do not fully contribute to the inversion and the plume is not recovered completely in P-velocity α, particularly its lowermost part. The isotropic S-velocity β (Fig. 5.2(d)) delineates the plume well throughout most of the mantle. The lowermost 500 km are recovered less well, but further iterations may improve the recovery in this part.

The most interesting result are the differences in the recovery of the plume for the horizontally (Fig. 5.2(e)) and vertically (Fig. 5.2(f)) polarised S-velocities β_{sh} and β_{sv} . Although the original plume perturbation is isotropic, the plume is recovered only partly by each of the components. There appears to be a reduced sensitivity to the plume in β_{sv} between 200 and 700 km, but the plume is well-recovered in β_{sv} below 700 km. In contrast, the upper 1300 km of the plume are well-recovered in β_{sh}, but the recovery is very weak below 1500 km depth. The recovery for the different polarisation directions appears to be complementary, since the gap in the upper part of the β_{sv}-plume is the part recovered best in β_{sh}.

5.4 Discussion and conclusions

The results show that the instantaneous phase misfit allows us to use the full seismogram, at least in the synthetic experiment where the synthetic and



(a) Input: α

(b) Input: β

(c) Output: α

(d) Output: β (Voigt average)

(e) Output: β_{sv}

(f) Output: β_{sh}

Figure 5.2: Original plumes and recovered plumes for the P-velocity α, the isotropic (Voigt average) S-velocity β, and the vertically and horizontally polarized S-velocities β_{sv} and β_{sh}.

observed seismograms are relatively similar. If different wave types with different amplitudes are used in the same inversion, the amplitude-dependent damping threshold b (see Section 4.5.3) should be defined individually for each wave type. Otherwise, part of the lower-amplitude wave information may be over-damped and will not contribute to the inversion. This was the case in the presented inversion, where part of the P-wave information was damped too strongly, resulting in a weakened recovery of the plume in P-velocity.

The important finding of this study is that the polarisation directions β_{sh} and β_{sv} recover the idealised mantle plume differently, each recovering only part of it. The most likely explanation is that the diffractions at the vertically-oriented boundary of the mantle plume depend on the polarisation direction. Tilmann et al. (1998) found a considerable asymmetric coupling between SH- and SV-waves for scattering from a vertical cylindrical anomaly. They noticed that scattered SH-waves produced a relatively strong vertical displacement, while the horizontal displacement from scattered SV-waves was much weaker. Other mechanisms may also be responsible for differences between β_{sh} and β_{sv}. For instance, the choice of damping of the instantaneous phase measurement may contribute, excluding diffracted waves below a certain amplitude. Furthermore, the iterative update of the model by the non-linear iterative scheme may proceed differently for β_{sh} and β_{sv}, particularly within the first iterations. In a full-waveform inversion for the Australian continent, Fichtner et al. (2010) noticed that radial anisotropy is increasingly well-explained by isotropic structure with an increasing number of iterations.

While our study is based on an idealised setup with a larger number of receivers than usually available in real-data inversions, differences in the resolution of the components may also be introduced by a different sampling of the model, for instance if more vertical- than horizontal-component seismograms are used.

The main conclusion is that care needs to be taken when interpreting differences between β_{sh} and β_{sv} in terms of radially anisotropic structure. This may be particularly problematic in non-linear full-waveform inversions, where diffracted waves play a significant role in constraining the small-scale structure. If the characteristics of diffraction depend strongly on the polarisation of the waves, this may lead to different β_{sh}- and β_{sv}-models even in an originally isotropic model. Clearly, more research is needed to further investigate a possible dependence of finite-frequency ef-

fects on wave polarisation. For the specific case of a cylindrical plume, our results indicate that within the upper $\sim$ 1300 km, it will likely be better-resolved in β_{sh} than in β_{sv}. For a full-mantle study, the isotropic S-velocity seems to provide the best overall resolution over the full depth range.

Chapter 6

The Iceland - Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion

6.1 Summary

We present a high-resolution S-velocity model of the North Atlantic region, revealing structural features in unprecedented detail down to a depth of 1300 km. The model is derived using full-waveform tomography. More specifically, we minimise the instantaneous phase misfit between synthetic and observed body- as well as surface-waveforms iteratively in a full three-dimensional, adjoint inversion. Highlights of the model in the upper mantle include a well-resolved Mid-Atlantic Ridge and two distinguishable strong low-velocity regions beneath Iceland and beneath the Kolbeinsey Ridge

The content of this chapter is under revision as: Rickers, F., Fichtner, A. and Trampert, J., 2013. The Iceland - Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion. Earth Planet. Sci. Lett.

west of Jan Mayen. A sub-lithospheric low-velocity layer is imaged beneath much of the oceanic lithosphere, consistent with the long-wavelength bathymetric high of the North Atlantic. The low-velocity layer extends locally beneath the continental lithosphere of the southern Scandinavian Mountains, the Danish Basin, part of the British Isles and eastern Greenland. All these regions experienced post-rift uplift in Neogene times, for which the underlying mechanism is not well understood. The spatial correlation between the low-velocity layer and uplifted regions suggests dynamic support by low-density asthenosphere originating from the Iceland and Jan Mayen hotspots. Our model further suggests a lower-mantle source for the Iceland and Jan Mayen hotspots. Two distinguishable low-velocity conduits are imaged, connecting the upper-mantle anomalies beneath Iceland and Jan Mayen into the lower mantle. Both conduits are tilted to the South-East, reflecting the westward motion of the Mid-Atlantic Ridge. The location of the imaged Iceland conduit is in agreement with the observation of a locally thinned transition zone south of Iceland from receiver function studies.

6.2 Introduction

The North American and Eurasian continental margins are drifting apart since the onset of ocean spreading in the North Atlantic about 55 Ma ago. The continental breakup went along with the eruption of large amounts of magma within a short geological time (White and McKenzie, 1989). Following the breakup, the magma production rate along the Mid-Atlantic Ridge remained locally unusually high. This resulted in the formation of Iceland (Fig. 6.1), which is part of an extensive bathymetric and gravimetric high observed over much of the North Atlantic (Jones et al., 2002). To the South-West of Iceland, the elevation decreases gradually along the Reykjanes Ridge towards the Charlie-Gibbs Fracture Zone. In contrast, the Kolbeinsey Ridge to the North of Iceland remains at a relatively constant, high elevation until it encounters the Jan Mayen Fracture Zone. The Jan Mayen Islands are another centre of increased magma production, possibly related to a separate hotspot. However, existing global and regional seismic models cannot resolve individual hotspots in this region (e.g. Ritsema et al., 1999; Ritsema and Allen, 2003; Bijwaard and Spakman, 1999; Grand, 2002; Pilidou et al., 2005; Legendre et al., 2012).

Considerable (kilometre-scale) post-rift uplift in Neogene times (Fig.

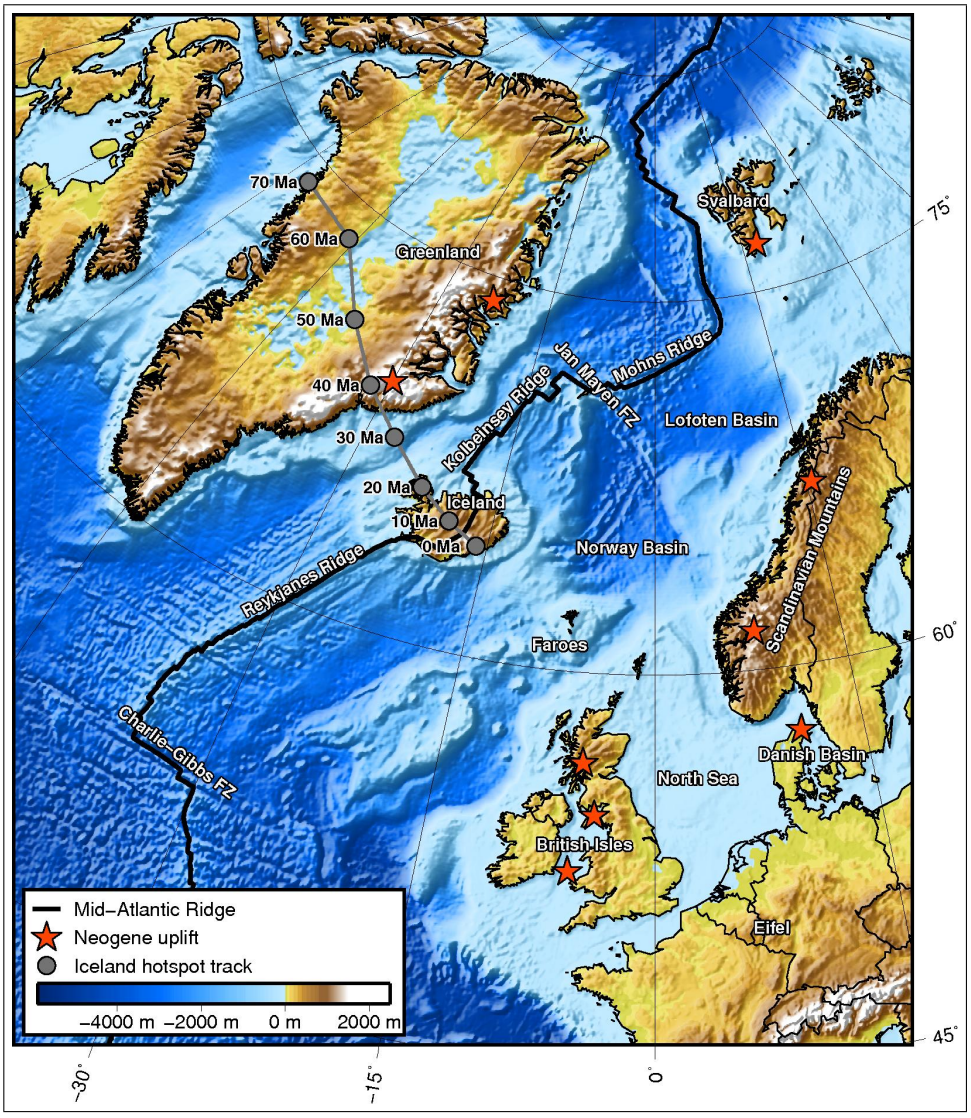


Figure 6.1: Bathymetry and topography of the North Atlantic region. The black line indicates the Mid-Alantic Ridge, the grey dots represent the reconstructed Iceland hotspot track between 70 Ma and today (Lawver and Müller, 1994). Red stars indicate a (non-complete) selection of regions where Neogene uplift is documented (after Japsen and Chalmers, 2000; Japsen et al., 2007; Holford et al., 2008, and references therein). FZ is used as abbreviation for Fracture Zone.

6.1) is documented on the continental shelves surrounding the North Atlantic (see summaries in e.g. Japsen and Chalmers, 2000; Doré et al., 2002; Carminati et al., 2009). Uplifted regions include, among others, the southern and northern Scandinavian Mountains in western Scandinavia (Rohrman et al., 1995; Redfield et al., 2005), part of the British Isles (George, 1966; Green, 1989; Japsen, 1997; Duncan et al., 1998; Hall and Bishop, 2002; Holford et al., 2008), the Danish Basin (Japsen et al., 2002, 2007), eastern Greenland (Mathiesen et al., 2000; Johnson and Gallagher, 2000) and Svalbard (Vågnes and Amundsen, 1993). Deep cratonic roots, which could isostatically balance the additional topography, are lacking beneath these regions. The mechanism for the uplift is debated (e.g. Rohrman and van der Beek, 1996; Ebbing and Olesen, 2005; Holford et al., 2008; Pascal and Olesen, 2009; Ebbing et al., 2012). A connection to the Iceland hotspot is supported by the tomographic study of Weidle and Maupin (2008), who image a low-velocity finger extending beneath the lithosphere to the southern Scandinavian Mountains. Beneath parts of the British Isles, Arrowsmith et al. (2005) image low velocities, and Davis et al. (2012) find a negative correlation between crustal thickness and topography. Both of these studies indicate dynamic support, which they suggest to be related to low-density material from the Iceland hotspot. Scaled long-wavelength gravity maps (Jones et al., 2002) indicate that parts of Britain, southern Scandinavia and eastern Greenland are presently experiencing dynamic support related to the Iceland hotspot.

Morgan (1971) proposed that the localised, long-lived magmatism of hotspots could be caused by mantle plumes. They are envisioned as narrow, hot upwellings originating from the core-mantle boundary. In the case of Iceland, a possible mantle plume would be interacting with a spreading ridge, and might have weakened the continental lithosphere prior to continental breakup. The mantle plume hypothesis found widespread acceptance, but unambiguous seismic evidence for the continuation of upper-mantle low-velocity anomalies into the lower mantle beneath Iceland is still sparse.

Some global tomographic models show broad and relatively weak low-velocity structures in the lower mantle below Iceland (e.g. Bijwaard and Spakman, 1999; Ritsema et al., 1999; Zhao, 2004), which cannot be interpreted with certainty as continuous plume structures. Several studies attempted to image the mantle below Iceland using array data recorded on Iceland (Tryggvason et al., 1983; Wolfe et al., 1997; Foulger et al., 2001;

Allen et al., 2002; Bjarnason et al., 2002; Delorey et al., 2007). While all these models agree on the presence of low velocities in the uppermost mantle, results are contradictory in deeper mantle regions. Keller et al. (2000) showed that the limited array aperture used in such studies does not permit unique constraints on the depth extent of the Iceland anomaly. Using receiver functions, Shen et al. (1998, 2002) imaged a locally thinned transition zone below Iceland, indicative for elevated temperatures which could possibly be caused by a hot plume. Other studies, however, imaged a flat transition zone using similar methods (Du et al., 2006).

Hwang et al. (2011) find that simple cross-correlation traveltime measurements are not suited to extract lower-mantle plume information from seismic data. Rickers et al. (2012) come to the same conclusion and explain it with the fact that such methods do not account for diffraction effects in seismic wave propagation, which are strong in the case of small-scale heterogeneities.

