Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods

Andreas Fichtner,[1] Brian L. N. Kennett,[2] Heiner Igel[1] and Hans-Peter Bunge[1]

[1]*Department of Earth and Environmental Sciences, Geophysics Section, Ludwig-Maximilian University, Munich, Germany.*
E-mail: andreas.fichtner@geophysik.uni-muenchen.de
[2]*Research School of Earth Sciences, Australian National University, Canberra, Australia*

SUMMARY

We present a full seismic waveform tomography for upper-mantle structure in the Australasian region. Our method is based on spectral-element simulations of seismic wave propagation in 3-D heterogeneous earth models. The accurate solution of the forward problem ensures that waveform misfits are solely due to as yet undiscovered Earth structure and imprecise source descriptions, thus leading to more realistic tomographic images and source parameter estimates. To reduce the computational costs, we implement a long-wavelength equivalent crustal model.

We quantify differences between the observed and the synthetic waveforms using time–frequency (TF) misfits. Their principal advantages are the separation of phase and amplitude misfits, the exploitation of complete waveform information and a quasi-linear relation to 3-D Earth structure. Fréchet kernels for the TF misfits are computed via the adjoint method. We propose a simple data compression scheme and an accuracy-adaptive time integration of the wavefields that allows us to reduce the storage requirements of the adjoint method by almost two orders of magnitude.

To minimize the waveform phase misfit, we implement a pre-conditioned conjugate gradient algorithm. Amplitude information is incorporated indirectly by a restricted line search. This ensures that the cumulative envelope misfit does not increase during the inversion. An efficient pre-conditioner is found empirically through numerical experiments. It prevents the concentration of structural heterogeneity near the sources and receivers.

We apply our waveform tomographic method to ≈1000 high-quality vertical-component seismograms, recorded in the Australasian region between 1993 and 2008. The waveforms comprise fundamental- and higher-mode surface and long-period *S* body waves in the period range from 50 to 200 s. To improve the convergence of the algorithm, we implement a 3-D initial model that contains the long-wavelength features of the Australasian region. Resolution tests indicate that our algorithm converges after around 10 iterations and that both long- and short-wavelength features in the uppermost mantle are well resolved. There is evidence for effects related to the non-linearity in the inversion procedure.

After 11 iterations we fit the data waveforms acceptably well; with no significant further improvements to be expected. During the inversion the total fitted seismogram length increases by 46 per cent, providing a clear indication of the efficiency and consistency of the iterative optimization algorithm. The resulting *SV*-wave velocity model reveals structural features of the Australasian upper mantle with great detail. We confirm the existence of a pronounced low-velocity band along the eastern margin of the continent that can be clearly distinguished against Precambrian Australia and the microcontinental Lord Howe Rise. The transition from Precambrian to Phanerozoic Australia (the Tasman Line) appears to be sharp down to at least 200 km depth. It mostly occurs further east of where it is inferred from gravity and magnetic anomalies. Also clearly visible are the Archean and Proterozoic cratons, the northward continuation of the continent and anomalously low *S*-wave velocities in the upper mantle in central Australia.

GJI Seismology

This is, to the best of our knowledge, the first application of non-linear full seismic waveform tomography to a continental-scale problem.

Key words: Numerical solutions; Inverse theory; Surface waves and free oscillations; Seismic tomography; Computational seismology; Australia.

1 INTRODUCTION

1.1 State of the art and summary of previous work

The crust and upper mantle of the Australasian region have been the object of seismological studies since the 1940s when de Jersey (1946) inferred crustal thickness values from Rayleigh wave dispersion and *PP/P* amplitude ratios. Observations of *P* and *S* arrivals from nuclear explosions indicated early on that the Precambrian central and western parts of the continent have fast seismic wave speeds, whereas the Phanerozoic east has slower wave speeds (Bolt *et al.* 1958; Cleary 1967; Cleary *et al.* 1972). Those results were confirmed and extended by the analysis of surface wave dispersion (Bolt 1957; Bolt & Niazi 1964; Goncz & Cleary 1976). Data from the SKIPPY portable array—operated between 1993 and 1996 by the Australian National University (van der Hilst *et al.* 1994)—enabled the construction of tomographic images with unprecedented resolution. Deep cratonic roots in the centre and west, a pronounced low-velocity zone around 140 km depth in the east and a sharp contrast between Precambrian and Phanerozoic lithosphere were clearly imaged using surface wave tomography (Zielhuis & van der Hilst 1996; Simons *et al.* 1999; Yoshizawa & Kennett 2004; Fishwick *et al.* 2005, 2008; Fishwick & Reading 2008). Significant azimuthal anisotropy has been shown to exist in the uppermost mantle under Australia by Debayle & Kennett (2000) and Simons *et al.* (2002). Constraints on the locations of seismic discontinuities and the attenuation structure were obtained through the analysis of body wave arrivals (e.g. Gudmundsson *et al.* 1994; Kaiho & Kennett 2000).

The quality of the tomographic images has improved continuously thanks to increasing data coverage and data quality, technological developments and advancements in theoretical and numerical seismology. While de Jersey (1946) analysed seismograms from three stations, broad-band data from several hundred recording sites in the Australasian region are available today. The theoretical developments of the past two decades have led to an evolution from high-frequency ray theory to more realistic models of seismic wave propagation in complex structures. Several authors have derived Fréchet kernels for measurements at finite frequencies (e.g. Yomogida 1992; Friederich 1999; Dahlen *et al.* 2000; Yoshizawa & Kennett 2005), which have been successfully incorporated into seismic inverse problems (e.g. Friederich 2003; Yoshizawa & Kennett 2004; Sigloch *et al.* 2008). Progress in numerical seismology enables us to simulate the propagation of seismic waves through realistic earth models with unprecedented accuracy (e.g. Faccioli *et al.* 1997; Komatitsch & Tromp 2002; Dumbser & Käser 2006). We could now, in principle, use 3-D numerical methods and complete waveforms for the solution of full waveform tomographic problems. However, a full waveform tomography on a global scale—as envisioned by Capdeville *et al.* (2005)—has not been achieved to date.

This is in contrast to applications in engineering and exploration seismology where full waveform inversion has been used, at least in 2-D, since the early 1980s (Bamberger *et al.* 1982; Crase *et al.* 1990; Igel *et al.* 1996; Pratt & Shipp 1999; Bleibinhaus *et al.* 2007). While the equations of motion are scale-independent, at least within the

macroscopic world, there are nevertheless fundamental differences between local and global tomography that explain this discrepancy: (1) engineering and exploration problems can often be reduced to dependence on just one or two dimensions. The computational costs are therefore comparatively moderate and (2) on smaller scales, seismograms are strongly affected by the scattering properties of the Earth. On larger scales, however, the transmission properties primarily determine the character of a seismogram in general, and the phases of seismic waves in particular.

The recognition that phase information must be extracted explicitly from seismograms is of fundamental importance for the success of a transmission-dominated waveform tomography. Luo & Schuster (1991) therefore proposed to quantify phase differences by cross-correlating observed and synthetic seismograms. A similar approach was taken by Gee & Jordan (1992) who introduced generalized seismological data functionals as frequency-dependent measures of waveform differences. Both concepts have been used recently for 3-D tomography in southern California (Chen *et al.* 2007a; Tape *et al.* 2009).

In this study, we apply time–frequency (TF) misfits as defined by Fichtner *et al.* (2008) using Gabor transforms of the observed and synthetic seismograms. These misfit measures allow us to quantify independent phase and envelope differences as a function of both time and frequency. The most notable advantages of the TF misfits in the context of regional- to global-scale tomography are the separation of phase and amplitude information, the applicability to any type of seismic wave and a quasi-linear relation to Earth structure.

The combination of the above developments—increased data quality and volume, efficient numerical wave propagation and the construction of suitable misfit measures—forms the basis of the full waveform tomography for the Australian upper mantle presented in this paper.

1.2 Objectives and outline

The primary objectives of this study are: (1) the implementation of an efficient algorithm for regional- to continental-scale full waveform tomography. This includes the accurate solution of the forward problem in realistic earth models, the physically meaningful quantification of waveform differences, the computation of Fréchet kernels and the iterative solution of the non-linear misfit minimization problem; (2) the assessment of the resolution capabilities of the full waveform tomography algorithm for our data set in the Australasian region and the identification of effects related to the non-linearity of the misfit functional; (3) the computation and interpretation of full waveform tomographic images of the Australian upper mantle and (4) comparisons of the results with previous studies based on more restrictive approximations.

Our development starts with the solution of the forward problem through a spectral-element method that operates in a spherical section. We emphasize the implementation of crustal structure and the design of long-wavelength equivalent crustal models that allow us to reduce the numerical costs substantially. We then provide a brief review of the definition of the TF misfit functions and two

examples that illustrate their application in the context of full waveform tomography. In Section 3.2, we address the computation of Fréchet kernels for the TF misfits via the adjoint method. To increase the numerical efficiency of the kernel calculations, we propose a simple data compression scheme and an adaptive scheme for the time increment for the integration of the combined forward and adjoint wavefields, designed to preserve suitable accuracy. The centrepiece of the misfit minimization is the pre-conditioned conjugate gradient (PCG) algorithm that we introduce in Section 3.3. Special features of this PCG algorithm are the design of an empirical pre-conditioner, and a restricted line search method that allows us to implicitly incorporate amplitude information that does not enter the inversion directly. The description of the technical and theoretical aspects is followed by specific applications. Section 4 is dedicated to data selection and processing, the construction of the initial model and the resolution analysis. In later sections, we present full waveform tomographic images of the upper mantle in the Australasian region that we obtained after 11 iterations with the PCG algorithm. Following the evaluation of the final waveform fit, we conclude with an interpretation of the tomographic images and a comparison with previous studies.

2 SOLUTION OF THE FORWARD PROBLEM

One of the principal advantages of our waveform inversion method is the accurate solution of the forward problem in laterally heterogeneous earth models. High accuracy is particularly important for surface waves that are mostly sensitive to strong material contrasts in the Earth's upper thermal boundary layer, the lithosphere. A precise solution ensures that the misfit between data and synthetics is primarily caused by imprecise source parameters and the differences between the mathematical model Earth and the real Earth. Since numerical errors can largely be neglected, we can hope to obtain tomographic images that are more realistic than those obtained from approximate solutions that do not take the complete physics of seismic wave propagation into account.

Analytical solutions of the wave equation are not available for realistic 3-D heterogeneous earth models. We therefore have to find fully numerical solutions, and we will place emphasis on spatial discretization and implementation of crustal structure.

2.1 Discretization of the equations of motion

Several methods have been developed for the solution of the 3-D elastic wave equation, each being well suited for particular problems. These methods include, but are not limited to, finite-difference schemes (e.g. Igel *et al.* 2002; Kristek & Moczo 2003), their optimal operator variants (e.g. Takeuchi & Geller 2000) and discontinuous Galerkin methods (e.g. Dumbser & Käser 2006). For seismic wave propagation on continental and global scales, the spectral-element method (SEM) has proven to be a working compromise between accuracy and computational efficiency (e.g. Faccioli *et al.* 1997; Komatitsch & Tromp 2002). The SEM requires a comparatively small number of gridpoints per wavelength, and the vanishing of traction at the free surface is automatically accounted for by solving the weak form of the equations of motion. The correct treatment of the free-surface condition ensures the accurate simulation of surface waves that make up more than 90 per cent of the waveforms in our data set.

We have implemented an SEM variant that operates in a spherical section, as shown on the left-hand panel of Fig. 1. Both, viscoelastic dissipation and anisotropy can be modelled. The unphysical boundaries of the spherical section are treated with the anisotropic perfectly matched layers technique (APML) proposed by Teixeira & Chew (1997) and Zheng & Huang (1997). To circumvent the inherent long-term instability of all PML variants, including the APML, we successively replace the perfectly matched layers by Gaussian tapers (Cerjan *et al.* 1985). An important aspect of our implementation is that we keep the natural spherical coordinates θ (colatitude), ϕ (longitude) and r (radius). This means, in mathematical terms, that the $N_e \in \mathbb{N}$ elements are defined by

$$G_e = [\theta_{e,\min}, \theta_{e,\min} + \Delta\theta] \times [\phi_{e,\min}, \phi_{e,\min} + \Delta\phi]$$
$$\times [r_{e,\min}, r_{e,\min} + \Delta r], \quad e = 1, \ldots, N_e, \quad (1)$$

with constant increments $\Delta\theta$, $\Delta\phi$ and Δr. The transformations from G_e to the reference cube $\Lambda = [-1, 1]^3$ are then defined by the equations

$$\theta = \theta_{e,\min} + \frac{1}{2}\Delta\theta_e(1 + \xi), \quad \phi = \phi_{e,\min} + \frac{1}{2}\Delta\phi_e(1 + \zeta),$$
$$r = r_{e,\min} + \frac{1}{2}\Delta r_e(1 + \eta), \quad (\xi, \zeta, \eta) \in \Lambda. \quad (2)$$

The geometry of the transformation $G_e \to \Lambda$ is visualized in Fig. 1. Note that this transformation is possible only because the spherical section excludes both the centre of the Earth and the poles. The principal advantages of this SEM approach are numerical efficiency and conceptual simplicity, which is of outstanding importance in the development stage of a new methodology. Inside the reference cube, the dynamic fields are approximated by Lagrange polynomials that are collocated at the Gauss–Lobatto–Legendre points. For the waveform tomography we use the polynomial degree 6 and nearly 150 000 elements ($\approx$30 × 10^6 gridpoints) that are 1.2° × 1.2° × 40 km. The minimum period is around 25 s. The time integration is based on an explicit second-order finite-difference scheme, and we solve the discrete equations in parallel on 126 processors (Oeser
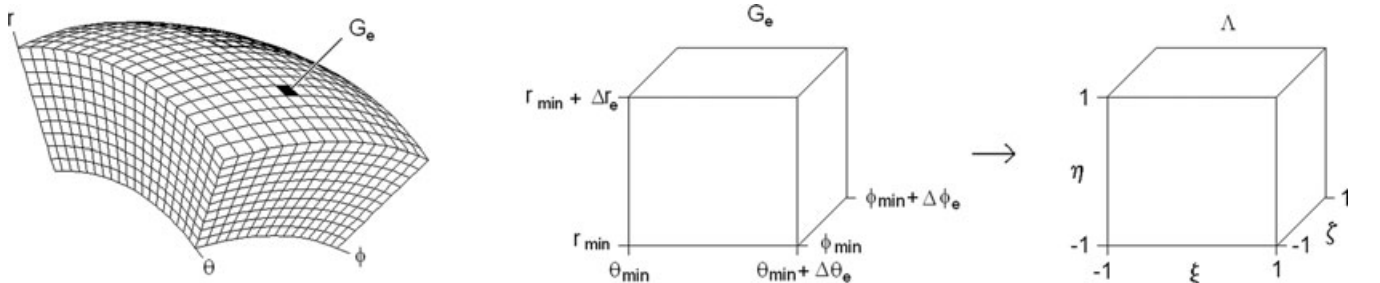

Figure 1. Left-hand panel: coordinate lines in a spherical section parametrized with the natural spherical coordinates r (radius), θ (colatitude), ϕ (longitude). Right-hand panel: illustration of the mapping from the element G_e to the reference cube Λ. See eq. (2).
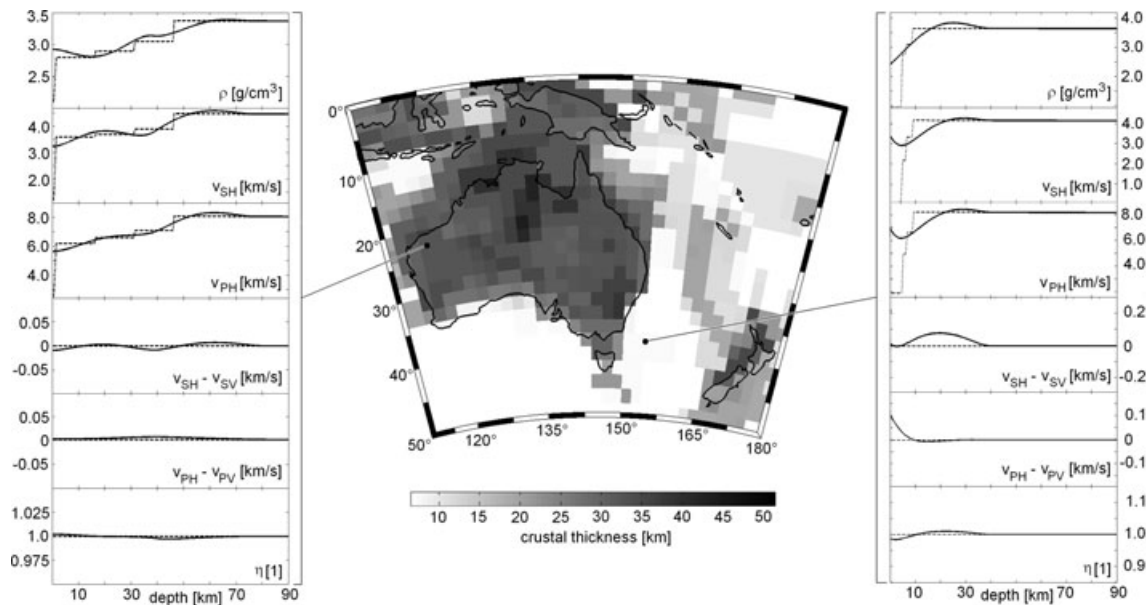
Figure 2. Centre panel: map of the crustal thickness in the Australasian region, according to the model crust2.0 (Bassin *et al.* 2000, http://mahi.ucsd.edu/Gabi/rem.html). Within the continent, the crustal thickness varies between 25 km and more than 50 km. Left-hand panel: profiles of ρ, v_{SH}, v_{PH}, $v_{SH} - v_{SV}$, $v_{PH} - v_{PV}$ and the dimensionless parameter η (Takeuchi & Saito 1972; Dziewonski & Anderson 1981) from the surface to 90 km depth for a location in NW Australia. The original profile from crust2.0 is plotted with dashed lines and the long-wavelength equivalent model with solid lines. The long-wavelength equivalent model is only mildly anisotropic. Right-hand panel: the same as on the left-hand side but for a location in the Tasman Sea where the water depth is approximately 4.5 km. A significant amount of anisotropy is required in the long-wavelength equivalent model in order to match both Love and Rayleigh waves.

et al. 2006). A validation of the solution accuracy can be found in Fichtner & Igel (2008).

