Tectonophysics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Review Article

Seismic imaging: From classical to adjoint tomography

Q. Liu [a,*], Y.J. Gu [b]

[a] Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7
[b] Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2E1

ARTICLE INFO

Article history:
Received 6 December 2011
Received in revised form 3 July 2012
Accepted 11 July 2012
Available online xxxx

Keywords:
Seismic tomography
Sensitivity kernels
Inverse problem
Adjoint methods
Adjoint tomography
Full waveform inversion

ABSTRACT

Seismic tomography has been a vital tool in probing the Earth's internal structure and enhancing our knowledge of dynamical processes in the Earth's crust and mantle. While various tomographic techniques differ in data types utilized (e.g., body vs. surface waves), data sensitivity (ray vs. finite-frequency approximations), and choices of model parameterization and regularization, most global mantle tomographic models agree well at long wavelengths, owing to the presence and typical dimensions of cold subducted oceanic lithospheres and hot, ascending mantle plumes (e.g., in central Pacific and Africa). Structures at relatively small length scales remain controversial, though, as will be discussed in this paper, they are becoming increasingly resolvable with the fast expanding global and regional seismic networks and improved forward modeling and inversion techniques.

This review paper aims to provide an overview of classical tomography methods, key debates pertaining to the resolution of mantle tomographic models, as well as to highlight recent theoretical and computational advances in forward-modeling methods that spearheaded the developments in accurate computation of sensitivity kernels and adjoint tomography. The first part of the paper is devoted to traditional traveltime and waveform tomography. While these approaches established a firm foundation for global and regional seismic tomography, data coverage and the use of approximate sensitivity kernels remained as key limiting factors in the resolution of the targeted structures. In comparison to classical tomography, adjoint tomography takes advantage of full 3D numerical simulations in forward modeling and, in many ways, revolutionizes the seismic imaging of heterogeneous structures with strong velocity contrasts. For this reason, this review provides details of the implementation, resolution and potential challenges of adjoint tomography. Further discussions of techniques that are presently popular in seismic array analysis, such as noise correlation functions, receiver functions, inverse scattering imaging, and the adaptation of adjoint tomography to these different datasets highlight the promising future of seismic tomography.

Contents

* Corresponding author. Tel.: +1 416 978 5434.
E-mail address: liuqy@physics.utoronto.ca (Q. Liu).

0040-1951/$ – see front matter.
http://dx.doi.org/10.1016/j.tecto.2012.07.006

1. Introduction

For good reasons, seismic tomography has been closely linked with computed tomography (CT) commonly performed on X-ray or ultrasound recordings. The similarities are undeniable, as both techniques aim to extract the internal properties of an object based on integrated data trajectories. The former method is predicated on seismic waves propagating through crustal or mantle rocks, and is therefore influenced by, and sensitive to, the medium elastic parameters. The latter approach mainly seeks a numerical description of tissue density as a function of position. When performed under ideal conditions, both approaches are able to reconstruct accurate two- or three-dimensional (3D) images of targeting structures through planar slices (or *tomos* in Greek connotation). The similarity goes beyond the obvious: in fact, the governing principles of seismic and computed, mainly radiological (Cormack, 1963), tomography are all rooted in the Abel transform of linear paths in a radially concentric sphere (Herglotz, 1907; Wiechert, 1910). Mathematically this approach was identical to the construction of a spherically symmetric quantum mechanical potential in 3D (Keller et al., 1956; Snieder and Trampert, 1999). A more general form of linear tomography was presented as the central slice theorem by Radon (1917), which was later shown to reduce to Abel transform for Earth-like geometry (Deans, 1983; Nowack, 1990; Vest, 1974). While the mathematics introduced by Radon (1917) are beyond the scope of this review, the simple concept of integrating and inverting a projection function between two generic points connected by a straight line (or *path*) was instrumental not only for seismic and radiological tomography, but also for modern-day data de-noising and reconstruction (Deans, 1983; Gu, 2010; Vest, 1974).

It is unclear whether the subsequent developments of CT and seismic tomography were contemporaneous or that the former inspired the early pioneering efforts of seismic data analysis. In 1971, the first practical CT brain scanner by Sir Hounsfield in England opened a new chapter in medical radiology. By recording X-ray transmission through patients, medical tomographers began to extract images of the X-ray attenuation coefficients via what was once termed *EMI scans* (Cunningham and Judy, 2000; Filler, 2009). This breakthrough pre-dated the first 2D seismic tomographic regional map produced by Aki and Lee (1976), though it could be legitimately argued that the fundamentals of geophysical inverse problems were already in place when Backus and Gilbert (1967, 1968, 1970) published their seminal studies on the formation, solution and non-uniqueness of linearized inversions (Wiggins, 1972). Rather than quantifying the inverse problem in abstract mathematical terms, Backus and Gilbert (1967) provided formulation and practical solutions to the inversion of normal mode observations based on Rayleigh's Principle (Rayleigh, 1877). The findings of this and subsequent studies by the same authors championed a new generation of studies pertaining to the Earth's free oscillations (Dziewonski, 1971; Dziewonski and Anderson, 1981; Gilbert and Dziewonski, 1975). Just as importantly, functional derivatives of seismic measurement with respect to structural model parameters (Backus and Gilbert, 1967; Fréchet, 1941; Woodhouse, 1974), commonly known as *Fréchet derivatives*, *Fréchet kernels* or *sensitivity kernels*, made their debut in geophysics and would soon become the backbone of subsequent studies of traveltime (Aki and Lee, 1976; Woodhouse, 1974, 1981) and waveform (Dziewonski and Anderson, 1981; Woodhouse and Dziewonski, 1984) tomography.

The era of *global* seismic tomography officially began in 1981 with the publication of Preliminary Reference Earth Model (PREM, Dziewonski and Anderson, 1981), the first well-accepted one dimensional (1D), radially anisotropic model of the Earth. This study combines normal mode, traveltimes, and attenuation observations with physical parameters such as Earth's mass and moment of inertia to establish an accurate reference frame, with typical errors less than 1% between predicted and observed traveltimes for teleseismic arrivals, for advanced 3D surveys. This seminal study eventually earned the two original authors well-deserved Crawford Award, the Nobel prize equivalence for all of geosciences. Over the next 10 years, a series of pioneering studies in linear (Dziewonski, 1984; Nakanishi and Anderson, 1982; Woodhouse and Dziewonski, 1984) and non-linear (Ho-Liu et al., 1988, 1989; Snieder and Romanowicz, 1988; Tarantola and Valette, 1982) inversions followed suit, combining mathematical elegance with artistry to propel seismic tomography to the forefront of Earth sciences. The need for greater computing power and memory storage also became apparent. For instance, to enable inversions for up to degree 6 spherical harmonics, the original Fortran 77 codes for key parts of the waveform analysis had to resort to the extended use

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

of 'equivalence' statement to circumvent computer memory limitations (Woodhouse and Dziewonski, 1984). It is fair to state that, after four decades from Backus and Gilbert's first studies, the pursuit of higher resolution and accuracy of the tomographic images continues to push computer software and hardware capacities to their allowable limits.

Increasing seismic data combined with improvements in measurement accuracy, model parameterization, and forward modeling theory have resulted in a number of published global tomographic models from different geophysics programs (e.g., Ekström and Dziewonski, 1998; Ekström et al., 1995; Gu et al., 2001; Houser et al., 2008; Kárason and van der Hilst, 2000; Kustowski et al., 2008; Masters et al., 2000; Montelli et al., 2004; Panning and Romanowicz, 2006). Captivating images of the mantle, which at times invoke inevitable comparisons with objects of little scientific connection (Fig. 1), are often considered compelling *observations* by researchers from various disciplines. Steady efforts were also made to go beyond simple ray theory and compute sensitivities of broadband seismic phases by mode coupling (e.g., Li and Romanowicz, 1995), body-wave ray theory (e.g., Dahlen et al., 2000), surface-wave ray theory (e.g., Zhou et al., 2004), normal-mode summation (e.g., Capdeville, 2005; Zhao and Jordan, 1998) and full 3D numerical simulations (e.g., Liu and Tromp, 2008). The improved Fréchet derivatives (Backus and Gilbert, 1967; Woodhouse, 1974) translates to more accurate design matrices for the tomographic inverse problem and, at least in principle, should improve the quality of the inverted images of the Earth's interior. However, different dataset, processing algorithms, sensitivity kernels and regularization manifested in global tomographic models that agree only in relatively long-wavelength scales (e.g., Dziewonski, 2005; Romanowicz, 2003, also see Section 3). Despite considerable efforts in quantifying and reconciling the model differences (e.g., Becker and Boschi, 2002; Becker et al., 2003; Fukao et al., 2001; Gu et al., 2003; Romanowicz, 2003), global tomography seems to have reached an impasse in resolution, which impedes its capability of being directly linked to tectonic and geodynamical interpretations beyond long wavelengths. The impact of sensitivity kernels to the resolution of global seismic tomography was highlighted by a contentious debate surrounding *banana-doughnut* sensitivity kernels (Dahlen et al., 2000) and its applications (e.g., Dahlen and Nolet, 2005; Montelli et al., 2004, 2006a; van der Hilst and de Hoop, 2005).

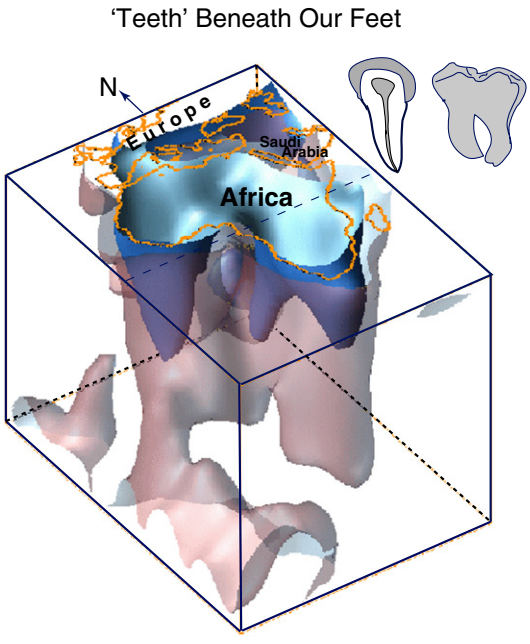

Fig. 1. Isotropic seismic velocity perturbations beneath Africa and surrounding regions down to 250-km depth. The model values are computed from *S20U7L5* (Ekström et al., 1995) and the iso-surfaces highlight shear velocity contours of 1% (exterior) and 2% (inner surface, blue). The continental roots suggested by the 2% shear velocity perturbations display striking similarities to human teeth. The tomographic rendering was courtesy of Misha Salganik.

Parallel to the development of global seismic tomography, structural inversions have long been formulated in the exploration context as *full-waveform inversion* problems, where the calculation of Fréchet derivatives of an objective function involves two numerical simulations of wave propagation. This formulation was applied initially to exploration seismic problems due to the heavy computational cost associated with numerical simulations two decades ago (e.g., Gauthier et al., 1986; Igel et al., 1996). However, with the advancement of numerical modeling techniques and improvement of computational resources, simulations of seismic wave propagation for complicated and realistic 3D heterogeneous media at both global and regional scales became practical in recent years (e.g., Castro and Käser, 2010; Igel, 1996; Komatitsch and Tromp, 2002a; Moczo et al., 2007; Tromp et al., 2010a). This naturally led to interest in calculating finite-frequency sensitivity kernels based on more accurate forward modeling methods (Boschi et al., 2007; Liu and Tromp, 2006, 2008; Zhao et al., 2005), and its ensuing applications to tomographic problems (Chen et al., 2007a; Fichtner et al., 2009b; Tape et al., 2010). This type of tomographic inversion method (which we refer to as *adjoint tomography* in this paper) takes advantage of accuracy 3D numerical simulations, and through iterations can resolve velocity perturbations up to 10–30%, far beyond those achievable by classical tomographic techniques. However, adjoint tomography is still in its infancy, where computational cost reduction (Akcelik et al., 2003; Tape et al., 2010), objective assessment of model resolution (Fichtner and Trampert, 2011a, 2011b), and applications to array data (e.g., via noise correlation and scattered wave field analysis) remain areas of ongoing research. Still, adjoint tomography holds promise for a new-generate high-resolution images of the Earth at local, regional and global scales.

While a chronological account of both exploration, crustal and mantle tomographic techniques and results may be tempting, a truly objective presentation of all relevant research is impractical due to space limitation. In the first part of this paper, we aim to revisit methods that had been the backbone of global seismic tomography (up to ~2003); key formulations are provided for completeness and easy comparison. We then discuss some of the recent debates associated with the resolution of tomographic images (up to ~2006), and carefully review the formulation and applications of adjoint tomography. Further information on broad subject of seismic tomography can be found in a number of informative reviews (e.g., Dziewonski, 1996; Dziewonski and Woodhouse, 1987, Fukao et al., 2001, 2009; Grand, 2002; Masters, 1989; Masters and Shearer, 1995; Masters et al., 2000; Montagner, 1994; Rawlinson et al., 2010; Ritzwoller and Lavely, 1995; Romanowicz, 1991, 2003; Trampert and van der Hilst, 2005; Virieux and Operto, 2009; Woodhouse and Dziewonski, 1989).

2. The tomographic problem and classical solutions

2.1. Inverse problem

According to Joseph Keller, a prominent mathematician and physicist, inverse and forward problems are tightly coupled where the 'formulation of each involves all or part of the solution of the other'. In other words, an inverse problem is only defined if the answer to 'the inverse of what?' is provided through an explicit expression for, and a solution of, the forward problem (Engl et al., 2000). For a forward operator $\mathbf{G}$ that maps model parameters $\mathbf{m}$ to observations $\mathbf{d}$ based on selected rules or principles of physics,

$$\boldsymbol{G}(\boldsymbol{m}) = \boldsymbol{d}. \tag{1}$$

The problem of recovering suitable parameters from a set of observations is the associated inverse problem (e.g., Aster et al., 2005; Iyer

and Hirahara, 1993; Menke, 1984; Nolet, 1993; Tarantola, 2005; Wiggins, 1972). The forward-inverse relationship is exemplified by integration and differentiation, two rudimentary mathematical operations that are critical in the inversion approach of Backus and Gilbert (1967, 1968, 1970). Simple solution of this inverse problem involves the following key steps (Aster et al., 2005; Menke, 1984; Snieder, 1998):

1. Writing the problem based on a set of discrete model coefficients.
2. Computing the predicted data based on the choice of model parameters for an a priori structure, the majority being known 1D model structures.
3. Defining an objective function and adjusting the model parameters to meet the pre-defined goodness-of-fit criteria.
4. Estimating the accuracy and resolution of the inversion outcome, repeating the above steps when necessary.

The first two steps govern the forward problem, which takes advantages of laws of physics and makes informed predictions of the observation. The last two steps aim to recover and refine the physical parameters, if properly defined, by minimizing the differences between observations and predictions while attaining certain desired properties in the model outcomes (Fig. 2; Aster et al., 2005; Menke, 1984; Tarantola, 2005). This process could be repeated for continued refinement of the model and objective function.

With a suitable choice of model parameters, Eq. (1) can often be linearized as a straight-forward multiplication of **G**, known as the *design matrix*, and a vector containing the unknowns (see Section 2.2). This process is often aided by a sound initial solution and perturbation theory, the foundation of Backus and Gilbert (1967) in the investigation of rock density and elastic moduli based on Rayleigh's principle (Pekeris and Jarosch, 1958; Rayleigh, 1877, 1906; Woodhouse, 1976). On the other hand, *existence*, *ill-posedness*, and *uniqueness* (Backus and Gilbert, 1967) must be critically assessed for any inversion, especially in view of limited observations and the nonlinearity of inverse problems pertaining to Earth's internal structure. The section below revisits key algorithms in classical tomography, especially linear inversions, in the order of increasing sophistication.

2.2. Traveltime inversions

Ray-based traveltime tomography has been the staple of seismic imaging largely due to its simplicity. The forward problem involving the computation of the traveltime t of a ray can be expressed as

$$t = \int_{\mathbf{x}_s}^{\mathbf{x}_r} \mathbf{s}\cdot\mathbf{dl}, \tag{2}$$

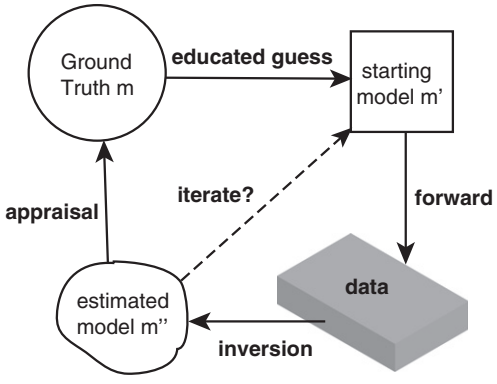

Fig. 2. A flow chart showing the key steps in the construction and evaluation of tomographic models. This figure is inspired by Fig. 2 of Snieder (1998).

where **dl** is a small segment of the ray path, **s** is the slowness vector tangential to the ray path with a magnitude equal to the inverse of the local seismic wave speed (Chernov, 1960), and $\mathbf{x}_s$ and $\mathbf{x}_r$ are the spatial coordinates of the source and receiver, respectively. Based on Snell's law (Descartes and Laurence, 1631; Khan and Lantz, 2002; Snellius, 1621), one could trace a given ray between the two end points and solve Eq. (2) as a boundary value problem either by *shooting* (Acton, 1997; Keller, 1968) or *bending* (Julian and Gubbins, 1977) methods; the latter approach was adopted due to faster convergence despite more complex formulations at discontinuous boundaries within the Earth.

In this formulation, the perturbation of arrival time relative to the unperturbed model prediction t is directly attributed to local perturbations in slowness. The main objective of the inverse problem is to determine the fractional change of velocity for all solvable segments along the ray path (Menke, 1984; Tarantola and Valette, 1982; Woodhouse, 1978).

2.3. Linearization and parameterization

Basis functions with desirable properties (e.g., orthogonality) are generally adopted to expand slowness perturbation as a set of linear equations with discrete unknown weights (as *model coefficients*). Early tomographic studies took advantage of a back-projection approach based on gridded *blocks* (e.g., Aki and Lee, 1976; Chou and Brooker, 1979; Hearn and Clayton, 1986). This parameterization was influential for global traveltime inversions in the 1990s using large catalogs of *P* wave traveltimes (e.g., Grand et al., 1997; van der Hilst et al., 1997) and, owing to its simplicity, remains a popular choice in regional and exploration seismic applications. The formulation is a natural extension of the discrete form of traveltime

$$\delta t_i = \sum_{j=1}^{N_p} \delta s_j l_{ij}, \tag{3}$$

where δt_i is the traveltime difference between observation and reference model prediction for ray i , δs_j is the slowness perturbation of the j-th block in the medium, and l_{ij} is the length of the i-th ray in the j-th block. The summation can be performed over the entire study region, though only a small number of blocks (N_p) for a given source–station pair are sampled and contribute to the traveltime residual (Aki and Lee, 1976; Ho-Liu et al., 1989). The objective of the inverse problem is to determine δs_j for the majority, if not all, of the blocks through inversions.

While uniform grids based on rectangular cells or blocks are an obvious choice due to the simplistic assumption of piece-wise straight-line path segments, the shapes and distribution of the blocks vary broadly in published studies. In fact, to circumvent the lack of sensitivity of surface wave 'great-circle' paths (Backus, 1964) to local heterogeneities, early tomographic inversions adopted prior regionalization schemes and partitioned the Earth's surface into a few irregular, continent-sized blocks (Dziewonski and Steim, 1982; Kanamori, 1970). Problems associated with a priori regionalization (Kawakatsu, 1983) were soon recognized and partially resolved by the path-average approximation and a global expansion based on spherical harmonics (Woodhouse and Dziewonski, 1984, see also next section).

The desirable properties of orthogonality and smoothness made spherical harmonics, which was first proposed as solutions to the Laplace's equation, a popular choice of horizontal basis functions in early global tomographic studies of traveltime and/or waveform data (Fig. 3a). The expansion takes the following form,

$$\frac{\delta v(r,\theta,\phi)}{v(r)} = \sum_{k=0}^{K}\sum_{l=0}^{L}\sum_{m=0}^{l} f_k(r) p_l^m(\cos\theta)\left[{}_kA_l^m \cos(m\phi) + {}_kB_l^m \sin(m\phi)\right], \tag{4}$$

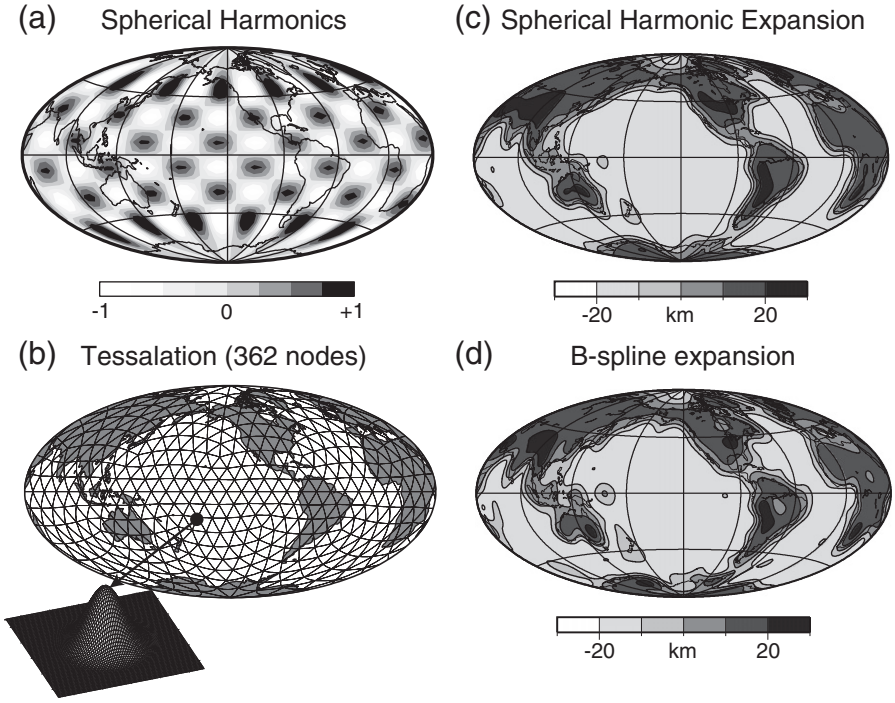

Fig. 3. (a) Normalized spherical harmonic function (Press et al., 2007) with angular order $l=18$ and azimuthal order $m=6$. The total number of model coefficients (i.e., *sin* and *cos* terms) up to degree 18 is 361. (b) Equal-area tessellation of the Earth's surface. The total number of vertices is 362. (c) Degree-18 spherical harmonic expansion of crustal thickness (CRUST 2.0, Bassin et al., 2000). (d) Expansion of crustal thickness using spherical B-splines (Gu et al., 2001) centered at each vertex in panel (b). With similar numbers of model coefficients, the results of the inversion-based expansions are highly consistent between the two parameterizations.

where θ is colatitude (90°-latitude), ϕ is longitude, r is radius, and $v(r)$ is the depth dependent velocity of the 1D reference Earth model. The term p_l^m is the associated Legendre polynomial of degree l and azimuthal order m, and $f_k(r)$ corresponds to a radial basis function that captures 1D variations of wave speeds along the depth axis (Dziewonski, 1984; Su et al., 1992; Woodhouse and Dziewonski, 1984). Finally, the constant multipliers ${}_kA_l^m$ and ${}_kB_l^m$ are the target model coefficients in the 3D expansion of seismic velocities. The above formulation deviates slightly from the original study by Woodhouse and Dziewonski (1984), in which a similar expression was detailed for squared velocity perturbations using Legendre polynomial up to the third power and spherical harmonics up to degree 8. In either form, the first two terms of the Taylor series of traveltimes establishes a linear dependence between perturbations in time (observation) and velocity (target of investigation) relative to a reference model such as PREM (Dziewonski and Anderson, 1981) or IASPEI91 (Kennett and Engdahl, 1991). In matrix form, a model vector $\mathbf{m}$ containing unknowns ${}_kA_l^m$ and ${}_kB_l^m$ in Eq. (4) could be estimated from the following set of linear equations

$$\mathbf{Gm} = \mathbf{d}, \tag{5}$$

where $\mathbf{G}$ contains assumed path sensitivity and $\mathbf{d}$ represents the set of traveltime or waveform observations. Frequently cited examples of tomographic models based on this approach are *S12WM13* (Su et al., 1992), *S20A* (Ekström and Dziewonski, 1998), *S20RTS* (Ritsema and van Heijst, 2000) and, more recently, *S40RTS* that contain more than 1600 horizontal model parameters (Ritsema et al., 2011).

While parameterization and nominal resolution (Kissling et al., 2001) often depend on a researcher's personal preference, judicious choices should reflect the merits of basis functions as well as data coverage and density. Unlike well-designed exploration seismic experiments where receiver distributions are dense and uniform, the global tomographic problem is under-determined due to uneven earthquake and station distributions. For this reason, variable resolutions or irregular global grids (see review by Sambridge and Rawlinson, 2005) have often been adopted to provide a closer association between model and data spaces. Widely accepted but inherently subjective criteria for model evaluation, such as smoothness, also play key roles in the selection of model parameters. For instance, spline functions (Gu et al., 2001, 2003; Kustowski et al., 2008; Mégnin and Romanowicz, 2000; Nettles and Dziewonski, 2008) have been widely adopted as an effective compromise between localization and smoothness. Still, robust outcomes from well constrained inversion problems should not be strongly influenced by the choice of model parameters. When data coverage is sufficient, the global tomographic problem is inherently similar to a simple interpolation problem. Fig. 3 compares an inversion of global Moho depth of CRUST2.0 (Bassin et al., 2000) based on different basis functions. For similar numbers of model parameters (361 and 362), the outcomes of 2D expansions of crustal thickness based on B-splines and spherical harmonics are virtually identical. The 'same earth' assumption will be further assessed through comparisons of published 3D models in Section 3.

2.4. Waveform inversions

2.4.1. Path average approximation (PAVA)

Due to the lack of consensus among published studies, we subjectively characterize waveform tomography as an inversion that 1) uses time series rather than traveltimes, amplitudes or other secondary attributes of the recorded data, and 2) depends on the approximation of full wave equation rather than ray approximation (Panning et al., 2009; Romanowicz, 2003). Hence, waveform inversions are more accurate than its traveltime counterpart despite more complex formulations and increased nonlinearity from waveform fitting.