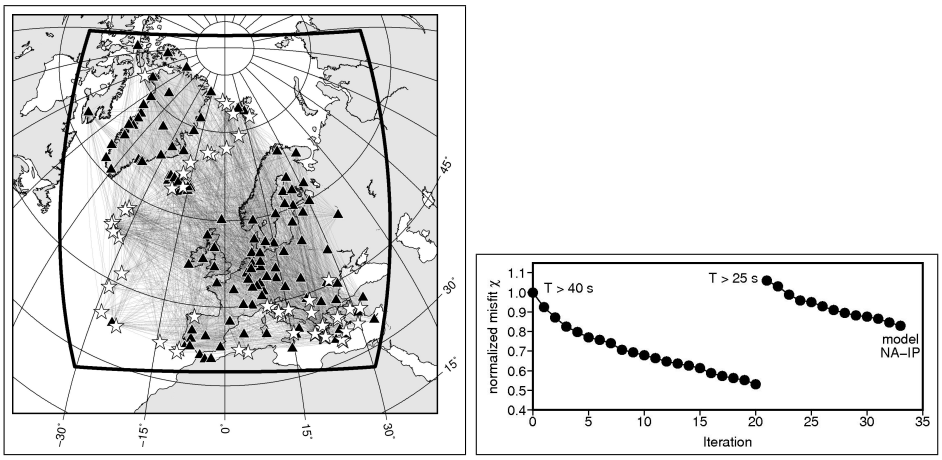
The lack of unambiguous evidence for a lower-mantle plume below Iceland leaves the possibility of a different mechanism producing the increased magma volumes. Such a mechanism could be fertile, old crust in the upper mantle, being overridden by the Mid-Atlantic Ridge (Foulger et al., 2001; Foulger and Anderson, 2005; Foulger, 2012).

The need for an improved tomographic model of the whole North Atlantic region, covering the upper and at least part of the lower mantle, motivated this study. Our model is based on non-linear full-waveform inversion techniques and measurements of the instantaneous phase misfit. This method has been shown to overcome limitations of classical linearised tomography and to potentially improve tomographic resolution in both the upper as well as the lower mantle (Fichtner and Trampert, 2011b; Rickers et al., 2012).

6.3 Method, Data and Inversion

6.3.1 Tomographic method

The inversion is based on a non-linear iterative conjugate gradient optimisation scheme (Fletcher and Reeves, 1964). Gradients are computed with the adjoint method (Tarantola, 1984; Tromp et al., 2005; Fichtner et al., 2006), and the misfit between synthetic and observed seismograms is quantified by their instantaneous phase difference (Bozdağ et al., 2011; Rickers et al.,



(a) Events (white) and stations (black) (b) Misfit vs. Iteration number

Figure 6.2: (a) Events (white stars) and receivers (black triangles) used for the inversion. The thin lines represent the great-circle ray paths of the data. The black outline indicates the model boundaries. (b) Normalised value of the misfit function χ as a function of the number of iterations. The normalisation, for the purpose of clarity of the plot, is with respect to the initial value of χ at iteration 0. Up to iteration 20, the minimum period of the seismograms is 40 seconds. From iteration 21, the minimum period is 25 seconds.

2012). This misfit is a time-dependent and amplitude-independent measurement of waveform differences. In contrast to traditional tomographic methods, it allows for the consideration of smaller-amplitude phases such as diffracted waves in an inversion. The proper treatment of diffracted waves improves resolution and is essential for the imaging of small-scale anomalies such as plumes, in particular in the lower mantle (Rickers et al., 2012). The aim of the inversion is the iterative minimisation of a misfit function χ, which for a current model m is defined as

$$\chi(m) = \frac{1}{2} \sum_{i=1}^{N} \int_0^T \|\phi_i^{synt}(t,m) - \phi_i^{obs}(t)\|^2 dt, \tag{6.1}$$

where $\phi_i^{synt}(t,m)$ and $\phi_i^{obs}(t)$ denote the instantaneous phase of the (windowed) synthetic and observed seismogram i at time t, respectively. The total number of seismograms used in the inversion is N, the end time of the synthetic seismograms is T. For a detailed derivation of the instantaneous

phase misfit, the reader is referred to Bozdağ et al. (2011).

The forward and adjoint synthetic wavefields are computed with the spectral element code SES3D (Fichtner and Igel, 2008), which accurately solves the elastic wave equation in strongly heterogeneous, three-dimensional media. Similar techniques have been successfully applied in recent regional full-waveform tomographic studies (e.g. Fichtner et al., 2009; Tape et al., 2010; Zhu et al., 2012). This study is the first real data application of the instantaneous phase misfit, which has been demonstrated to potentially resolve lower-mantle plumes in preceding synthetic studies (Rickers et al., 2012).

6.3.2 Model setup

The model covers all of Europe, Greenland, as well as parts of Canada and Russia, spanning a distance of 6000×6000 km in the lateral directions. The covered area is marked by the black outline in Fig. 6.2(a). In the vertical direction, the model extends down to a depth of 1300 km. While the actual region of interest is smaller and restricted to the North Atlantic and the adjacent regions, the large model area improves azimuthal ray coverage because events occurring in southern Europe can be included in the inversion, in addition to the events at the Mid-Atlantic Ridge. Furthermore, larger epicentral distances improve the illumination of the lower mantle by body waves.

The model is divided into blocks with a horizontal spacing of 100 km and a depth-dependent vertical spacing of 10 km in the uppermost 350 km, 20 km between 350 km and 700 km depth and 50 km between 700 and 1300 km depth. This honours the steeper vertical velocity gradients and the higher resolution in the uppermost mantle.

We invert for both the horizontally and vertically polarised S-wave velocity β_{sh} and β_{sv}, which is necessary to fit Rayleigh- as well as Love-waves. The P-wave velocity α, the density ρ and the attenuation Q are kept at the initial values of the starting model throughout the inversion. We base the discussion of the model on the β_{sh}-component of the S-velocity, which our resolution tests indicate to be better resolved.

6.3.3 Starting model and 1-D reference model

Phase measurements, such as the instantaneous phase difference used in this study to quantify waveform misfits, require synthetic and observed wave-

forms to be within half a period of each other. An accurate starting model is therefore important. It increases the number and length of measurement windows that meet this requirement at the beginning of the inversion and accelerates the convergence towards the optimum model. We used the European full-waveform model developed by Fichtner and Trampert (2011b) as starting model. It is based on time-frequency misfits of surface- and body-waves with a minimum period of 60 s. In the remainder of this paper, we refer to it as EU-TF (Europe - Time Frequency). Coherently, we refer to our final model as NA-IP (North Atlantic - Instantaneous Phase). Model slices presented throughout this paper show velocity perturbations with respect to a 1-D reference model. This reference model is a modified version of PREM (Dziewonski and Anderson, 1981), with a linear gradient replacing the 220 km discontinuity (Fichtner and Trampert, 2011b).

6.3.4 Data

We used a carefully chosen set of high-quality seismic waveform data covering the North Atlantic region, including a substantial amount of body waves sampling the lower mantle. We selected 60 events with magnitudes above Mw 5.4, aiming at a good azimuthal coverage. Data were obtained from the *Incorporated Research Institutions for Seismology (IRIS)* and the *Observatories and Research Facilities for European Seismology (ORFEUS)*. Station networks include the *Greenland Ice Sheet Monitoring Network (GLISN)*, the *Greenland Lithosphere Analysed Teleseismically on the Ice Sheet (GLATIS) Network* and the *Iceland HOTSPOT Network*, which together with stations in continental Europe provided a good coverage of the whole North Atlantic region. The distribution of events and receivers and the ray coverage are shown in Fig. 6.2(a).

Because of the natural distribution of earthquakes and the uneven coverage of the region with seismic stations, a bias in the ray coverage cannot be avoided. The iterative inversion scheme can balance a possible bias to a certain degree, but this comes at the cost of a slower convergence of the misfit function. Careful initial weighting of the data can therefore accelerate the inversion. We attempted to correct for the uneven data coverage by manually adjusting the relative weight of stations in the inversion. Based on a visual inspection of the ray coverage and the local station density, the weight of the individual stations was adjusted within a range of $\pm 50\%$ with respect to their initial weight.

6.3.5 Inversion

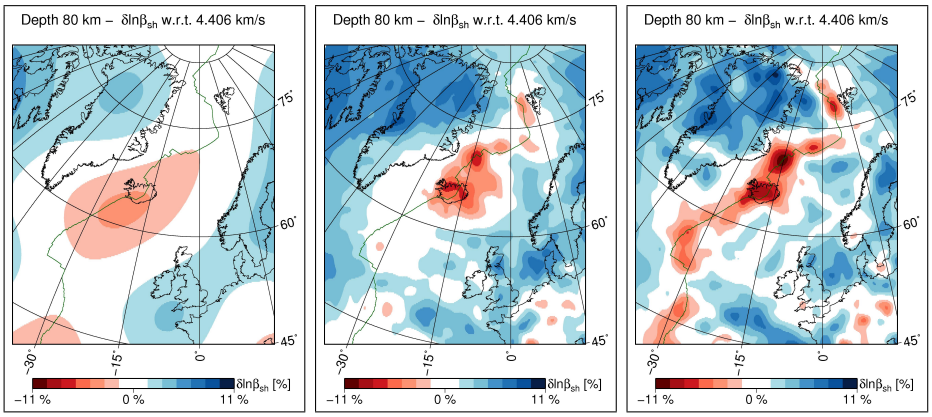
Prior to the start of the inversion, we filtered the observed and synthetic seismograms in a period range between 40 and 200 seconds. All three components were used, and we selected 4366 high-quality seismograms which allowed for meaningful measurements. We manually selected measurement windows where data and synthetics are within half a period of each other, and computed the instantaneous phase difference and corresponding adjoint sources (Tromp et al., 2005; Bozdağ et al., 2011; Rickers et al., 2012). The convergence of the misfit function χ as a function of the number of iterations is shown in Fig. 6.2(b). After 20 iterations, the initial misfit had decreased by almost 50%. At this stage of the inversion, we increased the frequency content of the seismograms to a minimum period of 25 s. We again selected measurement windows and performed 13 more iterations. While the misfit initially increased at iteration 21 due to the higher frequency content and changed measurement windows, it had decreased again by about 25% at the final iteration.

6.4 Validation

The inversion scheme is non-linear, which on the one hand allows for the inclusion of complex wave propagation effects leading to better-constrained models. On the other hand, it is difficult to assess the resolution of the resulting model because there is no simple operator relating inversion input and model output. We estimate the validity of the model in different ways. These are (1) a visual inspection of the global S-velocity model S20RTS (Ritsema et al., 1999), the initial model EU-TF (Fichtner and Trampert, 2011b) and the presented model NA-IP, (2) a comparison of waveform fits for these three models, (3) a linearised resolution analysis for specific regions of the model using Point Spread Functions (Fichtner and Trampert, 2011b).

6.4.1 Visual inspection

A visual inspection of horizontal model slices at a depth of 80 km is presented in Fig. 6.3. We show the global mantle model S20RTS (Ritsema et al., 1999) along with the full-waveform models because S20RTS served as starting model to construct EU-TF (Fichtner and Trampert, 2011b), which was then used as starting model to derive NA-IP. It is apparent that the imaged details increase from the smooth global model S20RTS



(a) S20RTS (b) EU-TF (c) NA-IP

Figure 6.3: Horizontal slices through three different models at 80 km depth beneath the North Atlantic region. (a) global S-velocity model S20RTS (Ritsema et al., 1999) (b) β_{sh}-component of full-waveform starting model EU-TF (Fichtner and Trampert, 2011b) (which itself was constructed using S20RTS as starting model) (c) β_{sh}-component of presented model NA-IP.

via the European full-waveform model EU-TF to the final North Atlantic model NA-IP. While in S20RTS, a smooth low-velocity zone covers the region around Iceland, in EU-TF it narrows down on Iceland and Jan Mayen and the two hotspots become distinguishable. Other features along the Mid-Atlantic Ridge are not well-defined, however. The higher frequency content, incorporation of regional data and possibly the choice of misfit in the present study eventually lead to a well-defined Mid-Atlantic Ridge in the final model NA-IP. High-velocity features, such as the Greenland and Baltic Shields as well as the North Sea, also become more pronounced. The correlation of imaged velocity perturbations with topographic and tectonic features (Fig. 6.1) indicates the geologic plausibility of the model, at least in the upper part.

An interesting observation is the much stronger perturbation amplitude of models derived with non-linear full-waveform methods (compare S20RTS and NA-IP in Fig. 6.3). While differences in regularisation and parametrisation certainly contribute to the difference in amplitude between the models, the occurrence of higher amplitudes in full-waveform models is also a result of the consideration of diffraction effects. While in classi-

cal linearised inversions, these effects are not properly accounted for and consequently do not contribute significantly to the model, diffracted waves are focused back onto their source region in non-linear full-waveform inversions, leading to a much better recovery of the perturbation amplitude (see Rickers et al., 2012). A higher frequency content enhances this effect for small-scale structures, explaining the difference in perturbation strength of, for instance, the Iceland hotspot between models EU-TF and NA-IP.

6.4.2 Waveform fits

Improved waveform fits are another indicator of an improved model. In Fig. 6.4, examples are presented for a Mw 6.0 event which occurred in Turkey on September 19, 2011. This event is one of the events that was used in the inversion. Synthetic waveforms computed in models S20RTS, EU-TF and NA-IP are compared to the corresponding data recorded at different stations. The waveforms are filtered at periods between 25 and 200 s and contain S-waves and surface-waves as well as their reflections and diffractions. The improved waveform fit of the presented model NA-IP is evident, and in many cases every wiggle within the measurement window could be closely fitted to the data. We note that the amplitude fit improves along with the phase, even though we did not explicitly invert the amplitude information.