2.2 Implementation of crustal structure

Seismograms at periods of 30 s and longer are dominated by surface waves that are sensitive to the structure of the crust while not being able to resolve its characteristic features: the strengths and locations of discontinuities. A realistic crustal structure is thus required as part of the initial model. The implementation of realistic crustal structures is, however, complicated by both geophysical and technical problems.

2.2.1 Geophysical aspects

The crustal structure in the Australasian region has mostly been estimated from reflection/refraction profiles (e.g. Lambeck *et al.* 1988; Klingelhoefer *et al.* 2007) and receiver functions (Shibutani *et al.* 1996; Clitheroe *et al.* 2000; Chevrot & van der Hilst 2000). 3-D crustal models can therefore only be obtained by interpolation that may not capture the strong lateral variations found along some isolated seismic lines (Lambeck *et al.* 1988). Receiver functions are most strongly influenced by discontinuities and strong gradient zones, and in some circumstances there is a significant trade-off between the depth of an interface and the average elastic properties above it (Ammon *et al.* 1990). Coincidence of refraction and receiver function studies in Australia is limited, but there is good agreement as to the nature of the crustal profile where it does exist.

The determination of 3-D crustal structure constitutes an inverse problem with non-unique solutions, and so we are forced to make a choice that includes subjective decisions. We have implemented the crustal structure from the model crust2.0 (Bassin *et al.* 2000, http://mahi.ucsd.edu/Gabi/rem.html), that is displayed in the centre

of Fig. 2. The principal advantage of this model is that statistical inference is used to provide plausible vertical profiles in regions without direct control. However, there are some notable discrepancies between crust2.0 and models obtained directly by the inversion of receiver functions (e.g. Clitheroe *et al.* 2000). These may arise from different parametrizations and inconsistent definitions of the seismological Moho in regions where there is a broad transition instead of a clear crust–mantle discontinuity (e.g. central Australia; see Collins *et al.* 2003). We try to reduce the limitations of the assumed crustal structure by allowing the inversion scheme to modify the shallow parts of the model.

2.2.2 Technical aspects

Most crustal models, including crust2.0, contain thin layers separated by discontinuities in the medium properties. A layer is *thin* when the dominant wavelength of the elastic waves is much larger than the layer thickness. Thus, the oceans and most crustal layers are thin for a wave with a period of 30 s which corresponds to a wavelength on the order of 100 km.

To achieve high numerical accuracy, thin crustal layers need to be honoured by the spectral-element mesh, meaning that element boundaries must coincide with the structural discontinuities. This results in elements that are several times smaller than the minimum wavelength. In practice, however, one element per minimum wavelength is sufficient to represent an elastic wave (Komatitsch & Tromp 1999). Thin layers therefore substantially decrease the numerical efficiency.

In order to reduce the computational costs, we substitute the original crustal structure from crust2.0 by a smooth long-wavelength equivalent (SLWE) crust. This allows us to employ larger elements because there are no discontinuities to be honoured. The details of the SLWE model construction are described in Fichtner & Igel

(2008); this approach is conceptually similar to the homogenization technique of Capdeville & Marigo (2007, 2008). Examples of SLWE profiles for a continental and an oceanic crust are shown in the left- and right-hand panels of Fig. 2, respectively.

The 3-D SLWE crustal model for the Australasian region is constructed as follows. (1) For the set of $5° \times 5°$ grid cells across the area of interest we compute a SLWE version of the original crustal profile. This is done by matching the dispersion curves of the original and the SLWE profiles. The fundamental and first three higher modes in the period range from 15 to 100 s are included in the fitting procedure. In general, the SLWE models are anisotropic with vertical symmetry axis (Backus 1962), and the degree of anisotropy depends on the layer thickness and the vertical velocity contrasts. For the continental profile in Fig. 2(left-hand panel) the anisotropy is weak. This means that the parameter η (Takeuchi & Saito 1972; Dziewonski & Anderson 1981) is close to 1 and that the wave speed differences $v_{SH} - v_{SV}$ and $v_{PH} - v_{PV}$ are several orders of magnitude smaller than the individual wave speeds v_{SH} and v_{PH}. In contrast to this, the anisotropy in the oceanic SLWE model, shown in the right-hand panel of Fig. 2, is comparatively strong because of the large velocity jump at the bottom of the thin ocean layer. (2) The SEM solutions for the individual SLWE profiles have been compared with semi-analytical solutions (Friederich & Dalkolmo 1995) for the corresponding original crustal profiles. For periods longer than 25 s we find that the numerical errors introduced by the SLWE approach are negligibly small compared to the differences between observed and synthetic seismograms. (3) A smooth 3-D crustal model is constructed by interpolating the individual SLWE profiles.

3 THEORETICAL ASPECTS OF THE NON-LINEAR INVERSE PROBLEM SOLUTION

3.1 Definition of phase and envelope misfits in the TF domain

One of the central questions in full waveform inversion is a suitable choice of the misfit functional that quantifies the differences between observed and synthetic waveforms. The misfit functional must extract as much information as possible while conforming to the specifics of the physical problem.

Some of the most important information about the Earth's structure is contained in the phases of waveforms with comparatively small amplitudes, a classical example is provided by P body waves. Phase differences between models are known to be quasi-linearly related to wave speed variations, and so are well suited for an iterative, gradient-based misfit minimization. In contrast, the dependence of amplitudes on variations in the medium properties is frequently highly non-linear. An iterative inversion algorithm may therefore converge slowly or not at all; see Gauthier *et al.* (1986) for an example. Amplitudes depend strongly on the local geology near receivers that may not be well controlled. Information about the deeper Earth can thus be masked by shallow structures such as hidden sedimentary basins. Source inversions for the properties of smaller tectonic earthquakes are often not sufficiently well constrained to warrant the use of amplitudes in structural inverse problems.

In the frequency range used for continental- and global-scale waveform tomography, seismograms are mostly affected by the transmission properties of the Earth and only to a lesser extent

by its diffraction properties. The transmission properties manifest themselves in the time- and frequency-dependent phases of seismic waveforms.

Thus, phases and amplitudes need to be separated and weighted depending on their usefulness for the solution of a particular tomographic problem. One approach that allows us to extract full waveform information while meeting the requirement of phase and amplitude separation consists in the construction of independent phase and envelope misfits, as suggested by Fichtner *et al.* (2008).

For seismograms recorded at the position $\mathbf{x} = \mathbf{x}^r$ we express the i-component of an observed record by $u_i^0(\mathbf{x}^r, t)$, and the corresponding synthetic seismogram by $u_i(\mathbf{x}^r, t)$. In the interest of notational brevity we will omit the index i and the dependencies on $\mathbf{x}^r$ wherever possible. We map the data into the TF domain using the windowed Fourier transform

$$\tilde{u}^0(t, \omega) = F_h[u^0](t, \omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^0(\tau) h(\tau - t) e^{-i\omega\tau} d\tau, \quad (3)$$

with $h(t) = (\pi\sigma)^{-1/4} e^{-t^2/2\sigma^2}$ and σ set approximately equal to the dominant period of the considered waveform. In analogy to (3), we define the TF representation of the synthetics, $u(t)$, as $\tilde{u}(t, \omega) = F_h[u](t, \omega)$. Both, $\tilde{u}^0$ and $\tilde{u}$ can be written in exponential form

$$\tilde{u}^0(t, \omega) = |\tilde{u}^0(t, \omega)| e^{i\phi^0(t,\omega)}, \quad \tilde{u}(t, \omega) = |\tilde{u}(t, \omega)| e^{i\phi(t,\omega)}. \quad (4)$$

Eqs (4) define the envelopes $|\tilde{u}^0(t, \omega)|$ and $|\tilde{u}(t, \omega)|$ and the corresponding phases $\phi^0(t, \omega)$ and $\phi(t, \omega)$. They allow us to introduce an envelope misfit, E_e, and a phase misfit, E_p, in the form of weighted L_2 norms of the envelope difference $|\tilde{u}| - |\tilde{u}^0|$ and the phase difference $\phi - \phi^0$, respectively

$$E_e^2(u^0, u) := \int_{\mathbb{R}^2} W_e^2(t, \omega)[|\tilde{u}(t, \omega)| - |\tilde{u}^0(t, \omega)|]^2 dt d\omega, \quad (5a)$$

$$E_p^2(u^0, u) := \int_{\mathbb{R}^2} W_p^2(t, \omega)[\phi(t, \omega) - \phi^0(t, \omega)]^2 dt d\omega. \quad (5b)$$

The positive weighting functions W_e and W_p act as filters in the TF domain. They allow us (1) to exclude phase discontinuities, (2) to select and weigh particular waveforms and (3) to reduce the influence of seismic noise. Functional forms of W_e and W_p that proved to be useful in both real and synthetic inversions are

$$W_e(t, \omega) = W_T(t) W_F(\omega), \quad (6a)$$

$$W_p(t, \omega) = W_T(t) W_F(\omega) \log[1 + |\tilde{u}^0(t, \omega)|]$$
$$/ \max_{t,\omega} \log[1 + |\tilde{u}^0(t, \omega)|]. \quad (6b)$$

In eqs (6) the symbols W_T and W_F denote Gaussian time and frequency windows, respectively. They are chosen manually for each recording. The time window, W_T, is used to select parts of the seismograms that we subjectively find to be useful for the inversion. We use the frequency window, W_F, to up weigh lower frequencies and, thus, to improve the resolution at greater depth. The effect of the logarithm in (6b) is to increase the relative weight of small-amplitude waveforms. Body wave arrivals may need to be up-weighted individually. More technical details concerning the measurements of phase and envelope misfits can be found in Fichtner *et al.* (2008).

Fig. 3 illustrates the measurement of the phase misfit for vertical-component waveforms from an event in the Loyalty Islands region (2007 March 25; latitude: $-20.60°$, longitude: $169.12°$, depth: 41 km) recorded at the stations CAN (latitude: $-35.32°$, longitude: $148.99°$) and BLDU (latitude: $-30.61°$, longitude: $116.71°$). The position of both stations is shown in Fig. 4. The advance of the
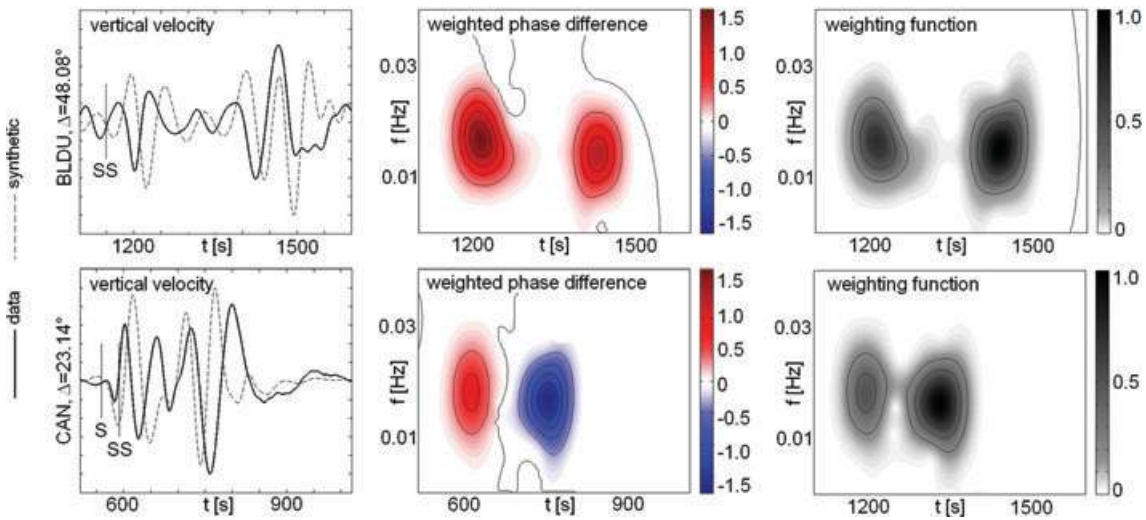
Figure 3. Left-hand panels: vertical-component velocity seismograms from an event in the Loyalty Islands region recorded at the stations BLDU (top panel; latitude: $-30.61°$, longitude: $116.71°$) and CAN (bottompanel; latitude: $-35.32°$, longitude: $148.99°$). The data are plotted as solid lines and the synthetics as dotted lines. Both data and synthetics are low-pass filtered with the cut-off frequency $f_c = 0.02$ Hz. Centre panels: Weighted phase differences, $W_p(\omega, t)\Delta\phi(\omega, t)$, corresponding to the seismograms on the left-hand side. Contour lines are plotted at multiples of 20 per cent of the maximum value. Right-hand panel: weighting functions $W_p(\omega, t)$ that were used to compute the weighted phase differences $\Delta\phi(\omega, t)$.
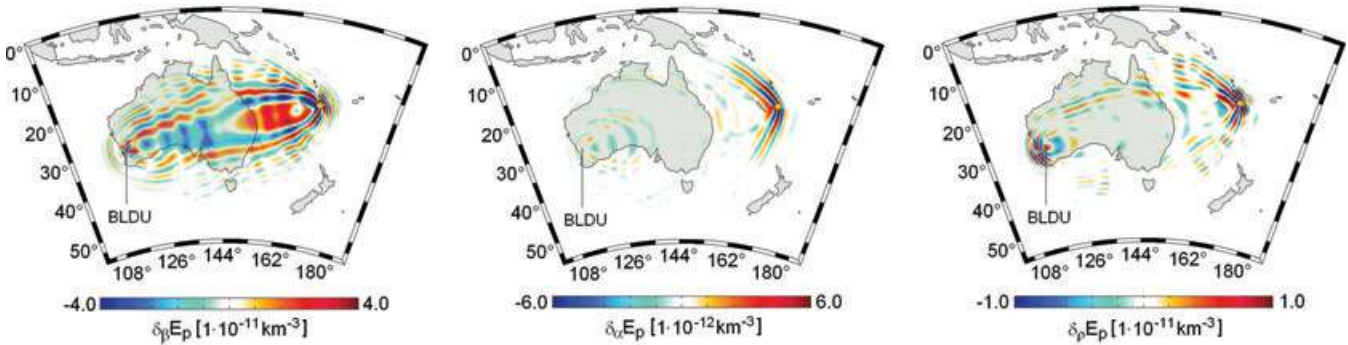

Figure 4. Horizontal slices at 100 km depth through the Fréchet kernels $\delta_\beta E_p$ (left-hand panel), $\delta_\alpha E_p$ (centre panel) and $\delta_\rho E_p$ (right-hand panel) for the waveforms shown in Fig. 3(top panel). A yellow circle indicates the position of the epicentre in the Loyalty Islands region. The kernels with respect to the *P* velocity, $\delta_\alpha E_p$, and density, $\delta_\rho E_p$, attain small absolute values, compared to the kernel with respect to the *S* velocity, $\delta_\beta E_p$. This implies that 3-D *P* velocity and density variations can not be constrained in the inversion. Note that the colour scales are different. The oscillatory nature of the kernels results from the presence of different seismic phases with in the wave train.

observed waveforms with respect to the synthetic waveforms at station (Fig. 3, top left-hand panel) maps to a phase difference that is positive throughout the TF range of interest (Fig. 3, top centre). Particular regions in the TF space are emphasized or suppressed through the weighting function plotted in the upper right-hand panel of Fig. 3. At station CAN, both a phase advance (positive phase difference) and a phase delay (negative phase difference) can be observed. The middle column of Fig. 3 highlights the fact that the phase misfit is a multifrequency measurement. It includes, in our case, frequencies between $\approx 5 \times 10^{-3}$ and $\approx 3 \times 10^{-2}$ Hz.

We note that the TF misfits can equally be defined for velocity or acceleration seismograms. Throughout this study, velocity seismograms are being used.

3.2 Computation of sensitivity kernels and gradients via the adjoint method

An iterative misfit minimization based on gradient methods requires the computation of misfit sensitivity kernels, or Fréchet kernels, with respect to the model parameters. This can be done either

through the scattering integral method (Chen *et al.* 2007a,b) or the adjoint method (e.g. Lions 1968; Tarantola 1988; Tromp *et al.* 2005; Fichtner *et al.* 2006). We decided to implement the adjoint method, which has been effective in as diverse fields as seismology (e.g. Sieminski *et al.* 2007a,b; Liu & Tromp 2008; Stich *et al.* 2009), geodynamics (e.g. Bunge *et al.* 2003) or meteorology (e.g. Talagrand & Courtier 1987). Details concerning the application of the adjoint method to the phase and envelope misfits defined in Section 3.1 can be found in Fichtner *et al.* (2008).

Horizontal slices through the sensitivity kernels corresponding to the phase misfit measured at station BLDU (Fig. 3) are shown in Fig. 4. Since the measurement is dominated by shear waves, the sensitivities with respect to the *P*-wave speed, α and density, ρ, are small compared to the sensitivity with respect to the *S*-wave speed, β. Both lateral *P*-wave speed and density variations are therefore unlikely to be resolvable. Significant sensitivity far from the geometrical ray between source and receiver can generally be observed. This, and the complexity of the sensitivity kernels, is due to the length of the analysed time window and the frequency band limitation of the waveforms.