Woodhouse and Dziewonski (1984) marks a milestone effort in early stages of global waveform tomography. Having recognized the merits of surface wave analysis (e.g., Brune et al., 1963; Knopoff, 1972) and its largely untapped potential on the global scale (Masters

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

et al., 1982), this study abandoned controversial a priori regionalization schemes (Dziewonski, 1971; Jordan, 1981; Kanamori, 1970; Kawakatsu, 1983; Souriau and Souriau, 1983; Toksoz and Anderson, 1966) and formally introduced the *path integral approximation* (or *path average approximation*, for short PAVA, see also Mosca and Trampert, 2009; Panza, 1985; Woodhouse and Dziewonski, 1986) in search of lateral heterogeneity in the mantle. This approximation, which is an effective WKBJ assumption (Bretherton, 1968; Woodhouse, 1974), essentially states that phase perturbations of surface waves are caused by lateral heterogeneities horizontally averaged over the source–receiver trajectory. Despite inaccurate assumptions of path sensitivities of body wave, triplications, and frequency-dependent surface waves (Li and Romanowicz, 1996; Snieder, 1996; Snieder and Nolet, 1987) along the great-circle arc (Backus, 1964), Woodhouse and Dziewonski (1984) is worthy of a brief review for its pioneering efforts in combining fundamental theories of linear inversion and free oscillations with data mining through global surface and body wave observations.

Adopting the original notation of Jordan (1978), Woodhouse and Dziewonski (1984) expresses total phase perturbations between the source and receiver after n orbits as

$$\delta\phi = \int_0^{T_x} \delta\omega_{\text{local}}\, dt + n\int_0^{T} \delta\omega_{\text{local}}\, dt, \tag{6}$$

where $\delta\omega_{\text{local}}$ denotes the local mantle wave frequency shift in response to lateral heterogeneities, and the integral is performed over the group traveltime along the source-station great circle path T_x for either a major or minor arc. The second term is a cumulative sum of the phase shift after n passages of the surface wave around the great-circle path defined by the source and receiver. One can define a fictitious frequency shift as

$$\delta\omega = \frac{1}{T}\int_0^{T_x} \delta\omega_{\text{local}}\, dt = \text{path average}. \tag{7}$$

The essence of PAVA is to rewrite the above phase perturbation in terms of this apparent frequency term and source–receiver distance perturbation $\delta\theta$ (see Eq. (11) of Woodhouse and Dziewonski, 1984). The apparent frequency and distance shifts are different for each mode, and their partial derivatives with respect to model parameters are used to compute the Fréchet kernels for the waveform inversion problem.

Based on PREM (Dziewonski and Anderson, 1981), synthetic seismograms can be computed by normal mode superposition (Gilbert, 1970; Gilbert and Dziewonski, 1975; Romanowicz, 2008; Tanimoto, 1984) and PAVA facilitates the formulation of a linear inverse problem of model perturbations. It was the building block for a series of global models of mantle heterogeneities (e.g., Ekström and Dziewonski, 1998; Gu et al., 2001, 2003, 2005; Kustowski et al., 2008; Liu and Dziewonski, 1998; Nettles and Dziewonski, 2008; Ritsema and van Heijst, 2000; Su et al., 1992, 1994). In theory, PAVA is sensitive to both the phase and the amplitude information of surface wave waveforms. In practice, however, perturbations of waveforms in connection with the reduction in data variance is dominated by phase shifts (e.g., Li and Romanowicz, 1996; Woodhouse and Dziewonski, 1984); consequently, waveform inversions based on PAVA falls short in properly accounting for the amplitudes of surface waves (e.g., Dahlen and Tromp, 1998; Li and Romanowicz, 1995; Panning et al., 2009; Zhao and Dahlen, 1995).

2.4.2. PAVA in classical traveltime tomography

The application of PAVA goes far beyond waveform tomography. While it has been suggested earlier (see review by Romanowicz, 2003), a recent study by Mosca and Trampert (2009) explicitly validated the notion of ray theory as the body-wave equivalence of PAVA to perturbations in modal frequencies. Assume ray bottom radius r_b and a depth-dependent velocity $v(r)$, the ray theoretical traveltime $t(p)$ of a seismic phase (Brune, 1966; Woodhouse, 1978) becomes,

$$t(p) = 2\int_0^{r_b} \left[\frac{1}{v^2(r)} - \frac{p^2}{r^2}\right]^{1/2} dr, \tag{8}$$

where r is the instantaneous position of a ray segment from the center of the Earth in kilometers and p represents the spherical ray parameter. The key is to recognize the inherent connection between the horizontal ray parameter p and the traveltime equivalence of the frequency–phase velocity relationship (Dahlen and Tromp, 1998; Mosca and Trampert, 2009)

$$\frac{\delta\omega_{\text{local}}}{\omega} = \frac{\Delta(p)/T(p)}{c}\frac{\delta c_{\text{local}}}{c}, \tag{9}$$

where $\Delta(p)$ and $T(p)$ correspond to ray-parameter dependent distance and traveltime, respectively, and phase velocity c is defined by the simple relationship $c = p^{-1}$. Eq. (9) is the traveltime equivalence of Eq. (6), and the total traveltime perturbation becomes an integral of local perturbations in the exact form of Eq. (7) (Mosca and Trampert, 2009),

$$\delta T(p) = \frac{1}{\Delta}\int_0^{\Delta} \delta t_{\text{local}}(p,\theta,\phi)d\Delta. \tag{10}$$

In essence, the equivalence between waveform PAVA and ray theory is the manifestation of mode-ray duality under the zeroth order asymptotic assumption (Mosca and Trampert, 2009; Zhao and Dahlen, 1995). For this reason, the application of PAVA is far more prevalent in seismic tomography than that suggested by waveform inversions.

2.4.3. Improved waveform approaches

PAVA detailed by Woodhouse and Dziewonski (1984) was initially applied to low-pass filtered (cutoff period 135 s) long-period *mantle waves*, named in consideration of their broad depth sensitivity to mantle structure, and body waves (cutoff period 45 s). Low computation cost was a crucial advantage for waveform tomography based on PAVA in 1980s due to the assumed dependence of synthetic seismogram on model perturbations averaged along source–receiver path. However, with vastly improved computer memory and processor speed in the early 1990s, tomographers began to place a greater emphasis on the accuracy of the inversion outcomes instead of on the savings on computational resources. Even for relatively 'smooth' variations, a condition that PAVA is based upon, there are well-documented caveats. First, in a laterally heterogeneous mantle, PAVA is inaccurate for body-wave phases with geometrically banana-shaped sensitivities (Dahlen et al., 2000). More accurate kernels were developed by considering mode-branch coupling (Li and Romanowicz, 1995; Marquering and Snieder, 1995), which reduces finite-frequency sensitivity to planar, 2-D ray paths. In the same year, Zhao and Dahlen (1995) derived the asymptotic expressions for the Fréchet kernels of modal eigenfunctions in the heterogeneous mantle. Differential waveforms (Passier and Snieder, 1995) have also been suggested as a partial solution to improve the accuracy of PAVA.

Inaccurate body wave path assumptions are only a part of the problem associated with PAVA. Even with improved geometrical path integrals, PAVA would remain a gross assumption due to the lack of consideration for finite frequency effects of both body (Nolet et al., 2005) and surface (Zhou et al., 2004) waves. This problem is highlighted by the use of 'wiry' ray theoretical kernel in traveltimes, the waveform equivalent of PAVA, and requires further discussions (see Section 3.2).

2.4.4. Partitioned waveform inversion and multi-stage inversion schemes

Based on Fréchet kernels, the inversion approaches outlined in the preceding sections aim to resolve perturbations in seismic velocities

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

in reference to a starting 1D (or known 3D) model directly from traveltime, phase, or waveform observations. While Woodhouse and Dziewonski (1984) suggested that a direct inversion for the coefficients is more reliable for their long period Love and Rayleigh wave datasets this one-step inversion process is, for most datasets, not a requirement but a choice. In fact, the earliest tomographic applications (especially in medical imaging based the central slice theorem of Radon, 1917) were predicated on two-step processes involving the construction of 3D volumes from a series of 2D sections. Multi-stage inversion approaches have also been adopted to recover mantle velocity structures, exemplified by the Partitioned Waveform Inversion (PWI) (Nolet, 1990) method.

The PWI method introduced by Nolet (1990, 2008) consists of two key steps: 1) fit each waveform observation with an average 1D velocity along the source–receiver great-circle path, and 2) solve for 3D structure via linearized inversions and the collection of path-averaged misfits from step 1 (e.g., Das and Nolet, 1998; Maggi and Priestley, 2005; Marquering et al., 1996; Nolet, 1990; van der Lee and Nolet, 1997). Step 1 is conceptually important and potentially computationally intensive due to nonlinear optimization (Nolet, 1990). Unlike Eq. (7), this study recasts the path integral to more confined model domain and defines the path-average shear velocity perturbation as

$$\delta v(r)^{(j)} = \frac{1}{\Delta_j}\int_{\text{source}}^{\text{receiver}} \left[v(r,\theta,\phi) - v^{(j)}\right] d\Delta, \tag{11}$$

where $v^{(j)}(r)$ represents the radially symmetric reference speed, $v(r,\theta,\phi)$ corresponds to 3D shear velocity variations along a given path j with a path length of Δ_j. Then the path-averaged perturbations to the wavenumber $\delta k_n(\omega)^{(j)}$ of a surface wave mode at frequency ω can be expressed as

$$\delta k_n(\omega)^{(j)} = \int_0^a \left[\frac{\partial k_n^{(j)}(\omega)}{\partial v^{(j)}(r)}\right] \delta v^{(j)}(r) dr, \tag{12}$$

where a is the radius of the Earth. The spectrum of the wavefield $S_j(\omega)$ at this station becomes an exponential function of the path integral (Nolet, 1990) and, based on the WKBJ approximation, the velocity term can be discretized and recovered from the waveform misfit. To construct 3D variations in shear velocities $\delta v(r,\theta,\phi)$ from model coefficients, Nolet (1990) defines a new data vector $\mathbf{d}$ based on the 1D model coefficients and links the 3D velocity perturbation $v(r,\theta,\phi)$ to partitioned 1D velocity model $v^{(j)}(r)$ for path j,

$$\frac{1}{\Delta_j}\int_0^{\pi}\int_{\text{path j}} \left[v(r,\theta,\phi) - v^{(j)}(r)\right] g_i^{-1}(r) dr d\Delta = d_i \pm \delta d_i, \tag{13}$$

where Δ_j corresponds to the source–receiver distance for the j-th path. This equation is in the form of $\mathbf{Gm} = \mathbf{d}$ (see Eq. (5)) with $g_i^{-1}(r)$ as the sensitivity kernel and $v(r,\theta,\phi)$ as the unknown model vector. Unlike step 1, the 3D velocity perturbations are solved directly via a linearized inversion approach (Nolet, 1990).

A key advantage of PWI over PAVA is flexibility, since multiple starting models can be simultaneously introduced during the 1D waveform fitting procedure. During the past 20 years, a number of continental-scale studies have benefited from PWI in the investigation of crust (Das and Nolet, 1998; Manaman et al., 2011) and mantle structure (Maggi and Priestley, 2005; van der Lee and Frederiksen, 2005; van der Lee and Nolet, 1997). Similar to PAVA (Woodhouse and Dziewonski, 1984), however, the original form of PWI (Nolet, 1990) approximated body-wave path sensitivity using the WKBJ assumption in connection with surface wave modes. While this inaccuracy was largely resolved by incorporating mode coupling (e.g., Marquering and Snieder, 1995; Marquering et al., 1996), and lack of considerations for wave amplitudes (Panning et al., 2009) and finite frequency effects (see Section 3.2) remains problematic for PWI.

Two-step inversions of an entirely different flavor have also been well established in surface wave analysis. These studies first seek solutions to surface wave phase velocity (Ekström and Dziewonski, 1998; Montagner and Nataf, 1986; Nakanishi and Anderson, 1984; Nataf et al., 1984; Ritzwoller et al., 2002; Tanimoto and Anderson, 1985; Trampert and Woodhouse, 1996; Yang and Forsyth, 2006), group velocity (e.g., Bensen et al., 2007; Ekström and Dziewonski, 1998; Yao et al., 2006) or arrival angle information (Larson et al., 1998; Laske and Masters, 1996) for multiple frequency ranges, then construct the 3D structures by inverting the resulting dispersion curves or phase/group velocity maps (e.g., Ekström and Dziewonski, 1998; Shapiro and Ritzwoller, 2002). Stutzmann and Montagner (1993) proposed a two-step waveform inversion scheme for surface wave fundamental and higher modes that incorporated amplitude information. Rather than utilizing full waveforms, as was in the case for PAVA or PWI, maps obtained by these studies generally emphasize a specific aspect or aspects of surface wave records.

2.5. Inversions for anisotropic structure

Since the early 1960s, the dependence of seismic wave speed variations on particle motion or direction of propagation (generally known as *seismic anisotropy*) has been documented through the discrepant observations of P_n (Hess, 1964; Raitt et al., 1971) and surface (Aki, 1968; Forsyth, 1975; Nataf et al., 1984; Schlue and Knopoff, 1977; Yu and Mitchell, 1979) waves, as well as anomalous particle motions associated with body waves (Anderson, 1966; Ando et al., 1980; Hirn, 1977). Tradeoffs between heterogeneity and anisotropy have long been recognized as a potential factor affecting the outcomes of tomographic models (e.g., Dziewonski and Anderson, 1981; Smith and Dahlen, 1973; Woodhouse and Dziewonski, 1984), but the isotropic assumption was frequently made for mathematical and computational conveniences (Anderson, 1989). For instance, inversions of general anisotropic bodies require the recovery of 21 independent elastic parameters according to Hooke's Law. Consider a coarsely sample mantle with spherical harmonic expansion up to degree 20 (horizontal) and Chebyshev polynomial up to degree 13 (radial) (e.g., Ekström and Dziewonski, 1998), a back-of-envelope calculation suggests a total number of unknowns to be $(20+1)^2\times(13+1)\times 21 \approx 130{,}000$. This implies that a non-gradient factorization method such as Cholesky decomposition (Ekström and Dziewonski, 1998; Gu et al., 2001; Su et al., 1994) would require the storage and manipulation of two upper-triangular inner product matrix with $130000\times 130001/2 \approx 8.5\times 10^9$ elements each, or with a total memory requirement of ~700 GB in single precision. Even though today's high-end supercomputers could meet this baseline memory requirement, a fully anisotropic inversion remains impractical due the extensive computation time for matrix inverse, the inherent tradeoffs between the daunting number of independent parameters and, more importantly, the insufficient data constraints.

Despite challenges in modeling general anisotropy, tomographic inversions targeting anisotropic wave speeds garnered considerable attention in the early 1980s (Dziewonski and Anderson, 1981; Nataf et al., 1984; Woodhouse, 1981) and became more practical in the mid 1990s following the increased affordability and accessibility of supercomputers. Simplifying considerations were introduced based on the approximate symmetry of anisotropic minerals (Anderson, 1961; Backus, 1962; Carlson and Christensen, 1979; Christensen and Crosson, 1968; Dziewonski and Anderson, 1981; Postma, 1955). By assuming hexagonal symmetry, general anisotropy reduces to transverse isotropy (or radial anisotropy) (e.g., Anderson and Dziewonski, 1982) with five independent elastic parameters A, C, N, L, and F (Love, 1927). The first four elastic constants can be determined by velocity measured in two orthogonal directions and F is a function of velocity at intermediate directions (Dziewonski and Anderson, 1981). In addition to mathematical convenience, transverse isotropy offered

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

appealing physical descriptions of laminated or finely layered solids, oriented cracks or melt zones, as well as preferentially oriented mantle minerals such as olivine and pyroxene under shear deformation (Anderson, 1989; Nishimura and Forsyth, 1989; Silver, 1996).

Linearized anisotropic inversion methods pertaining to transverse isotropic solids were introduced by Dziewonski and Anderson (1981) and Woodhouse (1981) for waveform and traveltime data, respectively. Based on the theory initially developed by Takeuchi and Saito (1966) and the definition of Fréchet derivatives (e.g., Backus and Gilbert, 1967), a relative change in the squared eigenfrequency of a normal mode can be written as (Anderson and Dziewonski, 1982; Dziewonski and Anderson, 1981)

$$\frac{\delta\omega^2}{\omega^2} = \int_0^1 r^2 dr\left(\tilde{A}\delta A + \tilde{C}\delta C + \tilde{N}\delta N + \tilde{L}\delta L + \tilde{F}\delta F + \tilde{R}\delta\rho\right), \tag{14}$$

where the symbols with tilde signs represent the differential kernels (see Takeuchi and Saito, 1966 for alternative formulations) expressible as functions of the eigenfunctions (or the derivatives of eigenfunctions) and the angular degree of spherical harmonics (Dziewonski and Anderson, 1981). For practical reasons, this expression was recast into Fréchet derivatives of more measurable quantities such as horizontal and vertical wave speeds ($V_{\chi H}$ where 'χ' can be either P or S), and the last term corresponds to perturbations to the density of the anisotropic structure. Simple scaling relationships between P, S and density have been adopted (Ekström and Dziewonski, 1998; Gu et al., 2005; Kustowski et al., 2008; Marone et al., 2007; Nettles and Dziewonski, 2008; Yuan and Romanowicz, 2010) to reduce the number of free parameters in the inversion process. Anisotropic parameters have also been retrieved from independent inversions of Rayleigh and Love wave datasets (e.g., Ekström and Dziewonski, 1998; Nishimura and Forsyth, 1989), though the assumed polarized kernels are inaccurate due to the dependence of higher-mode surface waves on both vertical and horizontal wave speeds (see below).

Anisotropic inversion of body-wave traveltime observations were introduced by Woodhouse (1981) as an addendum to Dziewonski and Anderson (1981). For simplicity we briefly outline the key results on traveltime integral of *SH* waves and refer the readers to the original study by Woodhouse (1981) for more complex formulations for *P–SV* waves. Assume q_r is the integrand of a one-way traveltime path integral (see Eq. (8)), then asymptotic ray theory enables the computation of partial derivatives of *SH*-wave traveltime with respect to all five elastic parameters (Woodhouse, 1981) for a radially anisotropic solid. To obtain explicit expressions to recover anisotropic seismic velocities from traveltime observations, one first needs to map the Fréchet derivatives of q_r with respect to the five elastic constants to horizontal and vertical velocities (Woodhouse, 1981) using the chain rule, then insert the appropriate terms into the path integral over radius,

$$\delta t = 2\int_0^{r_b} \frac{\partial q_r}{\partial V_{SH}}\delta V_{SH}\, dr + 2\int_0^{r_b} \frac{\partial q_r}{\partial V_{SV}}\delta V_{SV}\, dr, \tag{15}$$

where δV_{SH} and δV_{SV} are the unknown perturbations in horizontal and vertical shear wave velocities, respectively, along the ray path. For proper choices of basis functions (see Sections 2.2 and 2.3), this formulation is convenient for a linear, anisotropic inversion of the model coefficients using traveltime observations.

Fig. 4 shows examples of traveltime and normal-mode frequency sensitivities to anisotropic shear velocities. The ray-theoretical traveltime of *P1* (an *SH*-polarized first arrival, Fig. 4a) is highly sensitive to wave polarization and propagation directions (Fig. 4b; Gu et al., 2005). The

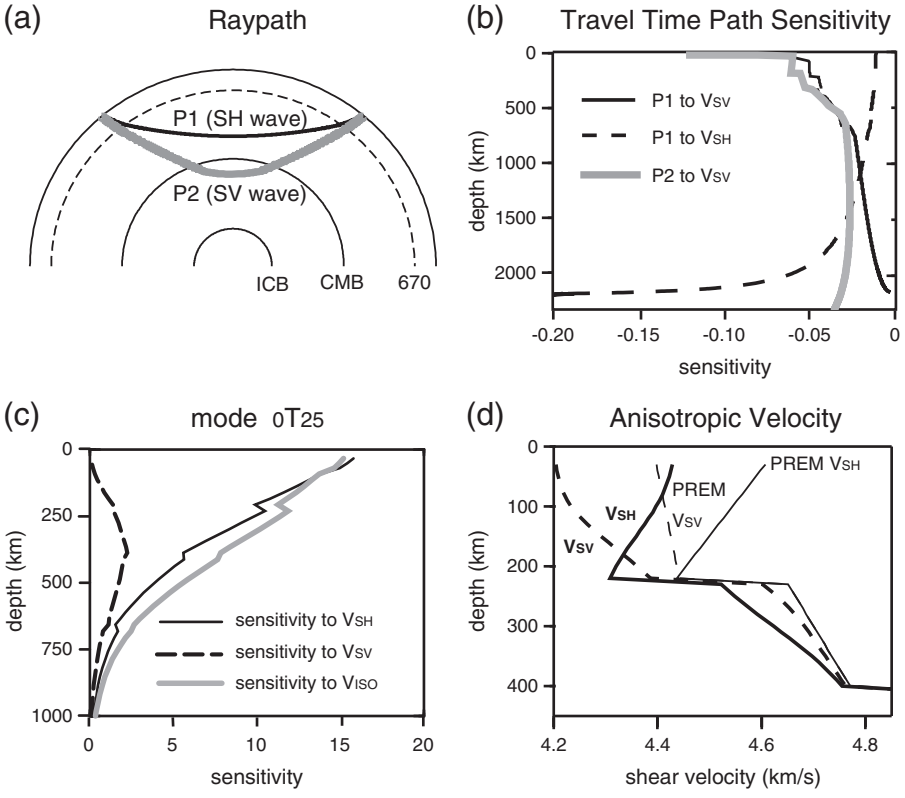

Fig. 4. (a) Ray paths of *S* (*SH*-polarized) and *SKS* (*SV*-polarized). (b) Traveltime depth sensitivities of the two phases shown in graph (a) to horizontal (V_{SH}) and vertical (V_{SV}) wave speeds. The sensitivity of *P2* is cut off at 2500 km. (c) Sensitivity of a toroidal normal mode frequency to horizontal and vertical wave speeds. The isotropic kernels (ISO) are computed using the Voigt average (see Ekström and Dziewonski, 1998; Gu et al., 2005). (d) Averaged 1D anisotropic velocities near the East Pacific Rise inverted by Gu et al. (2005), relative to PREM (Dziewonski and Anderson, 1981). Inverted 1D V_{SH} and V_{SV} structures are denoted by the thick solid and dashed lines, respectively, and the corresponding values of PREM V_{SH} and V_{SV} are represented by the thin solid and dashed lines, respectively. Significant anisotropy exists in the lithosphere and potentially extends to the top of the transition zone.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

consideration for anisotropy is vital near the source and receiver positions where the wave vectors are approximately parallel to the symmetry axis, or near the turning points where an *SH*-polarized wave is only sensitive to the horizontal velocity. On the other hand, *SV*-polarized waves are only sensitive to vertical shear speed, but the sensitivity varies as a function of propagation direction (see Fig. 4b). The dependence of mode frequencies on the anisotropic wave speeds are more complex than the traveltime counterpart, and contributions of both V_{SH} and V_{SV} are nonzero for toroidal (Fig. 4c) as well as spheroidal (not shown) modes. Fig. 4d shows the averaged horizontal and vertical shear wave speeds under the East Pacific Rise (Gu et al., 2005), retrieved using both traveltime and waveform information. The substantial differences between the inverted anisotropic wave speeds, their departures from PREM (Dziewonski and Anderson, 1981; Tanimoto, 1984) and the large depth extent (e.g., Beghein and Trampert, 2004; Chen and Brudzinski, 2003; Gu et al., 2005; Jung and Karato, 2001; Karato, 1998; Leveque et al., 1998; Panning and Romanowicz, 2006) clearly accentuate the need to properly account for seismic anisotropy in seismic tomography.

Finally, tomography based on the radially isotropic assumption generally considers minimal lateral variations in wave speeds in planes perpendicular to the symmetry axis. This is a gross approximation considering the abundance of evidence, dating back to early observations of P_n wave traveltimes (Hess, 1964; Raitt et al., 1971) and shear-wave birefringence (Long and Silver, 2009). Smith and Dahlen (1973) introduced the following formulation to characterize the azimuthal dependence of wave speeds with respect to a fixed direction (also known as azimuthal anisotropy),

$$\frac{\delta c(\mathbf{r},\phi)}{c} = a_0(\mathbf{r}) + a_1(\mathbf{r})\cos 2\phi + a_2(\mathbf{r})\sin 2\phi + a_3(\mathbf{r})\cos 4\phi + a_4(\mathbf{r})\sin 4\phi, \tag{16}$$

where $\mathbf{r}$ is a 2D positional vector, ϕ is the relative azimuth of the ray path and a_i $(i=1,\cdots,4)$ represent the unknown coefficients of the major (2ϕ) and secondary (4ϕ) terms (e.g., Tanimoto, 1984; Tanimoto and Anderson, 1985). This is a convenient linear equation where coefficients are readily invertible based on observed phase velocity perturbations $\delta c/c$. Similar to radial anisotropy, the strength and nature of azimuthal anisotropy vary broadly depending on the tectonic history and depth extent (Becker et al., 2003; Larson et al., 1998; Marone and Romanowicz, 2007; Montagner and Nataf, 1986; Montagner and Tanimoto, 1990; Park and Levin, 2002; Simons and van der Hilst, 2003; Tanimoto, 1984; Thybo and Perchuc, 1997; Trampert and van Heijst, 2002; Trampert and Woodhouse, 2003; Wookey et al., 2005; Yuan and Romanowicz, 2010). The broad range of outcomes from these studies were highly instrumental for the understanding of mineralogy and deformation below surface. At the present time, however, determination and interpretation of seismic anisotropy remain challenging, especially in view of its tradeoff with isotropic heterogeneity in anisotropic inversions and the apparent lack of depth sensitivity from *S*/*SKS* splitting observations.

2.6. Inversions with other geophysical or geochemical constraints

The accuracy of tomographic imaging is a strong function of data density and data sensitivity to the targeted structure. Despite a fast growing number of seismic stations worldwide and increasing diversity in seismic data types, high-resolution global seismic tomography remains an under-determined inverse problem (Dziewonski, 2005; Romanowicz, 2003) and could certainly benefit from additional data constraints on the geometry, density and composition of mantle thermochemical variations (see Section 3.1 for detailed discussions). One such constraint is the existence and undulation of known and postulated seismic interfaces (e.g., crust, upper mantle discontinuities, D″). Due to strong temperature dependence, mineral phase changes (Fig. 5a) of olivine to wadsleyite and ringwoodite to perovskite and magnesiowüstite (Ita and Stixrude, 1992; Katsura and Ito, 1989, see Helffrich, 2000; Shearer, 2000 for reviews) can occur at depths tens of kilometers away from their respective global averages (Deuss, 2009; Flanagan et al., 1998; Gu et al., 1998; Shearer, 1993). These steep depth variations (Fig. 5b) will impact the traveltimes and trajectories of direct and reflected/converted seismic body waves. Hence, rather than relying on the conventional treatment to discontinuity depths (i.e., assumption of flat interfaces; Fig. 5c), a growing number of studies are beginning to resolve the tradeoff between seismic wave speed and discontinuity depth (Fig. 5d) via simultaneous inversions (Chang et al., 2010; Gu et al., 2003; Houser et al., 2008; Lawrence and Shearer, 2006).