6.4.3 Point Spread Functions

A more quantitative method to assess the resolution of tomographic models derived with non-linear inversion techniques are Point Spread Functions (PSFs) (Fichtner and Trampert, 2011b). A PSF indicates how well a small, localised structure in the final model is resolved, and in particular it reveals if trade-offs with structure elsewhere in the model exist. The panels in Fig. 6.5(a) show two smoothed, localised perturbations in β_{sh}, which is the velocity of horizontally polarised S-waves. The diameter of the perturbations is $\sim$200 km, and they are centred at depths of 150 km and 900 km below Iceland. The perturbations are superimposed onto the final model NA-IP. The chosen location of the perturbations corresponds to locations where low velocities are imaged in model NA-IP. A confirmation that these structures are well-resolved and no strong trade-offs exist could justify an interpretation in terms of an Iceland plume. Trade-offs in the form of smearing can lead to apparent plume-like structure in tomo-

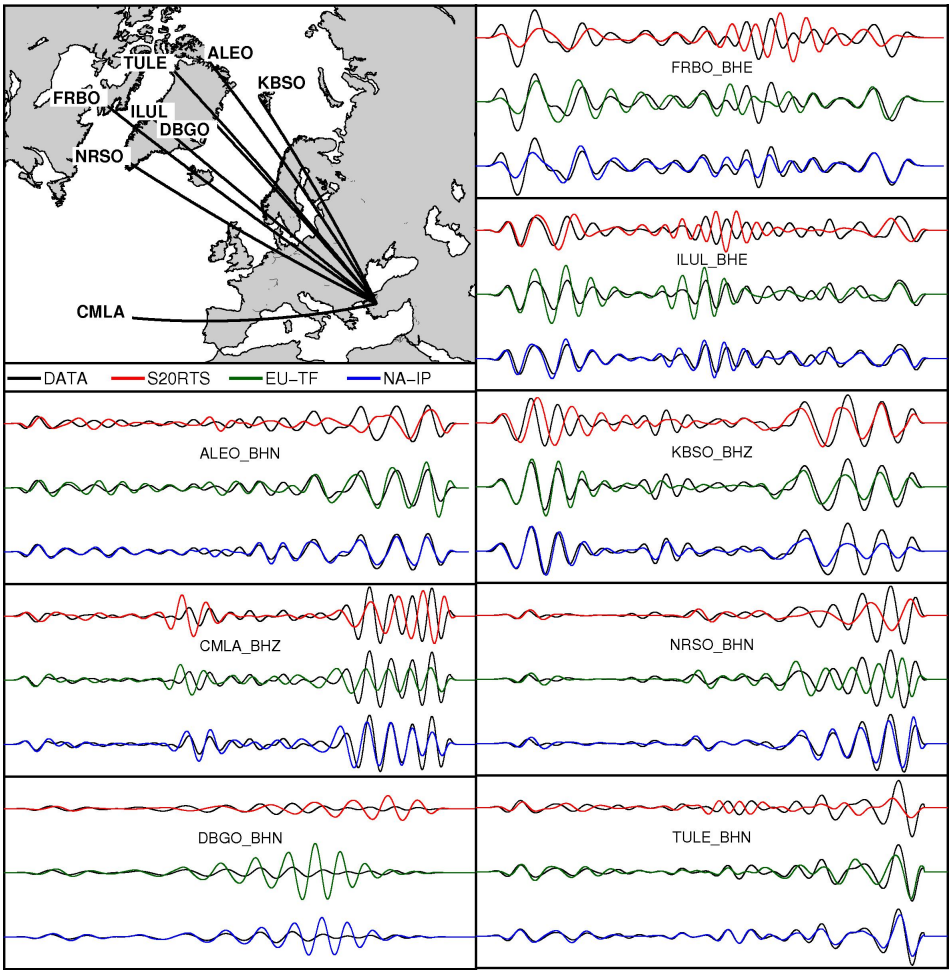
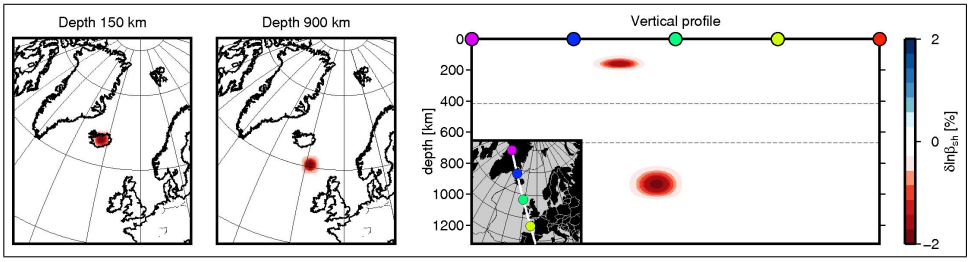
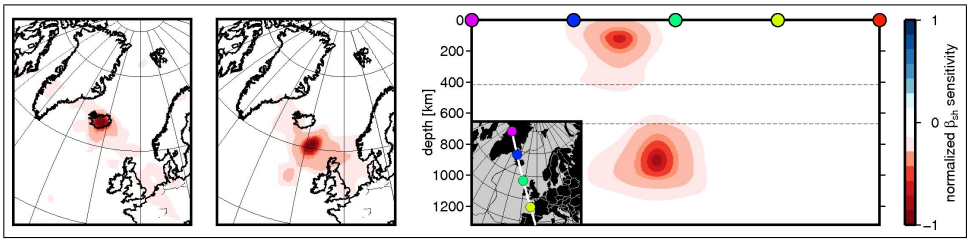


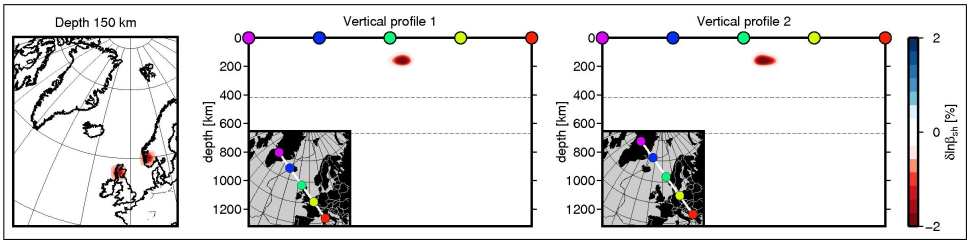
Figure 6.4: Waveform fit between synthetics and data for different models. Synthetics are computed in the global model S20RTS (Ritsema et al., 1999) (red), in the full-waveform starting model EU-TF (Fichtner and Trampert, 2011b) (green) and in the presented model NA-IP (blue). The data (black) are from a Mw 6.0 event in Turkey on September 19, 2011, recorded at different stations. The presented waveforms are S- and surface-waves, including their reflections and diffractions, at a minimum period of 25 seconds.



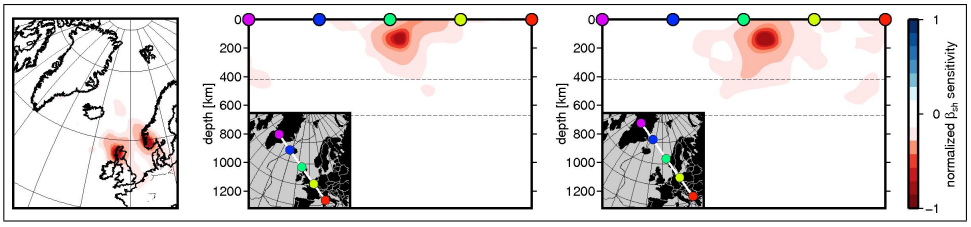
(a) Test 1 - Input: Smoothed point perturbations



(b) Test 1 - Output: Point Spread Functions



(c) Test 2 - Input: Smoothed point perturbations



(d) Test 2 - Output: Point Spread Functions

Figure 6.5: Two examples for local resolution tests with Point Spread Functions (PSFs) (a) Test 1 - Input: two smoothed point perturbations of the SH-velocity β_{sh} at depths of 150 km and 900 km below Iceland. The perturbations are superimposed onto the final model NA-IP. The two panels on the left show horizontal slices through the perturbations at depths of 150 km and 900 km, the panel on the right shows a vertical slice through both perturbations. (b) Test 1 - Output: PSFs showing the sensitivity of the β_{sh}-component as a response to the perturbations. (c) Test 2 - Input: two smoothed point perturbations in β_{sh} at a depth of 150 km. The panel on the left shows a horizontal slice through the perturbations at a depth of 150 km, the two panels on the right show vertical slices through each of the perturbations.(d) Test 2 - Output: PSFs showing the sensitivity of the β_{sh}-component as a response to the perturbations.

graphic images beneath hotspot regions (Keller et al., 2000). The panels in Fig. 6.5(b) show the resulting PSFs for the β_{sh}-component. The two perturbations are clearly distinguishable and the location of their maximum amplitude has not changed, indicating that imaged structures at these locations are well-resolved. In the two bottom panels of Fig. 6.5, we present a second test at two different locations at a depth of 150 km, corresponding to two prominent low-velocity structures in model NA-IP (see section 6.5.3). The diameter of the smoothed point perturbations in β_{sh} is $\sim$150 km (Fig. 6.5(c)). The resulting PSFs in Fig. 6.5(d) again indicate that the model is well-resolved at the tested locations. We calculated PSFs for a number of different locations and depths, consistently indicating that the β_{sh}-component of our model is well-resolved. In contrast, we find that the β_{sv}-component is less well-resolved in the deeper part of the model.

6.5 Model

In this section, the β_{sh}-component of model NA-IP is presented. The discussion is divided into four parts, each focusing on a particular feature of the model. These are (1) the cratons and ocean basins, (2) the Mid-Atlantic Ridge and the Iceland and Jan Mayen hotspots, (3) a sub-lithospheric low-velocity layer beneath oceanic and continental lithosphere, (4) low-velocity conduits beneath the Iceland and Jan Mayen hotspots, suggesting a lower-mantle source.

6.5.1 Cratons and ocean basins

In Fig. 6.6, horizontal slices of the model at depths between 60 and 300 km are presented. Prominent high-velocity features are the thick cratonic Greenland and Baltic Shields. They are imaged in numerous studies (e.g. Pilidou et al., 2005; Ritsema et al., 2011; Legendre et al., 2012), and are recovered sharply and with great detail in our model. The cratonic roots are deepest in northern and south-western Greenland, where they extend to about 160 to 200 km depth. Towards eastern Greenland, their thickness decreases to below 100 km, which is better visible from the western part of the vertical slices in Fig. 6.7. An East-West oriented corridor of reduced lithospheric thickness is imaged across Greenland, which is particularly apparent at depths between 120 and 200 km (Fig. 6.6). Its location approximately correlates with the reconstructed track of the Iceland hotspot between 70 and 40 Ma ago (see Fig. 6.1) (Morgan, 1983; Lawver and Müller, 1994),

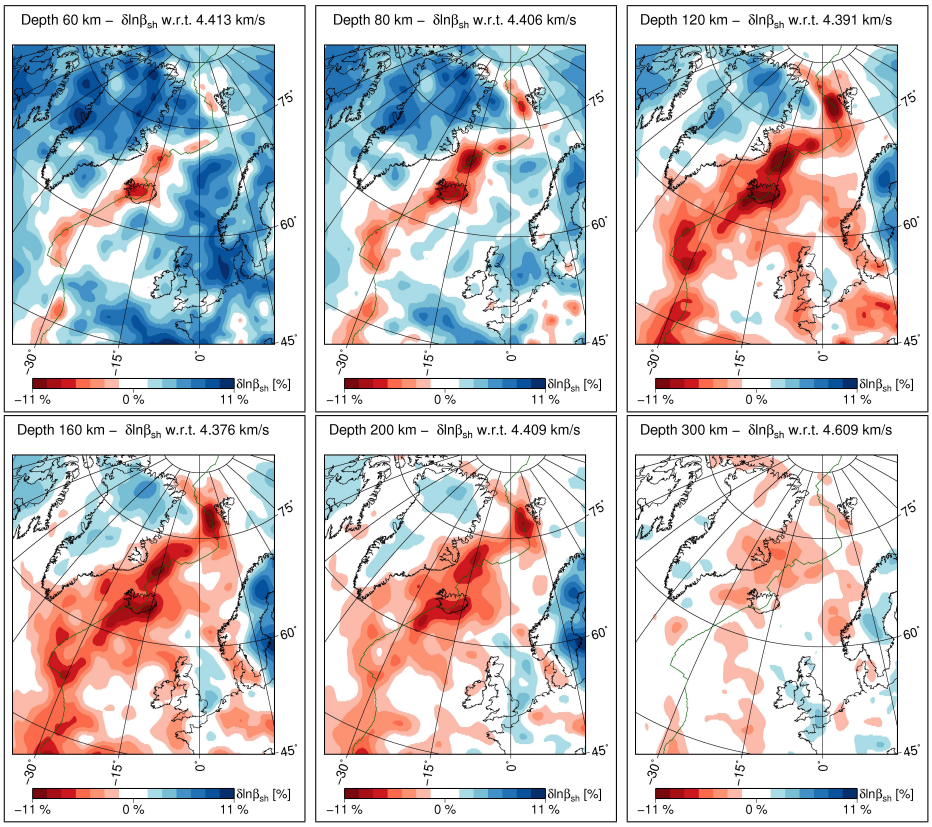


Figure 6.6: Horizontal slices through the β_{sh}-component of model NA-IP, at depths between 60 and 300 km. Perturbations are with respect to the reference velocity indicated in each panel.

which may have caused weakening and thinning of the lithosphere along its track.

The Baltic Shield is characterised by high velocities to depths of more than 200 km. Our model shows a sharp boundary between the deep Precambrian lithosphere of the Baltic Shield and the much shallower lithosphere of South-West Norway and the Danish Basin. This is particularly apparent from the eastern end of the vertical slice in Fig. 6.7(b), which cuts into the Baltic Shield of southern Sweden. High velocities are further imaged beneath the North Sea and, to a lesser extent, beneath the Lofoten and Norway Basins.