The gradient of the misfit functional is obtained by projecting the Fréchet kernels onto a set of basis functions. In this study, we use regular blocks that are $1° \times 1°$ wide and 10 km deep. This parametrization allows us to capture the details of the Fréchet kernels, and it reflects the expected maximum spatial resolution.

The computation of sensitivity kernels via the adjoint method requires that the regular wavefield—propagating forward in time—and the adjoint wavefield—propagating backward in time—be simultaneously available. In the presence of viscoelastic dissipation or absorbing boundaries, we are therefore forced to save the complete regular wavefield at intermediate time steps. The resulting amount of data can easily exceed conventional storage capacities. Checkpointing algorithms (e.g. Griewank & Walther 2000; Charpentier 2001) store the regular wavefield at a smaller number of time steps, called checkpoints, and solve the forward problem from there until the current time of the adjoint calculation is reached. The storage requirements are thus reduced at the expense of significantly increased computation time. To keep the computation time as short as possible while reducing the storage requirements we adopt the following two measures.

3.2.1 Data compression

In the spectral-element approach, used to solve the forward problem, the dynamic fields are represented in terms of Nth order Lagrange polynomials. Since neighbouring elements share gridpoints, the storage requirements are proportional to N^3. This can be reduced by storing lower-order polynomial representations of the forward dynamic fields that are then re-converted to Nth degree polynomials during the adjoint calculation. A reduction of the polynomial degree from 6 to 4 results in a compression ratio of 3.4 with no significant deterioration of the sensitivity kernels.

3.2.2 Accuracy-adaptive time increment

The regular wavefield needs to be stored at a sufficient number of intermediate time steps in order to ensure the accurate representation of the principal features of the sensitivity kernels., that is, of the sensitivity contributions in the different Fresnel zones. We derive an estimate of the storage time interval as a function of the dominant period T and the index of the highest-order Fresnel zone that we wish to take into account.

For this, let d_m be the width of the mth Fresnel zone. We can expect the sensitivity kernel to be accurate on length scales similar to d_m when the regular wavefield is stored at least once while propagating over the distance d_m. Hence, an optimistic estimate of the maximum possible storage interval is $t_s = d_m/\beta$, where β is a representative S-wave speed. This means that the regular wavefield needs to be stored at least every $n = t_s/\Delta t = d_m/(\Delta t \beta)$ time steps, where Δt is the time increment. The value of Δt is dictated by the stability condition $\Delta t \leq c\Delta x_{\min}/\alpha_{\max}$, with the Courant number c, the maximum P-wave speed in the model, $\alpha_{\max}$, and the minimum grid spacing, $\Delta x_{\min}$. An approximation of Δt in terms of β and the dominant period T is

$$\Delta t \leq c \frac{\Delta x_{\min}}{\alpha_{\max}} \approx \frac{c\,\lambda}{10\,\alpha_{\max}} \approx \frac{c\beta T}{10\,\alpha_{\max}}, \tag{7}$$

where λ is the dominant wavelength. Eq. (7) is based on the assumption that the minimum grid spacing, $\Delta x_{\min}$ is approximately equal to one tenth of the dominant wavelength, λ. Combining (7)

and the average width of the mth Fresnel zone

$$d_m \approx \frac{1}{4}\sqrt{\beta T \ell}(\sqrt{m} - \sqrt{m-1}), \tag{8}$$

yields the following estimate for n

$$n < \frac{5}{2}\frac{\alpha_{\max}}{c}\sqrt{\frac{\ell}{T\beta^3}}\,(\sqrt{m} - \sqrt{m-1}), \tag{9}$$

where ℓ denotes the length of the ray path. Using suitable parameter values: $\alpha_{\max} = 10\,\mathrm{km\,s}^{-1}$, $c = 0.1$, $\ell = 1000$ km, $T = 50$ s, $\beta = 4\,\mathrm{km\,s}^{-1}$ and $m = 2$ gives $n < 58$. For our inversion we use $n = 20$, meaning that the regular wavefield is stored every 20 time steps. This allows us to accurately represent sensitivity contributions in the higher-order Fresnel zones. Eq. (9) is a rule of thumb that is derived under simplistic assumptions. The resulting estimates for the storage interval, n, prove useful in the context of our particular application. Nevertheless we recommend that the Fréchet kernels for any $n \gtrsim 10$ be verified through comparison with the results for $n = 1$.

It is the combined effect of data compression and accuracy-adaptive time integration that can reduce the storage requirements by almost two orders of magnitude. This is essential for the applicability of the adjoint method for 3-D problems.

We note that the successive increase of the bandwidth in the course of the iterative inversion tends to suppress the higher Fresnel zones. While this effect is still weak in our application, it may in the future be used to increase the storage interval, that is, to increase n.

3.3 Optimization scheme

We cast our formulation of full waveform tomography in the form of a non-linear optimization problem. For the time being, we focus our attention on the minimization of the phase misfit E_p only. While surface wave amplitudes have been used for tomography (e.g. Yomogida & Aki 1987) both their structural information content and the nature of the coupled source/structure inverse problem still require further investigations (Ferreira & Woodhouse 2006). We impose, however, that the envelope misfit, E_e, does not increase from its value for the initial model during the inversion. In mathematical terms the optimization problem may be written as

find an earth model $\mathbf{m}^{\mathrm{opt}}$ such that $\quad E_p(\mathbf{m}^{\mathrm{opt}})$

$$= \min_{\mathbf{m}} E_p(\mathbf{m}), \quad \text{and} \quad E_e(\mathbf{m}^{\mathrm{opt}}) \leq E_e(\mathbf{m}^{(0)}), \tag{10}$$

where $\mathbf{m}^{(0)}$ denotes the initial model. The size of the model space and the computational costs of the forward problem solution prohibit the use of Monte Carlo methods for minimization, and we thus rely on a gradient-based algorithm as described below.

3.3.1 PCG algorithm

To solve the optimization problem (10) we employ a PCG method (Fletcher & Reeves 1964; Alt 2002; Quarteroni *et al.* 2002). Denoting by $\mathbf{m}^{(k)}$ and $P^{(k)}$ the earth model and the pre-conditioner in the kth iteration, the general iterative procedure is as follows.

(i) *Initialization:* Choose an initial model $\mathbf{m}^{(0)}$. Set the iteration index k to 0 and the initial search direction, $\mathbf{d}^{(0)}$, to

$$\mathbf{d}^{(0)} = -P^{(0)}\nabla E_p^{(0)}\left(\mathbf{m}^{(0)}\right). \tag{11a}$$

The gradient, ∇, is with respect to the model parameters, $\mathbf{m}$.

(ii) *Update:* Find an efficient step length, $\sigma^{(k)}$ (see Section 3.3.2 for details) and set

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \sigma^{(k)} P^{(k)} \mathbf{d}^{(k)}. \tag{11b}$$

(iii) *Search direction:* Compute the next search direction, $\mathbf{d}^{(k+1)}$, according to

$$\beta^{(k)} = \frac{\left|\left| P^{(k)} \nabla E_p^{(k+1)} \left[\mathbf{m}^{(k+1)} \right] \right|\right|^2}{\left|\left| P^{(k)} \nabla E_p^{(k)} \left[\mathbf{m}^{(k)} \right] \right|\right|^2},$$

$$\mathbf{d}^{(k+1)} = -P^{(k)} \nabla E_p^{(k)} \left[\mathbf{m}^{(k+1)} \right] + \beta^{(k)} \mathbf{d}^{(k)}. \tag{11c}$$

(iv) *Iteration* Set $k := k + 1$ and go to (ii).

The notation $E_p^{(k)}$ accounts for the dependence of the phase misfit on the iteration index, k, that is introduced through the progressively increasing frequency bandwidth. We elaborate on the construction of the pre-conditioner, $P^{(k)}$, in Section 3.3.3 The requirement $E_p(\mathbf{m}^{\text{opt}}) \leq E_p(\mathbf{m}^{(0)})$ from (10) is met via a restricted line search for the step length, $\sigma^{(k)}$, as described in Section 3.3.2

From a purely theoretical point of view, the iteration terminates when $\nabla E_p^{(k)}(\mathbf{m}^{(k)}) = 0$ for some iteration index k. In practice, however, this exact termination criterion is never met. We therefore stop the iteration when further updates do not lead to significant changes in the model or improvements of the data fit. This subjective choice for our particular case leads to termination after 11 iterations.

In the course of the iterative misfit minimization, we successively decrease the upper cut-off period from 100 s in the first iteration, to 50 s in the final, that is, the 11th, iteration. This introduces a dependence of the phase misfit, E_p, on the iteration index, k, that is accounted for in the notation of eqs (11a) and (11c). There are two positive effects of this well-known strategy (e.g. Pratt 1999). First, the quasi-linear dependence of the phase misfit on earth model perturbations can be maintained by choosing the current frequency such that the phase misfits are sufficiently small. Second, there is a larger number of waveforms for which a valid phase misfit can be computed. Starting immediately with the broadest possible frequency band would result in waveform discrepancies that are too large for a phase misfit to be meaningful (Fichtner *et al.* 2008).

We invert for the hypocentre locations and moment tensor components after the 1st, 4th, 7th and 10th iterations. Improved hypocentre-locations are found by a single-step steepest-descent optimization that is based on the adjoint method for source parameters (Tromp *et al.* 2005). For the moment tensor inversion we compute numerical Greens functions that are then used in a non-linear optimization by Simulated Annealing (Kirkpatrick *et al.* 1983).

Both the successive decrease of the upper cut-off period and the periodic source inversions interrupt the normal sequence of conjugate-gradient iterations. This leads to an effective restart of the optimization algorithm. As pointed out by Alt (2002), periodic restarts can help to accelerate the convergence rate of iterative gradient methods.

3.3.2 Restricted line search: acceptance and rejection criteria

To ensure a reasonably fast convergence of the PCG algorithm, an efficient step length, $\sigma^{(k)}$, needs to be determined for each iteration. The optimal step length satisfies the *acceptance criterion*

$$E_p \left[\mathbf{m}^{(k)} + \sigma^{(k)} P^{(k)} \mathbf{d}^{(k)} \right] = \min_{\sigma} E_p \left[\mathbf{m}^{(k)} + \sigma P^{(k)} \mathbf{d}^{(k)} \right]. \tag{12}$$

Since we can assume the misfit functional $E_p(\mathbf{m})$ to be approximately quadratic in $\mathbf{m}$ near the optimum $\mathbf{m}^{\text{opt}}$, we can estimate the optimal step length with a parabolic line search. For this search

we fit a quadratic polynomial in the trial step length σ through the phase misfit for the models $\mathbf{m}^{(k)}$, $\mathbf{m}^{(k)} + \sigma_1 P^{(k)} \mathbf{d}^{(k)}$ and $\mathbf{m}^{(k)} + \sigma_2 P^{(k)} \mathbf{d}^{(k)}$. The trial step lengths σ_1 and σ_2 are chosen based on physical intuition and experience from synthetic inversions. An approximation of the optimal step length is the value of σ where the parabola attains its minimum.

The algorithm described so far, constructs a monotonically decreasing sequence of phase misfits, $E_p(\mathbf{m}^{(k)})$, given that the acceptance criterion (12) is met for each iteration. To enforce the second optimization criterion from eq. (10) that the envelope misfit E_e does not increase from its initial value, we restrict the line search for an efficient step length. We therefore introduce the *rejection criterion*

$$E_e \left[\mathbf{m}^{(k)} + \sigma^{(k)} P^{(k)} \mathbf{d}^{(k)} \right] \leq E_e \left[\mathbf{m}^{(k)} \right], \tag{13}$$

and we require that $\sigma^{(k)}$ satisfies the acceptance criterion (12) while being subject to the rejection criterion (13). Thus, the rejection criterion may lead to the rejection of a choice of step length that is optimal in the unconstrained sense. The algorithm terminates when an optimal step length in the constrained sense can not be found, that is, when the phase misfit can not decrease without increasing the envelope misfit.

To implement the rejection criterion we extend the quadratic fitting procedure for the phase misfit to the envelope misfit. Moreover, we iteratively determine amplitude source and receiver corrections, closely following the approach taken by Tibuleac *et al.* (2003).

The acceptance/rejection criterion concept allows us to indirectly incorporate secondary data that may yield additional information. This strategy can be interpreted in a broader inverse problem context as a weak form of joint inversion.

We note that alternative approaches to the solution of the non-linear inverse problem may be found in Geller & Hara (1993), Pratt (1999) or Tape *et al.* (2007).

3.3.3 Pre-conditioning

In the vicinity of the global minimum, the convergence rate of a PCG algorithm is proportional to $[\sqrt{\text{cond}(PHP)} - 1]/[\sqrt{\text{cond}(PHP)} + 1]$, where cond (.) is the condition number and H is the Hessian of the misfit functional E_p (e.g. Quarteroni *et al.* 2002). We omit any dependence on the iteration index, in the interest of clarity. The objective of the pre-conditioner is therefore to force the product PHP as close as possible to the unit matrix, I. For $P = H^{-1/2}$ we have $[\sqrt{\text{cond}(PHP)} - 1]/[\sqrt{\text{cond}(PHP)} + 1] = 0$, meaning that the convergence is superlinear.

Pratt *et al.* (1998) demonstrated that the full Hessian can be computed efficiently when the forward problem is solved in the frequency domain. This approach is, however, computationally too expensive for 3-D elastic time-domain solvers, such as the spectral-element method described in Section 2.1. As an alternative, one may use an approximate Hessian, as in the Gauss–Newton method.

We base the construction of the pre-conditioner, P, mainly on experience from synthetic inversions in both 2-D and 3-D. Our objective is to design P such that it reduces the excessively large sensitivities in the vicinity of the sources and receivers. Our approach is conceptually similar to the one taken by Igel *et al.* (1996) who corrected the sensitivity kernels for the geometric spreading of the regular and adjoint wavefields.

For the seismic source with index s ($s = 1, \ldots, N_s$), we denote the regular wavefield by $\mathbf{u}_s(\mathbf{x}, t)$ and the adjoint wavefield by $\mathbf{u}_s^\dagger(\mathbf{x}, t)$. All adjoint sources corresponding to the event s contribute to $\mathbf{u}_s^\dagger(\mathbf{x}, t)$. We define the temporal maxima of $\mathbf{u}_s(\mathbf{x}, t)$ and $\mathbf{u}_s^\dagger(\mathbf{x}, t)$ as
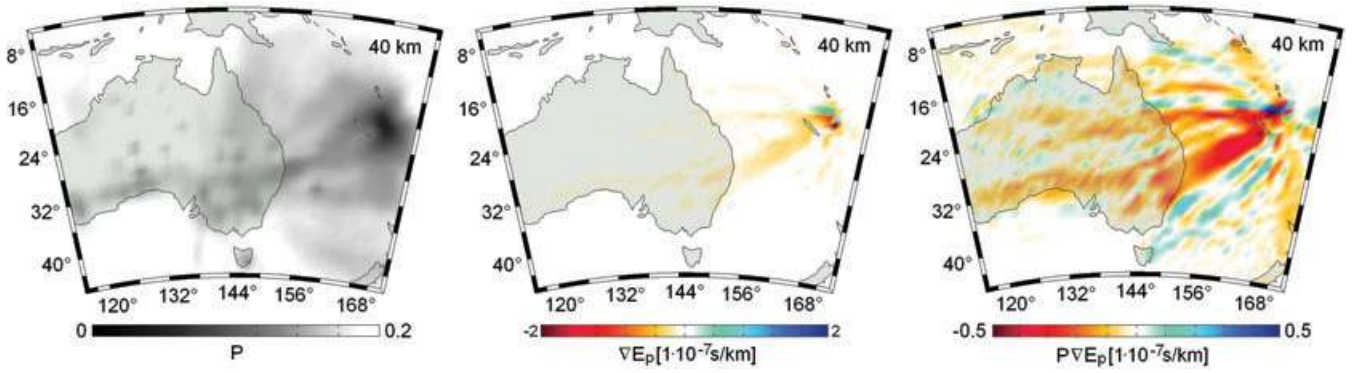
Figure 5. Horizontal slices at 40 km depth through the pre-conditioner P (left-hand panel), the gradient ∇E_p (centre panel) and the pre-conditioned gradient $P\nabla E_p$ (right-hand panel) for the Loyalty Islands event from Section 3.2. The pre-conditioner, P, is characterized by small values at the source and the various receivers on the Australian continent. Its effect on the misfit gradient, ∇E_p, is to emphasize sensitivity contributions between source and receiver, while reducing the sensitivity in the vicinity of the source and directly at the receivers.

functions of the position $\mathbf{x}$

$$M_s(\mathbf{x}) := \max_t |\mathbf{u}_s(\mathbf{x}, t)|, \qquad M_s^\dagger(\mathbf{x}) := \max_t |\mathbf{u}_s^\dagger(\mathbf{x}, t)|. \qquad (14)$$

Based on (14) we furthermore define

$$Q_s(x) := q \, \frac{M_s(\mathbf{x}) - \min_{\mathbf{x}} M_s(\mathbf{x})}{\max_{\mathbf{x}} M_s(\mathbf{x}) - \min_{\mathbf{x}} M_s(\mathbf{x})} + 1,$$

$$Q_s^\dagger(x) := q^\dagger \, \frac{M_s^\dagger(\mathbf{x}) - \min_{\mathbf{x}} M_s^\dagger(\mathbf{x})}{\max_{\mathbf{x}} M_s^\dagger(\mathbf{x}) - \min_{\mathbf{x}} M_s^\dagger(\mathbf{x})} + 1, \qquad (15)$$

where q and $q^\dagger$ are empirically determined scalars. For our specific problem we empirically find $q = 10$ and $q^\dagger = 1$ to be optimal. The function Q_s is a measure of the amplitude of the regular wavefield that ranges between 1 and $q + 1$. Similarly, $Q_s^\dagger$ is proportional to the amplitude of the adjoint wavefield, and it ranges between 1 and $q^\dagger + 1$. From (16) we compute

$$P_s(\mathbf{x}) := \frac{1}{Q_s(\mathbf{x}) + Q_s^\dagger(\mathbf{x})}. \qquad (16)$$

The action of the pre-conditioner P upon the gradient ∇E_p is now defined in terms of the gradients of the phase misfit for each individual event, $\nabla E_{p,s}$

$$P \nabla E_p := \sum_{s=1}^{N_s} P_s \, \nabla E_{p,s}. \qquad (17)$$

Fig. 5 illustrates the effect that P has on the pure gradient, ∇E_p. The pre-conditioner (Fig. 5, left-hand panel) attains relative small values in the source region and in the vicinity of the receivers that are visible as isolated grey areas. Sensitivity contributions between the source and the receivers are therefore enhanced at the expense of reduced sensitivity near the source and the receivers. This prevents the optimization scheme (11) from converging towards a local minimum where nearly all structural heterogeneities are located in the source or receiver regions.