A joint inversion approach represents a significant improvement over a priori corrections and assumptions of flat interfaces. Mathematically, the added constraints and new model parameters can be incorporated into the inversion procedure via

$$\begin{bmatrix}\mathbf{G}_0\\ \mathbf{G}_1\end{bmatrix}\delta\mathbf{v} = \begin{bmatrix}\mathbf{d}_0\\ \mathbf{d}_1\end{bmatrix}, \tag{17}$$

where $\boldsymbol{d}_0$ and $\boldsymbol{d}_1$ denote conventional and newly added data, respectively, and matrix on the left hand side of the above equation contains the combined data sensitivity. In this simple example, model vector $\delta\mathbf{v}$ is the seismic velocity perturbation, which is modified by the additional weighted equations of constraints during the regularization and minimization of data errors. Useful constraints include, but are not limited to, the depths of Moho (Chang et al., 2010; Xu and Song, 2010), core–mantle boundary (Boschi and Dziewonski, 2000; Morelli and Dziewonski, 1987) and transition zone discontinuities (Gu et al., 2003; Houser et al., 2008; Lawrence and Shearer, 2006), as well as those associated with magnetotelluric (Moorkamp et al., 2007, 2010; Roux et al., 2011), geodynamical (Forte, 2000; Forte and Mitrovica, 2001; Forte et al., 1994; Simmons et al., 2006, 2007; Trampert et al., 2004) and geodetic (Chambat and Valette, 2001; Khan et al., 2008) observations.

Data integration, which has been a driving force behind some of the recent tomographic efforts (e.g., Chang et al., 2010; Yuan and Romanowicz, 2010), is vital for the pursuit of a self-consistent mantle model (see Section 3). The accuracy and interpretation of these models could benefit further from improved imaging of anelasticity in crust and mantle rocks (e.g., Bhattacharyya et al., 1996; Dalton et al., 2008; Durek et al., 1993; Flanagan and Wiens, 1994; Gung and Romanowicz, 2004; Ho-Liu et al., 1988; Romanowicz, 1990, 1994, 1995; Roth and Wiens, 2000; Suda et al., 1991), an important observational constraint (see Romanowicz, 1998; Mitchell and Romanowicz, 1998 and references therein) that received unfair coverage in this review due to space limitation. The compatibility of the different data sets and the relative weights of vastly different datasets, however, remain critical issues to be resolved by more quantitative investigations.

2.7. Global optimization methods and probabilistic tomography

With the exception of PWI, methods documented in this section emphasize the retrieval of a single set of model parameters through the process of minimization. This deterministic approach is mathematically ill-conditioned due to noise and finite path coverage (Romanowicz, 2003; Snieder and Trampert, 1999), thus requiring regularization (e.g., Menke, 1984; Parker, 1994; Tarantola and Valette, 1982), and a single optimal solution for highly nonlinear problems such as waveform matching (Gauthier et al., 1986; Mosegaard, 1998; Mosegaard and Tarantola, 2002; Sambridge et al., 1991; Shaw, 1992) has also been questioned at the conceptual level (Mosegaard and Tarantola, 2002; Resovsky and Trampert, 2003; Scales and Snieder, 1997, 2000; Tarantola, 2005). A key alternative to this inversion paradigm is a

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

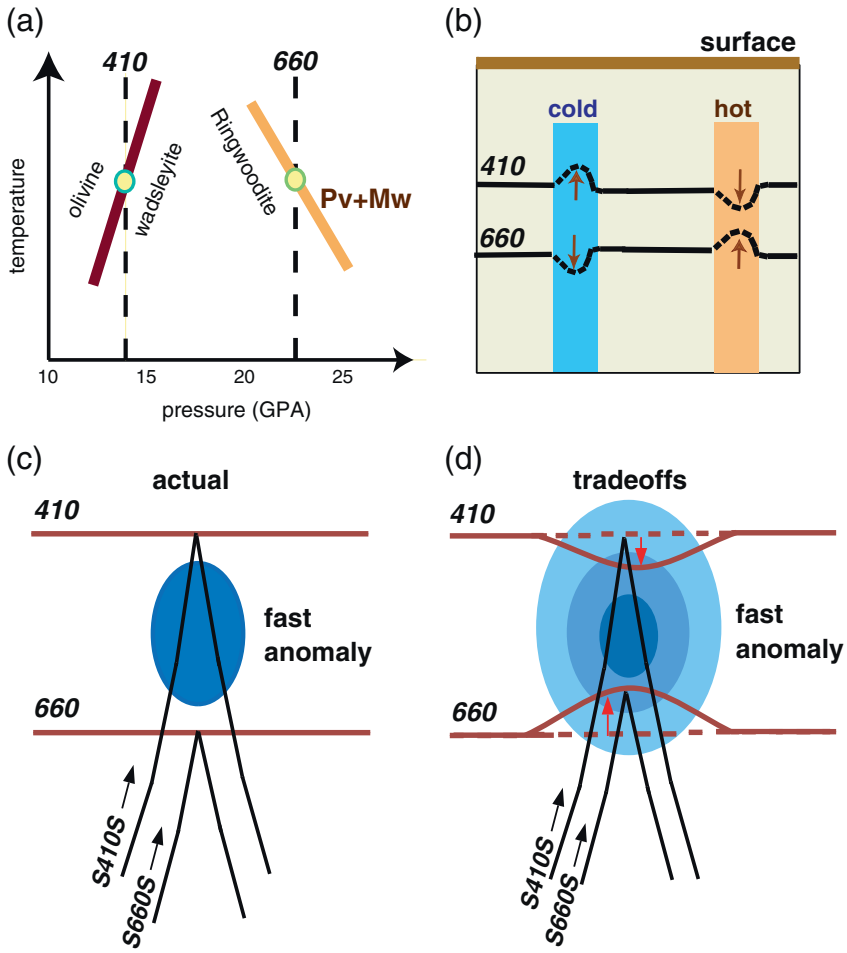

Fig. 5. (a) Phase diagrams of olivine phase transitions. The Clapeyron slopes of olivine to wadsleyite and ringwoodite to perovskite and magnesiowüstite transitions have opposite signs (Ita and Stixrude, 1992; Katsura and Ito, 1989, see Helffrich, 2000 for a review). Notation: *Pv* = perovskite, *Mw* = magnesiowüstite. (b) The expected phase transition behavior under low and high temperatures. A thick mantle transition zone is expected in a low temperature region such as an active subduction zone. (c) Results of conventional inversions for seismic velocity only. The velocity anomaly has no effective connection with, or effect on, the transition zone phase boundaries. (d) Illustration of how an outcome of a joint inversion of discontinuity topography and seismic velocity differs from (c). The tradeoff of the two quantities is properly considered in this case, i.e., the amplitude of the heterogeneity decreases due to changes in discontinuity depth (hence wave traveltime).

probabilistic solution based on the selection of a suitable ensemble of models that fits certain tolerance criterion (e.g., Metropolis et al., 1953; Press et al., 2007; Tarantola and Valette, 1982). Generally known as the Monte Carlo method (Metropolis and Ulam, 1949), nonlinear stochastic approaches take advantage of random numbers and statistical description of model spaces without resorting to gradients of objective functions (Mosegaard and Tarantola, 2002). After the pioneering study of Keilis-Borok and Yanovskaja (1967), applications of Monte Carlo methods became increasingly diverse in earthquake seismology (e.g., Bodin et al., 2012; Kennett et al., 1988; Koper et al., 1999; Menke and Schaff, 2004; Sambridge and Drijkoningen, 1992; Sambridge and Kennett, 1986; Scales et al., 1992; Shaw, 1992; Trampert et al., 2004). Rather than uniformly sampling the solution space (Sambridge and Kennett, 1986), which is computationally demanding and prone to 'rough' model outcomes (Sambridge, 1990), subsequent nonlinear global seismic inversion methods feature more sophisticated optimization methods that originated from ferromagnetic, thermodynamic and cellular systems; prime examples include simulated annealing (Iyer and Hirahara, 1993; Weber, 2000), genetic algorithm (Bodin et al., 2012; Koper et al., 1999; Sambridge and Drijkoningen, 1992) and neural network algorithm (Meier et al., 2007a,2007b). Through proper 'training' or 'tuning' (Mosegaard, 1998; Mosegaard and Tarantola, 2002), these algorithms are ideally suited for inverse problems involving high degrees of nonlinearity and/or multi-modal probability distributions (Koper et al., 1999; Mosegaard and Tarantola, 1995; Scales et al., 1992). Furthermore, the statistical description of the probabilistic model ensemble offers a more objective metric of model resolution than the standard resolution tests of a linear inversion (Fichtner and Trampert, 2011b; Sambridge and Mosegaard, 2002).

Like any other inversion algorithm, probabilistic inversion has its share of challenges. For instance, unbiased sampling of the solution space for a large number of model parameters remains a costly procedure even with fast improving computation speed and parallelization. A priori/posteriori probability distributions (Tarantola, 2005; Tarantola and Valette, 1982) could also be biased by the tomographer's subjective input (Scales and Snieder, 1997) in the presence of noise (Meier et al., 2007a). Hence, despite recent advances in the nonlinear analysis of velocity and density (e.g., Bodin et al., 2012; Fichtner and Igel, 2008; Fichtner and Trampert, 2011b; Maraschini and Foti, 2010; Mewes et al., 2010; Resovsky and Trampert, 2003; Trampert et al., 2004), probabilistic methods in regional and global seismic applications remain a promising but unfinished product.

3. Resolution of tomographic models

Results obtained from classical seismic tomography have been instrumental in the understanding of the large-scale structure and dynamics of the Earth's interior. Through sustained efforts and advances in technology and data acquisition, tomographic images have become

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

the de facto standard in assessing the dimension, motion and intrinsic properties of the lithosphere (e.g., Ekström and Dziewonski, 1998; Gu et al., 2005; Gung et al., 2003; Li and Romanowicz, 1996; Simons and van der Hilst, 2003; Su et al., 1992; Yuan and Romanowicz, 2010; Zhang and Tanimoto, 1992; Zhou and Clayton, 1990), plate boundary zones (e.g., Fukao et al., 1992; Nishimura and Forsyth, 1989; van der Hilst et al., 1991; Zhao, 2001a; Zhao et al., 1994), upper mantle transition zone (Becker et al., 2003; Gu et al., 2001; Mégnin and Romanowicz, 2000; Widiyantoro and van der Hilst, 1997, see Fukao et al., 2001, 2009 for reviews regarding regional subduction slab analysis), mid/lower mantle (Boschi and Ekström, 2002; Grand et al., 1997; Gu et al., 2003; Kárason and van der Hilst, 2000; Mégnin and Romanowicz, 2000; Ritsema and van Heijst, 2000; Zhao, 2001a), the lowermost mantle (Gu et al., 2001; Ishii, 2004; Kárason and van der Hilst, 2001; Liu and Dziewonski, 1998; Panning and Romanowicz, 2006) and the controversial outer core (Boschi and Dziewonski, 2000; Cao and Romanowicz, 2004; Vasco and Johnson, 1998). Impressive tomographic images of deep-penetrating slab (Fig. 6; e.g., Bijwaard et al., 1998; Fukao et al., 1992, 2001; Kárason and van der Hilst, 2000; van der Hilst et al., 1991) and ascending mantle plumes beneath Africa and the Pacific Ocean (Bijwaard and Spakman, 1999; Gu et al., 2001; Ritsema, 1999; Su et al., 1992; Zhao, 2001a) have served as popular reminders of the complex dynamic processes within the Earth's mantle. Many of these first-order mantle heterogeneities are well resolved by a wide spectrum of tomographic models, as evidenced by the general agreement between model predictions in the upper mantle transition zone (see Fig. 6).

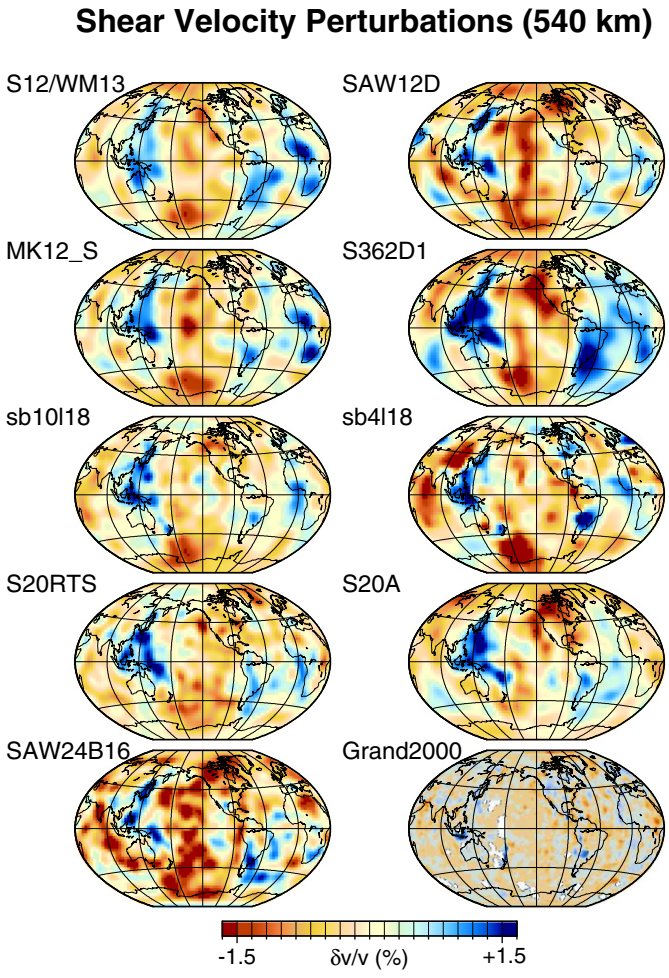

Fig. 6. Shear velocity perturbations in the upper mantle transition zone. Despite significant uncertainties under the oceans (see Section 3), high velocity perturbations from all models appear to agree on the presence of high velocity zones in the western Pacific and South America subduction zones. The references for the selected models are *S12/WM13* (Su et al., 1994), *SAW12D* (Li and Romanowicz, 1996), *MK12_S* (abbreviation of *S* model from *MK12WM13*, Su and Dziewonski, 1997), *S362D1* (Gu et al., 2001), *sb10l18* and *sb4l18* (Masters et al., 2000), *S20RTS* (Ritsema and van Heijst, 2000), *S20A* (Ekström and Dziewonski, 1998), *SAW24B1* (Mégnin and Romanowicz, 2000), and *Grand2000* (Grand, 2002).

3.1. Data constraints and general model resolution assessment

The resolution increase in seismic tomography has been steady, if not spectacular, in the past two decades. The road to higher resolution began with spherical harmonics (or basis functions with equivalent resolution) up to degrees 6–8 (Dziewonski, 1984; Woodhouse and Dziewonski, 1984) in the 1980s. With a resolving length of 3000–5000 km, the results of these early models established first-order connection between major tectonic regimes (i.e., continent and oceans) and seismic velocities. Following the development of broadband seismometers and the convenient data access provided by the Incorporated Research Institutions for Seismology (IRIS), different shear-wave phases were measured and the maximum spherical harmonic degree increased to 12–14 in the early 1990s (Bolton and Masters, 1991; Li and Romanowicz, 1996; Su et al., 1992), with radial basis functions to characterize the depth variations in shear velocity. In addition to offering 3–4 times the resolving power, this new generation of models was able to systematically quantify the spectral signatures, hence the characteristic scales, of mantle structure and deformation. In the late 1990s, the landscape of tomographic imaging changed significantly with the arrival of *high-resolution* models based on automatic traveltime picks from International Seismic Center (ISC) bulletins. While these large databases lack the quality control of individual phase picks (e.g., Bolton and Masters, 1991; Liu and Dziewonski, 1998; Ritsema, 1999; Su et al., 1992), the quantity of the data enables inversions with vastly increased nominal resolution. The output of these models (Grand et al., 1997; Kárason and van der Hilst, 2000) features spectacular descending slab-like structures with lateral length scales as small as a few hundred kilometers. While the realistic resolutions and their correlation with traditional long-wavelength models have been questioned (Boschi and Dziewonski, 1999), there is no denying that high-resolution block models became a catalyst in the debate on the style of mantle convection (e.g., Albarede and van der Hilst, 2002; Kellogg et al., 1999). The advent of high-resolution tomography and renewed focus on resolution improvements (Becker and Boschi, 2002; Ekström and Dziewonski, 1998; Gu et al., 2001, 2003; Houser et al., 2008; Mégnin and Romanowicz, 2000) tipped the perception scale clearly in favor of a partially layered convection system.

While it is impractical to assess the resolutions of individual global models in a short review, there are certain helpful tips from the past that could assist future construction of global and regional models. First of all, *more is not necessary better*. Due to the large number of pre-existing stations (e.g., Global Digital Seismic Network stations) and large number of earthquakes occurring on pre-existing fault lines, there is considerable overlap in the path coverage over time. Adding more data from similar paths would increase the redundancy, but not true resolution in the inversion process. On the other hand, *more variety is better*. Incorporation of different data types offers additional path coverage and constraints in spite of similar source-station locations for many of the phase picks (Fig. 7). The depth sensitivities of common data types for inversions of shear velocities are illustrated in Fig. 8. Surface wave dispersion measurements offer the best global constraint on the anisotropic mantle lithosphere (e.g., Ekström and Dziewonski, 1998; Gung et al., 2003; Larson et al., 1998; Laske and Masters, 1996; Montagner and Nataf, 1986; Nettles and Dziewonski, 2008; Nishimura and Forsyth, 1989; Yuan and Romanowicz, 2010). Since surface waves are shear waves in nature, global *S* wave models generally achieve greater spatial resolutions than *P* wave models at upper mantle depths (Romanowicz, 2003). The longest period surface waves also contribute to the resolution of mantle asthenosphere and

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

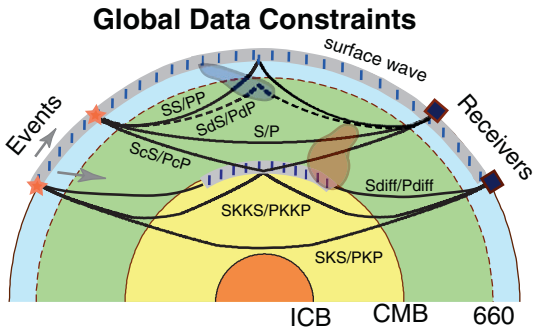

Fig. 7. Sample ray paths of commonly used seismic phases in global seismic tomography. The significant differences in penetrating depth and coverage among the phases highlight the importance of an integrated study. Notation: d = discontinuity depth.

transition zone (see Fig. 8), particularly with full waveform information from deep-penetrating, reverberating mantle waves. On the other hand, lower-mid mantle region is best resolved by body wave traveltimes (see Fig. 7), both for *P* (Antolik et al., 2003; Grand et al., 1997; van der Hilst et al., 1997; Widiyantoro and van der Hilst, 1997) and *S* (Gu et al., 2001, 2003; Houser et al., 2008; Kustowski et al., 2008; Masters et al., 2000; Mégnin and Romanowicz, 2000). Inclusions of direct and reflected *P* waves can provide coverage in otherwise undersampled regions such as the oceans. Comparable data coverage and sensitivity in this depth range render a combined *P* and *S* (or bulk and sound) analysis particularly informative (e.g., Antolik et al., 2003; Houser et al., 2008; Kennett et al., 1998; Manners, 2008; Su and Dziewonski, 1997). Relative to an average lower mantle depth, the data coverage and sensitivity increases exponentially near the core-mantle boundary. Traveltimes of reflected, diffracted and differential core phases (see Fig. 7; Boschi and Dziewonski, 2000; Gu et al., 2001, 2003; Kárason and van der Hilst, 2001; Liu and Dziewonski, 1998; Mégnin and Romanowicz, 2000; Obayashi and Fukao, 1997; Tkalčić et al., 2002) have been widely used in tomographic imaging to illuminate the base of the mantle.

It should be noted that differences between tomographic images could be both informative and challenging for isolating facts from artifacts (Dziewonski, 2005). Existing global *P* and *S* tomographic models differ in parameterization, data coverage and inversion strategies. Some of these differences reflect tomographer's personal preferences, e.g., choices of model parameterization and matrix factorization algorithms (e.g., Boschi and Dziewonski, 2000; Spakman and Nolet, 1988; van der Hilst, 1999), others result from informed decisions based on limited data coverage and/or finite computational resources. The inaccessibility of vast sub-crustal depths through direct means necessitates an objective assessment of *true* (as opposed to nominal) resolution in model appraisal. As a simple exercise we compute the bootstrapped average (Efron and Tibshirani, 1991) and standard deviation of 18 published global tomographic models from 1992 to 2004 (Fig. 9). Perturbations in *P* wave velocities have been amplified by a factor of 1.73 (Masters et al., 2000; Romanowicz, 2003) to be compatible with the *S*-wave models. Robust heterogeneous structures include strong continent–ocean contrast in the upper mantle lithosphere, high velocities along the western Pacific subduction zones in the upper mantle transition zone, a weak north–south orienting high velocity zone across the Americas, and anomalously low velocities associated with the Pacific and African 'plume groups' (Dziewonski et al., 1993) (see Fig. 9). The spectra of the entire mantle are clearly dominated by low-degree spherical harmonics (Gu et al., 2001; Masters, 1989; Su et al., 1992), or longer wavelengths (Dziewonski, 2005; Romanowicz, 2003), despite an apparent loss of overall amplitude due to averaging.

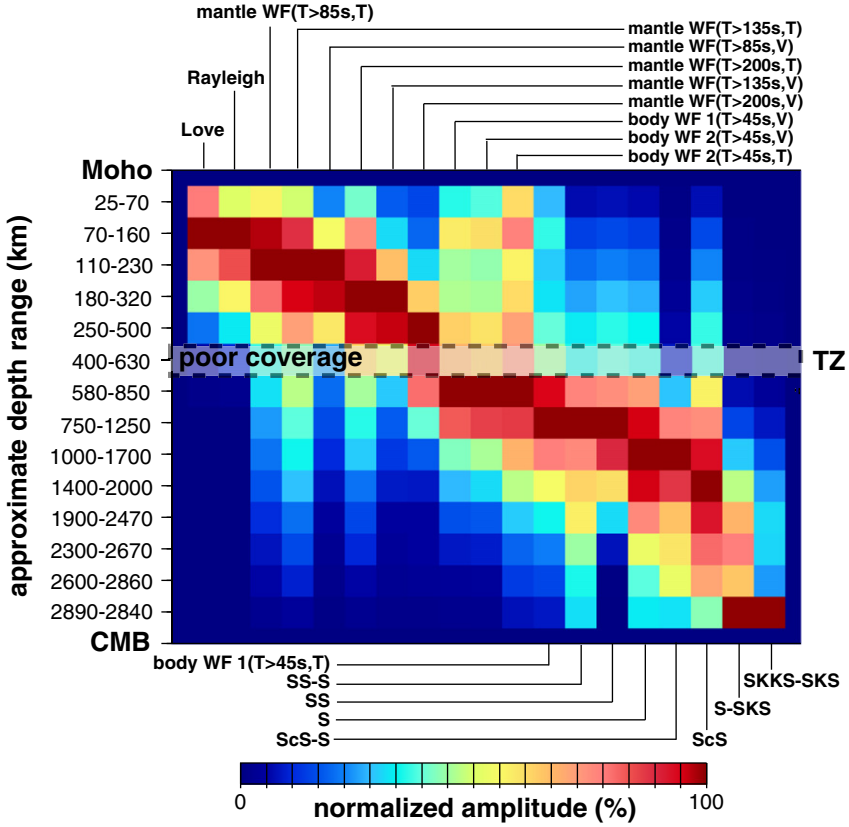

Fig. 8. A graphic representation of the lower half of the inner-product matrix used by Gu et al. (2001) in the inversion for mantle isotropic velocities. The vertical axis is approximated by the peak locations and half-widths of B-spline functions used in the inversion. WF1 represents a waveform dataset from recordings between 1991 and 1998; waveform records prior to this time period are represented by WF2. Relatively low resolution is observed in the upper mantle transition region (400–1000 km).
This figure is edited from Plate 1 of Gu et al. (2001).

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

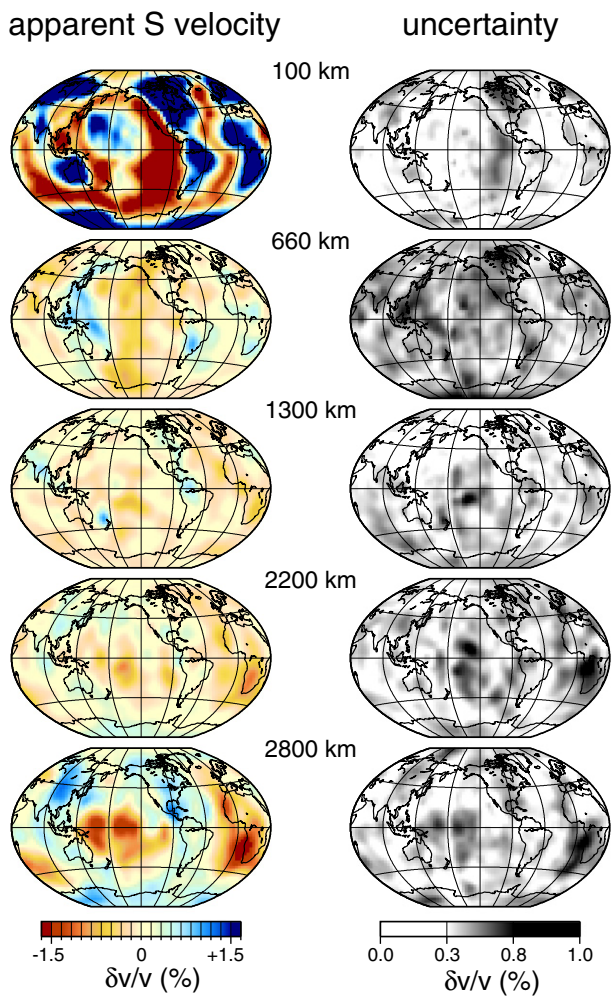

Fig. 9. Mean shear velocity perturbations (left column) and standard deviations (right column) from 18 published global tomographic (10 shear and 8 compressional, up to 2004) models. The lithosphere shows the best agreement among the global models, whereas the upper mantle transition zone, particularly in the vast oceanic regions, remains problematic. Models used in this analysis include all *S* velocity models presented in Fig. 6 as well as eight additional *P* and/or *S* models from Boschi and Dziewonski (1999); Kárason and van der Hilst (2001); Zhao (2001b); Antolik et al. (2003); Montelli et al. (2004).

Radially, the lithosphere appears to be best resolved, showing a 3+% average heterogeneity with merely ~1% maximum uncertainty near the East Pacific Rise and cratonic North America; dense surface wave coverage (e.g., Ekström and Dziewonski, 1998; Laske and Masters, 1996; Ritsema, 1999) is largely responsible for the enhanced resolution in this depth range (see also Figs. 7 and 8). Model differences at the remaining depths (see Fig. 9) accentuate the dichotomy in data sensitivity between continents and major oceans: structure beneath the oceanic regions, most notably the Pacific Ocean, show the largest disagreements among the models. Mantle compositional variations (e.g., near the base of the mantle), which would enhance the model difference between *P* and *S* velocities (Masters et al., 2000; Saltzer et al., 2001), are only partially responsible. Large model uncertainties clearly highlight the challenges in understanding the upper mantle transition zone, a well-documented under-sampled regime critical to the understanding of mantle circulation and mineralogical phase changes (see Fig. 8; Dziewonski, 2005; Fukao et al., 2001, 2009; Gu et al., 2001; Romanowicz, 2003), and the bulk of the lower mantle (Boschi and Dziewonski, 2000). A greater overall resolution in seismic tomography likely requires major improvements in ocean-bottom seismic recording and body-wave phase analysis.