6.5.2 The Mid-Atlantic Ridge and the Iceland and Jan Mayen hotspots

At shallow depths between 60 and 80 km (Fig. 6.6), low velocities are imaged within a narrow band to both sides of the Mid-Atlantic Ridge, where new ocean floor is being created and the oceanic lithosphere is thinnest. Two regions of anomalously low velocities are centred beneath Iceland and beneath the northern Kolbeinsey Ridge, just south of the Jan Mayen Fracture Zone. Velocities here are reduced by up to -11% with respect to the 1-D reference model. At depths between 100 and 200 km, these two low-velocity regions are still clearly distinguishable. At these depths, the Iceland anomaly is elongated in the direction of the Reyjkanes Ridge, with a slight tilt to the South. The Jan Mayen anomaly covers the whole length of the Kolbeinsey Ridge and extends beyond the Jan Mayen Fracture Zone to the southern Mohns Ridge. Velocity perturbations in those along-ridge regions reach between -7% and -11%. Below a depth of $\sim$250 km, perturbations are much weaker and no longer centered beneath the ridge.

The elevated magmatism in the region of the Jan Mayen Islands and the unusually high elevation of the Kolbeinsey Ridge along its whole length indicate the presence of a separate hotspot. This is supported by geochemical isotope studies (Schilling, 1985; Schilling et al., 1999), which suggest different sources of the basalts in the regions of Iceland and Jan Mayen. They propose that the boundary between the influence zones of the Iceland and the Jan Mayen hotspots is located about halfway along the Kolbeinsey Ridge between Iceland and Jan Mayen.

Most existing global and regional tomographic models image low velocities in the Iceland - Jan Mayen region (e.g. Bijwaard and Spakman, 1999; Ritsema et al., 2011; Pilidou et al., 2005; Legendre et al., 2012). The regional surface-wave model of Pilidou et al. (2005) appears to have the best resolution among these models. They image an elongated low-velocity zone centred on Iceland, extending $\sim$1700 km along-ridge and $\sim$600 km across at 100 km depth. However, the resolution of their model is not sufficient to constrain the anomaly further, and to make a statement regarding a possible separation of the Iceland and Jan Mayen hotspots.

Resolution tests using PSFs indicate that both hotspots are well-resolved. Our model favours the boundary between the hotspots to be located just to the North of Iceland.

To the South-West of Iceland, the deviation of the low-velocity anomaly with respect to the Reykjanes ridge may be related to radial anisotropy and

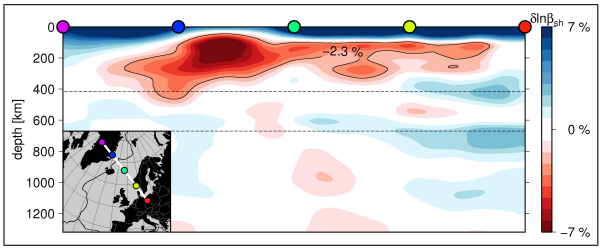
the fact that the presented model is a β_{sh}-model. Delorey et al. (2007) find an anomalously strongv(∼5%) and asymmetric (stronger on the European than on the North American side) $\beta_{sv} > \beta_{sh}$ - anisotropy pattern in the upper 100 km beneath the Reykjanes Ridge, peaking at a distance of ∼120 km parallel to the ridge axis on the European side. Their model represents an along-axis average over the length of the Reykjanes ridge and does not constrain along-ridge variations such as the increasing tilt away from Iceland that is imaged in our model. The resolution of their anisotropy model however does not extend below ∼100 km depth, where the observed pattern is strongest in our model.

6.5.3 Sub-lithospheric low-velocity layer beneath oceanic and continental lithosphere - connection to Neogene uplift?

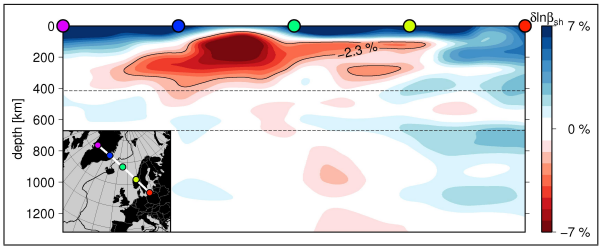
Velocity perturbations are strongest in the vicinity of the hotspots along the Mid-Atlantic Ridge, but a low-velocity layer is imaged beneath much of the oceanic lithosphere of the North Atlantic. It locally extends beneath the continental lithosphere. This suggests a flow of low-velocity hotspot material not only along-ridge, but also in directions away from the Mid-Atlantic Ridge over long distances.

Particularly intriguing are two low-velocity fingers at depths between 120 and 180 km (Fig. 6.6), reaching from the hotspot region far into the continental lithosphere. One finger extends towards South-West Norway and further beneath the Danish Basin. The other finger stretches along the British Isles to the tip of Brittany, roughly following the line defined by the Irish Sea between Ireland and the eastern British Isles. These two low-velocity fingers are separated by the North Sea, below which weakly positive velocities are imaged.

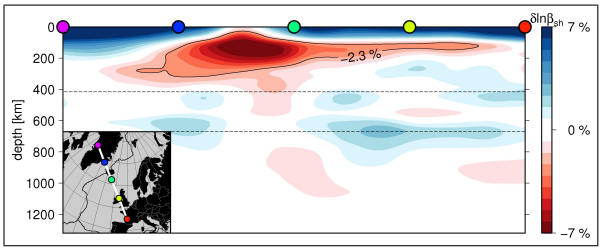
The vertical slices in Fig. 6.7 yield a clearer image of these sub-lithospheric low-velocity fingers and their connection to the Iceland and Jan Mayen hotspots. The slice in Fig. 6.7(a) cuts through the northern Kolbeinsey Ridge and follows the Scandinavian finger to Denmark. The low-velocity material forms a layer below the lithosphere. The layer thickness, defined by a velocity reduction of more than −2.3%, is approximately 60 to 80 km. Locally, in particular beneath South-West Norway, the layer is considerably thicker. This may be caused by a branch of the Jan Mayen plume (see Section 6.5.4 for details). A slice slightly further north (Fig. 6.7(b)) indicates that the low-velocity layer is confined to the shallow lithosphere of South-



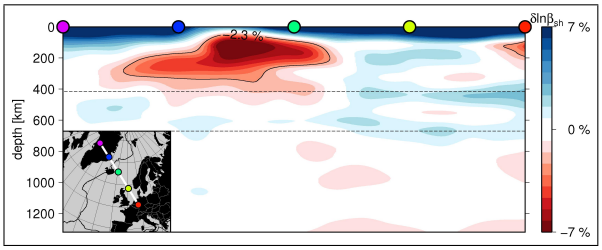
(a) Scandinavian low-velocity finger



(b) North of Scandinavian finger - Baltic Shield

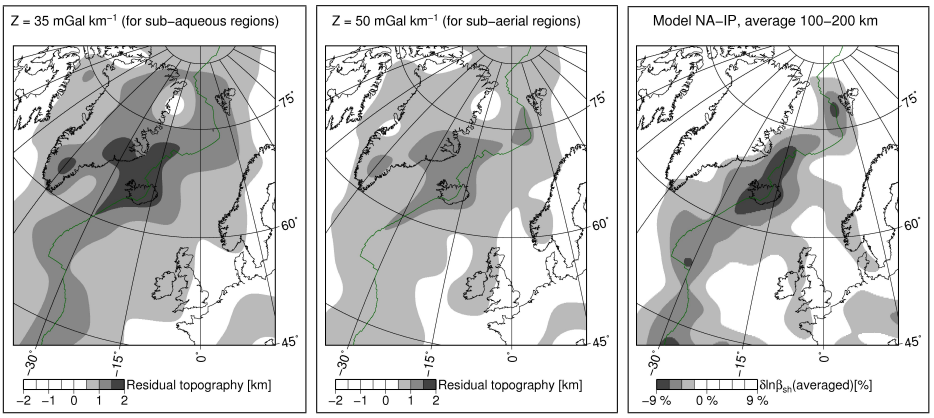


(c) British low-velocity finger



(d) In-between both low-velocity fingers - North Sea

Figure 6.7: Vertical slices through the β_{sh}-component of model NA-IP. The slices are roughly perpendicular to the Mid-Atlantic Ridge, the exact location is shown in the overview map in each panel. The black contours indicate perturbations of -2.3% with respect to the 1-D reference model. Slices (a) and (c) cut roughly along the sub-lithospheric low-velocity fingers extending into Scandinavia and the British Isles, respectively. Slice (b) cuts slightly north of the Scandinavian low-velocity finger and into to the Baltic shield. Slice (d) cuts inbetween the two low-velocity fingers through the (high-velocity) North Sea.



(a) Gravity (sub-aqueous) (b) Gravity (sub-aerial) (c) Model NA-IP (avg. 100-200 km)

Figure 6.8: (a) and (b) Estimates of present-day dynamic support in the North Atlantic region, calculated according to (Jones et al., 2002) through division of the long-wavelength free-air gravity anomaly field by a constant admittance Z. For estimates of dynamic support in sub-aqueous regions, Z = 35 mGal km^{-1} is considered appropriate, for sub-aerial regions Z = 50 mGal km^{-1}. (c) Long-wavelength average velocity perturbation between 100 km and 200 km depth of model NA-IP. To facilitate comparison with the estimated dynamic support, the average velocity is lowpass-filtered by convolution with a Gaussian of width 800 km.

West Scandinavia. It does not extend beneath the thicker Precambriam lithosphere of the Baltic Shield.

The vertical slice in Fig. 6.7(c) cuts through Iceland and along the finger beneath the British Isles. The low-velocity layer is slightly thinner and weaker than the layer below southern Scandinavia. It is not thickened locally. A vertical slice through Iceland and the North Sea (Fig. 6.7(d)), in-between the two fingers, shows that the low-velocity layer does not extend eastward beyond the Faroe Islands. Instead, weakly positive velocity perturbations are imaged beneath the lithosphere of the North Sea. The low-velocity region at the eastern end of the slice is related to the Eifel volcanic zone.

It appears that lithospheric thickness does not increase notably with growing distance from the Mid-Atlantic Ridge where the sub-lithospheric low-velocity layer is present. In contrast, the slice through the North Sea

shows an increase in lithospheric thickness away from the ridge. This suggests the existence of thin-lithosphere channels beneath the oceanic and continental lithosphere, along which the hot asthenosphere flows over long distances.

The presence of the sub-lithospheric low-velocity layer in our model beneath regions of major (kilometre-scale) post-rift Neogene uplift (Fig. 6.1) suggests dynamic support of the uplifted regions. A correlation is particularly evident for the British Isles and South-West Scandinavia.

Dynamic support of post-rift uplifted regions by hotspot-related low-density material has been proposed by various authors (e.g. Rohrman and van der Beek, 1996; Jones et al., 2002; Bott and Bott, 2004). This may explain the lack of deep crustal roots which normally isostatically balance elevated topography.

A number of smaller-scale tomographic studies confirm the presence of low velocities beneath uplifted regions. Weidle and Maupin (2008) image a sub-lithospheric low-velocity finger reaching from the region of the Iceland hotspot to southern Scandinavia. They are careful with their interpretation, however, because their model shows a continuation of the low-velocity anomaly into southern Sweden, where other models image the much thicker cratonic lithosphere of the Baltic Shield. Their imaged finger is consistent with our model, except that we do not image a continuation of the low-velocity layer beneath the Baltic Shield. Our model shows a continuation of the finger into the Danish Basin, along the southern boundary of the Baltic Shield. In the Danish Basin, Neogene uplift (and successive erosion) is documented (Japsen et al., 2002, 2007).

Beneath the northern and central British Isles, Arrowsmith et al. (2005) image a region of reduced compressional wave velocities which they attribute to a low-velocity finger from the Iceland hotspot. Davis et al. (2012) find a northward decrease in crustal thickness beneath Britain, which negatively correlates to average elevation. They argue for the presence of a low-velocity layer beneath the lithosphere, providing isostatic support of the elevated regions.

For the regions of the British Isles and Southern Scandinavia, the correlation of the sub-lithospheric low-velocity layer with Neogene uplift is most apparent. Low velocities are further imaged at depths between ∼100 km and 400 km beneath the cratonic lithosphere of eastern Greenland, possibly isostatically balancing the major Neogene uplift documented in this region. The Neogene uplift of the northern Scandinavian Mountains may

be caused by a different mechanism than the uplift of the southern Scandinavian Mountains, unrelated to mantle processes (Ebbing and Olesen, 2005). This would be consistent with our model, which does not show a low-velocity layer beneath the northern Scandinavian Mountains. Svalbard, where major Neogene uplift is also documented, is located at the edge of the well-sampled part of our model. A strong low-velocity anomaly is imaged to its West beneath the Mid-Atlantic Ridge. The anomaly appears to extend beneath the western part of Svalbard, but the resolution at the edge of the model may not be sufficient to confidently confirm a correlation.

The pattern of the low-velocity layer imaged in our model correlates well with estimates of present-day dynamic support by Jones et al. (2002). Their estimations are based on scaled versions of the long-wavelength (>800 km) free-air gravity anomaly field. In Fig. 6.8, we show an updated figure from Jones et al. (2002) based on the more recent GRACE gravity data (Tapley et al., 2007). As described in Jones et al. (2002), the appropriate scaling factors Z are different for sub-aqueous (Fig. 6.8(a), Z = 35 mGal km^{-1}) and sub-aerial (Fig. 6.8(b), Z = 50 mGal km^{-1}) regions. To facilitate a comparison with our model, Fig. 6.8(c) shows the averaged velocity perturbation of model NA-IP in the depth range between 100 km and 200 km. This depth range covers much of the sub-lithospheric low-velocity layer at the European side of the Mid-Atlantic ridge, as well as the low velocities beneath the Mid-Atlantic ridge. It does not cover the deeper low-velocity layer beneath Greenland. Similar to the gravity field, the averaged velocity has been low-pass filtered through convolution with a Gaussian of width 800 km.