An additional positive effect of the pre-conditioning is that it introduces a natural data weighting scheme. Data from areas with a comparatively high station density are down-weighted relative to data with a sparser station coverage.

4 SETUP OF THE WAVEFORM TOMOGRAPHY—DATA, INITIAL MODEL AND RESOLUTION ANALYSIS

4.1 Data selection and processing

Since the full waveform misfits are quite susceptible to noise and computational resources are finite, we have adopted the following data selection criteria: (1) The estimated signal-to-noise ratio in the frequency band from 7 to 20 mHz is required to be higher than 20. This ensures that the waveform misfit is dominated by the discrepancies between the mathematical model and the Earth. (2) The adjoint method for the computation of sensitivity kernels is efficient only when a large number of recordings per event is available; we therefore accept only those events with more than 10 high-quality recordings in the Australasian region. (3) The event magnitude is required to be smaller than Ms 6.9; this choice allows us to neglect finite-source effects in seismogram modelling. (4) After the first iteration we furthermore exclude any events where neither a relocation nor an additional full moment tensor inversion lead to satisfactory waveform fits.

Our final data set comprises 1075 vertical-component seismograms from 57 events that occurred between 1993 and 2008. Approximately, 70 per cent of the data were recorded at permanent stations operated by Geoscience Australia, IRIS and GEOSCOPE. The remaining 30 per cent originate from the SKIPPY, KIMBA, QUOLL and TASMAL experiments, undertaken by the seismology group at the Australian National University. All selected recordings contain prominent fundamental-mode surface waves that we used in the inversion. To this we added 306 waveforms from long-period S-body waves and higher-mode surface waves. We do not include P-body waves because they are rarely observable in the frequency range of interest (7–20 mHz). The ray coverage for the fundamental-mode surface waves is good throughout the eastern part of the continent and decreases towards the west, as shown in the left-hand panel of Fig. 6. We did not include events from the Ninety-East and Indian ridges to the west that would have improved the coverage in Western Australia at the expense of a substantially enlarged numerical model and the need for more extensive computational resources. The higher-mode and S-body wave coverage is acceptable in Eastern Australia but poor in the west (Fig. 6, right-hand panel). We therefore do not expect good resolution below 400 km depth.
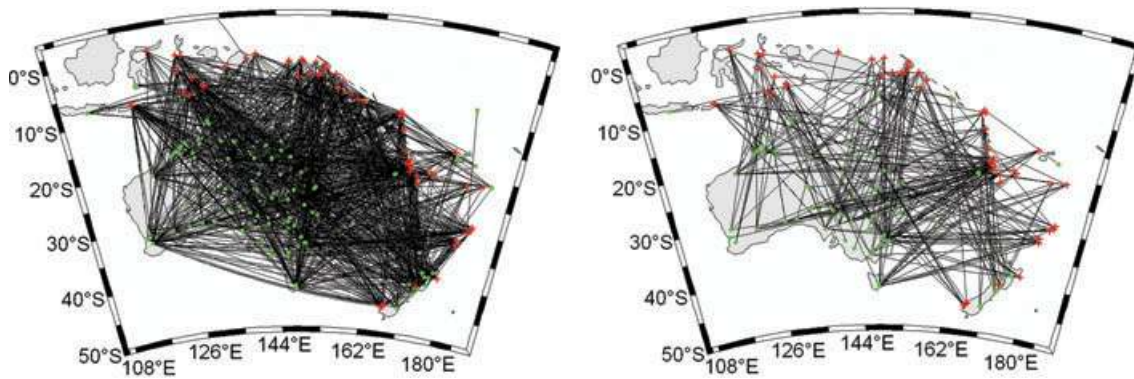
Figure 6. Left-hand panel: ray coverage for the 1075 fundamental mode surface waves. Epicentres and stations are marked by red plus signs and green dots, respectively. Right-hand panel: the same as on the left-hand side but for the 306 recordings of higher-mode surface and long-period *S* body waves.
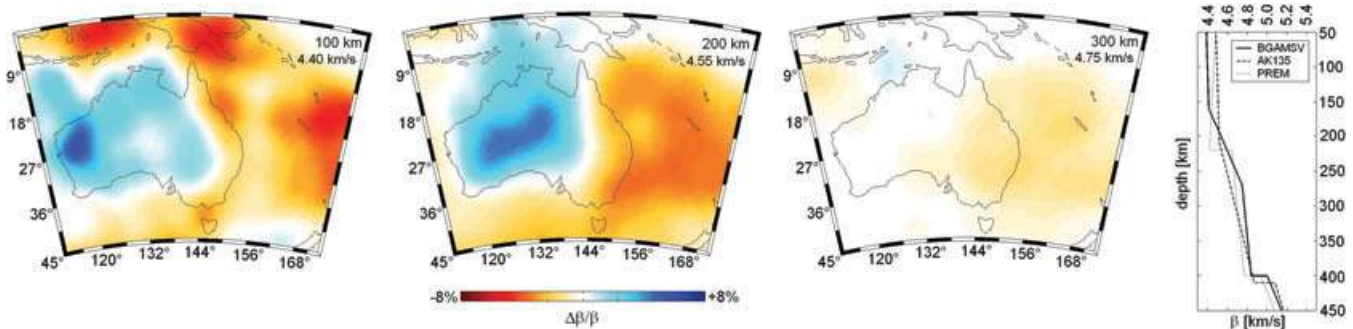

Figure 7. Horizontal slices through the initial *S* wave speed variations at 100, 200 and 300 km depth. The depth labels and 1-D background wave speeds are shown in the upper right-hand corner of the subfigures. Clearly visible are the very long wavelength features of the Australasian region: (1) higher than average velocities in western and central Australia and (2) lower than average velocities in eastern Australia and the adjacent Tasman and Coral Seas. The model is radially symmetric beyond 350 km depth. For easier comparison the colour scale is the same as for the tomographic images in Fig. 10. In the rightmost panel, we show our 1-D reference model, BGAMSV, compared to the global models AK135 (Kennett *et al.* 1995) and PREM (Dziewonski & Anderson 1981).

Currently, we disregard horizontal-component recordings. This allows us to work with an isotropic model. A 3-D radially anisotropic model is work in progress.

Our data set differs from those used in previous surface wave studies (e.g. Zielhuis & van der Hilst 1996; Simons *et al.* 1999; Debayle & Kennett 2000; Simons *et al.* 2002; Yoshizawa & Kennett 2004; Fishwick *et al.* 2005) in that it contains mostly recordings from permanent stations, many of which were not operational prior to 2006. The improved permanent station coverage allows us to invert a comparatively large number of high-quality waveforms with dominant periods >60 s.

To avoid potentially unstable time integrations, we work with velocity rather than with displacement seismograms. For the first iteration, all data are bandpass filtered between 100 and 1000 s. In the course of the inversion, as described in detail in Section 3.3, we successively decrease the lower cut-off period to 50 s. This procedure ensures that the quasi-linearity of the misfit functional with respect to earth model perturbations is maintained. Since absolute amplitude information does not enter the inversion, we do not normalize the recordings.

4.2 Initial model

To ensure the fast convergence of the iterative optimization algorithm and the efficiency of a parabolic line search (see Section 3.3.2), we implement an initial model that already contains the very long wavelength features of the Australasian upper mantle. The *S*-wave speed variations, shown in Fig. 7, are a smoothed version of the sur-

face wave tomogram by Fishwick *et al.* (2005). Based on the results obtained by Kaiho & Kennett (2000), who studied refracted body waves, we set the initial *P*-wave speed variations to 0.5 times the initial *S*-wave speed variations. There are no lateral density variations in the initial model, and density is not updated in the inversion. This restricted model is justified by the negligible sensitivity of our measurements to 3-D *P*-wave speed and density perturbations (see Fig. 4). The radially symmetric reference model, referred to as BGAMSV, is shown in the rightmost panel of Fig. 7. It was found by trial and error perturbations from an isotropic variant of PREM without the 220 km discontinuity. We use BGAMSV as reference for plots of 3-D wave speed variations (Figs 7 and 10–12) and as initial model for synthetic inversions (Section 4.3). The Q model is from PREM (Dziewonski & Anderson 1981). We present a brief discussion on the dependence of the tomographic images on the initial model in Section 7. Initial estimates of the hypocentre coordinates and the moment tensor components were obtained from the Centroid Moment Tensor catalogue (http://www.globalcmt.org).

4.3 Resolution analysis and identification of effects related to non-linearity

One of the major advantages of the use of the full waveform modelling scheme for 3-D varying media is that we are able to carry out resolution analysis incorporating genuine 3-D velocity structure. No approximations are required in the construction of the artificial data. The reconstructions of the input models therefore provide realistic estimates of the resolution capabilities of our full waveform
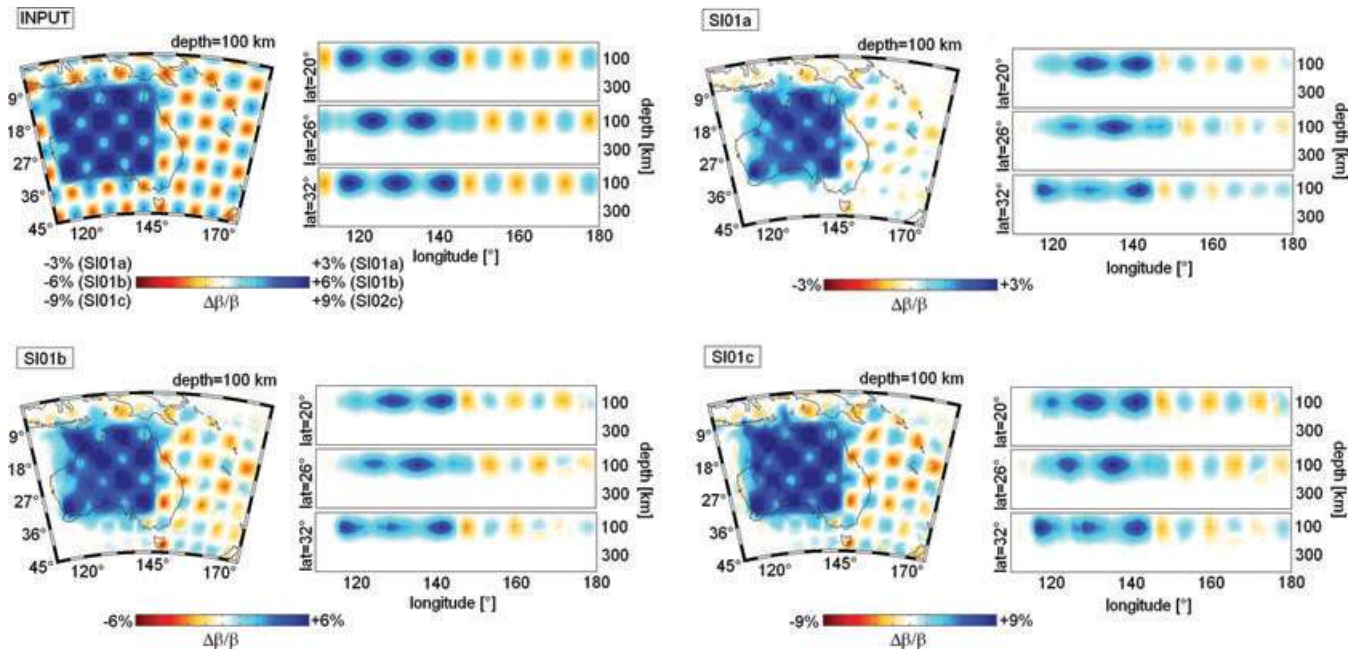
Figure 8. Summary of the resolution tests for shallow structures. Horizontal and vertical slices through the input structure are shown in the upper left-hand part of the figure. The maximum amplitudes of the perturbations are ±3 per cent for the synthetic inversion SI01a (top right-hand side), ±6 per cent for SI01b (bottom left-hand side) and ±9 per cent for SI01c (bottom right-hand side).

tomographic method. Note that each resolution test requires the same number of regular and adjoint simulations as the solution of the actual tomographic problem.

A set of anomalies are introduced into the 1-D background model (Fig. 7, rightmost panel), and full calculations are made using the realistic path distribution and the source mechanisms for the real events. The model is then attempted to be recovered using the algorithm described above. We do not use a 3-D initial model for the synthetic inversions in order to obtain a conservative resolution estimate, and to avoid biases introduced by potentially too accurate initial models. We also do not add noise to the synthetic data because it plays a minor role in the real data.

In a first series of tests the input model, shown in the upper left-hand panel of Fig. 8, consists of a $3° \times 3°$ chequer-board pattern superimposed onto a high-velocity patch in central and western Australia. This is intended to mimic the positive anomalies found in the Precambrian parts of the continent. The anomalies are centred around 100 km depth, and their maximum amplitudes are variable: ±3 per cent for model SI01a, ±6 per cent for model SI01b and ±9 per cent for model SI01c. We present the outcomes of the inversions after 10 iterations in the remaining panels of Fig. 8.

Both the long- and short-wavelength structures of the input model are well recovered in all three scenarios. This result highlights the advantages of the TF phase misfit as opposed to the L_2 norm used in diffraction tomography, where only short-wavelength features can be recovered (e.g. Gauthier *et al.* 1986). The amplitudes of positive chequer-board anomalies generally seem to be less well recovered than their neighbouring negative anomalies—a phenomenon for which we can currently not provide a definite explanation.

There is a clear non-linear effect in the sense that the quality of the reconstructions improves when the amplitudes of the perturbations are increased while keeping their geometry constant. The input pattern for SI01c, with maximum perturbations of ±9 per cent, can be reconstructed more accurately than for SI01a, where the maximum perturbations are ±3 per cent. This is in contrast to

linearized ray tomography. An intuitive explanation might be that small-amplitude perturbations essentially translate to phase shifts only. Larger-amplitude perturbations, however, lead to both phase shifts and substantial waveform distortions. In the latter case, the waveforms therefore seem to contain more information even though the geometry of the anomalies has not changed. This additional waveform information seems to improve the resolution.

After 10 iterations, the synthetic waveforms and the artificial data are practically identical. This implies that the imperfections seen in the reconstructions in Fig. 8 are mostly the result of insufficient data coverage and the frequency band limitation. Potential algorithmic problems such as an insufficient exploitation of waveform information or slow convergence do not seem to play a significant role.

To assess the resolution of deeper structures, we place a $8° \times 8°$ chequer board pattern around 280 km depth. The results are summarized in Fig. 9. The principal features can be recovered, though less optimally than in the case of shallower heterogeneities, presented in Fig. 8. This is a direct consequence of the relative sparcity of higher-mode surface wave and long-period body wave data. Due to vertical smearing, the amplitudes of the recovered heterogeneities are mostly too small.

We conclude that above 200 km depth the resolved length scale is about $3°$ horizontally and 50 km vertically. Below 250 km, this increases to about $10°$ and 100 km, respectively.