It is worth noting that the model comparisons detailed in this section are only a simple exercise in the spirit of a statistical description of the solution space. Practical and unbiased assessment of data resolution remains a key area of geophysical research (Sambridge and Mosegaard, 2002). The following list highlights key considerations and challenges during the resolution assessment of tomographic inversions:

1. High resolution does not automatically translate to smaller-scale seismic structures (e.g., Boschi and Dziewonski, 1999). The characteristic length scales of anomalies are, to first order, reflected by low order spherical harmonics reported over two decades ago (e.g., Romanowicz, 2003; Woodhouse and Dziewonski, 1984).
2. Nominal resolution is not equivalent to the true resolution of a given dataset. The term *nominal resolution*, which appeared in early discussions of source parameter inversions (Gilbert and Dziewonski, 1975), is a subjective choice of the tomographer based on his/her own assessment of the dataset. The true resolution is an intrinsic property of the dataset independent of the modeling technique.
3. Successful recovery of an input checkerboard in the deterministic inversion approaches is only a necessary condition, but not a sufficient condition, in the pursuit of the true structure. Furthermore, resolution assessment centered on a single optimal model likely underestimates the true uncertainties of the inversion results (e.g., Sambridge and Mosegaard, 2002).

In addition to continued deployment of more seismic stations (see more discussions in Section 6), seismic tomography could greatly benefit from multiscale inversions with enhanced clarity on well-covered regions (e.g., Boschi et al., 2004; Chiao and Kuo, 2001; Hung et al., 2011; Zhao, 2009) as well as data mining beyond main body and surface wave arrivals on existing waveforms (see Section 6). The ultimate goal is to be able to explain every wiggle and recover as much information about the structures of the Earth's interior as data permit.

3.2. Ray vs finite-frequency kernel debate

Potential improvement in the resolution of tomographic models can be also achieved through more accurate forward modeling theory. Fitting all wiggles on a seismogram is a very challenging task, and as is true to all inverse problems, the accuracy and efficiency of solutions to the forward problem (i.e., seismic wave equations) are critical. Due to the complications involved in solving the seismic equation, the majority of tomographers still rely on approximation theories (e.g., high-frequency ray approximation for body-wave modeling or asymptotic methods for surface-wave modeling; see more discussions in Section 2) or assume simple 1D Earth structure (e.g., in normal-mode theory) to compute Green's functions. The accuracy of sensitivity kernels for measurements, which form row vectors of the design matrix **G** in Eq. (5), is strongly influenced by the accuracy of these forward solutions.

The classical solution began with geometrical ray theory for high-frequency waves, which is effectively a zeroth order approximation to seismic wave equations emphasizing the onset time of seismic arrivals (see Section 2). The resulting sensitivity kernels of high-frequency arrival times are reduced to 2D geometrical ray paths connecting sources and receivers (see Section 2). The applicability of ray theory relies on the assumption of weak and relatively large-scale lateral heterogeneities, which is not valid for structures with sizes that are comparable to or less than the dominant wavelengths of the interrogated seismic waves. Instead, the sizes of *Fresnel zones*, i.e., regions in the neighborhood of geometrical ray paths that give rise to constructively interfering arrivals, may be significant for global finite-frequency arrivals measured from broadband recordings. For instance, the Fresnel zone width of a 5-s *P* wave with a wavelength of ~40 km is approximately 560 km at 60° epicentral distance, which is similar to the resolution limit of global tomographic inversions (Nolet, 2008, Chapter 2). The complicated

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

sensitivity kernels for reflected waves present further complications (Deuss, 2009; Liu and Tromp, 2008; Shearer, 1993). Therefore, finite-frequency effects need to be considered during the computation of body (at 5–40 s) and surface (40 s or longer, Ritzwoller et al., 2011) wave sensitivities in global tomographic inversions.

Decades of theoretical and computational development have aimed to go beyond ray theory by combining solutions to seismic wave equations with first-order Born approximation. Improved sensitivities of broadband seismic phases have been calculated by mode coupling (e.g., Li and Romanowicz, 1995; Li and Tanimoto, 1993; Marquering et al., 1998), body-wave ray theory (e.g., Dahlen et al., 2000; Hung et al., 2000), surface-wave ray theory (e.g., Zhou, 2009; Zhou et al., 2004, 2005), normal-mode summation (e.g., Capdeville, 2005; Zhao and Chevrot, 2011a; Zhao and Jordan, 1998; Zhao et al., 2006) and full 2D/3D numerical simulations (Liu and Tromp, 2006, 2008; Nissen-Meyer et al., 2007a). These sensitivity kernels generally follow geometrical ray paths for the selected phases, and occupy volumetric regions in the vicinity of the ray paths instead of being restricted to the center of the volume. In particular, Marquering et al. (1999) and Dahlen et al. (2000) showed that, counter-intuitively, sensitivity kernels for finite-frequency cross-correlation traveltime measurements are precisely zero along the geometrical paths and therefore acquires the shape of 'hollow bananas'. Based on the paraxial ray approximation that simplifies the calculations of traveltimes and geometrical spreading factors for body waves, these sensitivity kernels can be computed efficiently and have been applied to long-period body waves in both global (e.g., Montelli et al., 2004, 2006b; Tian et al., 2009) and regional (e.g., Hung et al., 2004, 2010, 2011; Obrebski et al., 2011, 2012; Sigloch et al., 2008) tomographic problems. The counter-intuitive 'banana-doughnut' shape of these sensitivity kernels (and the novelty of its name), as well as some striking global tomography images based on these kernels depicting the morphology of possible mantle plumes (e.g., Montelli et al., 2004, 2006b), drew considerable attention from a wide geophysical audience. This new set of mantle images also spurred significant controversies over its improvement in resolution (Fig. 10; Dahlen and Nolet, 2005; de Hoop and van der Hilst, 2004; Montelli et al., 2006a; van der Hilst and de Hoop, 2005, 2006). de Hoop and van der Hilst (2004) contended that the zero sensitivity of these banana-doughnut kernels on the ray paths is mainly the result of a combination of factors (e.g., the use of ray-theory, choice of measurement, assumption of background media, and model dimensionality), and does not hold for realistic media, different choices of measurements, and model parameterizations. Finite-frequency sensitivity kernels can be regarded as a form of regularization for infinitely thin rays, and thus exhibit similar effects in tomographic inversions as geometrical rays once model parameterization and damping are applied. Still, the main issues hindering the improvement of global tomographic resolution are the uneven distributions of events and stations, as well as the lack of ray coverage in parts of the mantle (see Section 3.1). As global tomographic problems remain under-determined, regularization is inevitable and hampers a *fair* comparison between finite-frequency- and ray-based tomographic images (Montelli et al., 2006b; Panning and Romanowicz, 2006). This provides a plausible explanation for the lack of discernible improvement in the image resolution of global tomographic models when approximate 2D and 3D sensitivity kernels were used in place of rays (e.g., Kárason and van der Hilst, 2000; Li and Romanowicz, 1995).

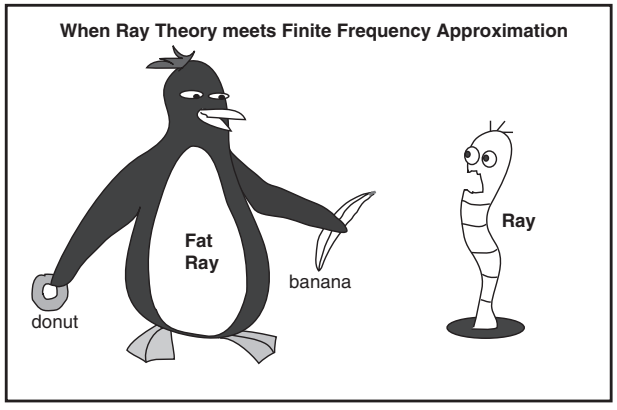

Fig. 10. A cartoon that underscores the heated debate centered on the resolution improvement of global tomographic models based on finite-frequency sensitivity kernels vs. simple ray approximation.

However, the divergent views presented above do not diminish the importance of choosing proper sensitivity kernels to map lateral heterogeneities, particularly with an increasing number of measurements from global seismic network and regional seismic arrays (Romanowicz, 2008). In principle, the use of appropriate sensitivity kernels accounts for the finite-frequency nature of observations, allows different datasets to be consistently combined in tomographic inversions, and makes it possible to image heterogeneities of sizes similar to the first Fresnel zone (e.g., Chevrot and Zhao, 2007; Hung et al., 2004). After all, consistent sensitivity kernels based on improved theory is paramount for the pursuit of more accurate solutions to tomographic problems.

4. Adjoint tomography: theory and implementations

With the advancement of numerical modeling techniques and expansion of computational resources, numerical simulations of seismic wave propagation for complex and realistic 3D heterogeneous media at both global and regional scales have now become routine practices (e.g., Castro and Käser, 2010; Komatitsch and Tromp, 2002a; Moczo et al., 2007; Tromp et al., 2010a). Factors such as gravity forcing, rotation, attenuation, anisotropy, topography of the free surface and internal discontinuities that are important for global and/or regional seismology can now be accounted for accurately. This naturally renders the possibility of calculating finite-frequency sensitivity kernels based on improved forward solver (Liu and Tromp, 2006, 2008; Zhao et al., 2005). At the present time, however, numerical calculations of kernels for all individual source–receiver pairs (i.e., row vectors of matrix **G**) in realistic regional and global tomographic problems remain computationally prohibitive.

Empowered by computational advancement, and potentially inspired by the success of adjoint methods in meteorological (e.g., Courtier and Talagrand, 1987; Talagrand and Courtier, 1987) and geodynamical (e.g., Bunge et al., 2003) applications, (Tromp et al., 2005) made important connections among time-reversal imaging, finite-frequency sensitivity kernels in seismic tomography and full waveform inversion popular in exploration seismology. Similar to 'time-reversal mirrors' used in medical imaging (Fink, 1992, 1997; Fink et al., 1989) where residuals of recorded data at receivers are physically time-reversed and transmitted to focus back at either original source locations or possible locations of model inadequacy, data residuals at seismic stations can now be time-reversed to obtain updates on earthquake source parameters based on 3D background models (Hjörleifsdóttir, 2007; Tape et al., 2007). The first step of this method is, in essence, a more sophisticated version of the back-projection technique used for imaging earthquake rupture processes (e.g., Ishii et al., 2005; Larmat et al., 2010). On the other hand, similar to results derived in full waveform inversions (e.g., Tarantola, 1984a), Fréchet derivatives of structural model parameters (e.g., density and wave speeds) can be computed by the interaction of the regular forward wavefield in the reference model and the adjoint wavefield generated by setting time-reversed data residual signals at receivers as simultaneous fictitious sources, i.e., two simulations of seismic wave equations. Applications of adjoint methods to seismological

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

problems have been similarly derived by Fichtner et al. (2006a,b); Fichtner (2010).

The seismic inverse problem may be treated as an optimization problem, where the optimal solution is sought to minimize a pre-defined misfit criterion that quantifies the difference between data and synthetics. When numerical solvers are used for the forward problem, the optimization/inverse problem is usually solved based solely on local gradients, i.e., Fréchet derivatives of the misfit function (e.g., Epanomeritakis et al., 2008; Fichtner et al., 2009b; Tape et al., 2010), without forming and accessing the *Hessian matrix* (i.e., the second-order derivatives of the misfit function, see also Section 4.4.3) due to computational and storage restrictions. Depending on the technical details of computational practices, this type of optimization method has appeared under many different names and caused some confusion in the literature. For the consistency of this paper, we adopt the name of *adjoint tomography*, and loosely define it as tomographic inversion techniques based on numerical simulations of seismic waves and local gradients computed by adjoint methods. The following sections provide a detailed account of the origin, derivation, merits, seismic applications and future outlook of adjoint tomography.

4.1. Full waveform inversions in exploration seismology

There is a natural overlap between adjoint tomography, a subject that recently became popular in earthquake seismology, and the *full waveform inversion* (FWI) techniques that have existed for nearly three decades in exploration seismology. Therefore it is impossible to review the recent development on adjoint tomography without referring to the voluminous literature in exploration seismology pertaining to FWI (see Virieux and Operto, 2009 for a recent review).

Based on the inverse theory developed by Tarantola and Valette (1982), Tarantola (1984a) and Lailly (1984) formulated the acoustic imaging problem as an optimization problem that seeks the best model parameters which minimize the generalized least-square sum of weighted data residual. Reflected wavefield is treated as a scattered wavefield generated by secondary sources associated with perturbations to the background velocity model. Based on single-scatter approximation, the gradient of the nonlinear misfit function is calculated by correlation of forward propagation of the actual source in the current model and a propagation backward in time of the weighted data residuals at the receivers as virtual sources. The backward wavefield was also referred to as 'missing diffracted wavefield' or the 'residual wavefield'. Tarantola (1984b) also showed that the first iteration of this optimization problem is equivalent to pre-stack migration techniques based on imaging principles (Claerbout, 1971). However, instead of obtaining reflectivity images with no direct physical interpretation from pre-stack migration, nonlinear waveform inversions make improvements of physical model parameters such as velocities and density through successive iterations. Diffractions, refractions and multiple reflections are also automatically accounted for through iterations when forward problems are calculated based on numerical simulations. To improve the well-posedness of the nonlinear inversions, this method seeks solutions that are close to a given a priori model and simultaneously minimizes the data residual. Formulations were later derived for elastic wave equations (Tarantola, 1986) and anelastic wave equations (Tarantola, 1988). Parameter selection is also critical as long-period waveform inversions are mostly sensitive to seismic velocities while short-period waveforms are strongly affected by impedance contrasts in exploration-scale problems. (Tarantola, 1986), observations that reflect the scattering properties of elastic waves (Wu and Aki, 1985).

Gauthier et al. (1986) and Mora (1987, 1988) demonstrated the feasibility of FWI methods based on finite-difference scheme. Monte Carlo inversion techniques (see Section 2.7) were dismissed at the time due to high computational cost, while the more efficient local-gradient techniques were favored, since only two (forward and backward) simulation are needed to compute the gradient direction and one (or few) simulation(s) are needed to estimate the step length of model update along the direction. Serious concerns were also raised on the nonlinearity of the waveform misfit function, where the final model may be trapped in local minimum (Gauthier et al., 1986). Initial solutions need to be close enough to the true model to keep the misfit function in the quasi-linear regime with respect to model perturbations (Mora, 1987). Numerical tests showed that surface reflected data (equivalent to scattered/converted waves in earthquake seismology, see more discussions to Section 6.2), due to their nonlinearity with respect to large wavelength structures, recover mostly the high-frequency component of the original model, while transmission data (equivalent to direct main arrivals in earthquake seismology) improve the resolvability of the low-frequency content of the model (Gauthier et al., 1986; Mora, 1988). It was noted that preconditioning of the gradient direction based on simple geometrical arguments may help reduce the number of iterations (Gauthier et al., 1986). Mora (1987) also derived the FWI problem based on matrix formulation, and suggested that the best preconditioner is the inverse Hessian matrix such that the optimization problem can be solved by one quasi-Newton step.

Moderate success has been achieved when applying FWI techniques to realistic industry data (e.g., Crase et al., 1990; Igel et al., 1996), and serious challenges regarding the nonlinearity of the waveform misfit function remain unsolved. Luo and Schuster (1991) remedied this problem by introducing misfit functions that quantify the *traveltime* difference between observed and synthetic seismograms and noted that no phase identification for ray-theoretical arrivals is required when numerical modeling is used. Traveltime misfit functions are much more linear compared to full-waveform misfit, though with reduced model resolution. Another successful approach is the frequency-domain iterative waveform inversion based on frequency-domain numerical solvers. Models are first inverted at a low frequency where the initial guess is sufficiently close to the true model to guarantee the linearity of the waveform function, and then successively inverted at higher frequencies using inversion results from the previous frequency as the initial model for the current iteration (e.g., Brenders and Pratt, 2007; Geller and Hara, 1993; Plessix, 2009; Pratt, 1999; Pratt and Worthington, 1990a,1990b; Sirgue and Pratt, 2004). As traveltime tomography based on full-wave equations resolves velocity anomalies of the size of the first Fresnel zone, reflected waves and scattered waves can resolve structures with sizes close to the dominant wavelength used in the inversion (Wu and Toksoz, 1987). It was suggested that a combination of traveltime and waveform inversion will be ideal to recover both the long and short wavelength structures of the subsurface while keeping the inverse problem in the quasi-linear regime (e.g., Luo and Schuster, 1991; Pratt and Shipp, 1999). Multi-scale inversions in the time-domain also proved to be promising based on similar principles (Virieux and Operto, 2009).

4.2. Applications of FWI to seismic imaging in the crust and mantle

Although the theory and techniques of FWI can be directly applied to earthquake seismology, several essential distinctions exist. First there are the obvious lack of control in source locations and uneven station distributions. Regional seismic imaging for the crust and mantle mainly utilizes teleseismic data, since abundant recordings of local/regional earthquakes are available only for a few tectonically active regions. Furthermore, earthquake seismologists still work most comfortably with traveltime data to map internal structures of the earth, and are more cautious towards utilizing amplitude anomalies (or entire waveform) that are prone to attenuation or surface-wave focusing effects. Measurements are usually taken from individual body-wave phases and may consist of secondary arrival-time observables from surface-wave waveforms (e.g., Cara and Leveque, 1987). Since windowing is difficult to implement in frequency-domain FWI techniques, traveltime inversions may be performed only in the

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

time-domain. Lastly, due to the computation cost of a single numerical simulation at global or regional scales, which ranges from less than an hour to tens of hours even on modern parallel computers, inversion schemes need to be designed to reduce the number of iterations and speed up convergence.

From the inversion setup point of view, individual sensitivity kernels of source–receiver pairs form row vectors of the design matrix **G** for tomographic inversions and are often computed to demonstrate the effects of underlying approximations and finite frequency in global seismology. However, as full waveforms are inverted in exploration applications, computing sensitivities for individual data points of the scattered waves (also defined as *waveform perturbation density*, Zhao and Chevrot, 2011a) is impractical and unnecessary due to the high-frequency nature of controlled sources. Instead, the gradient of the misfit function is computed in two simulations, equivalent to the weighted sum of sensitivity kernels for scattered waves between all source–receiver pairs.

4.3. Forward solvers for 3D velocity models

The efficiency and accuracy of the forward-modeling technique is crucial to the success of inverse methods. Therefore we will summarize some of the common numerical techniques used to solve seismic wave equations for 3D heterogeneous models as a prelude to our discussion on adjoint tomography. This section by no means includes all references in the related fields, though, it is our hope that further information could be obtained from the references provided by this review.

With the development of computer technology and numerical techniques, significant advances have been made in solving seismic wave equations numerically. Finite difference (FD) methods have gained popularity since the early 80s due to their relatively simple formulation (e.g., Igel, 1996; Igel et al., 2002; Kelly et al., 1976; Moczo et al., 2007; Olsen et al., 1995). Pseudo-spectral methods (e.g., Carcione, 1994; Carcione et al., 2005; Tessmer et al., 1992) have also been used, although they were mostly restricted to smooth models. Both methods exhibit considerable difficulties with implementing free-surface boundary conditions for the accurate propagation of surface waves. On the other hand, finite-element method (FEM) implicitly incorporates natural boundary conditions (e.g., Akcelik et al., 2002; Bao et al., 1998) despite being more programmatically involved. In particular, a specific type of finite-element method, *spectral-element method* (SEM), was introduced to seismology from computational fluid dynamics (Patera, 1984) for both global (Capdeville et al., 2003; Chaljub and Valette, 2004; Fichtner et al., 2009a; Komatitsch and Tromp, 2002a,2002b) and regional (Komatitsch et al., 2004; Lee et al., 2008; Peter et al., 2011a; Pilz et al., 2011; Stupazzini, 2006) seismic wave propagation in 3D heterogeneous media (see Chaljub et al., 2007 for a review). It combines the geometrical flexibility of FEMs with exponential convergence properties of spectral methods. Surface topography and internal discontinuities are naturally accounted for by a finite-element mesh design. Mesh doubling or tripling takes advantage of the velocity increase with depth to minimize computational cost. Unlike classical FEM wherein large linear systems need to be solved at each time step, the mass matrix of the SEM is diagonal due to the choice of Gauss–Lobbatto–Legrendre points as both interpolation and integration points within hexahedral elements (Komatitsch and Tromp, 1999). This makes SEM extremely efficient and scalable on parallel computers, especially when ported to take advantage of recent Graphics Processing Unit (GPU) hardwares (Komatitsch et al., 2009, 2010). SEM has been applied to automatically simulate seismic wave propagation for $M_w \geq 3.5$ events in southern California (Komatitsch et al., 2004, http://www.shakemovie.caltech.edu) and $M_w \geq 5.5$ events at the global scale (Tromp et al., 2010a, http://www.globalshakemovie.princeton.edu). However, despite the computational efficiency, it remains challenging to design a proper conforming hexahedral unstructured SEM mesh for a region with strong velocity variations that 1) honors predefined internal surfaces and material discontinuities, and 2) maintains a consistent number of points-per-wavelength throughout the entire model for accuracy considerations. External hexahedral meshing softwares are often used (Casarotti et al., 2008; Peter et al., 2011a; Stupazzini et al., 2009), and careful considerations on load balancing between individual processors (Peter et al., 2011a) are required to efficiently simulate wave propagation on such a complicated mesh in parallel.

One variant of SEM is proposed by Capdeville et al. (2003), wherein waves are simulated based on the coupling between a 1-D normal-mode solution for a spherically symmetric interior (such as the core) and SEM solution for a 3D heterogeneous outer shell (such as the crust and mantle) using a Dirichlet-to-Newman boundary-condition operator. This method speeds up simulations without a significant loss of accuracy, and has been adopted by various global and continental-scale tomographic studies (Lekić and Romanowicz, 2011; Yuan et al., 2011).

Discontinuous Gerlekin (DG) methods also gained popularity in seismology in recent years (e.g., Castro and Käser, 2010; Dumbser and Käser, 2006; Hermann et al., 2010; Käser and Dumbser, 2006). Being more general in formulation than SEM, the DG approach can handle material discontinuities across element boundaries and be operated on both tetrahedral and hexhedral meshes. In particular, it is extremely useful for dynamic rupture simulations where high frequency waves and super-shear near the fault can only be captured by a denser mesh and improved time stepping schemes (e.g., de la Puente et al., 2009). Despite being more computational costly, the use of tetrahedral elements makes it relatively easier to honor complicated fault systems and internal discontinuous surfaces in the mesh design for geologically complex regions, therefore lending itself well to local and regional-scale seismic wave simulations.

In the context of seismic imaging and inverse problem, most material interfaces beneath the surface are not precisely known. To first order, most geological regions are composed of large blocks with relatively smooth velocity variations within. In such cases, SEM may be a preferred approach that strikes a good balance between computational cost and accuracy of mesh design. Certainly at the global scale, where velocity variations are believed to be long-wavelength (100–1000 km) and smooth (<10% in V_s) in most regions, SEM is still the most viable solution to simulate wave propagation in 3D.

4.4. Theory and implementations

The basic formulation of adjoint tomography has been derived by numerous researchers, either based on Born approximation and reciprocity of the Green's function (Tarantola, 1984a; Tromp et al., 2005), Lagrange multiplier approach (Akcelik et al., 2003; Liu and Tromp, 2006), where the adjoint field is just the time reversal of the Lagrange multiplier, or more general functional analysis where no assumption on Green's functions and reciprocity is required (Fichtner et al., 2006a,2006b). The Lagrange multiplier approach also enabled (Liu and Tromp, 2008) to derive expressions of Fréchet derivatives for more complicated global seismic wave equations including factors such as ellipticity, anelasticity and gravity. Here we only symbolically outline expressions for local gradients based on frequency-domain wave equations, emphasizing on the differences (e.g., the choice of misfit function and the computation/storage challenges with utilizing numerical solvers) between adjoint and classical (either ray or finite-frequency) tomographic techniques.

4.4.1. Definition of misfit function

Tomographic inversion aims to utilize as much information of recorded seismograms as possible to map internal structures of the Earth. It can also be casted as an optimization problem, where an optimal model $\boldsymbol{m}^{opt}(\mathrm{x})$ is sought to minimize a misfit functional that quantifies the difference between observed seismograms and synthetics

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

calculated based on model parameters $\mathbf{m}$. For example, a least-square misfit functional that quantifies the weighted residual between a specific type of measurement a^o_{erp}, made on the p-th phase (or wave group) of the data recorded by the r-th receiver for the e-th event, and the corresponding model prediction $a_{erp}(\mathbf{m})$ is given by

$$\Phi = \sum_e W_e \Phi_e = \frac{1}{2}\sum_e W_e \sum_{rp} W_{rp}\left[a_{erp}(\boldsymbol{m}) - a^o_{erp}\right]^2, \tag{18}$$

where Φ_e and W_e are the misfit functional and weight associated with the e-th event, respectively, and W_{rp} is the weight assigned based on measurement error and the relative importance of this phase and receiver in the inversion.

In principle, a minimization problem involving Eq. (18) could be solved using global optimization methods based on Monte Carlo simulations or neural network (see Section 2.7). In addition to the optimal models achieved, the most probable region of model parameter space may be sampled and the posterior probability distributions of model parameters may be estimated, which provide a basis for probabilistic tomography (Mosca, 2010; Resovsky and Trampert, 2003). However, these methods often involve many evaluations of the forward problem, which are generally impractical despite moderate success (e.g., Valentine and Trampert, 2011), due to high computational cost of numerical solvers. Hence, local gradient methods which linearize the misfit function with respect to a reference model are favored, especially for high-resolution inversions at regional and global scales, but special care must be taken to reduce the nonlinearity of the misfit function and properly assess the non-uniqueness of the inverse problem (see Section 2.7; Snieder, 1998).

4.4.2. Gradient of misfit function

Computing the gradient of misfit function for a single event Φ_e is a key ingredient in tomographic inversions, although alternative data reduction schemes have been adopted to take advantage of the sparsity of individual sensitivity kernels by stacking multiple event recordings for a common station (Capdeville et al., 2005). For brevity, the index dependence on event e will be omitted and only the misfit function for one event will be examined until later sections. Let us assume that the variation of the measurement $\delta a(\mathbf{m})$ can be related to the variation of synthetics $\delta\mathbf{s}$ through a *connective function* (Luo and Schuster, 1991) $\mathbf{g}_{rp}$:

$$\delta a_{rp}(\mathbf{m}) = \int \mathbf{g}_{rp}(\tau)\cdot\delta\mathbf{s}(\mathbf{x}_r,\tau;\mathbf{m})\, d\tau, \tag{19}$$

where $\mathbf{x}_r$ is the receiver location. The applicability of adjoint tomography to a particular type of measurement hinges on our ability to derive the above analytical expression $g(\tau)$ (Fichtner et al., 2006a; Tape et al., 2010), which may be algebraically involved at times (P. Chen et al., 2010).