It is apparent that the estimations of present-day dynamic support correlate to the imaged sub-lithospheric low-velocity layer on the European side of the ridge. The correlation is most intriguing for the fingers beneath the British Isles and the southern Scandinavian Mountains (sub-aerial, Fig. 6.8(b) and 6.8(c)). The low-velocity pattern along the Mid-Atlantic Ridge and particularly in the region of Iceland and the Kolbeinsey Ridge also resembles the pattern of estimated present day support fairly well (sub-aqueous, Fig. 6.8(a) and 6.8(c)).

Our S-velocity model gives strong support to the inferences of Jones et al. (2002) of wide-spread dynamic support in the North-Atlantic region by the Iceland plume. A sub-lithospheric flow of hotspot material in directions perpendicular to the Mid-Atlantic ridge is also argued for by the tomographic study of Delorey et al. (2007).

6.5.4 A lower-mantle source for the Iceland and Jan Mayen hotspots

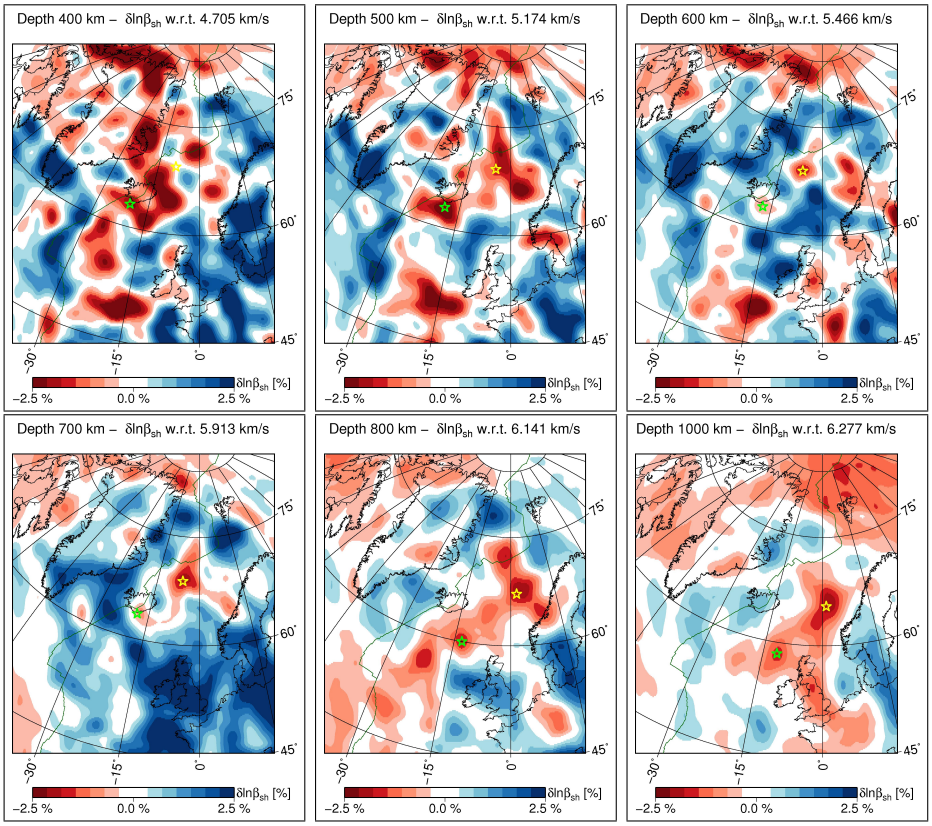
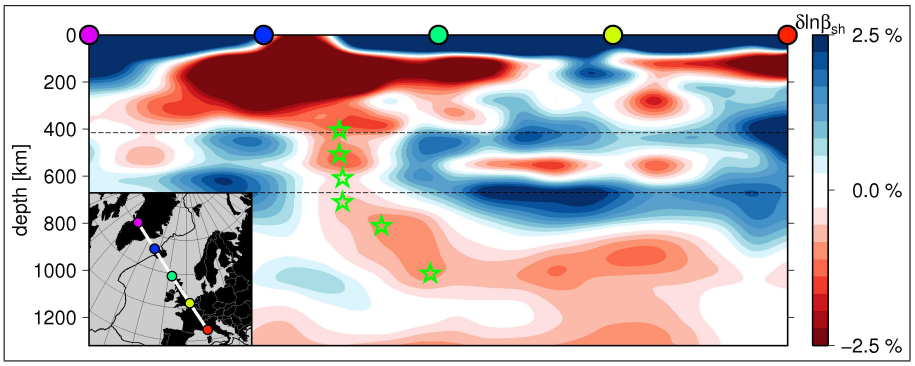
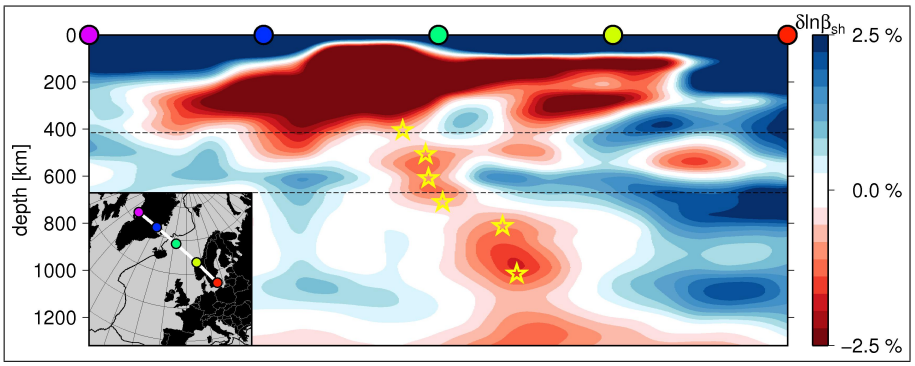


Figure 6.9: Horizontal slices through the β_{sh}-component of model NA-IP, at depths between 400 and 1000 km. Perturbations are with respect to the reference velocity indicated in each panel. The stars indicate the location of the Iceland (green) and Jan Mayen (yellow) plume conduits, corresponding to the stars in the vertical slices of Fig. 6.10.

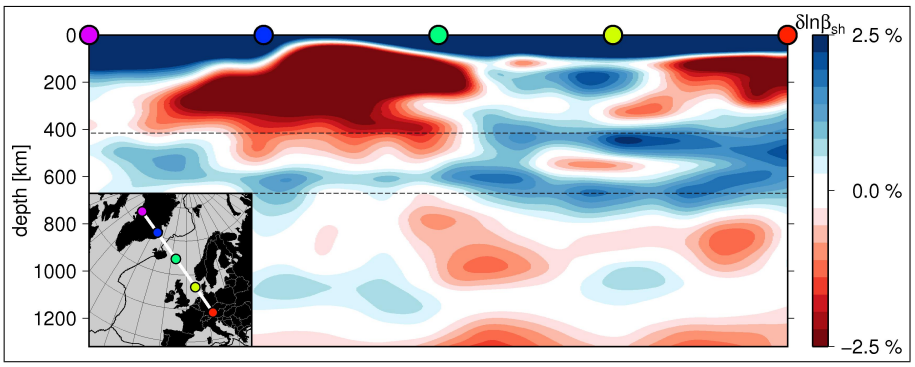
Our model shows separate low-velocity conduits connecting the upper mantle hotspots of Iceland and Jan Mayen into the lower mantle. This is best seen in the vertical slices in Fig. 6.10(a) and 6.10(b), which cut approximately perpendicular to the Mid-Atlantic Ridge through the Iceland and Jan Mayen hotspots. Below both hotspots, low velocities extend through the transition zone into the lower mantle. Between the two hotspots (Fig.



(a) Iceland plume conduit



(b) Jan Mayen plume conduit



(c) In-between the Iceland and Jan Mayen plumes conduits

Figure 6.10: Vertical slices through the β_{sh}-component of model NA-IP. The slices are roughly perpendicular to the Mid-Atlantic Ridge, the exact location is shown in the overview map in each panel. Perturbations are with respect to the 1-D reference model. The stars at different depths correspond to the stars in the horizontal slices in Fig.6.9. (a) Slice through the Iceland plume conduit. (b) Slice through the Jan Mayen plume conduit. (c) Slice in-between the two plume conduits.

6.10(c)), high velocities are imaged in the transition zone, indicating a separation of the conduits. The strength of the velocity perturbations in the conduits below a depth of 400 km varies between -0.5% and -2% with respect to the 1-D reference model. The conduits have a diameter of 300 to 400 km in the lower mantle and 100 to 200 km in the upper mantle. At depths above $\sim$300 km, the conduits expand laterally to form the broad low-velocity region below the North Atlantic lithosphere.

The mantle surrounding the low-velocity conduits shows strong positive velocity perturbations. These are more pronounced at those depths where the conduits are weaker, particularly around the 660 and 410 km discontinuities. The 1-D reference model appears to systematically underestimate the velocity at these depths, possibly because it is not accurate when used as reference for full-waveform models. Independent from the reference model, a strong velocity contrast between the low-velocity conduits and the surrounding mantle is imaged at all depths.

The conduits beneath both the Iceland and Jan Mayen hotspots are tilted to the South-East, possibly reflecting the westward motion of the Mid-Atlantic Ridge with respect to a fixed hotspot reference frame (Morgan, 1983). The location where the Iceland conduit is imaged to cross the mantle transition zone agrees with results of Shen et al. (2002). They constrain the variation in transition zone thickness below Iceland using receiver functions and find the centre of a thinned region, indicative of a hot transition zone, to be located just south of Iceland. Their location corresponds precisely to the location of the Iceland conduit at a depth of 600 km in our model (Fig. 6.9). The results of Pritchard et al. (2000), who analyse patterns in waveform data recorded in Scotland and Norway, are also consistent with a South-East tilted low-velocity conduit.

The vertical cross-section in Fig. 6.10(b) suggests that the Jan Mayen conduit splits into two branches. Besides the main conduit leading in the direction of Jan Mayen, a second branch ends in the thickened low-velocity zone beneath the lithosphere of the southern Scandinavian Mountains in Norway. The vertical cross-section suggests the branching to occur at a depth of about 500 km. However, an eastern branch closer to Norway is already visible at depths of 700 km and 800 km in Fig. 6.9. A branching plume could provide an explanation for the thickened low-velocity layer below the lithosphere of South-West Norway and the comparatively high elevation of the southern Scandinavian Mountains.

We constrained the temperature anomalies within the low-velocity con-

duits by converting the S-velocity model into a temperature model, based on mineral physics constraints (Stixrude and Lithgow-Bertelloni, 2005, 2011). With respect to the 1-D reference model, the excess temperature in both conduits exceeds 200 K throughout the lower mante. In parts of the transition zone, in particular within the Iceland conduit, it decreases to about 50 K. With respect to the (faster and colder than average) mantle surrounding the conduits in the transition zone, however, the excess temperature still exceeds 120 K throughout the transition zone. In the upper mantle above the transition zone, the temperature anomaly again increases to values exceeding 200 K. Within the uppermost 300 to 400 km of the mantle, the excess temperature increases steeply and reaches 700 K below the hotspots and large parts of the oceanic lithosphere. Such high excess temperatures suggest that partial melting strongly contributes to the extremely low S-velocities in the uppermost mantle.

6.6 Conclusions

We have derived a high-resolution S-velocity model of the North Atlantic region, covering the upper and part of the lower mantle to a depth of 1300 km. The model is derived with non-linear full-waveform methods, the misfit between synthetic and observed waveforms is quantified with the instantaneous phase difference. S-waves and surface-waves together with their reflected and diffracted waves constrain the model, the dataset consists of more than 4000 seismograms. Resolution tests using Point Spread Functions, an improved waveform fit and the correlation of the model with topographic, tectonic and gravimetric features confirm the geologic plausibility and high resolution of the model. The most important features of our model are:

(1) A well-resolved Mid-Atlantic Ridge and the presence of two separate hotspots beneath Iceland and beneath the northern part of the Kolbeinsey Ridge close to Jan Mayen. Separate hotspots are supported by geochemical (Schilling et al., 1999) and bathymetric evidence, but could not be identified in previous tomographic models. Velocities beneath both hotspots are reduced by up to -11%. A layer of low-velocity asthenosphere originating from the hotspots is imaged beneath much of the oceanic and part of the continental lithosphere of the North Atlantic region.

(2) A connection of the Iceland and Jan Mayen hotspots with post-rift uplift of South-West Scandinavia and parts of the British Isles is strongly

supported by our model. Gravity-based estimates of present-day dynamic support in the North Atlantic region (Jones et al., 2002) agree very well with the imaged sub-lithospheric low-velocity pattern on the European side of the ridge. A correlation between a sub-lithospheric low-velocity layer and Neogene uplift is also found for eastern Greenland.

Our model is consistent with, and places into a larger context, previous smaller-scale studies of the Scandinavian region (Weidle and Maupin, 2008) and the British Isles (Arrowsmith et al., 2005; Davis et al., 2012). These studies suggest the presence of hotspot-related low velocities beneath uplifted regions.

(3) A lower-mantle source for both the Iceland and the Jan Mayen hotspots is suggested by our model. We image separate South-East-tilted low-velocity conduits beneath both hotspots, connecting through the transition zone into the lower mantle. Their diameter is 300 km to 400 km in the lower mantle and 100 km to 200 km in the upper mantle. Velocity perturbations within the conduits reach -0.5% to -2.0% with respect to the 1-D reference model, corresponding to excess temperatures of approximately 50 K to 200 K. The 1-D reference model appears to underestimate velocities in the vicinity of the transition zone discontinuities, therefore the absolute velocity and temperature contrast between conduits and surrounding mantle is likely to be greater. The conduits are separated in the transition zone and the upper mantle, but a possible full separation in the lower mantle cannot be clearly inferred from our model.