5 RESULTS

We have applied the non-linear optimization scheme described in Section 3 to our waveform data set for the Australasian region. In the 11th iteration we observed improvements of the data fit of less than 5 per cent and maximum changes in the *SV*-wave speed of less than $0.05 \, \mathrm{km \, s^{-1}}$, and therefore terminated the iterative misfit minimization.
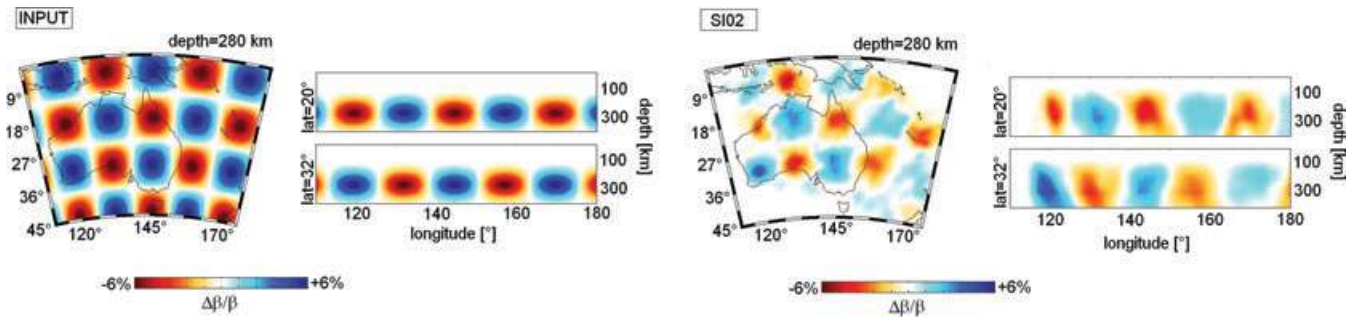
Figure 9. Resolution test for deeper structures. Horizontal and vertical slices through the input structure are shown in the left-hand part of the figure. The maximum amplitudes of the perturbations are ±6 per cent. The result is displayed in the right-hand panel.
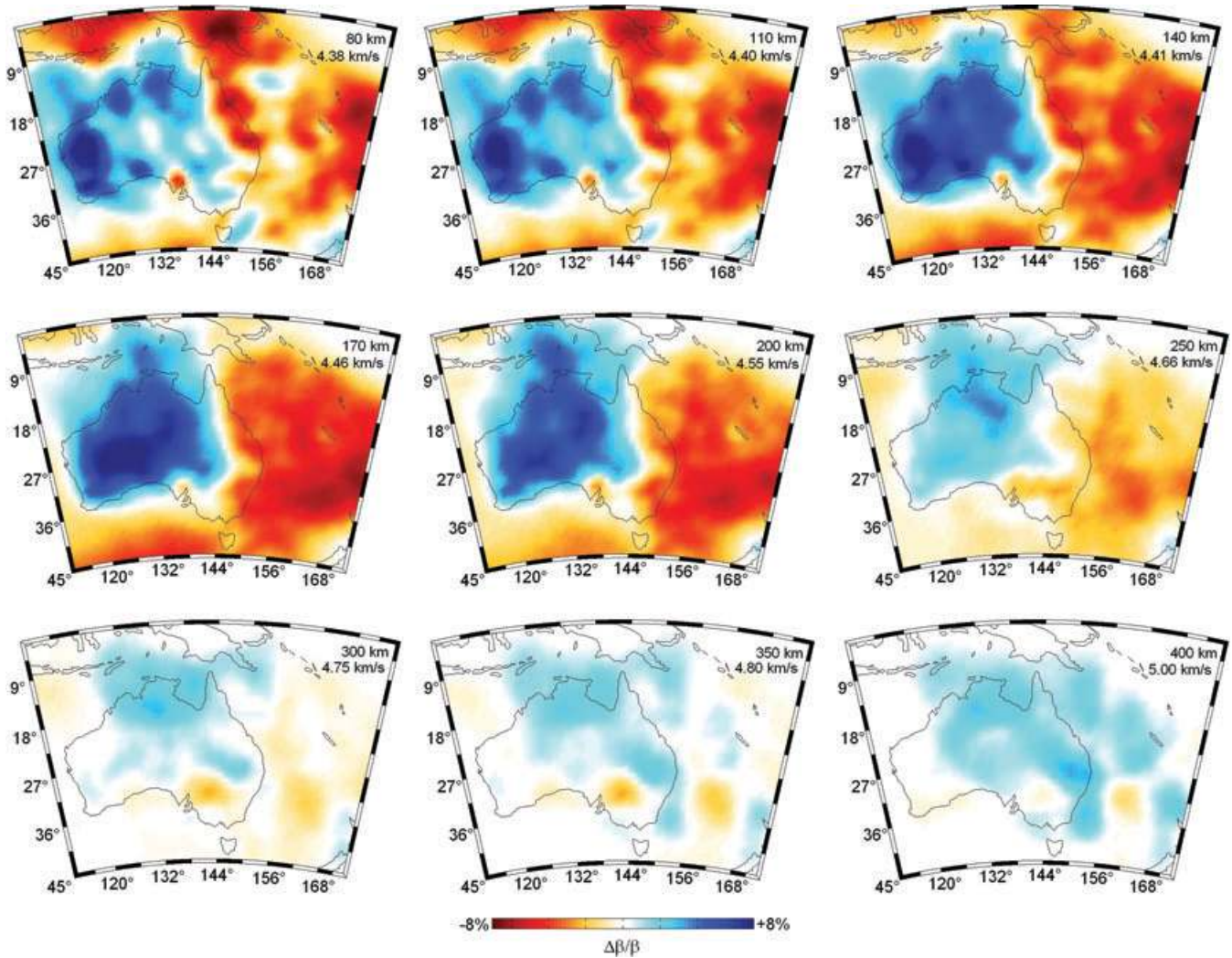

Figure 10. Horizontal slices through the tomographic model AMSV.11 at depths between 80 and 400 km. The depth labels and 1-D background wave speeds are shown in the upper right-hand corner of the subfigures. The colour scale saturates at *SV*-wave speed variations of ±8 per cent. The maximum perturbations at 80, 110, 140 and 170 km depth are ±10.4, ±9.6, ±8.8 and ±8.4 per cent, respectively. Below 200 km depth, lateral variations are generally less than ±8 per cent.

The resulting *SV* velocity model, AMSV.11, is presented in Figs 10–12. In the vertical slices we do not show the upper 50 km, where our use of the long-wavelength equivalent crustal model prohibits a direct interpretation of the images.

Above 250 km depth, the long-wavelength structure of AMSV.11 is dominated by the high-wave speeds of Precambrian Australia and the low-wave speeds of Phanerozoic Australia and the Tasman and Coral Seas. This is in agreement with previous studies (e.g. Zielhuis & van der Hilst 1996; Simons *et al.* 1999, 2002; Debayle & Kennett 2000; Yoshizawa & Kennett 2004; Fishwick *et al.* 2005). Notable discrepancies exist below 300 km depth and on length scales of less than 1000 km; these arise from the use of different data sets and
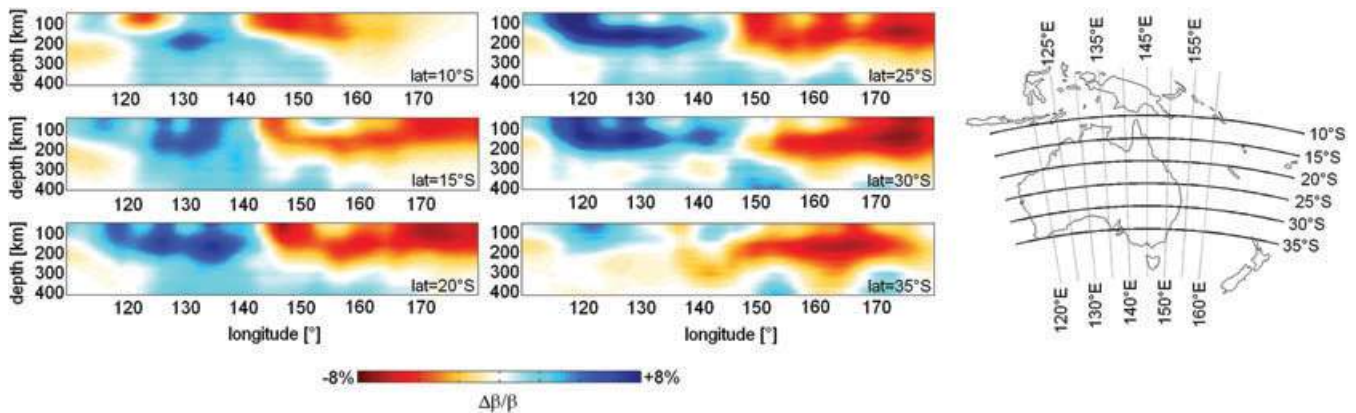
Figure 11. Vertical slices through the tomographic model AMSV.11 at different latitudes between 10°S and 35°S in steps of 5°. The depth interval between the surface and 50 km depth is not shown because the long-wavelength equivalent crustal model prohibits a direct interpretation of the heterogeneities in this interval.
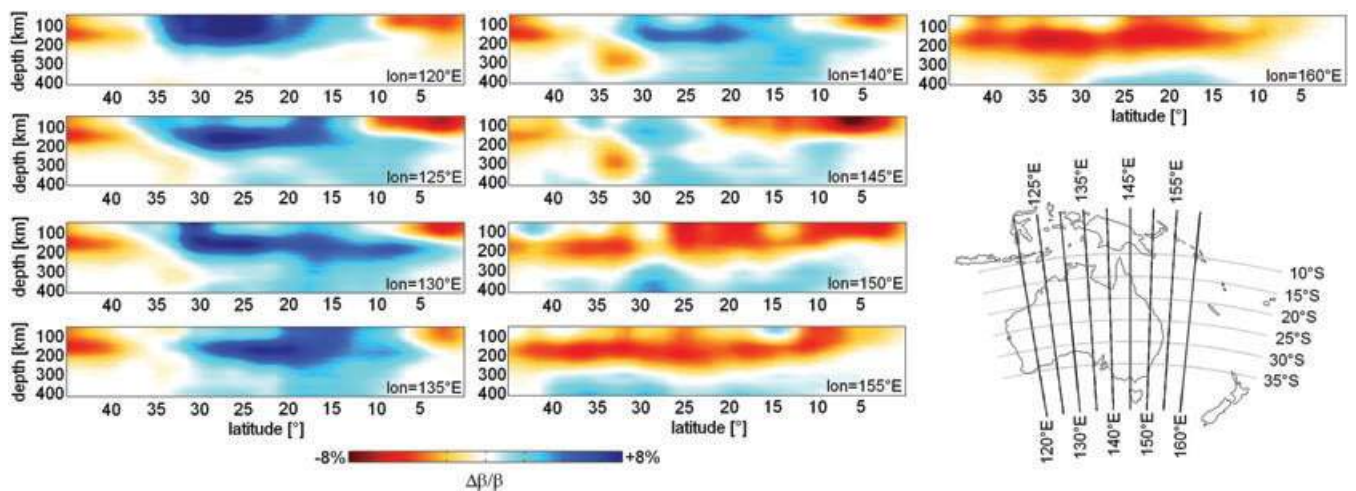

Figure 12. Vertical slices through the tomographic model AMSV.11 at different longitudes between 120°E and 160°E in steps of 5°. As in Fig. 11 we do not show the upper 50 km of the model.

methodologies. The maximum velocity perturbations of AMSV.11 relative to the 1-D reference model (rightmost panel in Fig. 7) are around ±10 per cent in the upper 100 km, compared to ±5 per cent in the initial model. Below 250 km depth the lateral variations reduce to at most ±5 per cent.

5.1 Waveform fit

A convenient measure of the global improvement achieved through the waveform fit is provided by the phase misfit distributions for the initial model and the final model after 11 iterations (Fig. 13).

At the first iteration we were able to use 865 time windows where a clear correspondence between data and synthetic waveforms was observable. The total time window length of the fitted portions was around 61 hr. A histogram of the phase misfit in the first iteration, normalized with respect to the product of the epicentral distance and the individual time window length, is shown in the upper left-hand panel of Fig. 13. Most normalized phase misfits are below 0.4, and the mean is 0.22. The lower left-hand panel of Fig. 13 displays the geographical distribution of the normalized phase misfit, with ray path segments plotted around their ray path centres for each of the 865 source–receiver pairs. The length of each segment is

proportional to the length of the ray and its colour indicates the corresponding normalized phase misfit. This figure provides a rough estimate of where the initial model deviates most from the real Earth—in Northern Australia and north of New Zealand.

In the course of the successive iterations of the inversion scheme, the phase misfit decreased continuously. The increasing similarities between data and synthetics allowed us to add new time windows and to increase the length of time windows that we had used in the first iteration already. The total time window length of the fitted segments in the final iteration is 90 hr (+46 per cent), distributed over 1100 individual time windows. The progressive improvement in the amount of data that can be explained by the updated models from successive iterations provides a clear indication of the efficacy and consistency of the inversion procedure.

In order to make a direct comparison of the phase misfits at the first iteration and after the final iteration comparable, we analyse the normalized phase misfits after the 11th iteration using just the time windows employed at the first iteration. The results are displayed in the central column of Fig. 13. The mean normalized phase misfit has dropped from 0.22 to 0.10, and there is no indication of any geographical region where the tomographic model deviates more than average from the true Earth. Plots of the normalized phase misfit distributions, computed with the increased number of time
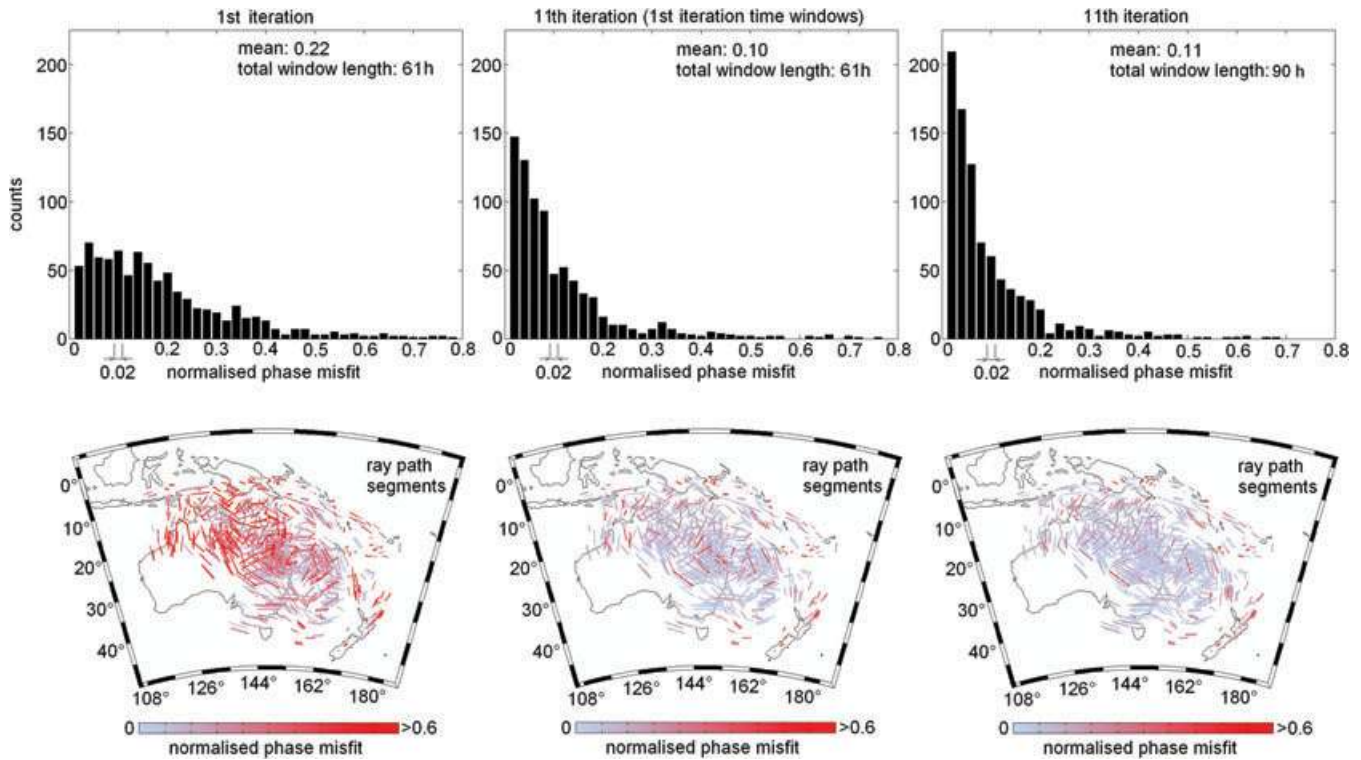
Figure 13. Top left-hand panel: distribution of phase misfits for the initial model, normalized with respect to the product of the individual time window length and the epicentral distance. The mean value is 0.22 and the total analysed time window length is 61 hr. Bottom left-hand panel: ray path segments are plotted around the ray path centres for each source/receiver pair that contributed to the 1st iteration. The length of each segment is proportional to the total ray path length and its colour indicates the corresponding normalized phase misfit. This figure provides a rough estimate of where the initial model is most inadequate. Centre panels: The same as in the left-hand column but after the 11th iteration and with the same time windows used for the 1st iteration. The mean normalized phase misfit has dropped to 0.10. From the ray segment plot we infer that the remaining phase misfit is geographically evenly distributed. This suggests that there is no region where the discrepancies between the real Earth and the tomographic model are anomalously large. Right-hand panel: the same as in the central column but including the new time windows that successively became usable in the course of the iteration. After 11 iterations the total time window length has increased by 46 per cent to 90 hr.

windows after 11 iterations are shown in the right-hand column of Fig. 13.

A more detailed view of the data fit is provided by Figs 14–16. Fig. 14 shows the data and synthetic waveforms for an event in the Maluku Island region. The data are plotted as solid black lines, the predictions from the initial model as dashed black lines and the predictions from the final model as red lines. All the synthetic waveforms corresponding to the initial model are late by 10 s and more, indicating that the initial model is too slow north of Australia. This is in agreement with the ray segment plot in the lower left-hand panel of Fig. 13. The final predictions generally agree well with the data.

Waveforms for an event south of Vanuatu, for which the ray paths are mostly oriented east–west, are shown in Fig. 15. At stations MSVF, NOUC and SNZO the data are well predicted by the initial model, suggesting that it is already adequate in the South Fiji Basin. As for the Maluku Islands event (Fig. 14), the final synthetics reproduce the data waveforms well, though any visual inspection is necessarily subjective. From a more complete analysis of the waveform fits we conclude that azimuthal anisotropy is not required in order to reproduce the data waveforms. This is supported by the examples given in Figs 14 and 15, where the data are fit equally well despite the different ray path orientations.

Since usable recordings of higher-mode surface waves and long-period *S* body waves are comparatively rare, we display a collection

of such time segments from a variety of events in Fig. 16, along with the synthetic waveforms after the 11th iteration.

There are many plausible contributions to the remaining differences between observed and synthetic seismograms: *P* velocity and density structure, source complexity, recording site effects, anisotropy and small-scale structures that can not be resolved or represented. We can currently not quantify the relative importance of these contributions. This is a necessary future research direction.