By substituting Eq. (19) into the variation of Eq. (18) and transforming the integration from time to frequency domain based on Parseval's theorem, the variation of misfit function can be rewritten as

$$\delta\Phi = \int\int \sum_{rp} W_{rp}\left[a_{rp}(\mathbf{m}) - a^o_{rp}\right]\delta(\mathbf{x}-\mathbf{x}_r)\mathbf{g}^*_{rp}(\omega)\cdot\delta\mathbf{s}(\mathbf{x},\omega;\mathbf{m})\, d\omega dV, \tag{20}$$

where $\mathbf{g}^*_{rp}(\omega)$ is the complex conjugate of $\mathbf{g}_{rp}(\omega)$, equivalent to a time reversal of $\mathbf{g}_{rp}(t)$ (Tarantola, 1984b).

Given seismic wave equations in the frequency domain (for simplicity, the effects of ellipticity, rotation and gravity are ignored)

$$-\rho\omega^2\mathbf{s} = \nabla\cdot(C:\nabla\mathbf{s}) + \mathbf{f}, \tag{21}$$

where $\rho(\mathbf{x})$ is the background density distribution, $\mathbf{C}(\mathbf{x})$ is elastic tensor relating stress to strain (note that C can be generalized to a complex number $C = C(\omega)$ for a visco-elastic medium, relating stress to strain by $T(\omega) = C(\omega):\nabla\mathbf{s}(\omega)$), and $\mathbf{f}$ is the force term associated with a seismic event. The dependence of synthetics $\mathbf{s}(\mathbf{x},\mathbf{m})$ on model $\mathbf{m}$ can then be linearized based on the Born approximation (i.e., single-scatter approximation, Nolet, 2008, Chapter 4):

$$-\rho\omega^2\delta\mathbf{s} = \nabla\cdot(C:\nabla\delta\mathbf{s}) + \left[\omega^2\mathbf{s}\,\delta\rho + \nabla\cdot(\delta C:\nabla\mathbf{s})\right], \tag{22}$$

i.e., the perturbation to the current wavefield $\delta\mathbf{s}$ is generated by secondary sources based on interactions of density and elastic tensor perturbations $(\delta\rho,\delta C)$ with the current wavefield $\mathbf{s}$. The Born approximation is only a first-order approximation excluding contributions from multiple-scattering, and the perturbed wave field calculated based on Eq. (22) will always arrive later than the unperturbed field (Fichtner et al., 2006b).

Suppose the Green's function $\mathbf{G}(\mathbf{x},\mathbf{x}_s,\omega)$ for the forward problem (21) can be computed for 1D and/or 3D reference models via approximate/semi-analytical methods or numerical simulations, the perturbed (scattered) wavefield then becomes

$$\delta\mathbf{s}(x,\omega) = \int \boldsymbol{G}(x,x',\omega)\cdot\left[\omega^2\mathbf{s}(x')\delta\rho(x') + \nabla\cdot(\delta C(x'):\nabla\mathbf{s}(x'))\right]dx'. \tag{23}$$

This expression can be substituted into Eq. (20) to obtain the relation between event misfit and material properties. In particular, if one defines an adjoint wavefield generated by injecting time-reversed weighted residuals at receivers as virtual adjoint sources as

$$\mathbf{s}^\dagger(\mathbf{x}',\omega) = \sum_{rp} W_{rp}\left[a_{rp}(\mathbf{m}) - a^o_{rp}\right]\mathbf{g}^*_{rp}(\omega)\cdot\mathbf{G}(\mathbf{x}',\mathbf{x}_r,\omega), \tag{24}$$

then based on the source–receiver reciprocity of Green's function (Aki and Richards, 2002, Chapter 2), the variation of misfit function (20) becomes

$$\begin{aligned}\delta\Phi &= \int\int \mathbf{s}^\dagger(\mathbf{x}',\omega)\cdot\left[\omega^2\mathbf{s}(\mathbf{x}',\omega)\delta\rho(\mathbf{x}') + \nabla\cdot(\delta C(\mathbf{x}'):\nabla\mathbf{s}(\mathbf{x}',\omega))\right]d\mathbf{x}'d\omega \\ &= \int\left[K^e_\rho(\mathbf{x})\frac{\delta\rho}{\rho} + K^e_c(\mathbf{x})\frac{\delta c}{c}\right]d\mathbf{x}.\end{aligned} \tag{25}$$

In this formulation, expressions for Fréchet derivatives of Φ with respect to relative perturbations of $\mathbf{m}$, also known as the *event kernels* (Tape et al., 2007), are given by

$$\begin{aligned}K^e_\rho(\mathbf{x}) &= \rho(\mathbf{x})\int\omega^2\mathbf{s}(\mathbf{x},\omega)\cdot\mathbf{s}^\dagger(\mathbf{x},\omega)\, d\omega \\ K^e_c(\mathbf{x}) &= -C(\mathbf{x})\int\nabla\boldsymbol{s}(\mathbf{x},\omega)\nabla\boldsymbol{s}^\dagger(\mathbf{x},\omega)\, d\omega,\end{aligned} \tag{26}$$

and involve the interaction of forward and adjoint wavefield. Detailed expressions of sensitivity kernels for isotropic and general anisotropic parameterization can be found in Tromp et al. (2005) and Sieminski et al. (2007a,b).

Note that the adjoint source used to generate the adjoint wavefield in Eq. (24) stems entirely from the definition of misfit function (18) and its variation Eq. (20). The choices of measurement and misfit function affect the gradient of the misfit function (25) through sensitivity kernels (26), both involving the adjoint field.

4.4.3. Hessian matrix

What has been computed so far is only the gradient of the misfit function. To solve the minimization problem, it may sometimes be desirable to obtain the second-order derivatives of the misfit function. For simplicity, the spatial function of model $\mathbf{m}(\mathbf{x})$ is first cast into discrete model parameter vector $\underline{m}$ through a set of basis functions (see Section 2.3). The misfit function can be linearized with respect to a reference model $\underline{m}$ as

$$\Phi\left(\underline{m} + \Delta\underline{m}\right) \sim \Phi\left(\underline{m}\right) + \mathbf{g}^T\delta\underline{m} + \frac{1}{2}\Delta\underline{m}^T\mathbf{H}\,\Delta\underline{m}, \tag{27}$$

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

where $\mathbf{g} = \frac{\partial \Phi}{\partial \underline{m}}$ is the gradient vector obtained from expanding Eq. (26) by the basis functions and $\boldsymbol{H} = \frac{\partial^2 \Phi}{\partial \underline{m}^2}$ is the Hessian matrix. It follows from the definition of event misfit (18) that

$$\frac{\partial^2 \Phi}{\partial \underline{m}^2} = \sum_{rp} W_{rp} \left\{ \left[a_{rp}\left(\underline{m}\right) - a_{rp}^o \right] \frac{\partial^2 a_{rp}}{\partial \underline{m}^2} + \frac{\partial a_{rp}}{\partial \underline{m}} \frac{\partial a_{rp}}{\partial \underline{m}} \right\}. \tag{28}$$

Under the assumption that the measurement function $a_{rp}(\mathbf{m})$ is quasi-linear with respect to $\mathbf{m}$, or that the reference model is sufficiently close to the true model such that the residual $\left[a_{rp}\left(\underline{m}\right) - a_{rp}^o \right]$ is generally small, the first term on the RHS of Eq. (28) can be ignored. Furthermore, the computation of Hessian requires access to derivatives of individual measurements $\frac{\partial a_{rp}}{\partial \underline{m}}$, which are row vectors of the design matrix in Eq. (5). Therefore, the approximate Hessian matrix is given by

$$\boldsymbol{H} = \frac{\partial^2 \Phi}{\partial \underline{m}^2} \sim \sum_{rp} W_{rp} \frac{\partial a_{rp}}{\partial \underline{m}} \frac{\partial a_{rp}}{\partial \underline{m}} = \boldsymbol{G}^T W \boldsymbol{G}, \tag{29}$$

where W is the weighting matrix with W_{rp} as diagonal elements. Neither $\mathbf{G}$ nor $\mathbf{H}$ involve actual data a_{rp}^0, and are fully determined by source–receiver geometry and known background model. The optimal model parameters that minimize the quadratic function (27) can be obtained by solving the linear system

$$\mathbf{H} \Delta \underline{m} = -\mathbf{g}, \tag{30}$$

although regularization has to be applied to the Hessian in most global tomographic problems involving insufficient data coverage.

In general, the misfit function can be approximated only locally by quadratic function (27), and successive iterations in which both $\mathbf{H}$ and $\mathbf{g}$ are recomputed based on updated velocity models $\underline{m} + \Delta \underline{m}$ are necessary to achieve final convergence. If the full Hessian matrix is used as in Eq. (28), the inverse method is referred to as the *Newton's method*; if the approximate Hessian matrix is used as in Eq. (29), it becomes *Gauss-Newton's method* (Nocedal and Wright, 2006).

For the case of $\Phi = a_{rp}(\mathbf{m})$ in Section 4.4.2, i.e., an individual measurement, the gradient is also given by Eq. (25) where the adjoint field is now generated by the time-reversed connectivity function $g_{rp}^*(\omega)$ as adjoint source. In addition to one forward calculation for $\mathbf{s}$, one adjoint calculation is also required to compute $\frac{\partial a_{rp}}{\partial \mathbf{m}}$. Therefore, the total number of adjoint calculations to assemble row vectors of G related to e-th event is $\sim N_R N_P$, where N_R is the average number of receivers used in an event and N_P is the average number of phases picked per receiver. In comparison, the computation of $\mathbf{g}$ as in Eq. (30) requires just one additional adjoint simulation where residuals for all phases and receivers are injected back as adjoint sources simultaneously. When Green's functions are obtained through computationally expensive numerical methods, assembling $\mathbf{H}$ becomes extremely costly and is generally avoided. Instead, optimization methods such as nonlinear conjugate-gradient techniques based only on gradient functions (Fichtner et al., 2009b; Fletcher, 1987; Tape et al., 2007) can effectively approximate the operation of $\mathbf{H}^{-1}$ by applying polynomial functions of $\mathbf{G}$ (the forward operator) and $\mathbf{G}^T$ (the adjoint operator) through successive iterations.

4.4.4. Time-domain implementation

In practice, when seismic wave equations are solved by numerical techniques, one simulation is performed to obtain the forward field $\mathbf{s}(\mathbf{x},t)$ and another targets the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{x},t)$. However, when Eq. (26) are transformed into the time-domain, e.g.,

$$K_\rho^e(\mathbf{x}) = -\rho(\mathbf{x}) \int_0^T \partial_t^2 \mathbf{s}(\mathbf{x}, T-t) \cdot \mathbf{s}^\dagger(\mathbf{x}, t) dt, \tag{31}$$

where T denotes the simulation end time, the forward field and the adjoint field need to be convolved at any spatial point $\mathbf{x}$ and then evaluated at time T. This requires access to $\mathbf{s}^\dagger$ and $\mathbf{s}$ at the respective times of t and $T-t$, which is generally unachievable when forward and adjoint simulations are run simultaneously. One apparent solution is to first carry out the forward simulation and save the entire wavefield as a function of time and space, then launch the adjoint simulation and at time t access the $T-t$ slice of $\mathbf{s}$ stored on disk. This procedure could be challenging due to high computer storage demand associated with saving the entire global/regional wavefields (typically, tens-to-hundreds of millions of grid points and tens of thousands of time steps). Having recognized that fewer number of grid points are required to accurately sample sensitivity kernels than to simulate seismic waves, Fichtner et al. (2009b) suggested a remedy by sub-sampling both in time and space based on the necessary spatial and temporal resolutions. An alternative solution, which reduces the storage requirement at the expense of computational cost, is to simulate wave propagation backward in time with the last frame of the wavefield as end condition by exploiting the invariance of the elastic seismic wave equation (where $C(\omega) = C$) under time reflection. In this case, while the forward field is reconstructed backward in time, the adjoint simulation is performed forward in time. At any given time t, the $T-t$ slice of $\mathbf{s}$ and the t slice of $\mathbf{s}^\dagger$ are simultaneously updated in computer memory and henceforth combined to obtain the contribution of time slice t to the sensitivity kernels in Eq. (26) (e.g., Akcelik et al., 2003; Liu and Tromp, 2006, 2008). Here the requirement on storing the entire forward wavefield is replaced by an extra simulation of reconstructing the forward field backward in time, which also doubles memory requirement for kernel calculations. Fig. 11 demonstrates the computation of a traveltime sensitivity kernel based on this approach for a single source–receiver pair in a 2D homogeneous model. If attenuation is included in the seismic wave equation (where $C(\omega)$ is a complex number), back reconstruction of $\mathbf{s}$ becomes unstable and a checkpointing scheme, hence extra computer storage, would be required to periodically reset the reconstructed wavefield (Liu and Tromp, 2008).

As indicated in Section 4.4.3, the gradient of individual measurements may be expensive to compute, while the overall gradient of the event misfit function (i.e., event kernels) can be retrieved based on the interaction of the forward and adjoint wavefields. The gradient of the misfit function of all events in Eq. (18) is then simply a weighted sum of all event kernels. Fig. 12 shows examples of event kernels generated for a synthetic checkerboard pattern based on a 2D uniform background velocity model. Each event kernel can be interpreted as the patterns of velocity anomalies 'seen' by the measurements pertaining to this particular event and skewed by its particular source–receiver geometry. When Hessian matrix is present, this 'skewed' view can be corrected or sharpened by applying the inverse Hessian as a preconditioner, and velocity model is simply updated in one iteration step as $\Delta \underline{m} = \mathbf{H}^{-1}\mathbf{g}$. However, when Hessian is not available, multiple (6–7 for 2D examples in Tape et al., 2007) nonlinear conjugate-gradient iterations are necessary to resemble the effect of one Hessian iteration (Akcelik et al., 2002; Chen et al., 2007b; Tape et al., 2007).

4.4.5. Comparisons with classical tomographic methods

Although derived under the framework of optimization methods, Eq. (30) is also essential in classical tomographic inversions (i.e., the normal equation of (5)). In this sense, classical ray-based and finite-

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

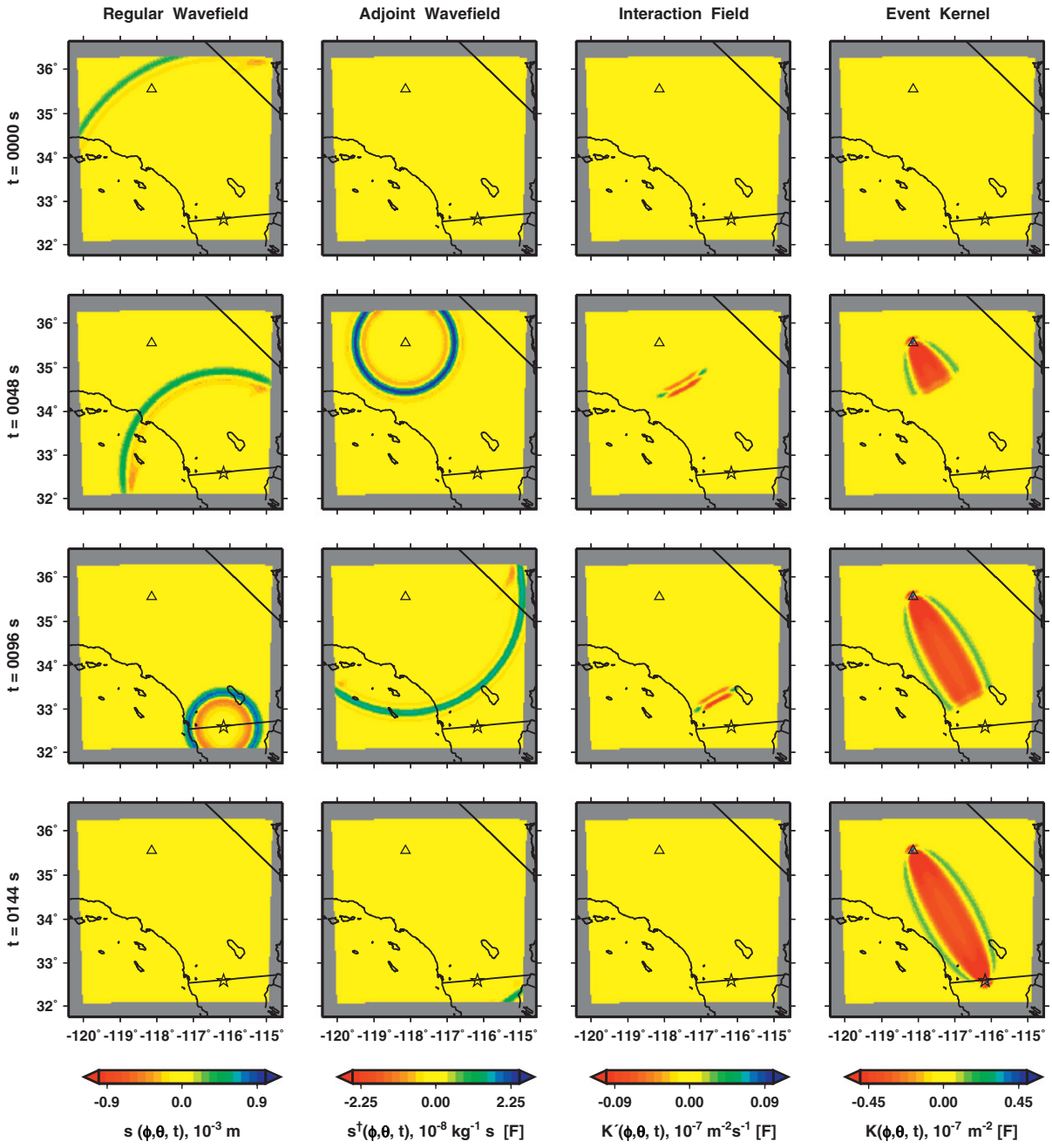

Fig. 11. Computation of sensitivity kernels by the interaction of the forward wavefield (first column) and the adjoint wavefield (second column) at four different time slices in a simple 2D homogeneous model. The contribution to sensitivity kernels by one particular time slice is shown in the third column, and the integrated value up to that time slice is shown in the fourth column. The forward field is constructed backwards in time while the adjoint field is constructed forward in time from top to bottom rows. Note the sensitivity kernel for a single traveltime measurement shows a 'cigar' shape in 2D instead of 'banana-doughnut' shape in 3D.
Courtesy of Tape et al. (2007).

frequency-based tomography only differ in how **H** and **g** are computed. In classical ray tomography, the design matrix **G** consists of effective lengths of source–receiver rays associated with model parameters, whereas different approximate or semi-analytical methods have been used to calculate Green's function $\mathbf{G}(\mathbf{x},\mathbf{x}_n,t)$ as well as the corresponding sensitivity kernels in the **G** matrix (see more discussions in Section 3.2) for the case of finite-frequency tomography. By adopting approximations or semi-analytical approaches, Green's functions can be computed relatively quickly and the **H** matrix can be assembled efficiently to obtain a model update. However, the approximate nature of these techniques restricts them to the application of specific wave types or simple (predominantly 1D) starting models. Consequently, only one Hessian-based inversion step can be taken for regions with strong velocity heterogeneities (e.g., the crust), even though the misfit function is highly nonlinear with respect to model parameters. Any successive iteration which requires recomputing the **H** matrix under the updated model will be infeasible or computationally expensive, and hence is less often attempted in classical tomography (e.g., Bijwaard and Spakman, 2000; Lekić and Romanowicz, 2011).

4.4.6. Sensitivity kernels for 1D reference velocity models

With the exception of lithosphere and portions of the D″ layer, the amplitude of 3D mantle heterogeneities on the global scale is relatively small with respect to 1D reference models. Therefore, sensitivity

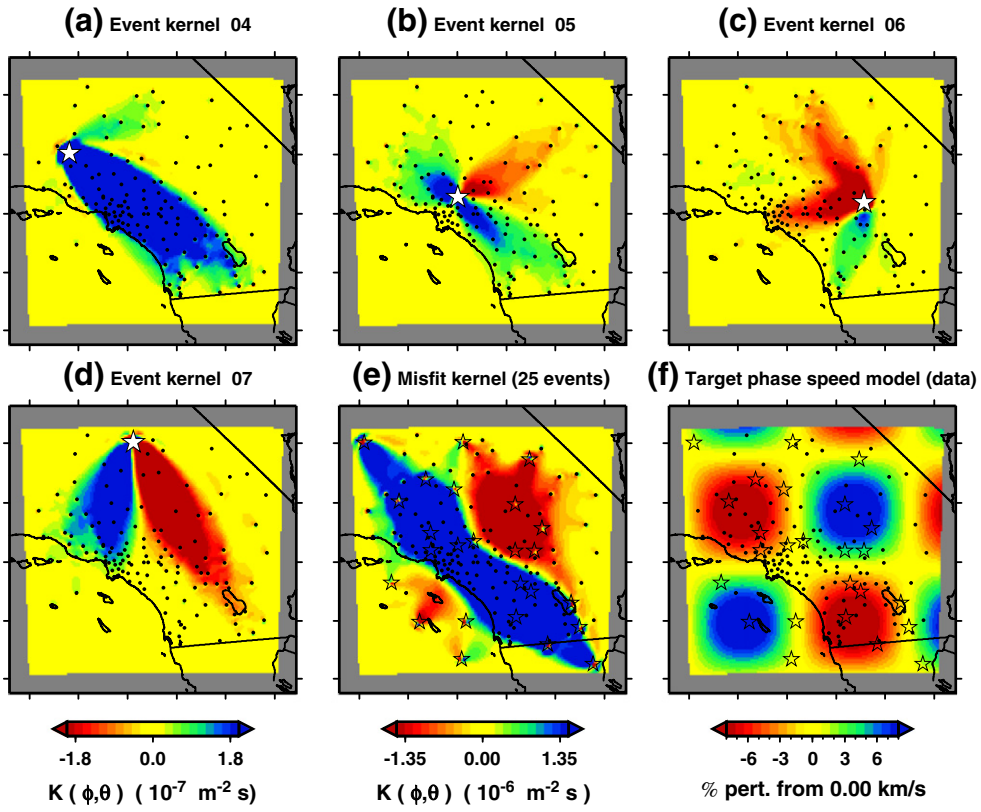

Fig. 12. Event kernels generated for a 2D synthetic example. The 'true' model shown in graph (f) is a checkerboard pattern of phase velocity perturbations from a uniform background model. The first four graphs (a–d) show event kernels for four different events (locations indicated by stars), illustrating their ability to 'see' the true model. The sum of all twenty-five event kernels (i.e., gradient vector **g**) is shown in graph (e), which indicates the direction of model update in the first iteration. Courtesy of Tape et al. (2007).

kernels for most long-period global body and surface wave phases may not differ significantly for 1D reference models and 3D tomographic models (Liu and Tromp, 2008; Zhou et al., 2011), i.e., classical tomographic inversions based on 1D reference models may still be an efficient and sufficient approach to provide velocity model updates at long wavelength scales (Capdeville et al., 2003). There has been great interest in efficient computation of sensitivity kernels for 1D velocity models based on accurate forward modeling tools such as normal-mode theory (Zhao and Chevrot, 2011a,2011b), frequency-domain algorithms (Friederich and Wielandt, 1995), Direct Solution Methods (Geller and Ohminato, 1994), algorithms that reduce 3D SEM calculations to 2D (Nissen-Meyer et al., 2007b, 2008), as well as full 3D SEM simulations (Liu, 2009). In particular, sensitivity kernels for up to 1-s teleseismic P waves can be efficiently obtained with modest computational resources (Fuji et al., 2011). This is particularly crucial for phases that cannot be treated accurately by ray theory or asymptotic methods, such as core diffracted phases that significantly deviate from assumed diffracted ray paths but are critical for resolving structures above the core–mantle boundary (Manners, 2008; Wysession, 1996; Zhao et al., 2000).

4.4.7. Scattering-integral methods

Another flavor of tomography methods based on full numerical simulations, known as *scattering-integral* (SI) methods (Chen, 2011; Chen et al., 2005, 2007a), differs from adjoint tomography by its approach in obtaining the adjoint wavefield in Eq. (24). When only a limited number of receivers are involved, Green's functions $G(\mathbf{x},\mathbf{x}_r,t)$ can be computed numerically for all receivers and stored on temporal and spatial grids. This amounts to a total of $3N_r$ number of simulations (Eisner and Clayton, 2001) with an additional N_s number of simulations for the forward wavefield $\mathbf{s}(\mathbf{x},\mathbf{x}_s,t)$. Sensitivity kernels for a single measurement pertaining to a source–receiver pair can hence be constructed based on Eq. (26), which provides extreme flexibility to compute both gradient vector **g** and individual kernels that are the building blocks for **G** and **H**; the inverse problem is then reduced to the classical case involving Eq. (30). SI methods also enable experimentation with inversion schemes, e.g., weighting schemes (W_e,W_{rp}) and exclusion/inclusion of certain sources, receivers or phases, without extra simulation for either **g** or **H**. The biggest gain is in the formal resolution analysis where checkerboard tests can now be performed to determine the resolution and error of inverted model.

SI methods are particularly appealing when the number of receivers are not excessive (i.e., $3N_r$ forward solutions can be attained in reasonable amount time), and the simulation domain is of relative small size such that $\mathbf{G}(\mathbf{x},\mathbf{x}_r,t)$ as a function of $(\mathbf{x},t)$ can be archived on hard disks for speedy access later. However, if the simulation domain is expanded, or shorter periods of seismic waves are inverted, storage may become a daunting issue even after interpolation schemes are introduced to project the simulation grid onto a coarser grid more reflective of the frequency content of the waves (Chen et al., 2007a; Zhao et al., 2006). The other disadvantage is that kernels of single measurements are computed based on the starting model, hence the Hessian matrix **H** is only valid for this model. When the actual structure deviates from the reference model significantly, e.g., up to 10 − 30% as in some local (Tape et al., 2009) and regional (Fichtner et al., 2009b) tomographic problems, one Hessian iteration may not be enough to fully eliminate the data residuals and successive iterations with Green's functions and Hessian matrices recomputed based on updated velocity models are necessary. For example, structural inversions for the Los Angeles basin in Chen et al. (2007a) showed a maximum 10% of shear-wave speed variations

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

in one quasi-Newton step, while adjoint tomographic inversions in Tape et al. (2010) displayed up to 30% perturbation after 16 adjoint iterations for a similar region. As one Hessian iteration may be equivalent to a few gradient-only iterations (usually related to the number of distinct eigenvalues of the Hessian, for example, approximately 6–7 for the 2D examples in Tape et al., 2007), the computation and storage demand for SI methods may outweigh the flexibility of Hessian inversions for regions with strong velocity contrasts. Overall, the optimal tomographic technique based on numerical simulations, either SI or adjoint methods, is problem specific, and may be subjected to change with the rapid advances in computer and storage technologies (Chen et al., 2007b).