The location of the imaged Iceland conduit is consistent with the receiver function study of Shen et al. (2002), who observe a locally thinned transition zone south of Iceland. The Jan Mayen plume appears to branch at a depth of about 500 km, or possibly deeper. The second branch is directed towards south-western Norway, where a thickened low-velocity layer beneath the continental lithosphere is imaged.

(4) The model resolves the Greenland and Baltic Shields in great detail. We image a East-West oriented corridor of thinned lithosphere beneath Greenland, approximately correlating with the reconstructed track of the Iceland hotspot between 70 Ma and 35 Ma (Lawver and Müller, 1994). The hotspot may have weakened the lithosphere while it was located beneath Greenland. A sharp boundary between the thick, Precambrian lithosphere of the Baltic Shield and the thinner continental lithosphere of South-West Norway and the Danish Basin is imaged in our model. The sub-lithospheric low-velocity layer cannot cross this boundary and is confined to the shal-

lower lithosphere of South-West Scandinavia.

Acknowledgements

This work was financed by the Netherlands Research Centre for Integrated Solid Earth Sciences under grants ISES-NorMar-2.6 and ISES-UU-PC-cluster. Use of the Huygens IBM p6 supercomputer at SARA Amsterdam was sponsored by the National Computing Facilities Foundation (N.C.F.) under the project SH-161-09 with financial support from the Netherlands Organisation for Scientific Research (N.W.O.). We are grateful to Theo van Zessen for maintaining the HPC clusters STIG and GRIT at the Department of Earth Sciences at Utrecht University. Nicky White and Stephen Jones kindly provided the long-wavelength gravity field data and gave inspiring comments. We thank Hans-Peter Bunge for helpful initial suggestions on the interpretation of the model, and Laura Cobden for providing the mineral physics constraints used for the temperature conversion. The constructive comments of Christophe Zaroli and Hejun Zhu greatly improved the manuscript. All figures were prepared with the Generic Mapping Tools (Wessel and Smith, 1998).

Chapter 7

Concluding remarks

Starting in the 1970's, increasingly detailed models of the Earth's interior have been constructed using seismic tomography. Their resolution has advanced to a point where approximations to the wave equation and the neglectance of much of the available waveform information limit further refinement. Non-linear full-waveform techniques currently allow us to compute the full seismic wavefield in realistic Earth models, making the inclusion of finite-frequency wave propagation effects in the construction of tomographic models possible. The numerical and theoretical techniques to account for the full complexity of wave propagation in large-scale Earth models are just established and further evolving. Although computationally costly, the construction of continental-scale full-waveform models at a reasonable resolution is now feasible on modern CPU clusters. First regional- and continental-scale full-waveform models have been published (Tape et al., 2007; Fichtner et al., 2009; Tape et al., 2010; Fichtner and Trampert, 2011b; Zhu et al., 2012). The publication of the first global model is a matter of time (Bozdağ et al., 2012).

The construction of full-waveform models involves a number of subjective choices, such as the misfit used to measure waveform differences and the regularisation of the gradient. Due to the non-linearity of full-waveform inversions, the effect of these choices on the resulting model is difficult to assess.

In this thesis, we investigated how full-waveform inversion can be efficiently used to increase the resolution of tomographic models. The synthetic and real-data experiments focused on the imaging of possible mantle plumes (Morgan, 1971), which are a challenging target for seismic tomog-

raphy due to their likely small lateral size and their extension deep into the lower mantle. Whether they really exist as a feature of mantle convection is a matter of intense debate in seismic tomography and in Earth sciences in general. The findings from this thesis are relevant not only to the imaging of mantle plumes, but to the imaging of small-scale structure in general. Our results demonstrate that the tomographic imaging of small-scale structure, even deep in the mantle, can be greatly improved by the use of full-waveform methods.

In **Chapter 3**, we discussed the principles and characteristics of finite-frequency wave propagation. We investigated the effect of an idealised mantle plume on the seismic wavefield and discussed and interpreted its characteristics. The main result was that the diffracted wavefield carries the major part of the information on small-scale anomalies. The diffracted wavefield is of relatively low amplitude and distributed over a wide azimuthal range.

In **Chapter 4**, we tested the capability of different misfit functions to recover an idealised mantle plume in synthetic experiments. These experiments were limited to the P-wave and its coda. We confirmed that the simple cross-correlation traveltime misfit does not constrain the lower mantle part of narrow plumes, even if inverted iteratively using adjoint methods. The problem is that the traveltime delays of transmitted waves are concealed by diffractive wavefront healing. Although the waveforms are still different at large distances behind the anomaly, the remaining information cannot be extracted with simple cross-correlation measurements. We identified the time-dependent and amplitude-independent instantaneous phase misfit as well-suited for extracting remaining waveform discrepancies. These are caused mainly by diffracted waves which distort the original waveform and, at larger receiver azimuths, arrive in the coda. The inversion of this information with full-waveform methods constrains the plume well, in amplitude and shape, even deep in the mantle. The high contribution of the diffracted waves to the total misfit, as a result of the amplitude-independence of the instantaneous phase, are important to constrain the shape of the plume. This is because an incorrectly placed or -sized plume leads to incorrect timing of the diffracted waves particularly at larger azimuths, strongly increasing the total misfit. We further found that the recovery of the plume is still possible in the presence of noise. The incoherent noise from the different measurements cancels out to a large degree in the misfit kernel. As a result, the kernel is dominated by the

coherent information due to real structure, which adds constructively.

In **Chapter 5**, we extended the experiment from Chapter 4 to a longer measurement window including S-waves, and by using an idealised plume perturbed in both S- and P-velocity. Radial anisotropy was included in the inversion by treating the horizontally and vertically polarised S-velocities β_{sh} and β_{sv} as independent parameters. We found that differently polarised waves have different sensitivities to the plume, each recovering the plume only in part. Vertically and horizontally polarised waves complement each other, and their isotropic average delineates the plume well. The different sensitivity may be due to the dependence of diffraction at a cylindrical anomaly on the direction of polarisation of the waves, which has been observed in an earlier study by Tilmann et al. (1998). This effect may be particularly present in models constructed with non-linear full-waveform techniques, where small-scale structure is constrained by diffracted waves. Care should therefore be taken when interpreting differences in β_{sh} and β_{sv} in terms of radial anisotropy. For plume-shaped, narrow, vertical structures, the synthetic test indicates a good recovery in β_{sh} in the upper $\sim$ 1300 km of the mantle. Between $\sim$ 200 and 500 km depth, the recovery in β_{sv} is weaker.

In **Chapter 6**, we presented a model of the North Atlantic and the surrounding continental margins down to a depth of 1200 km constructed with the instantaneous phase misfit. The β_{sh}-model, which we identified to be better-resolved compared to the β_{sv}-model, reveals features of the region in unprecedented detail. Highlights are separate hotspots beneath Iceland and Jan Mayen, connected by low-velocity conduits to the lower mantle. A sub-lithospheric low-velocity layer extends from the hotspots beneath much of the oceanic lithosphere, and locally beneath the continental lithosphere of the British Isles, South-West Scandinavia and eastern Greenland. These regions were uplifted in Neogene times, much later than the North Atlantic ocean opened. The imaged low-velocity layer suggests dynamic support of these regions by low-density asthenosphere originating at the hotspots. Such an interpretation is consistent with estimates of present-day dynamic support based on gravity data (Jones et al., 2002) and with a number of smaller-scale tomographic studies (Arrowsmith et al., 2005; Weidle and Maupin, 2008; Davis et al., 2012).

The better resolution of the β_{sh}-model compared to the β_{sv}-model may result from the higher sensitivity of β_{sh} to narrow vertical structures at shallow to intermediate mantle depths, as indicated by the synthetic tests

in Chapter 5.

Bibliography

Aki, K., Christoffersson, A., Husebye, E., 1977. Determination of the three-dimensional seismic structure of the lithosphere. J. Geophys. Res. 82, 277–296.

Aki, K., Kaminuma, K., 1963. Phase velocity in Japan. Part 1. Love waves from the Aleutian shock of March 9, 1957. Bull. Earthq. Res. Inst. 41, 243–259.

Alford, R., Kelly, K., 1974. Accuracy of finite-difference modeling of the acoustic wave equation. Geophysics 39, 834–842.

Allen, R., Nolet, G., Morgan, W., Vogfjord, K., Bergsson, B., Erlendsson, P., Foulger, G., Jakobsdottir, S., Julian, B., Pritchard, M., Ragnarsson, S., Stefansson, R., 2002. Imaging the mantle beneath Iceland using integrated seismological techniques. J. Geophys. Res. 107, 2325.

Alterman, Z., Karal, F., 1968. Propagation of elastic waves in layered media by finite difference methods. Bull. Seismol. Soc. Am. 58, 367–398.

Amaru, M., 2007. Global travel time tomography with 3-D reference models. Ph.D. thesis. Utrecht University.

Anderson, D., 1961. Elastic wave propagation in layered anisotropic media. J. Geophys. Res. 66, 2953–2963.

Arrowsmith, S., Kendall, M., White, N., van Decar, J., Booth, D., 2005. Seismic imaging of a hot upwelling beneath the British Isles. Geology 33, 345–348.

Babuska, V., Cara, M., 1991. Seismic anisotropy in the earth. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Backus, G., 1962. Long-wave elastic anisotropy produced by horizontal layering. J. Geophys. Res. 67, 4427–4440.

Baig, A., Dahlen, F., Hung, S., 2003. Traveltimes of waves in three-dimensional random media. Geophys. J. Int. 153, 467–482.

Barnes, A.E., 2007. A tutorial on complex seismic trace analysis. Geophysics 72, 33–43.

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

Bijwaard, H., Spakman, W., Engdahl, E., 1998. Closing the gap between regional and global travel time tomography. J. Geophys. Res. 103, 30055–30.

Birch, F., 1960. The velocity of compressional waves in rocks to 10 kilobars, Part 1. J. Geophys. Res. 65, 1083–1102.

Bjarnason, I., Silver, P., Rümpker, G., Solomon, S., 2002. Shear wave splitting across the Iceland hot spot: Results from the ICEMELT experiment. J. Geophys. Res. 107, 2382.

Boore, D.M., 1972. Finite–difference methods for seismic wave propagation in heterogeneous materials, in: Alder, B., Fernbach, S., Rotenberg, M. (Eds.), Methods of Computational Physics. Academic Press, New York.

Boschi, L., Becker, T.W., Soldati, G., Dziewonski, A.M., 2006. On the relevance of Born theory in global seismic tomography. Geophys. Res. Lett. 33, L06302–.

Bott, M., Bott, J., 2004. The Cenozoic uplift and earthquake belt of mainland Britain as a response to an underlying hot, low-density upper mantle. J. Geol. Soc. 161, 19–29.

Bozdağ, E., Trampert, J., Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophys. J. Int. 185, 845–870.

Bozdağ, E., Zhu, H., Peter, D., Tromp, J., 2012. Towards global adjoint tomography, in: Abbasi, A., Giesen, N. (Eds.), EGU General Assembly Conference Abstracts, p. 5603.

Bunks, C., Saleck, F., Zaleski, S., Chavent, G., 1995. Multiscale seismic waveform inversion. Geophysics 60, 1457–1473.

Carminati, E., Cuffaro, M., Doglioni, C., 2009. Cenozoic uplift of Europe. Tectonics 28, TC4016.

Chaljub, E., Komatitsch, D., Vilotte, J., Capdeville, Y., Valette, B., Festa, G., et al., 2007. Spectral-element analysis in seismology. Adv. Geophys. 48, 365–419.

Christiansen, R., Foulger, G., Evans, J., 2002. Upper-mantle origin of the Yellowstone hotspot. Geol. Soc. Am. Bull. 114, 1245–1256.

Dahlen, F.A., 2004. Resolution limit of traveltime tomography. Geophys. J. Int. 157, 315–331.

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

Davis, M.W., White, N.J., Priestley, K.F., Baptie, B.J., Tilmann, F.J., 2012. Crustal structure of the British Isles and its epeirogenic consequences. Geophys. J. Int. 190, 705–725.

Delorey, A., Dunn, R., Gaherty, J., 2007. Surface wave tomography of the upper mantle beneath the Reykjanes Ridge with implications for ridge–hot spot interaction. J. Geophys. Res. 112, B08313.

Deuss, A., Irving, J., Woodhouse, J., 2010. Regional variation of inner core anisotropy from seismic normal mode observations. Science 328, 1018–1020.

Doré, A.G., Cartwright, J.A., Stoker, M.S., Turner, J.P., White, N.J., 2002. Exhumation of the North Atlantic margin: introduction and background. Geol. Soc. London Spec. Publ. 196, 1–12.

Du, Z., Vinnik, L., Foulger, G., 2006. Evidence from P-to-S mantle converted waves for a flat '660-km' discontinuity beneath Iceland. Earth Planet. Sci. Lett. 241, 271–280.

Duncan, W., Green, P., Duddy, I., 1998. Source rock burial history and seal effectiveness: key facets to understanding hydrocarbon exploration potential in the East and Central Irish Sea Basins. AAPG Bull. 82, 1401–1415.

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

Dziewonski, A., Hager, B., O'Connell, R., 1977. Large-scale heterogeneities in the lower mantle. J. Geophys. Res. 82, 239–255.

Dziewonski, A.M., Anderson, D.L., 1981. Preliminary reference earth model. Phys. Earth Planet. In. 25, 297–356.

Ebbing, J., England, R., Korja, T., Lauritsen, T., Olesen, O., Stratford, W., Weidle, C., 2012. Structure of the Scandes lithosphere from surface to depth. Tectonophys. 536, 1–24.

Ebbing, J., Olesen, O., 2005. The Northern and Southern Scandes–structural differences revealed by an analysis of gravity anomalies, the geoid and regional isostasy. Tectonophys. 411, 73–87.