6 INTERPRETATION AND COMPARISON WITH PREVIOUS STUDIES

6.1 Tectonic setting

We briefly review the tectonic evolution of the Australasian region in order to provide the stage for the following interpretation of the tomographic images. For this we essentially follow Myers *et al.* (1996), Crawford *et al.* (2003) and Müller *et al.* (2000). The principal geologic features of Australasia are shown in Fig. 17.

Between 2500 and 1950 Ma, Archean continents were fragmented, leading to the formation of independent crustal blocks, such as the Pilbara and Yilgarn cratons that are well exposed in
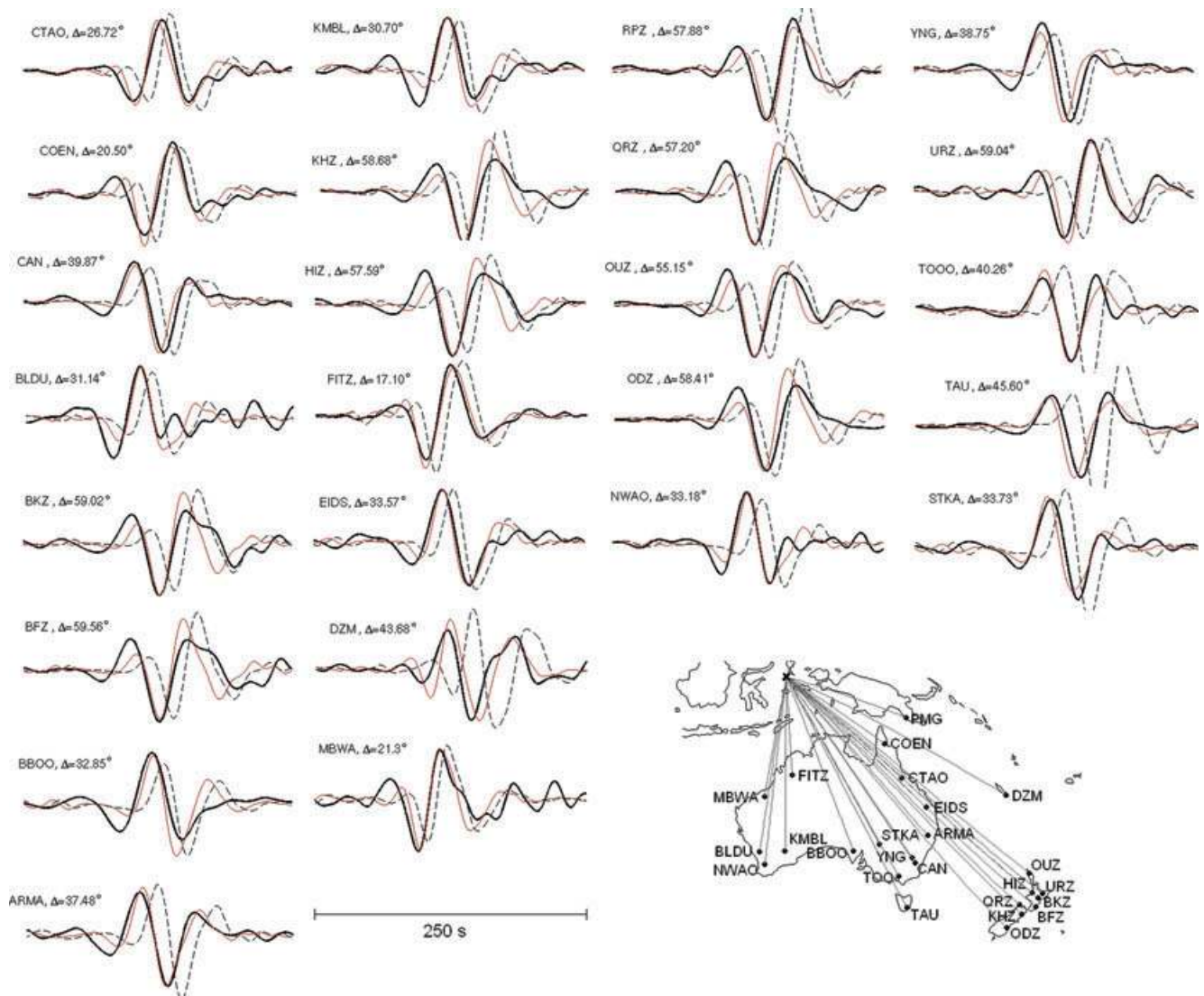
Figure 14. Waveform comparison. Data are plotted as thick black lines, final predictions as thin red lines and initial predictions as dashed black lines. Seismograms are for an $M_w = 6.7$ event that occurred 2007 February 20 in the Maluku Island region (latitude: $-0.91°$, longitude: $127.17°$). The lower cut-off period is 50 s. Ray paths and station locations are indicated in the lower right-hand corner. The comparison of initial and final fits reflects the velocity increase in northern Australia during the inversion.

Western Australia. Several older crustal fragments were assembled between 1950 and 1700 Ma, forming the North and South Australian cratons. The Yilgarn and Pilbara cratons joined to form the West Australian craton during the Capricorn Orogeny. After a period of accretion and intra-cratonic deformation, the North, South and West Australian cratons collided around 1300 to 1000 Ma and formed an early part of the supercontinent Rodinia. Subsequently, an intracratonic basin developed over the junction of the North, South and West Australian cratons. This basin was fragmented during the breakup of Rodinia in the Late Proterozoic between 750 and 540 Ma.

The Late Neoproterozoic and Phanerozoic evolution of eastern Australia was marked by repeated cycles of (1) the separation of continental ribbons during extensional events, (2) the subsequent initiation of subduction, (3) the formation of intra-oceanic arcs, (4) their return and accretion to the continental margin and (5) post-collisional extension and magmatism. This led to the successive formation of the Delamerian, Lachlan and New England Fold Belts, which together constitute the Tasman Fold Belt System in southeastern Australia. Present-day witnesses of this continuing sequence of processes are three marginal basins: the Tasman Sea, the New Caledonia Basin and the Loyalty Basin. They are separated by two microcontinental ribbons: the Lord Howe Rise and the New Caledonia-Norfolk ridge that separated from Australia between 120 and 95 Ma. The Coral Sea started to open around 64 Ma, and the spreading between the Lord Howe Rise and Australia ceased by 52 Ma.

In the course of the break-up of Pangea, East Antarctica detached from south Australia during the Cretaceous, opening the Southern Ocean. The Early Cretaceous separation of Greater India from southwestern Australia resulted in the opening of the central Indian Ocean. During most of the Cenozoic, Australia moved northward, approaching the Melanesian Arc system.

The current tectonic development of the Australasian region is marked by subduction and collision processes along its active boundaries. The continental regions of Australia are tectonically quiet are characterized by modest seismicity and diffuse small-volume Cenozoic volcanism.
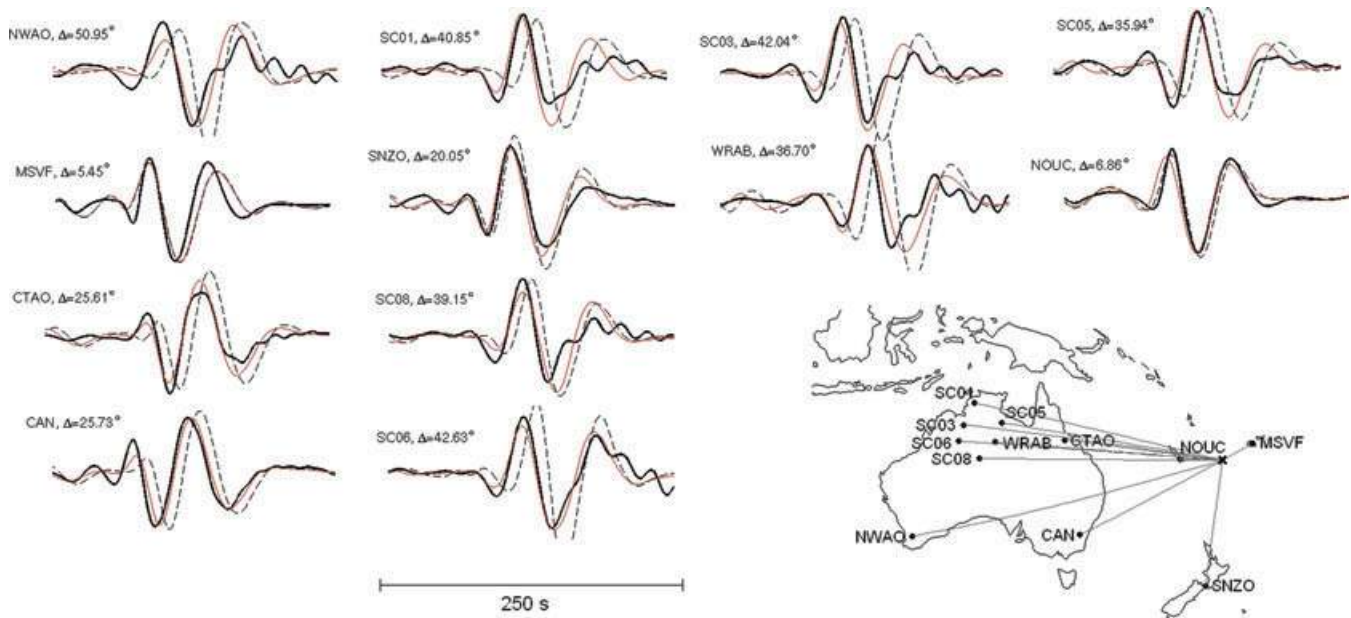
Figure 15. Waveform comparison. Data are plotted as thick black lines, final predictions as thin red lines and initial predictions as dashed black lines. Seismograms are for an $M_w = 6.2$ event that occurred 1994 September 3 south of Vanuatu (latitude: $-21.16°$, longitude: $173.80°$). The lower cut-off period is 50 s. Ray paths and station locations are indicated in the lower right-hand corner. The initial fit at stations MSVF, NOUC and SNZO is similar to the final fit, indicating that the initial model is already adequate in and around the South Fiji Basin.
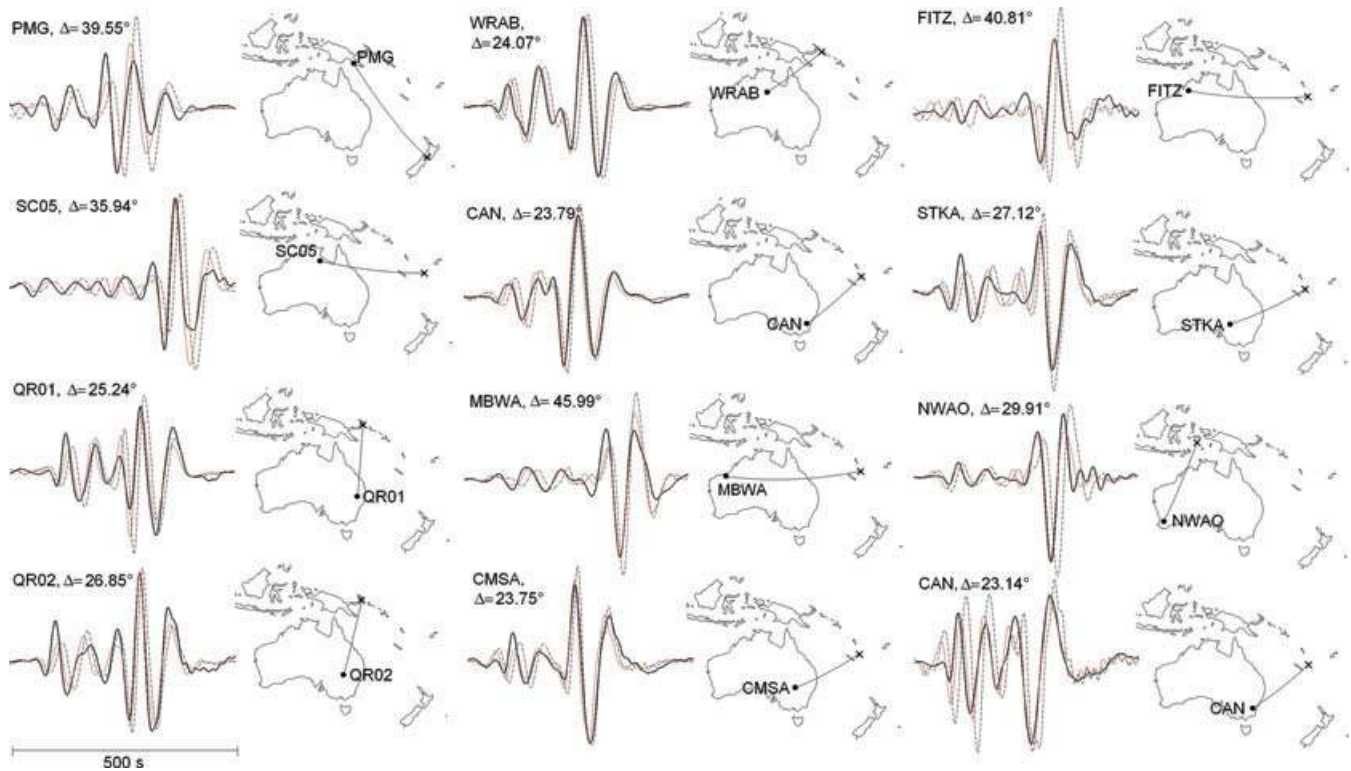

Figure 16. Waveform comparison for a selection of recordings where higher-mode surface waves or *S* body waves are visually distinguishable. Data are again plotted as thick black lines, final predictions as thin red lines and initial predictions as dashed black lines. The lower cut-off period is 50 s.

6.2 Structural elements of the upper mantle in the Australasian region

The dominant feature of the *S*-wave speed model is the relatively high velocities beneath the Precambrian regions of western and central Australia for depths below 130 km. High *S*-wave speeds in the uppermost mantle fit well with surface cratonic components. Easternmost Australia and the adjacent Coral and Tasman Seas are characterized by a pronounced low-velocity zone; this is centred around 140 km depth and is not present under the Precambrian components of the Australian lithosphere. Both the extent and the location of the low-velocity zone agree well with the results of
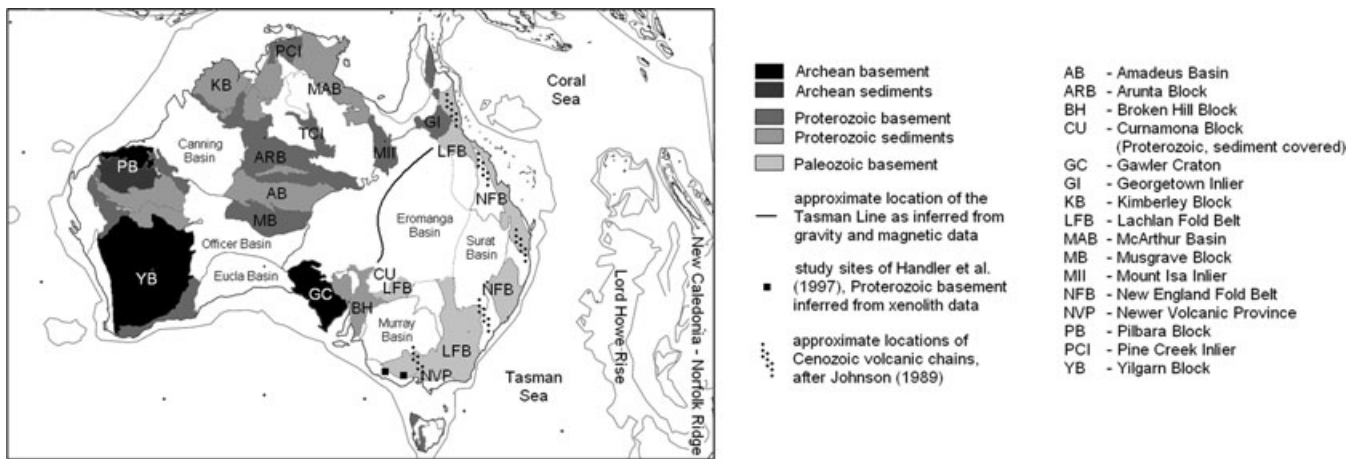
Figure 17. Map of major surface geologic features in the study area. Adapted from Myers *et al.* (1996).

previous surface wave studies (e.g. Goncz & Cleary 1976; Zielhuis & van der Hilst 1996; Fishwick *et al.* 2005). Above 140 km depth, the lower than average velocities under the Coral and Tasman Seas are interrupted by comparatively high *S*-wave speeds along the Lord Howe Rise and the New Caledonia-Norfolk ridge.

Below the Tasman Sea an approximately circular negative anomaly becomes clearly visible at depths greater than 200 km. This anomaly continues to exist as an isolated feature to depths of more than 350 km, where the interpretation of smaller-scale features becomes difficult. The coincidence of this low-velocity anomaly with the present-day location of the Tasmanid hotspot (McDougall & Duncan 1988) suggests a predominantly thermal origin. A similar, though less pronounced and extended feature exists in southeastern Australia slightly north of the Newer Volcanic Province.

Extended high-velocity anomalies are the dominant features below 300 km. The comparatively low resolution at those depths does not allow us to interpret the details of the lateral wave speed variations. Nevertheless, the broad high-velocity regions can be considered robust because they are clearly required to explain the long-period body and higher-mode surface wave data. We hypothesise that the elevated velocities under the Coral Sea and north of New Zealand are caused by the fossil slabs of the Melanesian and Tonga-Kermadec subduction systems (DiCaprio *et al.* 2009).

We now examine the structural elements revealed by the waveform inversion in more detail.