On the other hand, the storage of strain Green's function fields for seismic stations $\nabla \mathbf{G}(\mathbf{x}, \mathbf{x}_r, t)$ for SI methods also allows rapid seismic source determinations independent of, or as a part of, structural inversions (see also Section 5.6). Stand-alone source inversions based on adjoint methods and full numerical simulations are feasible (Hingee et al., 2011; Kim et al., 2011; Liu et al., 2004), but can be numerically expensive. Therefore, source inversions based on stored Green's function databases for 3D reference models in SI methods provide an efficient alternative for improving the accuracy of earthquake source parameters in strongly heterogeneous regions (Lee et al., 2011).

5. Adjoint tomography: inversion schemes

In this section, we will discuss several aspects of adjoint tomographic inversion schemes, particularly in consideration of the intense computation and storage requirements of numerical simulations. In principle, tomographic problems can be treated as an unconstrained optimization problem, with a vast amount of literature available from the field of applied and computational mathematics (Nocedal and Wright, 2006). This concept has proved successful to some degree as demonstrated by both the elegant mathematical derivations and synthetic examples in Akcelik et al. (2002, 2003). However, applications to realistic crustal and mantle imaging require more practical considerations, e.g., choices of misfit functions, weighting schemes, preconditioners used for nonlinear conjugate techniques, as well as parameterization and regularization. Careful considerations of these subjects pay great dividends in decreasing the ill-posedness of inverse problems and speeding up the convergence of velocity models.

5.1. Misfit functions and measurements

The choices of measurements made to data and synthetics may include typical cross-correlation traveltime and amplitude measurements of body waves (Hung et al., 2000; Luo and Schuster, 1991; Sigloch and Nolet, 2006), frequency-dependent phase and amplitude measurements of surface waves (Tape et al., 2010; Zhou et al., 2004) based on multi-taper techniques (Laske and Masters, 1996; Park et al., 1987; Thomson, 1982), instantaneous phase differences and envelope ratios (Bozda et al., 2011a), and other frequency-dependent functionals such as those based on short-time Fourier transform (Fichtner et al., 2008), or generalized seismological data functional (Chen, 2011). Waveform difference between data and synthetics prevalent in exploration problems (see also Section 4.1) has also been utilized in global tomographic settings (Li and Romanowicz, 1996; Panning and Romanowicz, 2006). However, it is recognized that the relation between waveform residual $\Delta u = s - d$ and model perturbations can be highly nonlinear in the case of large phase delays (Panning et al., 2009), and the validity of Born approximation (see Eq. (22)) requires the initial velocity model to be close to the true model. In comparison, the Rytov approximation relating traveltime delay ΔT to model perturbation is quasi-linear, which is valid for traveltime delays up to a quarter of the dominant period (Chen et al., 2007b). Although attempts have been made recently to calculate derivatives $\frac{\partial \mathbf{s}}{\partial \mathbf{m}}$ more accurately by including ad-hoc path-averaged phase delays (Woodhouse and Dziewonski, 1984) for global surface waves (Panning et al., 2009; Romanowicz, 2008), most tomographic applications opt to invert traveltime or frequency-dependent phase measurements (see Section 2). Amplitude and waveform, which are also strongly affected by earthquake source parameters, are generally used to verify final inverted models (Chen et al., 2007a; Fichtner et al., 2009b; Tape et al., 2010).

Selecting proper phases and wavegroups with reasonable similarity in waveform and relatively small traveltime delays is crucial for the successful application of adjoint tomography. However, except for problems involving relatively small datasets, where manual selection and screening are possible (e.g., Chen et al., 2007a; Fichtner et al., 2009b), an automatic or semi-automatic scheme is needed to mine through the ever-growing broadband datasets. Automatic phase picking algorithms have been applied to body-wave phase selection for decades (e.g., Bai and Kennett, 2000; Earle and Shearer, 1994; Houser et al., 2008; Sigloch and Nolet, 2006). The use of 3D simulation made it possible to select windows based solely on waveform similarity and phase delays without a priori knowledge of associated phases. Maggi et al. (2009) scan seismograms for 'pulse' or 'wavegroup'-like features, and select windows that satisfy criteria on cross-correlation, amplitude ratio, and time shift between data and synthetics within the windows. This automatic scheme has been successfully applied to both global (Bozda et al., 2011b) and regional (Kim et al., 2011; Peter et al., 2011b; Zhu et al., 2012) datasets, and was a key component of the adjoint tomography for southern California crust (Tape et al., 2010).

Weights used in the misfit function (18) can be assigned based on measurement errors such as those estimated from similar measurements (e.g., *P* arrival differences at vertical and radial seismograms, Chen et al., 2007a). Alternatively, they can be associated with the remaining differences between data and synthetics after measurement corrections are applied (Tape et al., 2010), which further reinforces the requirement on window similarity. Weight functions can be also selected based on wave amplitudes within windows in order to balance contributions from different wave types (Fichtner et al., 2009b; Tape et al., 2010).

5.2. 3D initial models and parameterization

One of the main advantages of adjoint tomography compared to classical tomography is that it allows the use of 3D reference models and accurate computation of corresponding sensitivity kernels. Due to the equivalent computational cost of numerical simulations, 3D reference models could readily replace 1D reference models as the starting point of adjoint tomographic inversions. The selected 3D starting models could either originate from existing tomographic studies (Fichtner et al., 2009b) or represent an integration of existing results from tomography, local geology, and seismic reflection/refraction profiles (Komatitsch et al., 2004; Plesch et al., 2011). This approach is dubbed *3D–3D* tomographic inversion, where the two references of 3D correspond to the background velocity model and physical simulation domain, respectively. By this definition, classical tomographic studies based on 1D reference models can be generally considered as *1D–3D* inversions (Tape et al., 2007). The use of 3D initial models usually produces synthetics that more closely resemble data, which not only increases the number of windows or phases selected for iterative inversions, but also improves the applicability of Born or Rytov approximation on nonlinear misfit functions.

For global and continental scale applications, surface-wave waveforms are strongly affected by the 3D heterogeneous crust (Bozda and Trampert, 2008; Lekić et al., 2010). The simulation accuracy of surface waves is directly linked to the ability to mesh the crust accurately, even for relatively long period waves (Capdeville and Marigo, 2007). Unfortunately, honoring the Moho such as those provided by CRUST2.0 (Bassin

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

et al., 2000), e.g., adapting to the thin oceanic crust, may require small mesh sizes and significant computation cost for the forward simulation. Possible remedies are proposed by introducing a smooth and effective medium for the crust that is equivalent to the original model by applying corrections based on homogenization of varying coefficients (Capdeville and Marigo, 2007), matching observations such as local dispersion curves (Fichtner and Igel, 2008), or inverting for a smooth crustal model based on short-period group velocity dispersion maps (Lekić and Romanowicz, 2011).

Depending on the dataset and measurements used (see Section 5.1), either isotropic (Chen et al., 2007a) or anisotropic (Fichtner et al., 2010, also see Section 2.5) material properties can be targeted for the inversion procedure. Tarantola (2005) (section 6.3.3 thereof) indicated that the bulk-sound and shear-wave speeds are independent, therefore constitute a better set of parameters than the conventional compressional and shear velocities. Traveltime or phase delay measurements are less sensitive to density variations for inversions of relatively long-period waves (Tarantola, 1986). However, density kernels proved to be instructive in identifying internal sharp velocity contrasts at shorter periods (Liu and Tromp, 2006; Luo et al., 2009; Zhu et al., 2009). Attenuation structures have also been inverted in combination with elastic structures for 2D synthetic cases based on waveform misfit functions (Askan et al., 2007). However, attenuation has not been inverted for extensively in adjoint tomography due to its significant tradeoff with focusing/de-focusing on wave amplitudes. Therefore, we mainly focus on inversions of elastic velocity structures in the following sections.

5.3. Regularization

Due to the non-uniform distribution of sources and receivers and the uneven spatial coverage of sensitivities of phases or wavegroups, the inverse problem, i.e., solving linear system (30), is generally under-determined and ill-posed, while the associated solution are non-unique (Tarantola and Valette, 1982, also see Section 3). Some form of regularization is needed to stabilize the formal action of inverting the Hessian matrix. For Newton type of methods, a damping parameter is often used and its value is selected from L-curves that capture the tradeoff between model norm and reduction in misfit function value (Aster et al., 2005; Tape et al., 2007). In principle, basis functions for inversion parameters could be built on the 3D numerical simulation grids in adjoint tomography (Akcelik et al., 2002; Tape et al., 2010). However, grid spacings used in simulations (dictated by computation accuracy requirements) are generally much smaller than the actual model resolvability (dictated by the bandwidth of given dataset and source–receiver coverage), and using all grid points increases the ill-posedness of the inverse problem. Regularization may be explicitly introduced into the misfit function by providing constraints on model variations such as norm or roughness (Tarantola, 2005). Although L_2 norms are ubiquitously used in regularizing tomographic problems due to its simple formulation, L_1 norms of model variations also prove to be effective in preserving discontinuities in material properties (Akcelik et al., 2002; Askan et al., 2007; Aster et al., 2005, Chapter 7).

On the other hand, adjoint tomography relies mainly on the gradient vector of misfit function for model update at each iteration, and iterates successively towards models with improved data fits at the expense of increased model norm (or roughness). Spurious model updates in regions of inadequate source–receiver coverage are less likely to occur in the iteration process. The number of iterations is selected based on convergence criterion, which is similar to the discrepancy principle used in selecting the 'best' solution on an L-curve (Aster et al., 2005, Chapter 4). Still, implicit regularization may be introduced by simple local smoothing of gradient vectors, which also eliminates excessively large amplitudes of sensitivity kernels at sources and receivers due to singularities introduced by point moment-tensors or forces in the forward and adjoint simulations (Fichtner et al., 2009b; Liu and Tromp, 2008; Tape et al., 2007). It is worth noting that when smoothing is applied to gradient vectors, a similar smoothing operation should be applied to inverted models at each iteration in order to conform to the same data rule (i.e., Eq. (5), Chiao et al., 2010).

5.4. Step length and preconditioner

It is desirable to design inversion schemes that reduce the number of 3D numerical simulations of the seismic wave equation. When only gradient vectors are available to minimize the misfit function (18), nonlinear conjugate gradient (CG) methods could be the most natural solutions (Fletcher, 1987). In nonlinear CG methods, successive search (i.e., conjugate-gradient) directions are defined as linear combinations of the current gradient directions and previous search directions. Model update requires the computation of a step length that minimizes the misfit function in the given search direction, i.e., a line search (Nocedal and Wright, 2006). Quadratic and cubit interpolations, which require one (forward) and two (forward and adjoint) extra simulations, respectively, can be performed to obtain an optimal step length (Tromp et al., 2005). Since both schemes offer similar convergence rates (Tape et al., 2007), the quadratic approach is favored in line searches of most adjoint-type inversions when the quasi-linear misfit function could be approximated locally by a quadratic function of step length along the search direction (Fichtner et al., 2009b; Gauthier et al., 1986; Vigh et al., 2009). Alternative line search algorithms can be applied (Fletcher, 1987; Nocedal and Wright, 2006), as long as a balance is struck between extra computational cost and misfit reduction (Akcelik et al., 2002). Practical constraints, e.g., envelope misfit reduction (Fichtner et al., 2009b), may be used to guide the selection of step length. To reduce the computational cost at each iteration, a representative subset of events may be chosen for the determination of optimal step length (Bozdağ et al., 2011b; Zhu et al., 2012).

Weights defined in the misfit function (18) are effective preconditioners to the Hessian matrix that emphasize certain measurements while down-weighing others. More generally, a preconditioner $\mathbf{P}$ can be applied to Eq. (30) to give an equivalent linear system

$$\mathbf{P}\,\mathbf{H}\Delta\underline{m} = \mathbf{P}\,\mathbf{g}. \tag{32}$$

$\mathbf{P}$ is usually a symmetric positive definite matrix that helps reduce the condition number of this linear system and improve the convergence rate. In particular, the ideal form of $\mathbf{P}$ is the inverse Hessian matrix $\mathbf{H}^{-1}$ (Pratt et al., 1998), which reduces 6–7 steps of nonlinear CG iterations to a simple one-step Hessian inversion for some synthetic experiments (Chen et al., 2007b; Tape et al., 2007). As mentioned in Section 4.4.3, the $\mathbf{H}^{-1}$ preconditioner acts as a 'lens' that scales and sharpens the image of a gradient vector into an actual model update. The diagonal terms of $\mathbf{H}$ also provide natural scaling of different types of parameters as shown by Tape et al. (2007) and Kim et al. (2011). In adjoint tomography, $\mathbf{H}$ is not formulated and applying $\mathbf{H}^{-1}$ as a preconditioner is impractical. However, when sensitivity kernels of measurements can be relatively cheaply calculated by approximate methods (Hung et al., 2000; Li and Tanimoto, 1993; Zhou et al., 2004) or based on 1D background velocity models (Zhao et al., 2000, 2006), it is conceivable to apply an approximate H^{-1} as a preconditioner to speed up the convergence (Shin et al., 2001). Having realized that the Hessian matrix essentially captures information on the source–receiver geometry and measurement bandwidth, Igel et al. (1996) and Fichtner et al. (2009b) used less-elaborate preconditioners based on simple geometrical arguments to minimize the excessively large sensitivity at sources and receivers. This approach is closely linked to the diagonal of Hessian (Virieux and Operto, 2009), and may have similar effects as regularization (Tape et al., 2007). Alternatively, Tape et al. (2009, 2010) computed the 'best' model update in each iteration as a linear combination of event kernels based on a source subspace method

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

(Sambridge et al., 1991), which exploited the common features of different event kernels and effectively preconditioned the gradient vectors. In this case, no extra test solution needs to be evaluated for quadratic interpolations in the gradient direction, and the number of equivalent iterations to one Hessian-based step in the nonlinear inverse problem may be further reduced.

On the other hand, Hessian matrix and the local curvature of the misfit function can also be approximated based on previous iterations (Akcelik et al., 2003). An extra adjoint simulation used in cubic interpolation for step length provides additional local curvature information that can be incorporated into the updates of the approximate $\mathbf{H}^{-1}$ in BFGS-type quasi-Newton optimization algorithms (Epanomeritakis et al., 2008; Nocedal and Wright, 2006), also known as variable metric methods (section 3.4.3 of Tarantola, 2005). Applications of these approximate $\mathbf{H}^{-1}$ as preconditioners to gradient vectors speed up the convergence and are more advantageous than nonlinear CG methods in practice (Nocedal and Wright, 2006). In particular, in the limited-memory BFGS method (L-BFGS, Byrd et al., 1995), local curvature information is stored as a set of vectors saved over iterations without the need to form actual $\mathbf{H}^{-1}$ matrices; this approach is more adaptable to the procedures of adjoint tomography. Most importantly, the final approximate $\mathbf{H}^{-1}$ (i.e., the posterior covariance operator) provides a possible means to estimate model errors (Tarantola, 2005, more discussions in the next section).

Akcelik et al. (2002, 2003) approached the nonlinear optimization problem with Gauss–Newton methods. **H** was not formulated explicitly (for the same reasons discussed in Section 4.4.3), however, the matrix–vector product of **H** with any model update vector **q** can be calculated by two extra simulations, one forward and one adjoint. The ability to compute **Hq** allows the linear system (30) to be solved iteratively based on matrix-free Gauss–Newton–Krylov iterations. The number of inner iterations (Akcelik et al., 2002, Table 4.2) for every Gauss–Newton update is consistent with those demonstrated by Tape et al. (2007) and Chen et al. (2007b). However, the total number of outer iterations drastically exceeds those in Tape et al. (2009) and Fichtner et al. (2009b), due to the significantly larger velocity perturbation (~100%) used in their synthetic examples as well as the choice of simple unweighted waveform misfit function. In a practical tomographic problem, a reasonable 1D or 3D model from other studies can always be sought as an initial model for which velocity perturbations will be much less (e.g., <30% in regional-scale and <10% in continental-scale studies), and convergence can be achieved in significantly fewer (typically <20) iterations.

Another common practice to reduce the non-uniqueness of the inverse problem is to adopt a multi-scale approach (e.g., Akcelik et al., 2002; Askan et al., 2007) or iterative bootstrap inversion process (Pratt and Shipp, 1999). To avoid being trapped in local minima, initial solutions need to be sufficiently close to true solutions. Datasets are first filtered at longer periods and inverted for model parameters on a coarser grid corresponding to long-wavelength structures. The new model is then used as the initial model in a subsequent inversion of shorter-period data for finer-scale structures, where data residuals (such as traveltime or phase delay) will presumably have been reduced compared to those based on the original reference model. Iterative procedures therefore help resolve increasingly finer scale structures in the close vicinity of the 'true' model. Practical variations of this technique include successively decreasing the upper cutoff period over iterations (Fichtner et al., 2009b) or combining measurements from different period bands, where the longer period data dominate the first few iterations while more measurements from shorter periods are successively incorporated in later iterations (Tape et al., 2010). Fig. 13 shows the application of adjoint tomography to the inversion of southern California crustal structure (Tape et al., 2009) based on an automatic windowing scheme (Maggi et al., 2009). Large velocity variations are successfully resolved after 16 iterations as shown in the final model update (up to ±30% in the last column of Fig. 13a) and detailed 3D heterogeneous structures depicted by the final shear velocity model at shallow depth (2nd column of Fig. 13a). Waveform fits are significantly improved in successive iterations such that an extra usable window is automatically selected from the transverse component seismogram in the final iteration (see Fig. 13b).

5.5. Model verification, appraisal and resolution test

With the ability to numerically simulate wave propagation in 3D, adjoint tomography (as well as scattering-integral methods) can verify the final velocity model in terms of both reduction of measurement residuals and waveform fit itself. Tape et al. (2010) computed synthetics for a set of independent 'extra' events that were not used in the inversions for their final tomographic model. Reductions in both phase delay times and waveform misfits are statistically similar to those for events used in inversions. For inversions based on traveltime or phase delay measurements, amplitude measurements also provide independent datasets to validate the final velocity models (Chen et al., 2007a; Fichtner et al., 2009b; Tape et al., 2010). Usually more phases and windows are selected as iteration progresses, reflecting that the inverted models successively approach the true model, and the final model may produce synthetics that offer significantly more usable windows or phases under the same selection criteria (see Fig. 13b). This practice effectively improves data coverage.

The diagonal terms of the **H** matrix provide information on how well a domain is sampled based on the given source–receiver geometry and selected phases/wavegroups. Similar to the ray density maps prevalent in classical tomography, volumetric coverage kernels can be constructed by omitting measurement residuals in the adjoint sources defined in Eq. (24). It provides an ad hoc way of understanding model resolvability and offers guidance to the interpretation of final models (Tape et al., 2010). However, formal model error analysis and resolution tests common in classical tomography (as discussed in Section 3.1) are much more difficult for adjoint tomography. As the formal resolution matrix of model update in each Newton step of a Hessian-based inversion is given by

$$R = \left(\mathbf{G}^T\mathbf{G} + \Gamma\right)^{-1}\left(\mathbf{G}^T\mathbf{G}\right), \tag{33}$$

where Γ is a matrix representing regularization applied, the lack of individual sensitivity kernels (i.e., row vectors of **G**) in adjoint tomography makes it impossible to compute the full resolution matrix. Even checkerboard tests common in classical tomography require the full adjoint inversion machinery when only gradient vectors are available, i.e., one checkerboard test is numerically as costly as one full 3D adjoint inversion. Nevertheless, attempts were made to design checkerboard tests to evaluate model resolutions at both long and short length scales. Fig. 14 illustrates checkerboard tests performed by Fichtner et al. (2009b) for adjoint tomographic inversions of the Australasian region. The input model ('true' model) is a checkerboard pattern ($3^\circ\times3^\circ$) onto a high-velocity patch (~$25^\circ\times25^\circ$) mimicking positive velocity anomalies in the central and western part of the region. When velocity perturbations of the checkerboards are set to be 3%, 6% and even as large as 9%, both the long and short wavelength structures are well recovered by adjoint inversions except in regions with limited ray coverage. In classical tomography, the checkerboard test is simply a machinery to gauge the resolution of the design matrix (**G**), where the linearization of measurements with respect to model parameters has already been made. In contrast, these checkerboard tests compute synthetic data based on full 3D numerical simulations, thus providing metrics on the effect of source–receiver and kernel coverages as well as the nonlinearity of the misfit function, particularly in the presence of strong velocity variations.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

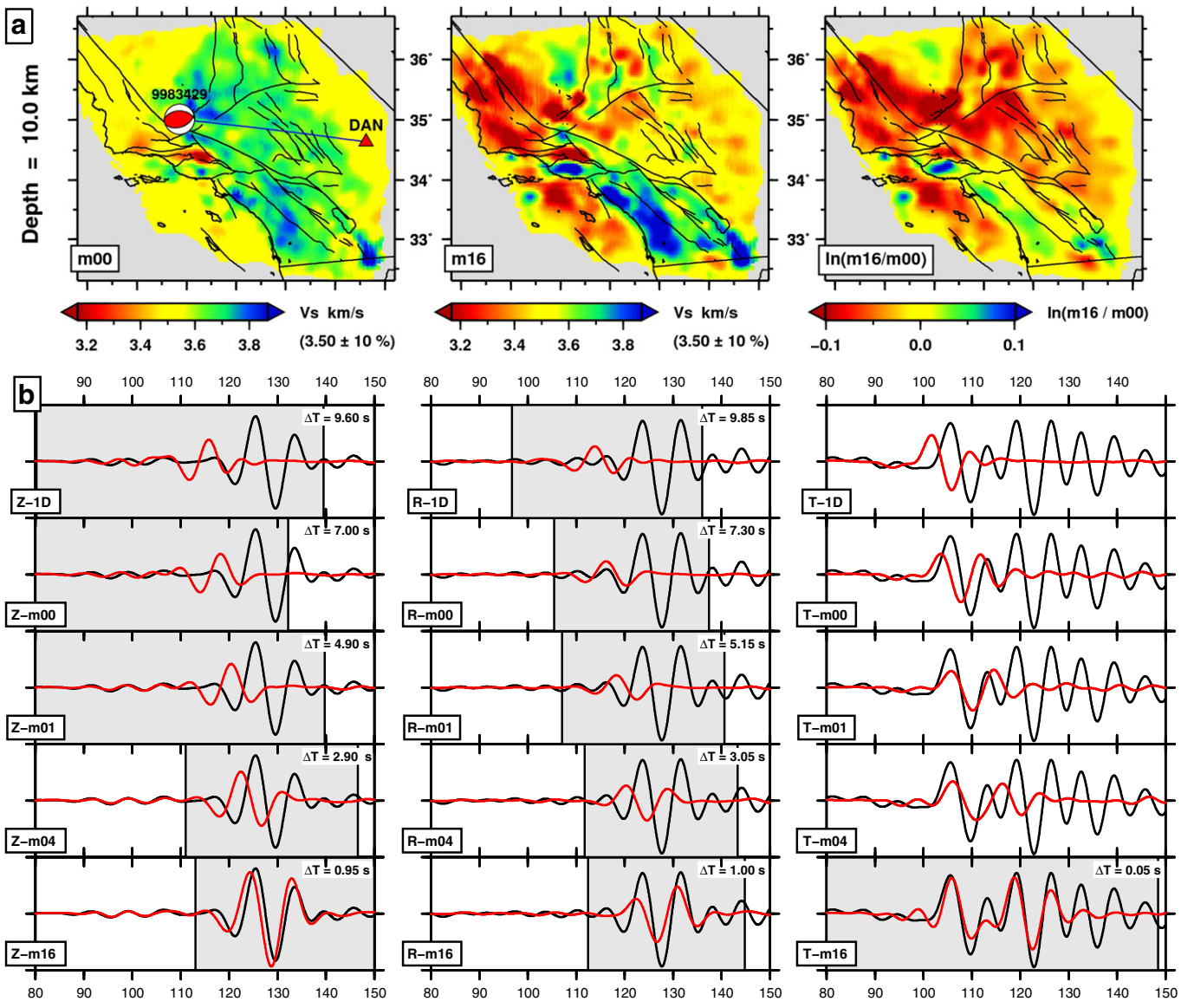

Fig. 13. Adjoint tomographic inversions of southern California crust. (a) Comparisons of the initial model, final model after 16 iterations, and model update of shear velocities at the depth of 10 km. The final model exhibits velocity perturbations up to ±30% and depicts geological features such as sedimentary basins, exhumed batholiths and lithological contrast across faults. Event 9983429 is denoted by a beach ball and station DAN is denoted by a red triangle. (b) Evolution of waveform fits between data and synthetics at station DAN for event 9983429 over iterations. Both are filtered between 6 and 30 s. Windows automatically picked by FLEXWIN algorithm (Maggi et al., 2009) are indicated by the gray boxes. Note the successive improvement in waveform fits from the initial model m_{00} to the final model m_{16}, and the extra window picked on the transverse component after 16 iterations. Courtesy of Tape et al. (2009).

Realizing that the inverse Hessian matrix $\mathbf{H}^{-1}$ provides an approximation to the covariance matrix of inverted model parameters assuming individual data are uncorrelated (Tarantola, 2005), Stadler (2010) made efforts to estimate significant eigenvalues and the corresponding eigenvectors of the **H** matrix through numerical methods (e.g., Lancos algorithm) based solely on matrix–vector product of **H** with a given directional vector **q** (Saad, 2011). In this case, the computation of **Hq** can be achieved by an extra forward and adjoint simulation (Akcelik et al., 2002). However, hundreds of such operations may be required to obtain significant eigenvalues of **H**, which is still impractical for most realistic applications. Alternatively, the approximate $\mathbf{H}^{-1}$ matrix saved over iterations can be naturally used as an approximate posterior model covariance matrix if a cubic interpolation scheme is used for step-length selection (Tromp et al., 2005) and BFGS-type of inversion schemes are selected (see also Section 5.4). The diagonal elements of the approximate $\mathbf{H}^{-1}$ provide error-bars of model parameters, while the off-diagonal elements are associated with correlations between different model parameters. Only one extra adjoint simulation is required for each iteration to construct the approximate $\mathbf{H}^{-1}$ matrix, and the ultimate computation cost may even be less than those of nonlinear CG methods due to improved convergence of BFGS-type of inversion algorithms and potentially more accurate line-search based on cubic interpolation. Therefore BFGS-type of inversion schemes may hold promise to both improved convergence and model resolution analysis in the future. To effectively sample the posterior probability distribution of model parameters, access to the square-root of $\mathbf{H}^{-1}$ is required (Tarantola, 2007) which prompts the use of square-root variable metric inversion techniques (Hull and Tapley, 1977).