Ekström, G., Dziewonski, A.M., 1998. The unique anisotropy of the Pacific upper mantle. Nature 394, 168–172.

Ekström, G., Tromp, J., Larson, E., 1997. Measurements and global models of surface wave propagation. J. Geophys. Res. 102, 8137–8157.

Fichtner, A., Bunge, H., Igel, H., 2006. The adjoint method in seismology: I. Theory. Phys. Earth Planet. In. 157, 86–104.

Fichtner, A., Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models - a technique based on non-linear dispersion curve matching. Geophys. J. Int. 173, 519–533.

Fichtner, A., Kennett, B., Igel, H., Bunge, H., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle. Earth Planet. Sci. Lett. 290, 270–280.

Fichtner, A., Kennett, B.L.N., Igel, H., Bunge, H.P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. Geophys. J. Int. 179, 1703–1725.

Fichtner, A., Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques. Geophys. J. Int. 185, 775–798.

Fichtner, A., Trampert, J., 2011b. Resolution analysis in full waveform inversion. Geophys. J. Int. 187, 1604–1624.

Fletcher, R., Reeves, C., 1964. Function minimization by conjugate gradients. Comput. J. 7, 149–154.

Foulger, G., Anderson, D., 2005. A cool model for the Iceland hotspot. J. Volcanol. Geotherm. Res. 141, 1–22.

Foulger, G., Natland, J., Anderson, D., 2005. A source for Icelandic magmas in remelted Iapetus crust. J. Volcanol. Geotherm. Res. 141, 23–44.

Foulger, G., Pritchard, M., Julian, B., Evans, J., Allen, R., Nolet, G., Morgan, W., Bergsson, B., Erlendsson, P., Jakobsdottir, S., et al., 2001. Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle. Geophys. J. Int. 146, 504–530.

Foulger, G.R., 2012. Are 'hot spots' hot spots? J. Geodyn. 58, 1–28.

Fukao, Y., Obayashi, M., Inoue, H., Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone. J. Geophys. Res. 97, 4809–4822.

Fukao, Y., Widiyantoro, S., Obayashi, M., 2001. Stagnant slabs in the upper and lower mantle transition region. Rev. Geophys. 39, 291–324.

George, T., 1966. Geomorphic evolution in Hebridean Scotland. Scott. J. Geol. 2, 1–34.

Giardini, D., Li, X., Woodhouse, J., 1987. Three-dimensional structure of the Earth from splitting in free-oscillation spectra. Nature 325, 405–411.

Grand, S., 1994. Mantle shear structure beneath the Americas and surrounding oceans. J. Geophys. Res. 99, 11591–11622.

Grand, S., 2002. Mantle shear–wave tomography and the fate of subducted slabs. Philos. Trans. R. Soc. Lond. A 360, 2475–2491.

Green, P., 1989. Thermal and tectonic history of the East Midlands shelf (onshore UK) and surrounding regions assessed by apatite fission track analysis. J. Geol. Soc. 146, 755–773.

Gung, Y., Panning, M., Romanowicz, B., et al., 2003. Global anisotropy and the thickness of continents. Nature 422, 707–711.

Hall, A., Bishop, P., 2002. Scotlands denudational history: an integrated view of erosion and sedimentation at an uplifted passive margin. Geol. Soc. London Spec. Publ. 196, 271–290.

He, X., Tromp, J., 1996. Normal-mode constraints on the structure of the Earth. J. Geophys. Res. 1012, 20053–20082.

van der Hilst, R., Widiyantoro, S., Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography. Nature 386, 578–584.

van der Hilst, R.D., de Hoop, M.V., 2005. Banana-doughnut kernels and mantle tomography. Geophys. J. Int. 163, 956–961.

van der Hilst, R.D., de Hoop, M.V., 2006. Reply to comment by R. Montelli, G. Nolet and F. A. Dahlen on 'Banana-doughnut kernels and mantle tomography'. Geophys. J. Int. 167, 1211–1214.

Holford, S., Green, P., Turner, J., Williams, G., Hillis, R., Tappin, D., Duddy, I., 2008. Evidence for kilometre-scale Neogene exhumation driven by compressional deformation in the Irish Sea basin system. Geol. Soc. London Spec. Publ. 306, 91–119.

de Hoop, M.V., van der Hilst, R.D., 2005. On sensitivity kernels for 'wave-equation' transmission tomography. Geophys. J. Int. 160, 621–633.

Hudson, J., 1977. Scattered waves in the coda of P. J. Geophys. 43, 359–374.

Hung, S., Dahlen, F., Nolet, G., 2001. Wavefront healing: a banana–doughnut perspective. Geophys. J. Int. 146, 289–312.

Hwang, Y.K., Ritsema, J., van Keken, P.E., Goes, S., Styles, E., 2011. Wavefront healing renders deep plumes seismically invisible. Geophys. J. Int. 187, 273–277.

Japsen, P., 1997. Regional Neogene exhumation of Britain and the western North Sea. J. Geol. Soc. 154, 239–247.

Japsen, P., Bidstrup, T., Lidmar-Bergström, K., 2002. Neogene uplift and erosion of southern Scandinavia induced by the rise of the South Swedish Dome. Geol. Soc. London Spec. Publ. 196, 183–207.

Japsen, P., Chalmers, J., 2000. Neogene uplift and tectonics around the North Atlantic: overview. Global Planet. Change 24, 165–173.

Japsen, P., Green, P., Nielsen, L., Rasmussen, E., Bidstrup, T., 2007. Mesozoic–Cenozoic exhumation events in the eastern North Sea Basin: a multi-disciplinary study based on palaeothermal, palaeoburial, stratigraphic and seismic data. Basin Res. 19, 451–490.

Johnson, C., Gallagher, K., 2000. A preliminary Mesozoic and Cenozoic denudation history of the North East Greenland onshore margin. Global Planet. Change 24, 261–274.

Jones, S., White, N., Clarke, B., Rowley, E., Gallagher, K., 2002. Present and past influence of the Iceland Plume on sedimentation. Geol. Soc. London Spec. Publ. 196, 13–25.

Kárason, H., van der Hilst, R., 2000. Constraints on mantle convection from seismic tomography. Geophys. Monogr. Ser. 121, 277–288.

Kárason, H., van der Hilst, R., 2001. Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, P diff). J. Geophys. Res. 106, 6569–6587.

Käser, M., Dumbser, M., 2006. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes–I. The two-dimensional isotropic case with external source terms. Geophys. J. Int. 166, 855–877.

Keller, J., 1962. Geometrical Theory of Diffraction. J. Opt. Soc. Am. 52, 116–130.

Keller, W., Anderson, D., Clayton, R., 2000. Resolution of tomographic models of the mantle beneath Iceland. Geophys. Res. Lett. 27, 3993–3996.

Komatitsch, D., Tromp, J., 2002. Introduction to the spectral element method for three-dimensional seismic wave propagation. Geophys. J. Int. 139, 806–822.

Komatitsch, D., Tsuboi, S., Tromp, J., 2005. The spectral-element method in seismology. Geophys. Monogr. Ser. 157, 205–227.

Lawver, L., Müller, R., 1994. Iceland hotspot track. Geology 22, 311–314.

Lebedev, S., Van Der Hilst, R., 2008. Global upper-mantle tomography with the automated multimode inversion of surface and S-wave forms. Geophys. J. Int. 173, 505–518.

Legendre, C., Meier, T., Lebedev, S., Friederich, W., Viereck-Götte, L., 2012. A shear wave velocity model of the European upper mantle from automated inversion of seismic shear and surface waveforms. Geophys. J. Int. 191, 282–304.

Lehmann, I., 1936. P'. Bur. Cent. Seismol. Int. Trav. Sci. A 14, 3–31.

Lekić, V., Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method. Geophys. J. Int. 185, 799–831.

Li, C., van der Hilst, R.D., Engdahl, E.R., Burdick, S., 2008. A new global model for P wave speed variations in Earth's mantle. Geochem. Geophys. Geosyst. 9, Q05018–.

Li, X., Giardini, D., Woodhouses, J., 1991. Large-scale three-dimensional even-degree structure of the Earth from splitting of long-period normal modes. J. Geophys. Res. 96, 551–577.

Li, X., Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. J. Geophys. Res. 101, 22245–22.

Luo, Y., Schuster, G.T., 1991. Wave-equation traveltime inversion. Geophysics 56, 645–653.

Maggi, A., Tape, C., Chen, M., Chao, D., Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography. Geophys. J. Int. 178, 257–281.

Mainprice, D., 1997. Modelling the anisotropic seismic properties of partially molten rocks found at mid-ocean ridges. Tectonophys. 279, 161–179.

Malcolm, A.E., Trampert, J., 2011. Tomographic errors from wave front healing: more than just a fast bias. Geophys. J. Int. 185, 385–402.

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

Masters, G., Johnson, S., Laske, G., Bolton, H., Davies, J., 1996. A shear-velocity model of the mantle [and discussion]. Philos. Trans. R. Soc. Lond. A 354, 1385–1411.

Masters, G., Jordan, T., Silver, P., Gilbert, F., 1982. Aspherical Earth structure from fundamental spheroidal-mode data. Nature 298, 609–613.

Mathiesen, A., Bidstrup, T., Christiansen, F., 2000. Denudation and uplift history of the Jameson Land basin, East Greenland–constrained from maturity and apatite fission track data. Global Planet. Change 24, 275–301.

McEvilly, T., 1964. Central US crustupper mantle structure from Love and Rayleigh wave phase velocity inversion. Bull. Seismol. Soc. Am. 54, 1997–2015.

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

Montelli, R., Nolet, G., Dahlen, F.A., 2006a. Comment on 'Banana-doughnut kernels and mantle tomography' by van der Hilst and de Hoop. Geophys. J. Int. 167, 1204–1210.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., 2006b. A catalogue of deep mantle plumes: New results from finite-frequency tomography. Geochem. Geophys. Geosyst. 7, Q11007–.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., Engdahl, E.R., Hung, S.H., 2004. Finite-frequency tomography reveals a variety of plumes in the mantle. Science 303, 338–343.

Morgan, J., 1983. Hotspot tracks and the early rifting of the Atlantic. Tectonophys. 94, 123–139.

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

Nataf, H., Nakanishi, I., Anderson, D., 1984. Anisotropy and shear-velocity heterogeneities in the upper mantle. Geophys. Res. Lett. 11, 109–112.

Nolet, G., Dahlen, F., 2000. Wave front healing and the evolution of seismic delay times. J. Geophys. Res. 105, 19043–19054.

Oldham, R., 1906. The Constitution of the interior of the Earth, as revealed by earthquakes. J. Geol. Soc. 62, 456–475.

Panning, M., Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle. Geophys. J. Int. 167, 361–379.

Pascal, C., Olesen, O., 2009. Are the Norwegian mountains compensated by a mantle thermal anomaly at depth? Tectonophys. 475, 160–168.

Pilidou, S., Priestley, K., Debayle, E., Gudmundsson, O., 2005. Rayleigh wave tomography in the North Atlantic: high resolution images of the Iceland, Azores and Eifel mantle plumes. Lithos 79, 453–474.

Pratt, G., Shin, C., et al., 2002. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. Geophys. J. Int. 133, 341–362.

Pritchard, M.J., Foulger, G.R., Julian, B.R., Fyen, J., 2000. Constraints on a plume in the mid-mantle beneath the Iceland region from seismic array data. Geophys. J. Int. 143, 119–128.

de la Puente, J., Dumbser, M., Käser, M., Igel, H., 2008. Discontinuous Galerkin methods for wave propagation in poroelastic media. Geophysics 73, T77–T97.

Redfield, T., Osmundsen, P., Hendriks, B., 2005. The role of fault reactivation and growth in the uplift of western Fennoscandia. J. Geol. Soc. 162, 1013–1030.

Resovsky, J., Ritzwoller, M., 1998. New and refined constraints on three-dimensional Earth structure from normal modes below 3 mHz. J. Geophys. Res. 103, 783–810.

Rickers, F., Fichtner, A., Trampert, J., 2012. Imaging mantle plumes with instantaneous phase measurements of diffracted waves. Geophys. J. Int. 190, 650–664.

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

Ritsema, J., Deuss, A., van Heijst, H., Woodhouse, J., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. Geophys. J. Int. 184, 1223–1236.

Ritsema, J., van Heijst, H., Woodhouse, J., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland. Science 286, 1925–1928.

Ritsema, J., van Heijst, H., Woodhouse, J., 2004. Global transition zone tomography. J. Geophys. Res. 109, B02302.

Ritzwoller, M., Masters, G., Gilbert, F., 1988. Constraining aspherical structure with low-degree interaction coefficients: application to uncoupled multiplets. J. Geophys. Res. 93, 6–369.

Rohrman, M., van der Beek, P., 1996. Cenozoic postrift domal uplift of North Atlantic margins: An asthenospheric diapirism model. Geology 24, 901–904.

Rohrman, M., van der Beek, P., Andriessen, P., Cloetingh, S., 1995. Meso-Cenozoic morphotectonic evolution of southern Norway: Neogene domal uplift inferred from apatite fission track thermochronology. Tectonics 14, 704–718.

Romanowicz, B., Gung, Y., 2002. Superplumes from the core-mantle boundary to the lithosphere: Implications for heat flux. Science 296, 513–516.

Schilling, J., 1985. Upper mantle heterogeneities and dynamics. Nature 314, 62–67.

Schilling, J., Kingsley, R., Fontignie, D., Poreda, R., Xue, S., 1999. Dispersion of the Jan Mayen and Iceland mantle plumes in the Arctic: A He-Pb-Nd-Sr isotope tracer study of basalts from the Kolbeinsey, Mohns, and Knipovich Ridges. J. Geophys. Res. 104, 10543–10.

Shapiro, N., Ritzwoller, M., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle. Geophys. J. Int. 151, 88–105.