6.2.1 Low-velocity zone beneath the eastern margin of the continent

A prominent feature in the tomographic images is a low-velocity band along the eastern margin of the Australian continent, which is most pronounced above ≈130 km depth where it can be clearly distinguished against the higher velocities of the microcontinental Lord Howe Rise. The lowered seismic wave speeds disappear below 250 km. The lowest velocities, −7 per cent, occur between 20°S and 32°S. Around 250 km depth and 35°S, the low-velocity band connects eastwards to the Newer Volcanic Province in southern Victoria. This connection then appears as an isolated low-velocity anomaly below 300 km depth. In remarkable agreement with Zielhuis & van der Hilst (1996), this band is interrupted around 30°S, 152°E by a region of zero velocity perturbation that develops into a high-velocity anomaly below ≈250 km depth. This location corresponds

to the southern part of the New England Fold Belt, primarily composed of Permian forearcs and accretionary wedges, and it is also characterized by strong negative magnetic and gravity anomalies (e.g. Wellman 1998).

The low-velocity band coincides with a zone of increased seismicity (Leonard 2008) and a variety of other independent geophysical phenomena that suggest a thermal influence and the presence of partial melt: Cenozoic volcanism (Johnson 1989), increased seismic attenuation (Mitchell *et al.* 1998; Abdulah 2007) and steep xenolith palaeogeotherms (Cull *et al.* 1991). It is, however, only in southeastern Australia where the low velocities correlate with an increased electrical conductivity (Lilley *et al.* 1981) and heat flow values around 80 mW m^{-2} that are in excess of the global average for Palaeozoic basement, which is close to 60 mW m^{-2} (Cull 1982; Pollack *et al.* 1993). The absence of a broad heat flow anomaly along much of the eastern continental margin is consistent with the model by Finn *et al.* (2005) who propose that subduction along the East Gondwana margin caused a metasomatical alteration of the subcontinental lithospheric mantle and therefore a decrease of its melting point. This would allow melt generation at relatively low temperatures. The seismic signature may thus be influenced by melting and the presence of volatiles rather than by temperature alone.

6.2.2 Precambrian to Phanerozoic transition (Tasman Line)

The Tasman Line is defined as the boundary between Precambrian and Phanerozoic Australia. Its surface location is constrained by outcrops in northeastern and southeastern Australia. In central eastern Australia there is no general consensus on the nature and location of the Tasman Line. It may coincide with the western limit of the Early Palaeozoic (450–340 Ma) Thomson Orogen as inferred from gravity and magnetic anomaly maps (Wellman 1998) or it may consist of several lineaments associated to various tectonic events (Direen & Crawford 2003). In our tomographic images, the transition from high *S*-wave speeds in western and central Australia to low *S*-wave speeds in eastern Australia generally occurs east of any of the recent Tasman Line definitions, at least above 200 km. This is in agreement with previous results (e.g. Zielhuis & van der Hilst 1996; Debayle & Kennett 2000; Kennett *et al.* 2004; Yoshizawa & Kennett 2004; Fishwick *et al.* 2005, 2008b). The sharpest lateral contrast appears around 200 km depth, where

the east–west velocity anomaly varies from +7 to −7 per cent over a horizontal interval of around 500 km. Below 200 km depth, the contrast is less pronounced, and the low-velocity region extends further west in southern Australia. One possible explanation for the offset between the Tasman Line and the fast-to-slow transition in the uppermost mantle is that Proterozoic basement—characterized by anomalously fast wave speeds—extends further east than indicated by outcrops and magnetic and gravity data that are mostly sensitive to shallow crustal structure. For southeastern Australia, this interpretation is supported by Os isotopic data from xenoliths which require that Proterozoic basement extends up to 400 km east of the Tasman Line (Handler *et al.* 1997). However, in other cratonic areas of the world there is a closer correspondence between the inferred Precambrian/Phanerozoic surface boundary and the high-to-low velocity contrast in the uppermost mantle. Examples are the Baltic Shield (e.g. Zielhuis & Nolet 1994), the Canadian Shield (e.g. Frederiksen *et al.* 2001) or the Kaapvaal craton (e.g. Chevrot & Zhao 2007).

We note that our 3-D starting model (Fig. 7) already contains a smooth transition from fast to slow that is centred east of the surface Tasman Line. Since the transition sharpens during the inversion, we are confident that it is not an artefact.

6.2.3 Archean and Proterozoic cratons

Above 120 km depth the following Proterozoic and Archean structural elements are clearly distinguished by *S*-velocity perturbations in excess of +8 per cent: the Yilgarn and Pilbara Blocks (West Australian craton), the Kimberley block, the Pine Creek inlier, the McArthur basin and the Mount Isa inlier (North Australian craton), the Gawler craton and the sediment-covered Curnamona block (South Australian craton). The velocities in central Australia increase to approximately +7 per cent below 140 km depth. This creates a comparatively homogeneous high-velocity region under central and western Australia where the association of smaller high-velocity patches to surface-geological features is not possible. Below depths around 250 km, the high-velocity perturbations decrease to less than +5 per cent, suggesting that the subcontinental lithospheric mantle, in the seismological sense, is confined to the upper 250 km. Both, the depth extent of the subcontinental lithospheric mantle (≈250 km) and the velocity perturbations found within it (>8 per cent) are consistent with results from the Baltic Shield (e.g. Zielhuis & Nolet 1994), the Canadian Shield (e.g. Frederiksen *et al.* 2001) and the Kaapvaal craton (e.g. Chevrot & Zhao 2007).

The high-velocities below Archean and Proterozoic continents are commonly interpreted as low-temperature regions where the thermally induced negative buoyancy is compensated by basalt-depletion (Jordan 1975, 1978). A compositional contribution to the high S velocities was also suggested by van Gerven *et al.* (2004) who studied relative density-to-shear velocity profiles derived from surface wave tomography and gravity anomalies.

Substantial differences exist between surface wave tomographic images at depths greater than about 250 km (e.g. Zielhuis & van der Hilst 1996; Simons *et al.* 1999; Debayle & Kennett 2000; Yoshizawa & Kennett 2004; Fishwick *et al.* 2005). While, for example, Simons *et al.* (1999) do not observe higher than average velocities in the Kimberley region, clear positive anomalies of the order of 4 per cent appear in the images of Debayle & Kennett (2000), Fishwick *et al.* (2005) and Fishwick & Reading (2008) and in those presented in Fig. 10. There is also no general consensus concerning the exact depth of different Precambrian elements. To some extent, the

lack of agreement at greater depth can be explained by differences in the forward problem solutions and the inversion strategy. The biggest factor is, however, the higher-mode surface and long-period body wave coverage, which is generally poor. An interpretation of smaller-scale structures below 250 km depth should therefore be done with caution until the tomographic images from different research groups converge.

6.2.4 Northward-continuation of the continent

In the tomographic images shown in Figs 10 and 12 we observe an extension of the high velocities of the North Australian craton off-shore under the Arafura shelf. A broad high-velocity region under the Timor and Arafura Seas is present to depths of at least 300 km. This suggests that the North Australian craton has a northward continuation that locally reaches into New Guinea, and that it influences the seismic properties below the subcontinental lithospheric mantle.

The high-velocity region in northern Australia is consistent with observations of refracted body waves by Kaiho & Kennett (2000) who found *S*-velocity variations of around 3 per cent at depths between 260 km and 410 km. Furthermore, a series of body wave studies (Dey *et al.* 1993; Gudmundsson *et al.* 1994; Kennett *et al.* 1994) consistently revealed a strongly attenuative zone ($Q_s \approx 100$) at depths between 200 and 400 km in northern Australia, that is not accounted for in the initial model. Since increased dissipation leads to additional phase delays at longer periods, the high velocities under northern Australia are likely to be underestimated.

6.2.5 Lowered S-wave speeds in the uppermost mantle in central Australia

A consistent feature of recent models for *S*-wave speed beneath Australia is a zone of slightly lowered seismic wave speed in a band across central Australia linking to the zone between the Pilbara and North Australian cratons (e.g. Kaiho & Kennett 2000; Fishwick *et al.* 2005). This result is confirmed with the waveform inversion. By 130 km depth the wave speeds in this region are fast compared with the reference model, which implies a strong gradient in seismic wave speeds that is difficult to reconcile with petrological models. The last tectonic activity in this region was associated with the Alice Springs orogeny ending around 300 Ma. Crustal structure is complex with abrupt changes in thickness associated with pronounced gravity anomalies (Wellman 1998).

7 DISCUSSION

7.1 Forward problem solutions

The correct solution of the equations of motion in realistically heterogeneous earth models is both the most significant advantage and disadvantage of the full waveform tomography method that we have used.

The numerical solution of the wave equation ensures that the differences between observed and synthetic seismograms are indeed the result of yet undiscovered Earth structure and erroneous source parameters. The occurrence of artefacts in the tomographic images that are due to simplifications of the wave propagation process can thus be avoided. The advantages of the full waveform method are particularly important for strongly heterogeneous regions of

the Earth, such as the thermal boundary layers or ocean–continent transitions.

The number of events we have used in the inversion was limited by the available computational resources. However, the comparatively small number of seismograms are in part compensated by extracting as much waveform information as possible, for example, through the application of the TF misfits.

7.2 TF misfits

The TF misfits as defined by Fichtner *et al.* (2008) and reviewed in Section 3.1 have several advantages in the context of full waveform inversion: (1) the separation phase and envelope information, (2) the applicability to any type of seismic wave, (3) a quasi-linear relation to Earth structure and, most importantly and (4) the use of complete waveform information.

There are, however, two disadvantages that the TF misfits share with any other measure of full waveform differences, such as the classical L_2 norm or the generalized seismological data functionals (Gee & Jordan 1992). First, the comparatively high susceptibility to noise, and second the difficulty of assessing noise effects on the misfit measures and the tomographic images. Since the amount of data in full waveform tomography is small, we can reduce the influence of noise only by choosing data of exceptional quality.

The phase misfit is meaningful only when the observed and synthetic waveforms are sufficiently close to avoid phase jumps. This criterion is usually satisfied when the phase differences are less than $\pi/2$. To ensure that no phase jumps occur, useful time windows need to be chosen in each seismogram. We currently pick and weight the time windows manually—a process that is unavoidably subjective, but efficient in the sense that it allows us to incorporate a trained seismologist's experience that can greatly accelerate the convergence of the minimization algorithm. An alternative to manual window selection has recently been proposed by Maggi *et al.* (2009).

7.3 Dependence on the initial model

Tomographic images generally depend on the choice of the initial model, and our full waveform tomographic method is no exception. One can, however, argue that this dependence can be weaker in a non-linear iterative inversion than in a linearized inversion: When the initial model, $\mathbf{m}^{(0)}$, is outside the basin of attraction of the true solution, then both an iterative non-linear and a linearized inversion will produce incorrect results that depend on $\mathbf{m}^{(0)}$. Otherwise, the non-linear iterative inversion closely approaches the true solution irrespective of the actual location of $\mathbf{m}^{(0)}$ within the basin of attraction. It therefore depends, at least in this case, less on the initial model than the linearized inversion that can not approach the true solution arbitrarily closely due to the non-linearity of the problem.

7.4 Resolution

Thanks to the accurate modelling of seismic wave propagation in a 3-D heterogeneous Earth, the tomographic images presented in Section 5 can be considered more realistic than similar images obtained using more restricitve methods. The effect of full wavefield modelling on the resolution of the tomographic images is less certain. Concerning the comparison between ray tomography and finite-frequency tomography, arguments both in favor (e.g. Yoshizawa & Kennett 2004; Montelli *et al.* 2004; Chen *et al.* 2007a) and against (e.g. Sieminski *et al.* 2004; van der Hilst & de Hoop 2005; Trampert & Spetzler 2006) a higher resolution of the latter have been presented.

There are several reasons for the absence of a general consensus on this issue. Most importantly, the resolution increase of finite-frequency and full waveform tomography, if present, is not always visually obvious. This is a subjective but nevertheless powerful impression. An objective comparison of resolution capabilities is difficult because realistic tomographies are computationally expensive, and because there is no universally valid definition of resolution for deterministic non-linear inverse problems. Our limited computational resources merely allow us to consider a small number of synthetic inversions—typically chequer-board tests—that are unlikely to be representative (Lévêque *et al.* 1993). A further complication is due to the fact that ray theoretical sensitivities are distributed along infinitesimally thin rays. The regularization in ray tomography smears the sensitivity into a region around the ray path. The width of this region is mostly chosen as a function of data coverage, and it may therefore be thinner than the actual influence zone of a wave with a finite frequency content. Thus, the apparent resolution in ray tomography may be higher than the resolution that is physically possible. This phenomenon becomes most apparent in the form of the central slice theorem (Cormack 1963) which ensures perfect resolution in the case of sufficient ray coverage.

In the light of those difficulties, we can currently not make a quantitative and objective statement concerning the comparative resolution capabilities of our full waveform tomography. We conjecture, however, that the advantages of our approach will become more apparent as the frequency band broadens and as the amount of exploitable information in individual seismograms increases.

We suggest to make a clear distinction between *realistic* and *well resolved*. The major factors that determine the resolution of a tomographic model are data coverage and data quality. Given a specific data set, the model can be made more realistic by more accurately accounting for the true physics of wave propagation. At this stage of its development, more realistic and physically consistent earth models are the principal advantage of our full waveform tomographic method. This improvement is crucial on our way towards a more quantitative interpretation of tomographic images in terms of geodynamic processes (Bunge & Davies 2001; Schuberth *et al.* 2009a,b).

7.5 Regularization

The regularization in our waveform inversion differs from the regularization in linearized tomographies where it is needed to improve the conditioning of the partial derivative matrix. We most explicitly regularize our inversion through the choice of the inversion grid, the design of the pre-conditioner and the termination of the misfit minimization after a finite number of iterations. The inversion grid—consistent of blocks that are $1° \times 1° \times 10$ km—is intended to prevent the occurrence of smaller-scale features that we would subjectively classify as not being trustworthy. Through the specific design of our pre-conditioner (see Section 3.3.3) we regularize by adding the prior information that heterogeneities are not confined to small volumes around the sources and receivers. A similar effect can be achieved by convolving the Fréchet kernels with a smoothing operator, at the cost of loosing the detailed structure of the kernels. In the case where heterogeneities are indeed confined to the

immediate vicinity of the sources and receivers, the prior information introduced by the pre-conditioner would be incorrect and the algorithm would converge slowly. The regularization introduced by the termination of the inversion after a relatively small number of iterations is currently not quantifiable. This is because we lack information about the behaviour of the optimization algorithm as the number of iterations tends to infinity. We conjecture that a more explicit regularization, that is, through convolutional smoothing, may then become necessary to prevent instabilities.

8 CONCLUSIONS AND OUTLOOK

We have been able to demonstrate the feasibility of 3-D full waveform inversion on continental scales. Our approach rests on the interplay of several components that have specifically been designed for this purpose: a spectral-element solver that combines accuracy, speed and algorithmic simplicity; long-wavelength equivalent crustal models that allow us to increase the numerical grid size; an accuracy-adaptive integration scheme for the regular and adjoint wavefields; TF misfits that extract as much useful information as possible and a rapidly converging conjugate-gradient algorithm with an empirical pre-conditioner. The numerical solution of the elastic wave equation ensures that the tomographic images are free from artefacts that can be introduced by simplifying approximations of seismic wave propagation in strongly heterogeneous media. Our results agree with earlier studies on length scales greater than 500 km. Both long- and short-wavelength structures can be interpreted in terms of regional tectonic processes.

While full waveform tomography is now in principle feasible, there are several important problems that need to be solved in the years to come: Most importantly, we need to develop strategies to better assess the resolution of non-linear tomographic problems that are too expensive to be solved probabilistically. Closely related are questions concerning the still somewhat conjectural higher resolution of full waveform inversion as compared to classical ray tomography. We must furthermore find quantitative estimates of waveform errors in order to evaluate the robustness of the full waveform tomographic images. To ensure that waveform differences are solely due to structure, inversions for potentially finite sources will need to be incorporated into future full waveform tomographies.

We finally wish to note that full waveform inversion is not an automatic procedure, and certainly not a black box that can be applied at will. From our experience with real-data applications, semi-automatic waveform inversions are likely to result in unreasonable earth models, a slow convergence of the optimization scheme or both. A successful application requires a balancing of different types of waveforms based on physical insight, careful inspection of waveform differences after each iteration and intelligent guidance of the inversion scheme. In this sense, the optimization algorithm described in Section 3.3 should not be taken as a dogma from which no deviations through human intervention are allowed.

REFERENCES

Abdulah, A., 2007. Seismic body wave attenuation tomography beneath the Australasian region, *PhD thesis,* The Australian National University.

Alt, W., 2002. *Nichtlineare Optimierung,* Friedr. Vieweg & Sohn Verlagsgesellschaft, Braunschweig, Wiesbaden.

Ammon, C.J., Randall, G.E. & Zandt, G., 1990. On the nonuniqueness of receiver function inversions, *J. geophys. Res.,* **95,** 15 303–15 318.

Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67**(11), 4427–4440.

Bamberger, A., Chavent, G., Hemon, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics,* **47**(5), 757–770.

Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.,* **81,** F897.

Bleibinhaus, F., Hole, J.A., Ryberg, T. & Fuis, G.S., 2007. Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. geophys. Res.,* **112**(B6), B06315, doi:10.1029/2006JB004611.

Bolt, B.A., 1957. Velocity of the seismic waves Lg and Rg across Australia, *Nature,* **180,** 495.

Bolt, B.A. & Niazi, M., 1964. Dispersion of Rayleigh waves across Australia, *Geophys. J. R. astr. Soc.,* **9,** 21–35.