An apparent advantage of scattering-integral methods over adjoint methods is the ability to perform a formal resolution analysis, thanks to the availability of the entire design matrix **G** (Chen et al., 2007a). The effect of nonlinearity of the misfit functions will also be reflected in checkerboard tests as synthetics are calculated numerically for 3D synthetic models. However, as indicated in Section 4.4.7, more than one scattering-integral iteration may be necessary to achieve further

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

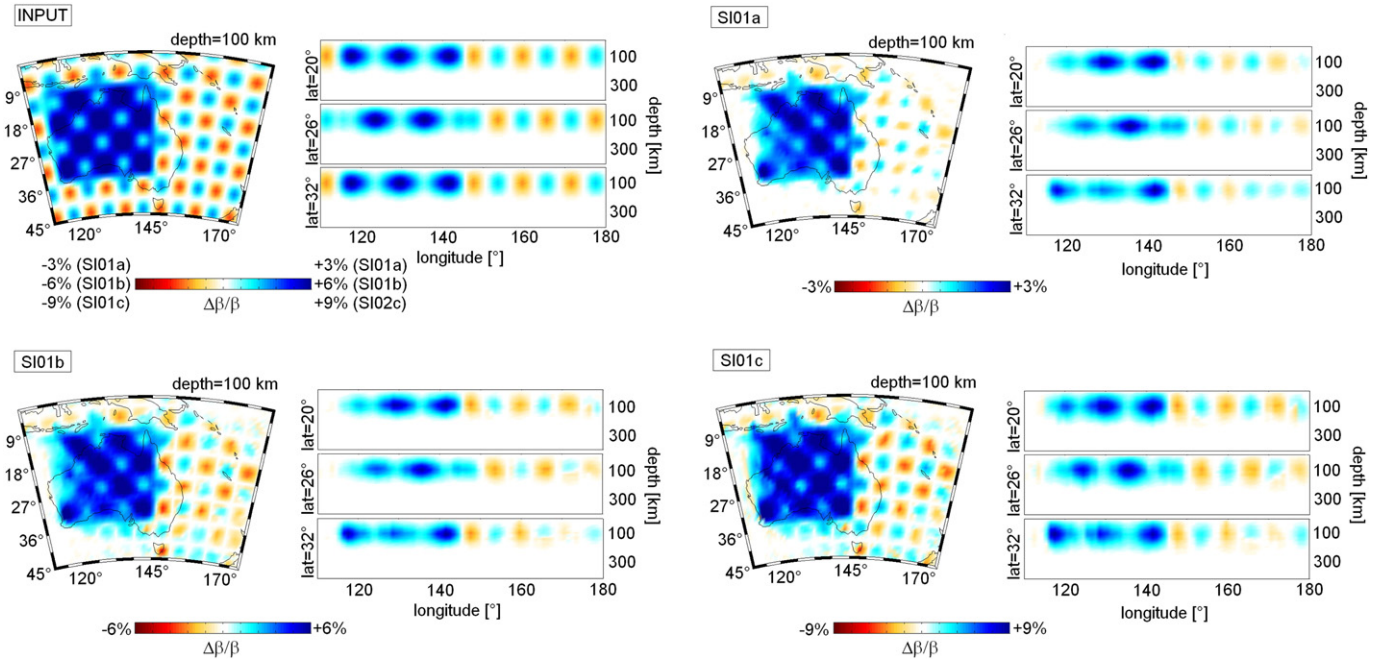

Fig. 14. Resolution tests for adjoint tomographic inversions of the Australasian region. Initial models are 3°×3° checkerboard patterns of various velocity perturbations (3%,6%,9%) superposed onto a 25°×25° high-velocity patch. The 1D background velocity model suitable for this regional is given by Fig. 7 of Fichtner et al. (2009b). Resolution tests undergo the same adjoint tomographic procedures as the actual 3D inversions and recover both the high velocity patch and checkerboard patterns of three different perturbations in regions with sufficient source–receiver coverage.
Courtesy of Fichtner et al. (2009b).

misfit reduction and convergence for regions with strong velocity contrasts.

5.6. Source inversions

Because the misfit between data and synthetics is caused by the intertwined effects of unknown source and structural variations, accurate knowledge of earthquake source parameters is critical in achieving unbiased model updates in tomography iterations. Similar to classical tomographic inversions (e.g., Manners, 2008; Roecker et al., 2006), adjoint tomographic inversions may explicitly include source parameters as part of the inversion parameter set (Chen et al., 2007a). However, for most cases, alternate source and structure inversions are performed where updates in one help reduce the bias in the other (Fichtner et al., 2009b; Tape et al., 2009). Since model updates are moderate between conjugate gradient iterations, source inversions for location and/or mechanisms only need to be performed at selected iterations (Tape et al., 2010). Even though Akcelik et al. (2003) inverted both a simple finite-fault source model and structural parameters from displacement recordings of a 2D synthetic example, source-time history of events are not inverted in most practical tomographic problems due to the severe nonlinearity of the misfit functions. Events are carefully selected based on magnitudes and source durations are generally much smaller than the main period bands of the data used for structural inversions (Fichtner et al., 2009b; Tape et al., 2010). In practice, finite-fault inversions based on adjoint techniques (Kremers et al., 2011) can be sufficiently decoupled from structural inversions and represent simple extensions of the popular back-projection techniques for fault rupture imaging (e.g., Ishii et al., 2005; Larmat et al., 2008).

6. Adjoint tomography: discussions and future directions

Here we summarize major advantages of adjoint tomography methods. The use of accurate Green's function through 3D full numerical simulations of seismic wave propagation offers accurate kernels of selected windows on seismograms. No prior knowledge of phases related to these time windows is assumed, which allows all possible time windows (either as impulses or wave groups) to be included in adjoint inversions. Compared to classical tomography, in which only few known phases are selected and measured, more usable windows selected for adjoint tomography effectively increase the ray-path coverage under the same source–receiver geometry. Another important consequence is that 3D models, inverted by ray- or finite-frequency-based classical tomographic techniques, can now be exploited as initial models for adjoint tomography since the simulation time of 3D models is similar to that of 1D models. Successive linearization of the inverse problem at updated models and the ability to recalculate gradients of misfit functions in every iteration are critical for the ultimate model convergence, particularly for regions with large velocity variations. Synthetics for the final models can be computed again by 3D numerical simulations, thus enabling true validation on final models.

At regional and global scales where velocity variations are moderate, hybrid quasi-Newton methods have been proposed whereby misfit functions are evaluated by accurate 3D numerical simulations and an approximate partial derivative matrix **G** is computed based on asymptotic methods (Lekić and Romanowicz, 2011; Yuan et al., 2011). The accurate 3D Green's functions reduce the effect of heterogeneous crust on surface waves and systematic errors associated with measurements, while the approximate **G** matrix exonerates the inversion from costly gradient calculations required by adjoint tomography. However, in the case of strong velocity heterogeneities, the asymptotic approximation of the **G** matrix may break down and a less accurate **G** matrix will require a significantly greater number of iterations for the final convergence. Despite the high computation cost, adjoint tomography may be the ultimate tool to produce new generation of high-resolution images for strongly heterogeneous regions such as the crust with large Moho and/or surface topography variations (Tape et al., 2010), or tectonically active regions such as subduction zones (M. Chen et al.,

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

2007), hotspots (Montelli et al., 2004) and continental collision zones (M. Chen et al., 2010).

Adjoint tomography also has its own share of challenges in addition to high computational cost. Localized gradient methods require initial solutions to be sufficiently close to the ground truth so that final solutions are not trapped in local minima of model space. This condition in turn requires either well-educated initial guesses of the media or well-designed misfit functions that reside within the linear regime. Convergence can be extremely slow if the gradient scheme is not properly preconditioned. Designing the appropriate preconditioner for seismic inverse problems involving millions of model parameters can also be a challenging task, although the end result could lead to significant reduction in the required number of iterations. Empirical damping in classical tomography is replaced by implicit regularization of the gradient vector. The sum of sensitivity kernels of residual measurements designates regions of models that require updates, rendering the inversion process stable. However, without the Hessian matrix (or its inverse), formal resolution analysis becomes difficult. Standard checkerboard test which provides information on the resolvability of anomalies of various sizes requires solving the full optimization problem again, and therefore is computationally infeasible unless numerical simulations become much less costly with hardware improvement in the future (Komatitsch et al., 2009, 2010). The BFGS-type of inversion schemes with approximated inverse Hessian matrix may be a viable remedy at this stage. Formal checkerboard test and resolution analysis for adjoint tomography remain areas of active research (Fichtner and Trampert, 2011a, 2011b).

With ever-expanding computing power, applications of adjoint tomographic inversions based on full 3D numerical simulations flourished in the past five years. They generated and are in the process of generating robust and high-resolution images of various regions of the Earth's interior at local (Chen et al., 2007a; Tape et al., 2010), regional (Fichtner et al., 2009b, 2010; Peter et al., 2011b; Zhu et al., 2012) and global (Bozda et al., 2011b) scales. Applications to other geologically and tectonically interesting regions are also on the horizon (M. Chen et al., 2010), especially with the continued improvement in numerical techniques, computational infrastructure and inversion schemes adapted to local-gradient based inversions. Combined with geodynamical modeling, the interpretation of these improved high-resolution images will further our understanding of the temperature, composition and dynamics of the Earth's interior.

In addition, with the proliferation of both permanent (e.g., Chen et al., 2006; Gee et al., 1996; Kanamori et al., 1991; Okada et al., 2004) and transportable arrays (e.g., Levander et al., 1999; Nábelek et al., 2009; Velasco et al., 2007), the concept of adjoint tomography could be highly valuable to the analysis of various datasets in array seismology (Gu, 2010; Levander and Nolet, 2005; Rost and Thomas, 2002, 2009). Modeling of ambient noise correlation functions and scattered/converted waves of teleseismic phases are two prime examples that will benefit from the framework of adjoint tomography. The following sections briefly review these two types of measurements and discuss how they may be naturally incorporated into adjoint tomography.

6.1. Noise correlation tomography

Since the late 1990s, techniques based on the cross-correlation of ambient noise fields have been overwhelmingly successful for a broad spectrum of problems in physics, geophysics (both global and exploration-based), engineering, acoustics and medical imaging (see reviews by Weaver and Lobkis, 2004; Snieder et al., 2009). Data and applications vary substantially among the various disciplines, but the vast majority of the studies aim to complete one or both of the following tasks: 1) to extract the Green's function between two receivers by cross-correlating their noise records (Aki, 1957; Shapiro and Campillo, 2004), and 2) to locate and characterize the source of dominant noise signals (e.g., Gu et al., 2007; Sabra et al., 2005; Yang et al., 2008). When the noise sources are uniformly distributed (or, *ambient*), it has been shown that the cross-correlation (or its derivative) of noise records at two nearby receivers is symmetric about the zero lag time and equivalent to the Green's function (Snieder, 2004). This result forms the basis for an increasing number of high-resolution, noise-based images of regional seismic structure; successful examples include California (e.g., Sabra et al., 2005; Shapiro et al., 2005), North America (e.g., Lin et al., 2008, 2010; Moschetti et al., 2007, 2010; Yang et al., 2008, 2011), China (e.g., Yao et al., 2006; Zheng et al., 2008), Northeast Asia (e.g., Cho et al., 2007; Nishida et al., 2008; Zheng et al., 2011), Africa (e.g., Yang et al., 2008), Europe (e.g., Villaseñor et al., 2007; Yang et al., 2007) and Oceania (e.g., Saygin and Kennett, 2012).

The theory behind noise correlation underwent stages of development since Aki published his seminal studies on the theory and auto/cross-correlation of microtremors (e.g., from traffic signals in Tokyo, Aki, 1957) over a half century ago (Aki 1956, 1957). The concept of correlation-Green's function equivalence was reiterated through autocorrelations of plane waves at different depths in a layered medium, later known as the *Clarebout conjecture* (e.g., Claerbout, 1968; Rickett and Claerbout, 1999), and the presence of seismic response in the time-domain cross-correlation function was validated both in theory (Lobkis and Weaver, 2002) and in observations of unfiltered seismic coda waves (Campillo and Paul, 2003). Subsequent studies using reciprocity and scattered waves (Snieder, 2004; Wapenaar, 2004) led to increased confidence in tomographic maps under the assumption of ambient noise sources.

Technically, for a spatially uniform broadband noise distribution and uniform medium with velocity v, the time derivative of the noise correlation function (NCF) of two receivers separated by a distance L can be approximated by the sum of the time-domain Green's functions from the forward and time-reversed wavefields (Roux et al., 2005; Sabra et al., 2005; Snieder, 2004). Mathematically, this is expressible as (Sabra et al., 2005; Weaver and Lobkis, 2003),

$$\frac{dC_{ij}}{dt} = -G_{ij}(r_1; r_2, t) + G_{ji}(r_2; r_1, t), \tag{34}$$

where C_{ij} corresponds to the NCF for the i-th station and j-th component. The time-domain Green's function $G_{ij}(r_1; r_2, t)$ is the displacement response in direction j at location r_2 to a unit concentrated impulse input in direction i at location r_1 (Sabra et al., 2005). The above approximation presents an effective means to extract the Green's function based on the derivative of the correlation function, or occasionally by the correlation function between two stations (e.g., Bensen et al., 2007; Campillo and Paul, 2003; Shapiro et al., 2005). The lag times of the correlation peaks offer reliable constraints on regional shear/Rayleigh wave speeds (e.g., Bensen et al., 2007; Shapiro et al., 2005; Yang et al., 2008; Yao et al., 2006), anisotropy (e.g., Lin et al., 2009; Moschetti et al., 2010; Yao and van der Hilst, 2009; Yao et al., 2010) and attenuation (Lawrence and Prieto, 2011; Levshin et al., 2010; Lin et al., 2011) down to ~100 km depth.

Aside from uncertainties associated with the 'ambient noise' assumption (e.g., Brzak et al., 2009; Stehly et al., 2006; Yang and Ritzwoller, 2008), noise correlation tomography is, in principle, no different from earthquake tomography. For this reason, both classical and adjoint tomographic approaches pertaining to earthquake observations are equally valid in the extraction of structural parameters from ambient noise. Furthermore, Tromp et al. (2010b) computed sensitivity kernels for ensemble-averaged noise correlation measurements based on the interaction of an ensemble forward wavefield and an ensemble adjoint wavefield, similar to methods detailed in Section 4.4.4. By injecting a signal with the power spectrum of the noise at the first station, the ensemble forward calculation first simulates and stores a generating field at the source region of the noise (e.g., the surface of the Earth), and then uses the generating field as virtual sources for another forward calculation. The ensemble adjoint field is simulated

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

by injecting the time-reversed measurement between synthetics and observed ensemble-averaged correlation functions at the second station as fictitious sources. Kernels calculated based on the interaction of these two ensemble wavefields for surface-wave cross-correlation functions sometimes display 'extruding parabolic jets' in addition to the typical 'banana-doughnut' shape (see also Section 3.2), which may be explained by sensitivities of multi-orbit surface waves to both minor and major arcs. Though the kernel calculation for ensemble-averaged noise measurements require an extra forward simulation and the storage of the generating wavefield, it also allows the implementation of nonuniform noise sources. Therefore tomographic inversions based on properly calculated sensitivity kernels for ensemble-averaged correlation functions may provide a greater understanding of structures embedded in the ambient noise as well as the nature of the noise sources.

6.2. Receiver functions and inverse scattering methods by passive seismic arrays

Regional-scale, high-resolution imaging has become popular in the past two decades, thanks to the proliferation of dense seismic arrays deployed around the world. In order to take advantage of dense station spacing, which varies from tens to hundreds of kilometers, seismic imaging methods have been adapted to map crust and mantle structures at passive receiver arrays. For example, traditional body-wave tomographic techniques have been modified for passive seismic array analysis by measuring multi-channel cross-correlation traveltimes of teleseismic waves between stations (e.g., Vandecar and Crosson, 1990), and attribute traveltime residuals solely to velocity anomalies underneath the observing array; surface-wave phase velocity maps are more frequently determined in combination with incident wavefields based on two- or multi-plane wave assumptions for passive seismic arrays (e.g., Forsyth et al., 1998; Friederich and Wielandt, 1995; Pollitz, 1999; Yang and Forsyth, 2006). These techniques led to hundreds of regional tomography studies and greatly improved knowledge on regional velocity structures (e.g., Chang and van der Lee, 2011, Dando et al., 2011; Hung et al., 2010; Mercier et al., 2009; Obrebski et al., 2011, Waldhauser and Tolstoy, 2011, Yang and Forsyth, 2006).

In addition to seismic tomography, amplitudes, arrival-times and waveforms of reflected, converted and scattered seismic waves from interfacial or heterogeneous structures are effective measures of fine-scale structures and changes in rock elastic properties beneath seismic arrays. As stated in Section 2.4.1, inclusion and modeling of these secondary phases are a necessity, rather than a luxury, for a greater accuracy in crustal and upper-mantle imaging/inversion. The list of popular secondary phases that may benefit from seismic arrays includes *ScS* reverberations (e.g., Gaherty and Jordan, 1995; Revenaugh and Jordan, 1989), *SS* and *PP* precursors (Shearer, 1991; Shearer and Masters, 1992), *P*-to-*S*/*S*-to-*P* converted waves (Langston, 1977; Vinnik, 1977) and *P*/*S* wave codas (Langston, 1989; Langston and Hammer, 2001). The first two phase groups are particularly well suited for global mapping of mantle stratification (see Section 2.6; Deuss, 2009; Flanagan et al., 1998; Gu et al., 1998; Shearer, 1993), though their dominant frequencies (<0.1 Hz) are less than ideal for resolving structural variations with characteristic length scales less than 300 km (Neele et al., 1997). This section will emphasize the key properties and future prospect of imaging based on scattered and converted waves, a phase group that exhibits broad frequency spectra and strong sensitivities to both crust and mantle structural variations at variable length scales.

Studies of converted waves have greatly benefited from the *receiver-function* techniques, which was introduced in the late 70s (e.g., Langston, 1977; Vinnik, 1977) and gradually became a routine method in seismic array analysis (e.g., Chen et al., 2009; Levander et al., 2005; Rondenay, 2009). Converted teleseismic *P* waves at the Moho (such as *Ps* phase) and later surface reverberations and multiple converted waves (such as *PpPs*, *PpSs* + *PsPs*, etc.) are primarily observed on the radial component. Therefore, in principle, deconvolution of the vertical (*Z*) from the radial (*R*) component will remove the source signature and isolate the receiver function response due to mode conversions; a rotation from *Z*–*R* to the *P*–*SV* ray coordinate system is often applied, and the *SV* component is subsequently deconvolved from the *P* component to obtain the receiver function. In reality, *P* coda waves (i.e., vertical receiver functions) may also contain information about the subsurface scatterers (e.g., Langston and Hammer, 2001; Pavlis, 2005), and the effective removal of source signature remains an open-ended topic of discussion (e.g., Li and Nfibelek, 1999; Langston and Hammer, 2001; Bostock, 2004; Escalante et al., 2007; C.W. Chen et al., 2010).

Based on a similar approach, teleseismic *S* receiver functions can be calculated for the analysis of *S*-to-*P* conversions below station (or source) locations (Farra and Vinnik, 2000; Vinnik and Farra, 2002; Wilson et al., 2006). Unlike *P*-receiver functions, *S*-receiver functions are not prone to interference from multiple *P* waves, hence are highly informative and complementary in mapping mantle stratifications such as the lithosphere–asthenosphere boundaries (LAB; e.g., Rychert and Shearer, 2011; Rychert et al., 2005) and transition zone discontinuities. However, *S*-receiver functions can be extracted only for a limited range of epicentral distances (Yuan et al., 2006), and therefore provide less data than *P*-receiver functions.

Local crustal thickness and Poisson's ratio can be estimated from receiver functions at individual stations (e.g, Niu and James, 2002; Yan and Clayton, 2007; Zhu and Kanamori, 2000), while 2D/3D subsurface image sections can be constructed through migration and stacking in an array setting (e.g, Dueker and Sheehan, 1997; Levander et al., 2005; Wilson and Aster, 2003). Transverse component seismograms could also be utilized to constrain the seismic anisotropy in the crust and mantle (Levin et al., 2008). In addition to arrival times and amplitude information, receiver functions may be inverted for physical structural parameters (velocity and density perturbations) based on inverse scattering techniques widely used in exploration seismology, e.g., the generalized Radon transform (e.g., Bostock and Rondenay, 1999; Bostock et al., 2001; Rondenay et al., 2005; Shragge et al., 2001).

Other minor phases surrounding the main arrivals, e.g., coda waves and precursors, provide further information on potential scatterers and interfaces at depth. For instance, van der Hilst et al. (2007) and Wang et al. (2008) used scattered *ScS* precursor and *SKKS* coda waves to detect multiple piecewise continuous interfaces in the D″ layer beneath Central and North America, while Cao et al. (2011) and Gu et al. (2012) used *SS* precursors to image the upper-mantle discontinuities beneath the central and western Pacific ocean, respectively.

Adjoint tomography can be readily adapted to robust high-resolution imaging of localized structures based on converted or scattered waves. Isochrons (e.g., Bostock et al., 2001) may be replaced by sensitivity kernels that are constructed based on the interaction of forward teleseismic wavefield corresponding to the main phase and adjoint wavefield generated by time-reversing converted or scattered waves. Constructing the forward teleseismic wavefield at the receiver region by propagating the field from the earthquake source to the receiver array is too costly, as numerical simulations of wave propagation accurate to typical *P* (1–2 s) and *S* (3–6 s) receiver-function periods are far from practical (Komatitsch et al., 2010). However, if one assumes that scattered waves are only caused by local structures directly underneath receiver (for receiver functions) or reflection points (for *PP* and *SS* reverberations), then a significantly smaller section of the teleseismic wavefield centered on the area of imaging interest would be needed in the kernel computations. Hybrid methods can be applied to simulate wave propagation in the 3D local media in which initial and boundary conditions are exchanged with rapidly computed forward wavefield for 1D background models (Bielak and Christiano, 1984; Bielak et al., 2003; Bouchon and Sanchez-Sesma, 2007; Godinho et al., 2009; Moczo et al., 1997; Opršal et al., 2009; Robertsson and Chapman,

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

2000). For receiver function studies of local arrays where incident teleseismic wavefield can be well approximated by plane waves (Liu and Chen, 2011), forward wavefield for a 1D background medium can be easily computed by semi-analytical methods for layered media (Aki and Richards, 2002; Zhu and Rivera, 2002). On the other hand, fast and accurate semi-analytical methods for 1D global background models have to be employed for large aperture arrays where the flat-earth approximation is no longer valid (Chevrot et al., 2011).

Similar to adjoint tomographic methods in global and regional studies, the advantages of inverting converted/scattered waves based on adapted adjoint tomography for passive arrays are many fold. First, the use of full 3D numerical simulations, i.e., accurate solutions to the full wave equations with plane wave or teleseismic wave inputs, enables the incorporation of all possible phases into the same time windows. For example, when interfaced with 1D global models, sensitivity kernels for time windows including scattered and converted *P* waves properly account for the incoming multiple *P* phases, thereby their use reduces artifacts in the interpretation of reflecting interfaces and scatterers. Furthermore, these sensitivity kernels are three-dimensional by construction and, as discussed in Section 4.1, not only exhibit a similar effect as Kirchhoff migration in the standard receiver function analysis (e.g., Levander et al., 2005), but also provide directions of actual physical model updates as in the inverse scattering techniques (e.g., Bostock et al., 2001; Rondenay et al., 2005). In addition, based on hybrid methods, sensitivity kernels can also be computed for 3D background models produced by previous tomographic studies based on measurements from local earthquakes as well as noise/multi-channel cross correlations. With 3D background models, interfaces may be mapped to more precise depths and the resulting image may appear sharper. Other advantages such as model iterations and validations are also carried over.

In short, adjoint tomography for converted, reflected and scattered-waves at passive arrays may become a vital tool in high-resolution imaging of tectonically and geodynamically interesting regions in the near future.

7. Conclusions

Seismic tomography has made significant strides since its practical applications in the late 70s. The increasingly detailed images of the crust and mantle greatly improved our knowledge of the Earth's internal structures and dynamics, as well as their potential connections to surface tectonics. The continued improvement in image resolution can be largely attributed to advances in instrumentation, proliferation of permanent regional seismic networks, and numerous deployments of transportable seismic arrays. While uneven spatial data coverage remains a major challenge in further improving seismic imaging resolution, theoretical and computational advances, the focus of this review, made it possible to interpret the fast-expanding datasets with greater accuracy. Among the various classical and current methods, we highlighted adjoint tomography as a promising tool for a new generation of regional and global imaging efforts. By taking advantage of accurate 3D forward simulations, this inversion method could resolve strongly heterogeneous structures with improved resolution.

The future of seismic imaging also rests on our abilities to develop and promote innovative approaches to extract more information from seismic recordings and reliably interpret them. While only a handful of such new and exciting datasets were highlighted in this review, it should be recognized that inversions based on a wide range of data could benefit from the framework of adjoint tomography and 3D numerical simulations. The outcomes of these inversions could provide further insights into regional and global processes at resolution and precision far exceeding those achievable by standard tomographic techniques.

Acknowledgment

First and foremost, we would like to thank Timothy Horscroft and Hans Thybo from Tectonophysics for the invitation to contribute to this review. They have shown extraordinary patience and understanding in the past 1.5 years during the drafting and reviewing of this manuscript. We thank Po Chen and Andreas Fichtner for their detailed and encouraging reviews. We thank Carl Tape, Andreas Fichtner for kindly providing some of the figures. We also would like to thank Jeroen Tromp, Guy Masters, Shu-huei Hung, Yingjie Yang, Dimitri Komatitsch, Li Zhao, Chin-wu Chen, Ling-yun Chiao, Huajian Yao, Adam Dziewonski and the late Albert Tarantola for their inspiring and fruitful discussions. Many figures were originally made using the GMT software (Wessel and Smith, 1991). Q. Liu is supported by the Discovery Grants of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Toronto Startup Fund. Y. J. Gu is currently supported by NSERC and Alberta Geological Survey.