Shen, Y., Solomon, S., Bjarnason, I., Nolet, G., Morgan, W., Allen, R., Vogfjörd, K., Jakobsdóttir, S., Stefánsson, R., Julian, B., et al., 2002. Seismic evidence for a tilted mantle plume and north–south mantle flow beneath Iceland. Earth Planet. Sci. Lett. 197, 261–272.

Shen, Y., Solomon, S.C., Bjarnason, I.T., Wolfe, C.J., 1998. Seismic evidence for a lower-mantle origin of the Iceland plume. Nature 395, 62–65.

Sigloch, K., Mcquarrie, N., Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography. Nat. Geosci. 1, 458–462.

Sigloch, K., Nolet, G., 2006. Measuring finite-frequency body wave amplitudes and traveltimes. Geophys. J. Int. 167, 271–287.

Simmons, N., Myers, S., Johannesson, G., Matzel, E., 2012. LLNL-G3Dv3: Global P wave tomography model for improved regional and teleseismic travel time prediction. J. Geophys. Res. 117, B10302.

Sirgue, L., Pratt, R., 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. Geophysics 69, 231–248.

Snieder, R., 1993. Global inversions using normal modes and long-period surface waves, in: Hirahara, H.I.K. (Ed.), Seismic tomography. Prentice-Hall, London, pp. 23–63.

Stixrude, L., Lithgow-Bertelloni, C., 2005. Thermodynamics of mantle minerals–I. Physical properties. Geophys. J. Int. 162, 610–632.

Stixrude, L., Lithgow-Bertelloni, C., 2011. Thermodynamics of mantle minerals–II. Phase equilibria. Geophys. J. Int. 184, 1180–1213.

Su, W., Woodward, R., Dziewonski, A., 1994. Degree 12 model of shear velocity heterogeneity in the mantle. J. Geophys. Res. 99, 6945–6980.

Taner, M.T., Koehler, F., Sheriff, R.E., 1979. Complex seismic trace analysis. Geophysics 44, 1041–1063.

Tape, C., Liu, Q., Maggi, A., Tromp, J., 2009. Adjoint tomography of the southern California crust. Science 325, 988–992.

Tape, C., Liu, Q., Maggi, A., Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods. Geophys. J. Int. 180, 433–462.

Tape, C., Liu, Q., Tromp, J., 2007. Finite-frequency tomography using adjoint methods - Methodology and examples using membrane surface waves. Geophys. J. Int. 168, 1105–1129.

Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F., Poole, S., 2007. The GGM03 mean earth gravity model from GRACE, in: AGU Fall Meeting Abstracts, p. 03.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics 49, 1259–1266.

Tilmann, F., McKenzie, D., Priestly, K., 1998. P and S wave scattering from mantle plumes. J. Geophys. Res. 103, 21.

To, A., Romanowicz, B., 2009. Finite frequency effects on global S diffracted traveltimes. Geophys. J. Int. 179, 1645–1657.

Trampert, J., van Heijst, H., 2002. Global azimuthal anisotropy in the transition zone. Science 296, 1297–1299.

Trampert, J., Spetzler, J., 2006. Surface wave tomography: finite-frequency effects lost in the null space. Geophys. J. Int. 164, 394–400.

Trampert, J., Woodhouse, J., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds. Geophys. J. Int. 122, 675–690.

Tromp, J., Tape, C., Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophys. J. Int. 160, 195–216.

Tryggvason, K., Husebye, E., Stefánsson, R., 1983. Seismic image of the hypothesized Icelandic hot spot. Tectonophys. 100, 97–118.

Vågnes, E., Amundsen, H., 1993. Late Cenozoic uplift and volcanism on Spitsbergen: Caused by mantle convection? Geology 21, 251–254.

Virieux, J., 1984. SH-wave propagation in heterogeneous media; velocity-stress finite-difference method. Geophysics 49, 1933–1942.

Visser, K., Trampert, J., Lebedev, S., Kennett, B., 2008. Probability of radial anisotropy in the deep mantle. Earth Planet. Sci. Lett. 270, 241–250.

Weidle, C., Maupin, V., 2008. An upper-mantle S-wave velocity model for Northern Europe from Love and Rayleigh group velocities. Geophys. J. Int. 175, 1154–1168.

Wessel, P., Smith, W., 1998. New, improved version of generic mapping tools released. EOS, Trans. Am. Geophys. Un. 79, 579–579.

White, R., McKenzie, D., 1989. Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. J. Geophys. Res. 94, 7685–7729.

Wielandt, E., 1987. On the validity of the ray approximation for interpreting delay times, in: Nolet, G. (Ed.), Seismic Tomography. Reidel, pp. 85–98.

Wolfe, C., Solomon, S., Laske, G., Collins, J., Detrick, R., Orcutt, J., Bercovici, D., Hauri, E., 2009. Mantle shear-wave velocity structure beneath the Hawaiian hot spot. Science 326, 1388–1390.

Wolfe, C.J., Th. Bjarnason, I., van Decar, J.C., Solomon, S.C., 1997. Seismic structure of the Iceland mantle plume. Nature 385, 245–247.

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

Woodhouse, J., Girnius, T., 1982. Surface waves and free oscillations in a regionalized Earth model. Geophys. J. Int. 68, 653–673.

Wu, R., Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity. Geophysics 50, 582–595.

Zhang, Y., Tanimoto, T., 1993. High-resolution global upper mantle structure and plate tectonics. J. Geophys. Res. 98, 9793–9823.

Zhao, D., 2004. Global tomographic images of mantle plumes and subducting slabs: insight into deep earth dynamics. Phys. Earth Planet. In. 146, 3–34.

Zhou, H., 1996. A high-resolution P wave model for the top 1200 km of the mantle. J. Geophys. Res. 101, 27791–27.

Zhou, Y., Dahlen, F.A., Nolet, G., Laske, G., 2005. Finite-frequency effects in global surface-wave tomography. Geophys. J. Int. 163, 1087–1111.

Zhu, H., Bozdag, E., Peter, D., Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography. Nat. Geosci. 5, 493–498.

Summary

Within the last decade, increasing computational power and theoretical advances have initiated the transition from approximation-based tomography to iterative full-waveform tomography in regional and global seismology. Such full-waveform methods allow for the exploitation of complete seismograms to constrain the 3-D structure of the Earth, which potentially leads to much higher resolution compared to approximation-based methods. The method used in this thesis relies on numerical wavefield simulations with the spectral-element method, which honours the full complexity of seismic wave propagation in realistic 3-D media. Gradients are constructed with the adjoint method. Full-waveform models are affected by a number of subjective inversion choices, such as the type of misfit and the regularisation of the gradient. The complexity of seismic wave propagation together with the non-linearity of the problem make it a challenging task to determine optimum inversion strategies.

For this thesis, we first performed synthetic experiments to investigate and optimise the capability of full-waveform tomography to constrain small-scale structures in the Earth's mantle. These experiments were focused on the tomographic detection of possible mantle plumes, which are a challenging target for seismic tomography due to their small lateral extent and their extension deep into the lower mantle. Using a misfit based on the time-continuous and amplitude-independent instantaneous phase difference, we succeeded to fully recover an idealised narrow mantle plume. The accurate measurement and inclusion of diffracted waves arriving in the body-wave coda was identified as essential for the recovery in the lower mantle. We further confirmed that simple cross-correlation traveltime misfits do not permit the recovery of plumes in the lower mantle, wavefront healing largely conceals the acquired time delays.

We applied these results to the construction of a high-resolution 3-D

model of the S-wave speed beneath the North Atlantic region, extending to a depth of 1200 km. Unprecedented details are revealed, giving new insight into the complex tectonics and dynamics of the North Atlantic region. The resolution of particularly interesting features was validated with resolution tests based on Hessian kernels. Striking features of the new model include individual plume conduits beneath the Iceland and Jan Mayen hotspots, extending into the lower mantle. This observation strongly supports the existence of mantle plumes as a feature of mantle convection, which has been debated passionately over the last decades due to a lack of conclusive seismological evidence. Our model is furthermore the first to clearly resolve two separated hotspots associated with Iceland and Jan Mayen. A further highlight is the observation of a wide-spread layer of low-velocity plume material beneath the lithosphere of much of the North Atlantic ocean. Two fingers of this layer extend beneath the continental lithosphere of parts of the British Isles and of southern Norway, providing an explanation for the considerable post-rift uplift that these regions experienced in Neogene times. Gravity-based estimates of present-day dynamic support agree very well with the location of the low-velocity layer.

Samenvatting (Summary in Dutch)

Centraal in dit proefschrift staat een nieuw 3-D model met hoge resolutie van de S-snelheid onder de Noord-Atlantische Oceaan en omgeving, tot een diepte van 1300 km. Het model is verkregen met een recent ontwikkelde tomografische techniek die gebaseerd is op de inversie van volledige golfvormen, wat geleid heeft tot een sterk verbeterde resolutie vergeleken met eerdere modellen van deze regio. Het model bevat ongekende details die een verbeterd inzicht geven in de complexe tektoniek en dynamica van het Noord-Atlantische gebied.

Ons model laat duidelijk zien dat zowel de IJslandse hotspot als de Jan Mayen hotspot oppervlakte-uitingen zijn van mantelpluimen. Dit zijn smalle stromen van heet materiaal vanuit de ondermantel naar de oppervlakte. Hiermee wordt een langdurige discussie over het al dan niet bestaan van mantelpluimen opgelost, waarvoor, door de afwezigheid van seismologische evidentie, reden was. Een ander belangrijk aspect van dit nieuwe model is een wijdverspreide laag van heet pluimmateriaal onder de lithosfeer, die onder grote delen van het Noord-Atlantische Oceaan in kaart is gebracht. Dit hete materiaal verheft de bovenliggende korst, wat tot een abnormaal ondiepe bathymetrie in het gebied leidt. Uitlopers van deze hete laag bereiken zelfs de Britse eilanden en Zuid-Noorwegen, wat de verheffing van de bergen in deze gebieden verklaart.

Voorafgaand aan de presentatie van het nieuwe model bevat dit proefschrift een aantal experimenten om de mogelijkheden van de inversie van volledige golfvormen te onderzoeken en optimaliseren. Met deze experimenten hebben we aangetoond dat mantelpluimen en andere kleinschalige structuren door de inversie van volledige golfvormen afgebeeld kunnen worden.

Acknowledgements

I had a very nice time living in Utrecht and being part of the Utrecht seismology group during the past four years. I am grateful to all the people that I got to know, who contributed to this thesis in different ways, and who made me feel at home in the beautiful city of Utrecht.

First of all, I want to thank my promotor Jeannot Trampert, who gave me the opportunity to pursue this PhD project. I was very glad having him as my supervisor. His experience and knowledge were of great help, and so were his encouraging words and belief in this project. His office door was always open, and he always took the time to help with all questions and problems.

I consider myself very lucky having Andreas Fichtner as my co-promotor. I greatly benefited from his knowledge in the field of seismology and particularly full-waveform inversion. He patiently explained to me many complex details and always had a useful idea when things did not work out or make sense. I am very grateful for all his support along the way, this thesis would look very different without his guidance.

Of much help for the success of this thesis was also Ebru Bozdağ. She greatly helped me with the first steps as a computational seismologist, and the numerous discussions about full-waveform inversion, first in Utrecht and later at conferences, always gave me new perspectives and ideas.

This thesis would not have been possible without the great work of Theo van Zessen. He maintained the computer clusters and network perfectly, and every computer problem was quickly solved (and there were many).

I want to thank Hanneke Paulssen for nice discussions and always quick help, whenever I had a question or a favour to ask.

Thanks to Jacqueline Landsheer, who did a great job organising the PhD and helping with paperwork.

I am very grateful to the dissertation committee (Jeroen Ritsema, Bar-

bara Romanowicz, Christine Thomas, Rinus Wortel, Sergei Lebedev), who took the time and effort to read this thesis. Their constructive comments were very much appreciated.

Working on this thesis in Utrecht would not have been the same without my friends and colleagues at the Seismology group and within the Earth Sciences department.

An essential part of every morning was the seismology group coffee break. Thank you Henk, for all the effort and the great coffee! I very much enjoyed the many nice discussions and the relaxed atmosphere.

I had a lot of fun spending time and being at the numerous dinners, borrels, and sometimes parties with my group. For the great time I want to thank Ebru, Ilaria (great coffee, also!), Benoit, Laura, Sonja, Mei, Tedi, Denise, Suzan, Andreas, Andrew, Paul, Ralph, Moritz, Paula, Arie, Henk, Theo, Kabir and Joop. Many thanks go to Wouter, my office mate during the first years!

Thanks also to the former and present members of the other Utrecht geophysics groups, particularly to Walter, Jeroen, Bram, Ali, Steven, Robin and Cedric. Just around the corner were the nice people from the stratigraphy and paleomagnetism groups. For great parties and dinners, running in the woods, and for everything else, I would like to thank Christian, Joyce, Vittoria, Helen, Tiuri, Nathan and Stephanie.

I had very nice people to spend time with in Utrecht. I very much enjoyed all the day- or weekend-trips, evenings out, afternoons in the park or at the gracht. I can't remember feeling bored any time during these years (but then again being bored is difficult when you are finishing a PhD). Thank you Henriette, Hannes, Sonja, Leida, Christian, Alex, Anke, Harald, Katleen, Filipo, Akshay, Kevin, Shari, Laura and everybody that I may have forgotten to mention. I had (and will have in the future) a great time with you.

Last but not least, I thank my family and my friends in Münster and Germany!

Curriculum vitae

June 27, 1981	**Born in Münster, Germany**
1991-2000	**High School (Abitur)** Gymnasium Wolbeck Münster, Germany
2000-2001	**Civil Service** Westfalenfleiss GmbH Münster, Germany
2001-2008	**MSc in Physics** Department of Physics University of Münster Münster, Germany
2008-2013	**PhD in Seismology** Department of Earth Sciences Utrecht University Utrecht, the Netherlands