Bolt, B.A., Doyle, H.A. & Sutton, D.J., 1958. Seismic observations from the 1956 atomic explosions in Australia, *Geophys. J. R. astr. Soc.,* **1**(2), 135–145.

Bunge, H.-P. & Davies, J.H., 2001. Tomographic images of a mantle circulation model, *Geophys. Res. Lett.,* **28,** 77–80.

Bunge, H.-P., Hagelberg, C.R. & Travis, B.J., 2003. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.,* **152,** 280–301.

Capdeville, Y. & Marigo, J.-J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media, *Geophys. J. Int.,* **170,** 823–838.

Capdeville, Y. & Marigo, J.-J., 2008. Shallow layer correction for spectral element like methods, *Geophys. J. Int.,* **172,** 1135–1150.

Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.,* **162,** 541–554.

Cerjan, C., Kosloff, D., Kosloff, R. & Reshef, M., 1985. A non reflecting boundary condition for discrete acoustic and elastic wave calculations, *Geophysics,* **50,** 705–708.

Charpentier, I., 2001. Checkpointing schemes for adjoint codes: application to the meteorological model Meso-NH, *SIAM J. Sci. Comp.,* **22,** 2135–2151.

Chen, P., Zhao, L. & Jordan, T.H., 2007a. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97**(4), 1094–1120.

Chen, P., Jordan, T.H. & Zhao, L., 2007b. Full three-dimensional tomography: a comparison between the scattering-integral and the adjoint-wavefield methods, *Geophys. J. Int.,* **170,** 175–181.

Chevrot, S. & van der Hilst, R.D., 2000. The Poisson ratio of the Australian crust: geological and geophysical implications, *Earth planet. Sci. Lett.,* **183,** 121–132.

Chevrot, S. & Zhao, L., 2007. Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton, *Geophys. J. Int.,* **169,** 201–215.

Cleary, J.R., 1967. P times to Australian stations from nuclear explosions, *Bull. seism. Soc. Am.,* **57**(4), 773–781.

Cleary, J.R., Simpson, D.W. & Muirhead, K.J., 1972. Variations in the Australian upper mantle structure, from observations of the Cannikin explosion, *Nature,* **236,** 111–112.

Clitheroe, G., Gudmundsson, O. & Kennett, B.L.N., 2000. The crustal thickness of Australia, *J. geophys. Res.,* **105,** 13 697–13 713.

Collins, C.D. N., Drummond, B.J. & Nicoll, M.G., 2003. Crustal thickness patterns in the Australian continent, *Geol. Soc. Australia Spec. Publ. 22 and Geol. Soc. America Spec. Pap.,* **372,** 121–128.

Cormack, A.M., 1963. Representation of a function by its line integrals, with some radiological applications, *J. appl. Phys.,* **34**(9), 2722–2727.

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic nonlinear waveform inversion—application to real data, *Geophysics,* **55**(5), 527–538.

Crawford, A.J., Meffre, S. & Symonds, P.A., 2003. 120 to 0 Ma tectonic evolution of the southwest Pacific and analogous geological evolution of the 600 to 220 Ma Tasman Fold Belt System, *Geol. Soc. Australia Spec. Publ. 22 and Geol. Soc. America Spec. Pap.,* **372,** 383–403.

Cull, J.P., 1982. An appraisal of Australian heat-flow data, *Bur. Min. Res. J. Aust. Geol. Geophys.,* **7,** 11–21.

Cull, J.P., O'Reilly, S.Y. & Griffin, W.L., 1991. Xenolith geotherms and crustal models in eastern Australia, *Tectonophysics,* **192,** 359–366.

Dahlen, F.A., Hung, S.H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes, I: theory, *Geophys. J. Int.,* **141**(1), 157–174.

Debayle, E. & Kennett, B.L.N., 2000. The Australian continental upper mantle: structure and deformation inferred from surface waves, *J. geophys. Res.,* **105,** 25 423–25 450.

de Jersey, N.J., 1946. Seismological evidence bearing on the crustal thickness in the South-West Pacific, University of Queensland, Brisbane, Paper v. 3, no. 2.

Dey, S.C., Kennett, B.L.N., Bowman, J.R. & Goody, A., 1993. Variations in upper mantle structure under northern Australia, *Geophys. J. Int.,* **114,** 304–310.

DiCaprio, L., Müller, R.D. & Gurnis, M., 2009. A dynamic process for drowning carbonate reefs on the northeastern Australian margin, *Geology,* in press.

Direen, N.G. & Crawford, A.J., 2003. The Tasman Line: where is it, and is it Australia's Rodinian breakup boundary? *Aust. J. Earth Sci.,* **50,** 491–502.

Dumbser, M. & Käser, M., 2006. An arbitrary high order discontinuous Galerkin method for elastic waves on unstructured meshes. II: the three-dimensional isotropic case, *Geophys. J. Int.,* **167**(1), 319–336.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Int.,* **25,** 297–356.

Faccioli, E., Maggio, F., Paolucci, R. & Quarteroni, A., 1997. 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method, *J. Seismol.,* **1**(3), 237–251.

Ferreira, A.M.G. & Woodhouse, J.H., 2006. Long-period seismic source inversions using global tomographic models, *Geophys. J. Int.,* **166,** 1178–1192.

Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimisation of discrete crustal models—a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.,* **173**(2), 519–533.

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology—I. Theory, *Phys. Earth planet. Int.,* **157,** 86–104.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental and global scale full waveform inversion in the time-frequency domain, *Geophys. J. Int.,* **175**(2), 665–685.

Finn, C.A., Müller, R.D. & Panter, K.S., 2005. A Cenozoic diffuse alkaline magmatic province (DAMP) in the southwest Pacific without rift or plume origin, *Geochem. Geophys. Geosyst.,* **6,** Q02005, doi:10.1029/2004GC000723.

Fishwick, S. & Reading, A.M., 2008. Anomalous lithosphere beneath the Proterozoic of western and central Australia: a record of continental collision and intraplate deformation? *Precambrian Res.,* **166,** 111–121.

Fishwick, S., Kennett, B.L.N. & Reading, A.M., 2005. Contrasts in lithospheric structure within the Australian craton—insights from surface wave tomography, *Earth planet. Sci. Lett.,* **231,** 163–176.

Fishwick, S., Heintz, M., Kennett, B.L.N., Reading, A.M. & Yoshizawa, K., 2008. Steps in lithospheric thickness within eastern Australia, evidence from surface wave tomography, *Tectonics,* **27**(4), TC0049, doi:10.129/2007TC002116.

Fletcher, R. & Reeves, C.M., 1964. Function minimisation by conjugate gradients, *Comput. J.,* **7,** 149–154.

Frederiksen, A.W., Bostock, M.G. & Cassidy, J.F., 2001. S-wave velocity structure of the Canadian upper mantle, *Phys. Earth planet. Int.,* **124,** 175–191.

Friederich, W., 1999. Propagation of seismic shear and surface waves in a laterally heterogeneous mantle by multiple forward scattering, *Geophys. J. Int.,* **136,** 180–204.

Friederich, W., 2003. The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms, *Geophys. J. Int.,* **153**(1), 88–102.

Friederich, W. & Dalkolmo, J., 1995. Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of the Green's function in the frequency domain, *Geophys. J. Int.,* **122,** 537–550.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51**(7), 1387–1403.

Gee, L.S. & Jordan, T.H., 1992. Generalised seismological data functionals, *Geophys. J. Int.,* **111,** 363–390.

Geller, R.J. & Hara, T., 1993. Two efficient algorithms for iterative linearized inversion of seismic waveform data, *Geophys. J. Int.,* **115,** 699–710.

Goncz, J.H. & Cleary, J.R., 1976. Variations in the structure of the upper mantle beneath Australia from Rayleigh wave observations, *Geophys. J. R. astr. Soc.,* **44,** 507–516.

Griewank, A. & Walther, A., 2000. Revolve: an implementation of checkpointing for the reverse of adjoint mode of computational differentiation, *Trans. Math. Software,* **26,** 19–45.

Gudmundsson, O., Kennett, B.L.N. & Goody, A., 1994. Broadband observations of upper-mantle seismic phases in northern Australia and the attenuation structure in the upper mantle, *Phys. Earth planet. Int.,* **84,** 207–226.

Handler, M., Bennett, V.C. & Esat, T.M., 1997. The persistence of off-cratonic lithospheric mantle: Os isotopic systematics of variably metasomatised southeast Australian xenoliths, *Earth planet. Sci. Lett.,* **151,** 61–75.

Igel, H., Djikpéssé, H. & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.,* **124,** 363–371.

Igel, H., Nissen-Meyer, T. & Jahnke, G., 2002. Wave propagation in 3D spherical sections: effects of subduction zones, *Phys. Earth planet. Int.,* **132,** 219–234.

Johnson, R.W., 1989. *Intraplate volcanism in Eastern Australia and New Zealand,* Cambridge University Press, New York.

Jordan, T.H., 1975. The continental tectosphere, *Rev. Geophys. Space Phys.,* **13**(3), 1–12.

Jordan, T.H., 1978. Composition and development of the continental tectosphere, *Nature,* **274,** 544–548.

Kaiho, Y. & Kennett, B.L.N., 2000. Three-dimensional structure beneath the Australasian region from refracted wave observations, *Geophys. J. Int.,* **142,** 651–668.

Kennett, B.L.N., Gudmundsson, O. & Tong, C., 1994. The upper-mantle S-velocity and P-velocity structure beneath northern Australia from broadband observations, *Phys. Earth planet. Int.,* **86,** 85–98.

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.,* **122,** 108–124.

Kennett, B.L.N., Fishwick, S., Reading, A.M. & Rawlinson, N., 2004. Contrasts in mantle structure beneath Australia: relation to Tasman Lines? *Aust. J. Earth Sci.,* **51,** 563–569.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., 1983. Optimization by Simulated Annealing. *Science,* **220,** 671–680.

Klingelhoefer, F., Lafoy, Y., Collot, J., Cosquer, E., Géli, L., Nouzé, H. & Vially, R., 2007. Crustal structure of the basin and ridge system west of New Caledonia (southwest Pacific) from wide-angle and reflection seismic data, *J. geophys. Res.,* **112,** doi:10.1029/2007JB005093.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.,* **139,** 806–822.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.,* **149**(2), 390–412.

Lambeck, K., Burgess, G. & Shaw, R.D., 1988. Teleseismic travel-time anomalies and deep crustal structure in central Australia, *Geophys. J. Int.,* **94,** 105–124.

Leonard, M., 2008. One hundred years of earthquake recording in Australia, *Bull. seism. Soc. Am.,* **98**(3), 1458–1470.

Lévêque, J.-J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic images, *Geophys. J. Int.,* **115,** 313–318.

Lilley, F.E. M., Woods, D.V. & Sloane, M.N., 1981. Electrical conductivity profiles and implications for the absence or presence of partial melting beneath central and southeast Australia, *Phys. Earth planet. Int.,* **25,** 419–428.

Lions, J.L., 1968. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles,* Dunod Gauthier-Villars.

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

Luo, Y. & Schuster, G.T., 1991. Wave equation traveltime inversion, *Geophysics,* **56**(5), 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.

McDougall, I. & Duncan, R.A., 1988. Age progressive volcanism in the Tasmanid Seamounts, *Earth planet. Sci. Lett.,* **89,** 207–220.

Mitchell, B.J., Baqer, S., Akinci, A. & Cong, L., 1998. Lg Coda Q in Australia and its relation to crustal structure and evolution, *Pure appl. Geophys.,* **153,** 639–653.

Kristek, J. & Moczo, P., 2003. Seismic-wave propagation in viscoelastic media with material discontinuities: a 3D fourth-order staggered-grid finite-difference modeling, *Bull. seism. Soc. Am.,* **93**(5), 2273–2280.

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.

Müller, R.D., Gaina, C., Tikku, A., Mihut, D., Cande, S.C. & Stock, J.M., 2000. Mesozoic/Cenozoic tectonic events around Australia, in *The History and Dynamics of Global Plate Motions, Geophysical Monograph,* **121,** 161–188.

Myers, J.S., Shaw, R.D. & Tyler, I.M., 1996. Tectonic evolution of Proterozoic Australia, *Tectonics,* **15,** 1431–1446.

Oeser, J., Bunge, H.-P. & Mohr, M., 2006. Cluster design in the Earth sciences: Tethys, *High Perform. Comput. Commun., Proc.,* **4208,** 31–40.

Pollack, H.N., Hurter, S.J. & Johnson, J.R., 1993. Heat flow from the Earth's interior: analysis of the global data set, *Rev. Geophys.,* **31**(3), 267–280.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

Pratt, R.G. & Shipp, R.M., 1999. Seismic waveform inversion in the frequency domain. Part 2: fault delineation in sediments using crosshole data, *Geophysics,* **64**(3), 902–914.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Quarteroni, A., Sacco, R. & Saleri, F., 2002. *Numerische Mathematik 1,* Springer-Verlag, Berlin, Heidelberg.

Schuberth, B.S.A., Bunge, H.-P., Steinle-Neumann, G., Moder, C. & Oeser, J., 2009a. Thermal versus elastic heterogeneity in high-resolution mantle

circulation models with pyrolite composition: high plume excess temperatures in the lowermost mantle, *Geochem. Geophys. Geosys.,* **10,** Q01W01, doi:10.1029/2008GC002235.

Schuberth, B.S.A., Bunge, H.-P. & Ritsema, J., 2009b. Tomographic filtering of high-resolution mantle circulation models: can seismic heterogeneity be explained by temperature alone? *Geochem. Geophys. Geosys.,* **10,** Q05W03, doi:10.1029/2009GC002401.

Shibutani, T., Sambridge, M. & Kennett, B.L.N., 1996. Genetic algorithm inversion for receiver functions with application to crust and uppermost mantle structure beneath Eastern Australia, *Geophys. Res. Lett.,* **23**(14), 1829–1832.

Sieminski, A., Lévêque, J.-J. & Debayle, E., 2004. Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.,* **31**(24), L24614.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007a. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **168**(3), 1153–1174.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007b. Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **171**(1), 368–389.

Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.,* doi:10.1038/ngeo231.

Simons, F.J., Zielhuis, A. & van der Hilst, R.D., 1999. The deep structure of the Australian continent from surface wave tomography, *Lithos,* **48,** 17–43.

Simons, F.J., van der Hilst, R.D., Montagner, J.-P. & Zielhuis, A., 2002. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophys. J. Int.,* **151,** 738–754.

Stich, D., Danecek, P., Morelli, A. & Tromp, J., 2009. Imaging lateral heterogeneity in the northern Appennines from time reversal of reflected surface waves, *Geophys. J. Int.,* **177,** 543–554.

Takeuchi, H. & Saito, M., 1972. Seismic surface waves, in *Methods in Computational Physics,* Vol. 11, pp. 217–295, ed. Bolt, B.A., Academic Press, New York.

Takeuchi, N. & Geller, R.J., 2000. Optimally accurate second order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media, *Phys. Earth planet. Int.,* **119,** 99–131.

Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory, *Q. J. R. Met. Soc.,* **113,** 1311–1328.

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

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the Southern California crust, *Sci,* **325,** 988–992.

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.,* **128,** 365–399.

Teixeira, F.L. & Chew, W.C., 1997. Systematic derivation of anisotropic pml absorbing media in cylindrical and spherical coordinates, *IEEE Microw. Guided Wave Lett.,* **7**(11), 371–373.

Tibuleac, I.M., Nolet, G., Michaelson, C. & Koulakov, I., 2003. P wave amplitudes in a 3-D Earth, *Geophys. J. Int.,* **155,** 1–10.

Trampert, J. & Spetzler, J., 2006. Surface wave tomography: finite-frequency effects lost in the null space, *Geophys. J. Int.,* **164**(2), 394–400.

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

van der Hilst, R.D., Kennett, B.L.N., Christie, D. & Grant, J., 1994. Project SKIPPY explores the lithosphere and mantle beneath Australia, *EOS, Trans. Am. geophys. Un.,* **75,** 177–181.

van der Hilst, R.D. & de Hoop, M.V., 2005. Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.,* **163**(3), 956–961.

van Gerven, L., Deschamps, F. & van der Hilst, R.D., 2004. Geophysical evidence for chemical variations in the Australian continental mantle, *Geophys. Res. Lett.,* **31,** L17607, doi:10.1029/2004GL020307.

Wellman, P., 1998. Mapping of geophysical domains in the Australian continental crust using gravity and magnetic anomalies, in *Structure and*

Evloution of the Australian Continent, AGU Geodynamics Series, **26,** 59–71.

Yomogida, K. & Aki, K., 1987. Amplitude and phase data inversions for phase velocity anomalies in the Pacific Ocean basin, *Geophys. J. R. astr. Soc.,* **88,** 161–204.

Yomogida, K., 1992. Fresnel zone inversion for lateral heterogeneities in the Earth, *Pure appl. Geophys.,* **138**(3), 391–406.

Yoshizawa, K. & Kennett, B.L.N., 2004. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects, *J. geophys. Res.,* **109,** doi:10.1029/2002JB002254.

Yoshizawa, K. & Kennett, B.L.N., 2005. Sensitivity kernels for finite-frequency surface waves, *Geophys. J. Int.,* **162,** 910–926.

Zielhuis, A. & Nolet, G., 1994. Shear-wave velocity variations in the upper mantle beneath central Europe, *Geophys. J. Int.,* **117,** 695–715.

Zielhuis, A. & van der Hilst, R.D., 1996. Upper-mantle shear velocity beneath eastern Australia from inversion of waveforms from SKIPPY portable arrays, *Geophys. J. Int.,* **127,** 1–16.

Zheng, Y. & Huang, X., 1997. Anisotropic perfectly matched layers for elastic waves in cartesian and curvilinear coordinates, MIT Earth Resources Laboratory, Consortium Report.