References

Acton, F.S., 1997. Numerical Methods that Usually Work. The Mathematical Association of America.
Akcelik, V., Biros, G., Ghattas, O., 2002. Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation. Proceedings of the 2002 ACM/IEEE Conference on Supercomputing, pp. 1–15.
Akcelik, V., Bielak, J., Epanomeritakis, I., Biros, G., Ghattas, O., Fernandez, A., Kim, E.J., Lopez, J., O'Hallaron, D., Tu, T., Urbanic, J., 2003. High resolution forward and inverse earthquake modeling on terascale computers. Proceedings of the 2003 ACM/IEEE Conference on Supercomputing, p. 52.
Aki, K., 1956. Correlogram analyses of seismograms by means of a simple automatic computer. Journal of Physics of the Earth 4, 71–79.
Aki, K., 1957. Space and time spectra of stationary stochastic waves with special reference to microtremors. Bulletin of the Earthquake Research Institute 35, 415–456.
Aki, K., 1968. Seismological evidence for the existence of soft thin layers in the upper mantle beneath Japan. Journal of Geophysical Research 73, 585–594.
Aki, K., Lee, W.H.K., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes, 1. A homogeneous initial model. Journal of Geophysical Research 81, 4381–4399.
Aki, K., Richards, P.G., 2002. Quantitative Seismology, 2nd edition. University Science Books.
Albarede, F., van der Hilst, R.D., 2002. Zoned mantle convection. Philosophical Transactions of the Royal Society of London, Series A 360, 2569–2592.
Anderson, D.L., 1961. Elastic wave propagation in layered anisotropic media. Journal of Geophysical Research 66, 2953–2963.
Anderson, D.L., 1966. Recent evidence concerning the structure and composition of the Earth's mantle. Physics and Chemistry of the Earth 6, 1–131.
Anderson, D.L., 1989. Theory of the Earth. Blackwell Scientific Publications.
Anderson, D.L., Dziewonski, A.M., 1982. Upper mantle anisotropy: evidence from free oscillation. Geophysical Journal of the Royal Astronomical Society 69, 383–404.
Ando, M., Ishikawa, Y., Wada, H., 1980. S-wave anisotropy in the upper mantle under a volcanic area in Japan. Nature 286, 43–46.
Antolik, M., Gu, Y.J., Ekström, G., Dziewonski, A.M., 2003. J362D28: a new joint model of compressional and shear velocity in the Earth's mantle. Geophysical Journal International 153, 443–466.
Askan, A., Akcelik, V., Bielak, J., Ghattas, O., 2007. Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures. Bulletin of the Seismological Society of America 97, 1990–2008.
Aster, R.C., Borchers, B., Thurber, C.H., 2005. Parameter Estimation and Inverse Problems. Elsevier Academic Press.
Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering. Journal of Geophysical Research 67, 4427–4440.
Backus, G.E., 1964. Geographical interpretation of measurements of average phase velocities of surface waves over great circular and great semi-circular paths. Bulletin of the Seismological Society of America 54, 571–610.
Backus, G.E., Gilbert, F., 1967. Numerical applications of a formalism for geophysical inverse problems. Geophysical Journal of the Royal Astronomical Society 13, 247–276.
Backus, G.E., Gilbert, F., 1968. The resolving power of gross earth data. Geophysical Journal of the Royal Astronomical Society 16, 169–205.
Backus, G.E., Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data. Philosophical Transactions of the Royal Society of London, Series A 266, 123–192.
Bai, C.Y., Kennett, B.L.N., 2000. Automatic phase-detection and identification by full use of a single three-component broadband seismogram. Bulletin of the Seismological Society of America 90, 187–198.
Bao, H., Bielak, J., Ghattas, O., Kallivokas, L.F., O'Hallaron, D.R., Shewchukb, J.R., Xu, J., 1998. Large-scale simulation of elastic wave propagation media on parallel computers. Computer Methods in Applied Mechanics and Engineering 152, 85–102.
Bassin, C., Laske, G., Masters, G., 2000. The current limits of resolution for surface wave tomography in North America. EOS Transactions AGU 81, F897.
Becker, T.W., Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models. Geochemistry, Geophysics, Geosystems 3. http://dx.doi.org/10.1029/2001GC000168.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Becker, T.W., Kellogg, J.B., Ekström, G., O'Connell, R.J., 2003. Comparison of azimuthal seismic anisotropy from surface waves and finite strain from global mantle-circulation models. Geophysical Journal International 155, 696–714.

Beghein, C., Trampert, J., 2004. Probability density functions for radial anisotropy: implications for the upper 1200 km of the mantle. Earth and Planetary Science Letters 217, 151–162.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F.C., Moschetti, M.P., Shapiro, N.M., Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. Geophysical Journal International 169, 1239–1260.

Bhattacharyya, J., Masters, G., Shearer, P.M., 1996. Global lateral variations of shear wave attenuation in the upper mantle. Journal of Geophysical Research 101, 22,273–22,289.

Bielak, J., Christiano, P., 1984. On the effective seismic input for non-linear soil-structure interaction systems. Earthquake Engineering and Structural Dynamics 12, 107–119.

Bielak, J., Loukakis, K., Hisada, Y., Yoshimura, C., 2003. Domain reduction method for three-dimensional earthquake modeling in localized regions, part I: theory. Bulletin of the Seismological Society of America 93, 817–824.

Bijwaard, H., Spakman, W., 1999. Tomographic evidence for a narrow whole mantle plume below Iceland. Earth and Planetary Science Letters 166, 121–126.

Bijwaard, H., Spakman, W., 2000. Non-linear global P-wave tomography by iterated linearized inversion. Geophysical Journal International 141, 71–82.

Bijwaard, H., Spakman, W., Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography. Journal of Geophysical Research 103, 30,055–30,078.

Bodin, T., Sambridge, M.S., Tkalčić, H., Arroucau, P., Gallagher, K., Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion. Journal of Geophysical Research 117. http://dx.doi.org/10.1029/2011JB008560.

Bolton, H., Masters, G., 1991. Long period absolute P times and lower mantle structure. EOS Transactions AGU 72, 339.

Boschi, L., Dziewonski, A.M., 1999. High- and low-resolution images of the Earth's mantle: implications of different approaches to tomographic modeling. Journal of Geophysical Research 104, 25,567–25,594.

Boschi, L., Dziewonski, A.M., 2000. Whole Earth tomography from delay times of P, PcP and PKP phases: Lateral heterogeneities in the outer core or radial anistropy in the mantle. Journal of Geophysical Research 105, 13,675–13,696.

Boschi, L., Ekström, G., 2002. New images of the Earth's upper mantle from measurements of surface wave phase velocity anomalies. Journal of Geophysical Research 107. http://dx.doi.org/10.1029/2000JB000059.

Boschi, L., Ekström, G., Kustowski, B., 2004. Multiple resolution surface wave tomography: the Mediterranean basin. Geophysical Journal International 157, 293–304.

Boschi, L., Ampuero, J.P., Peter, D., Mai, P., Soldati, G., Giardini, D., 2007. Petascale computing and resolution in global seismic tomography. Physics of the Earth and Planetary Interiors 163, 245–250.

Bostock, M.G., 2004. Green's functions, source signatures, and the normalization of teleseismic wave fields. Journal of Geophysical Research 109. http://dx.doi.org/10.1029/2003JB002783.

Bostock, M.G., Rondenay, S., 1999. Migration of scattered teleseismic body waves. Geophysical Journal International 137, 732–746.

Bostock, M.G., Rondenay, S., Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves, 1. Theory for oblique incidence. Journal of Geophysical Research 106, 30,771–30,782.

Bouchon, M., Sanchez-Sesma, F.J., 2007. Boundary integral equations and boundary elements methods in elastodynamics. Advances in Geophysics 48, 157–189.

Bozda , E., Trampert, J., 2008. On crustal corrections in surface wave tomography. Geophysical Journal International 172, 1066–1082.

Bozda , E., Trampert, J., Tromp, J., 2011a. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophysical Journal International 185, 845–870.

Bozda , E., Zhu, H.J., Peter, D., Tromp, J., 2011b. Towards global adjoint tomography, abstract S41A-2167. 2011 Fall Meeting. AGU.

Brenders, A.J., Pratt, R.G., 2007. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. Geophysical Journal International 168, 133–151.

Bretherton, F.P., 1968. Propagation in slowly varying waveguides. Proceedings of the Royal Society of London Series A 302, 555–576.

Brune, J., 1966. P and S wave travel times and spheroidal normal modes of a homogeneous sphere. Journal of Geophysical Research 71, 2959–2965.

Brune, J., Espinosa, A., Oliver, J., 1963. Relative excitation of surface waves by earthquakes and underground explosions in the California–Nevada region. Journal of Geophysical Research 68, 3501–3513.

Brzak, K., Gu, Y.J., Ökeler, A., Steckler, M., Lerner-Lam, A., 2009. Migration imaging and forward modeling of microseismic noise sources near southern Italy. Geochemistry, Geophysics, Geosystems 10. http://dx.doi.org/10.1029/2008GC02234.

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. Geophysical Journal International 152, 280–301.

Byrd, R.H., Lu, P.H., Nocedal, J., Zhu, C.Y., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16, 1190–1208.

Campillo, M., Paul, A., 2003. Long-range correlations in the diffuse seismic coda. Science 299, 547–549.

Cao, A.M., Romanowicz, B., 2004. Constraints on density and shear velocity contrasts at the inner core boundary. Geophysical Journal International 157, 1146–1151.

Cao, Q., van der Hilst, R.D., de Hoop, M.V., Shim, S.H., 2011. Seismic imaging of transition zone discontinuities suggests hot mantle west of Hawaii. Science 332, 1068–1071.

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

Capdeville, Y., Marigo, J.J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media. Geophysical Journal International 170, 823–838.

Capdeville, Y., Vilotte, J.P., Montagner, J.P., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global Earth models. Geophysical Journal International 152, 34–67.

Capdeville, Y., Romanowicz, B., Gung, Y., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking. Geophysical Journal International 162, 541–554.

Cara, M., Leveque, J.J., 1987. Waveform inversion using secondary observables. Geophysical Research Letters 14, 1046–1049.

Carcione, J.M., 1994. The wave equation in generalized coordinates. Geophysics 59, 1911–1919.

Carcione, J.M., Helle, H.B., Seriani, G., Plasencia Linares, M.P., 2005. Simulation of seismograms in a 2-D viscoelastic Earth by pseudospectral methods. Geofisica International 44, 123–142.

Carlson, R.L., Christensen, N.I., 1979. Velocity anisotropy in semi-indurated calcareous deep sea sediments. Journal of Geophysical Research 84, 205–211.

Casarotti, E., Komatitsch, D., Stupazzini, M., Lee, S.J., Tromp, J., Piersanti, A., 2008. CUBIT and seismic wave propagation based upon the Spectral-Element Method: an advanced unstructured mesher for complex 3D geological media. Proceedings, 16th International Meshing Roundtable. Springer, pp. 1–14.

Castro, C.E., Käser, M., 2010. Seismic waves in heterogeneous material: subcell resolution of the discontinuous Galerkin method. Geophysical Journal International 182, 250–264.

Chaljub, E., Valette, B., 2004. Spectral element modelling of three-dimensional wave propagation in a self-gravitating Earth with an arbitrarily stratified outer core. Geophysical Journal International 158, 131–141.

Chaljub, E., Komatitsch, D., Vilotte, J.P., Capdeville, Y., Valette, B., Festa, G., 2007. Spectral-element analysis in seismology. Advances in Geophysics 48, 365–418.

Chambat, F., Valette, B., 2001. Mean radius, mass, and inertia for reference Earth models. Physics of the Earth and Planetary Interiors 124, 237–253.

Chang, S.J., van der Lee, S., 2011. Mantle plumes and associated flow beneath Arabia and East Africa. Earth and Planetary Science Letters 302, 448–454.

Chang, S.J., van der Lee, S., Flanagan, M.P., Bedle, H., Marone, F., Matzel, E.M., Pasyanos, M.E., Rodgers, A.J., Romanowicz, B., Schmid, C., 2010. Joint inversion for three-dimensional S velocity mantle structure along the Tethyan margin. Journal of Geophysical Research 115. http://dx.doi.org/10.1029/2009JB007204.

Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances. Pure and Applied Geophysics 168, 1527–1552.

Chen, W.P., Brudzinski, M.R., 2003. Seismic anisotropy in the mantle transition zone beneath Fiji-Tonga. Geophysical Research Letters 30. http://dx.doi.org/10.1029/2002GL016330.

Chen, P., Zhao, L., Jordan, T.H., 2005. Finite-moment tensor of the 3 September 2002 Yorba Linda earthquake. Bulletin of the Seismological Society of America 95, 1170–1180.

Chen, Q.F., Chen, Y., Li, L., 2006. China digital seismic network improves coverage and quality. EOS Transactions AGU 87, 4–7.

Chen, M., Tromp, J., Helmberger, D.V., Kanamori, H., 2007. Waveform modeling of the slab beneath Japan. Journal of Geophysical Research 112. http://dx.doi.org/10.1029/2006JB004394.

Chen, P., Zhao, L., Jordan, T.H., 2007a. Full 3D tomography for the crustal structure of the Los Angeles region. Bulletin of the Seismological Society of America 97, 1094–1120.

Chen, P., Zhao, L., Jordan, T.H., 2007b. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. Geophysical Journal International 170, 175–181.

Chen, C.W., Rondenay, S., Evans, R.L., Snyder, D.B., 2009. Geophysical detection of relict metasomatism from an Archean (3.5 Ga) subduction zone. Science 326, 1089–1091.

Chen, C.W., Miller, D.E., Djikpesse, H.A., Haldorsen, J.B.U., Rondenay, S., 2010a. Array-conditioned deconvolution of multiple-component teleseismic recordings. Geophysical Journal International 182, 967–976.

Chen, M., Huang, H., Yao, H., van der Hilst, R.D., 2010b. Adjoint tomography using Green's functions from ambient noise, Abstract S31B-07. 2010 Fall Meeting. AGU.

Chen, P., Jordan, T.H., Lee, E.J., 2010c. Perturbation kernels for generalized seismological data functionals (GSDF). Geophysical Journal International 183, 869–883.

Chernov, L.A., 1960. Wave Propagation in a Random Medium. McGraw-Hill.

Chevrot, S., Zhao, L., 2007. Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton. Geophysical Journal International 169, 201–215.

Chevrot, S., Monteiller, V., Komatitsch, D., Fuji, N., Martin, R., 2011. A hybrid technique for 3-D waveform modeling and inversion of high frequency teleseismic body waves, Abstract S11D-05. 2011 Fall Meeting. AGU.

Chiao, L.Y., Kuo, B.Y., 2001. Multiscale seismic tomography. Geophysical Journal International 145, 517–527.

Chiao, L.Y., Fang, H.Y., Gung, Y., Chang, Y.H., Hung, S.H., 2010. Comparative appraisal of linear inverse models constructed via distinctive parameterizations (comparing distinctly inverted models). Journal of Geophysical Research 115. http://dx.doi.org/10.1029/2009JB006867.

Cho, K.H., Herrmann, R.B., Ammon, C.J., Lee, K., 2007. Imaging the upper crust of the Korean Peninsula by surface-Wave tomography. Bulletin of the Seismological Society of America 97, 198–207.

Chou, C.W., Brooker, J.R., 1979. A Backus–Gilbert approach to inversion of travel-time data for three-dimensional velocity structure. Geophysical Journal of the Royal Astronomical Society 59, 325–344.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Christensen, N.I., Crosson, R.S., 1968. Seismic anisotropy in the upper mantle. Tectonophysics 6, 93–107.
Claerbout, J.F., 1968. Synthesis of a layered medium from its acoustic transmision response. Geophysics 33, 264–269.
Claerbout, J.F., 1971. Toward a unified theory of reflector mapping. Geophysics 36, 467–481.
Cormack, A.M., 1963. Representation of a function by its line integrals, with some radiological applications. Journal of Applied Physics 34, 2722–2727.
Courtier, P., Talagrand, O., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: numerical results. Quarterly Journal of the Royal Meteorological Society 113, 1329–1347.
Crase, E., Pica, A., Noble, M., McDonald, J., Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: application to real data. Geophysics 55, 527–538.
Cunningham, I.A., Judy, P.F., 2000. Computed tomography, In: Bronzino, J.D. (Ed.), The Biomedical Engineering Handbook, Chapter 62, 2nd edition. CRC Press LLC.
Dahlen, F.A., Nolet, G., 2005. Comment on 'On sensitivity kernels for 'wave-equation' transmission tomography' by de Hoop and van der Hilst. Geophysical Journal International 163, 949–951.
Dahlen, F.A., Tromp, J., 1998. Theoretical Global Seismology. Princeton University Press, New Jersey.
Dahlen, F.A., Nolet, G., Hung, S.H., 2000. Fréchet kernels for finite-frequency traveltimes — I. Theory. Geophysical Journal International 141, 157–174.
Dalton, C.A., Ekström, G., Dziewonski, A.M., 2008. The global attenuation structure of the upper mantle. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2007JB005429.
Dando, B.D.E., Stuart, G.W., Houseman, G.A., Hegedüs, E., Brückl, E., Radovanović, S., 2011. Teleseismic tomography of the mantle in the Carpathian–Pannonian region of central Europe. Geophysical Journal International 186, 11–31.
Das, T., Nolet, G., 1998. Crustal thickness map of the western United States by partitioned waveform inversion. Journal of Geophysical Research 103, 30,021–30,038.
de Hoop, M.V., van der Hilst, R.D., 2004. On sensitivity kernels for 'wave-equation' transmission tomography. Geophysical Journal International 160, 621–633.
de la Puente, J., Ampuero, J.P., Käser, M., 2009. Dynamic rupture modeling on unstructured meshes using a discontinuous Galerkin method. Journal of Geophysical Research 114. http://dx.doi.org/10.1029/2008JB006271.
Deans, S.R., 1983. The Radon Transform and Some of its Applications. Wiley-Interscience, New York.
Descartes, R., Laurence, J.L., 1631. Discourse on method and meditations (trans.). Dover Publications, 2003.
Deuss, A., 2009. Global observations of mantle discontinuities using SS and PP precursors. Surveys in Geophysics 30, 301–326.
Dueker, K., Sheehan, A.F., 1997. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track. Journal of Geophysical Research 102, 8312–8327.
Dumbser, M., Käser, M., 2006. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes — II. The three-dimensional isotropic case. Geophysical Journal International 167, 319–336.
Durek, J.J., Ritzwoller, M.H., Woodhouse, J.H., 1993. Constraining upper mantle anelasticity using surface wave amplitude anomalies. Geophysical Journal International 114, 249–272.
Dziewonski, A.M., 1971. Overtones of free oscillations and the structure of the Earth's interior. Science 172, 1336–1338.
Dziewonski, A.M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6. Journal of Geophysical Research 89, 5929–5952.
Dziewonski, A.M., 1996. Earth's mantle in three dimensions. Seismic modelling of earth structure, Vol. 25. Editrice Compositoii, pp. 507–572.
Dziewonski, A.M., 2005. The robust aspects of global seismic tomography. GSA Special Papers 388, 147–154. http://dx.doi.org/10.1130/0-8137-2388-4.147.
Dziewonski, A.M., Anderson, D.L., 1981. Preliminary reference Earth model. Physics of the Earth and Planetary Interiors 25, 297–356.
Dziewonski, A.M., Steim, J.M., 1982. Dispersion and attenuation of mantle waves through waveform inversion. Geophysical Journal of the Royal Astronomical Society 70, 503–527.
Dziewonski, A.M., Woodhouse, J.H., 1987. Global images of the Earth's interior. Science 236, 37–48.
Dziewonski, A.M., Forte, A.M., Su, W.J., Woodward, R.L., 1993. Seismic tomography and geodynamics. In: Aki, K., Dmowska, R. (Eds.), Relating Geophysical Structures and Processes: The Jeffreys Volume. : Geophysical Monograph, 76. AGU, pp. 67–105.
Earle, P.S., Shearer, P.M., 1994. Characterization of global seismograms using an automatic-picking algorithm. Bulletin of the Seismological Society of America 84, 366–376.
Efron, B., Tibshirani, R., 1991. Statistical data analysis in the computer age. Science 253, 390–395.
Eisner, L., Clayton, R.W., 2001. A reciprocity method for multiple-source simulations. Bulletin of the Seismological Society of America 91, 553–560.
Ekström, G., Dziewonski, A.M., 1998. The unique anisotropy of the Pacific upper mantle. Nature 394, 168–172.
Ekström, G., Dziewonski, A.M., Su, W.J., Smith, G.P., 1995. Elastic and anelastic structure beneath Eurasia. Technical Report. Contract No. F49620-92-J-0392, sponsored by AFOSR.
Engl, H.W., Hanke, M., Neubauer, A., 2000. Regularization of Inverse Problems. Kluwer Academic Publishers, Netherland.
Epanomeritakis, I., Akcelik, V., Ghattas, O., Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Problems 24. http://dx.doi.org/10.1088/0266-5611/24/3/034015.
Escalante, C., Gu, Y.J., Sacchi, M., 2007. Simultaneous iterative time-domain sparse deconvolution to teleseismic receiver functions. Geophysical Journal International 171, 316–325.
Farra, V., Vinnik, L., 2000. Upper mantle stratification by P and S receiver functions. Geophysical Journal International 141, 699–712.
Fichtner, A., 2010. Full Seismic Waveform Modelling and Inversion. Springer-Verlag, Heidelberg.
Fichtner, A., Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models — a technique based on non-linear dispersion curve matching (DCM). Geophysical Journal International 173, 519–533.
Fichtner, A., Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques. Geophysical Journal International 185, 775–798.
Fichtner, A., Trampert, J., 2011b. Resolution analysis in full waveform inversion. Geophysical Journal International 187, 1604–1624.
Fichtner, A., Bunge, H.P., Igel, H., 2006a. The adjoint method in seismology — I. Theory. Physics of the Earth and Planetary Interiors 157, 86–104.
Fichtner, A., Bunge, H.P., Igel, H., 2006b. The adjoint method in seismology — II. Applications: traveltimes and sensitivity functionals. Physics of the Earth and Planetary Interiors 157, 105–123.
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. Geophysical Journal International 175, 665–685.
Fichtner, A., Igel, H., Bunge, H.P., Kennett, B.L.N., 2009a. Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method. Journal of Numerical Analysis, Industrial and Applied Mathematics 4, 11–22.
Fichtner, A., Kennett, B.L.N., Igel, H., Bunge, H.P., 2009b. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. Geophysical Journal International 179, 1703–1725.
Fichtner, A., Kennett, B.L.N., Igel, H., Bunge, H.P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle. Earth and Planetary Science Letters 290, 270–280.
Filler, A., 2009. The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI. Nature Precedings. http://dx.doi.org/10.1038/npre.2009.3267.
Fink, M., 1992. Time reversal of ultrasonic fields — Part I: basic principles. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 39, 555–566.
Fink, M., 1997. Time-reversed acoustics. Physics Today 34–40 (March).
Fink, M., Prada, C., Wu, F., Cassereau, D., 1989. Self focusing in inhomogeneous media with time reversal acoustic mirrors. IEEE Ultrasonics Symposium, pp. 681–686.
Flanagan, M.P., Wiens, D.A., 1994. Radial upper mantle attenuation structure of inactive back arc basin from differential shear wave measurements. Journal of Geophysical Research 99, 15,469–15,485.
Flanagan, M.P., Shearer, P.M., Cecil, H., 1998. Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors. Geophysics 103, 2673–2692.
Fletcher, R., 1987. Practical Methods of Optimization, 2nd edition. Wiley-Interscience Publication.
Forsyth, D.W., 1975. The early structural evolution and anisotropy of the oceanic upper mantle. Geophysical Journal of the Royal Astronomical Society 43, 103–162.
Forsyth, D.W., Webb, S.C., Dorman, L.M., Shen, Y., 1998. Phase velocities of Rayleigh waves in the MELT experiment on the East Pacific Rise. Science 280, 1235–1238.
Forte, A.M., 2000. Seismic-geodynamic constraints on mantle flow: implications for layered convection, mantle viscosity, and seismic anisotropy in the deep mantle. In: Karato, S. (Ed.), Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale. AGU, pp. 3–36.
Forte, A.M., Mitrovica, J.X., 2001. Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. Nature 410, 1049–1056.
Forte, A.M., Woodward, L., Dziewonski, A.M., 1994. Joint inversions of seismic and geodynamic data for models of three-dimensional mantle heterogeneity. Journal of Geophysical Research 99, 21,857–21,877.
Fréchet, M.R., 1941. Sur la loi de répartition de certaines grandeurs géographiques. Journal de la Société de Statistiques de Paris 82, 114–122.
Friederich, W., Wielandt, E., 1995. Interpretation of seismic surface waves in regional networks: joint estimation of wavefield geometry and local phase velocity. Method and numerical tests. Geophysical Journal International 120, 731–744.
Fuji, N., S., C., Zhao, L., Geller, R.J., Kawai, K., 2011. Efficient calculation of complete 3D Fréchet derivatives for high frequency compressional teleseismic body waves, Abstract S44A-01. 2011 Fall Meeting. AGU.
Fukao, Y., Obayashi, M., Inoue, H., Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone. Journal of Geophysical Research 97, 4809–4822.
Fukao, Y., Widiyantoro, S., Obayashi, M., 2001. Stagnant slab in the upper and lower mantle transition region. Reviews of Geophysics 39, 291–323.
Fukao, Y., Obayashi, M., Nakakuki, T., 2009. Stagnant slab: a review. Annual Review of Earth and Planetary Sciences 37, 19–46.
Gaherty, J.B., Jordan, T.H., 1995. Lehmann discontinuity as the base of an anisotropic layer beneath continents. Science 268, 1468–1471.
Gauthier, O., Virieux, J., Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophysics 51, 1387–1403.
Gee, L.S., Neuhauser, D.S., Dreger, D., Pasyanos, M.E., Uhrhammer, R.A., Romanowicz, B., 1996. Real-time seismology at UC Berkeley: The Rapid Earthquake Data Integration project. Bulletin of the Seismological Society of America 86, 936–945.
Geller, R.J., Hara, T., 1993. Two efficient algorithms for iterative linearized inversion of seismic waveform data. Geophysical Journal International 115, 699–710.
Geller, R.J., Ohminato, T., 1994. Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the Direct Solution Method. Geophysical Journal International 116, 421–446.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Gilbert, F., 1970. Excitation of the normal modes of the Earth by earthquake sources. Geophysical Journal of the Royal Astronomical Society 22, 223–226.

Gilbert, F., Dziewonski, A.M., 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. Philosophical Transactions of the Royal Society of London, Series A 278, 187–269.

Godinho, L., Amado, P.A., Tadeu, A., Cadena-Isaza, A., Smerzini, C., Sanchez-Sesma, F.J., Madec, R., Komatitsch, D., 2009. Numerical simulation of ground rotations along 2D topographical profiles under the incidence of elastic plane waves. Bulletin of the Seismological Society of America 99, 1147–1161.

Grand, S.P., 2002. Mantle shear-wave tomography and the fate of subducted slabs. Philosophical Transactions of the Royal Society of London, Series A 360, 2475–2491.

Grand, S.P., van der Hilst, R.D., Widiyantoro, S., 1997. Global seismic tomography: a snapshot of convection in the Earth. GSA Today 7, 1–7.

Gu, Y.J. (Ed.), 2010. Arrays and Array Methods in Global Seismology. Springer.

Gu, Y.J., Dziewonski, A.M., Agee, C.B., 1998. Global de-correlation of the topography of transition zone discontinuities. Earth and Planetary Science Letters 157, 57–67.

Gu, Y.J., Dziewonski, A.M., Su, W.J., Ekström, G., 2001. Models of the mantle shear velocity and discontinuties in the pattern of lateral heterogeneities. Journal of Geophysical Research 106, 11,169–11,199.

Gu, Y.J., Dziewonski, A.M., Ekström, G., 2003. Simultaneous inversion for mantle shear velocity and topography of transition zone discontinuities. Geophysical Journal International 154, 559–583.

Gu, Y.J., Lerner-Lam, A., Dziewonski, A.M., Ekström, G., 2005. Deep structure and seismic anisotropy beneath the East Pacific Rise. Earth and Planetary Science Letters 232, 259–272.

Gu, Y.J., Dublanko, C., Lerner-Lam, A., Brzak, K., Steckler, M., 2007. Probing the sources of ambient seismic noise near the coasts of southern Italy. Geophysical Research Letters 34. http://dx.doi.org/10.1029/2007GL031967.

Gu, Y.J., Okeler, A., Schultz, R., 2012. Tracking slabs beneath northwestern Pacific subduction zones. Earth and Planetary Science Letters 331–332, 269–280.

Gung, Y., Romanowicz, B., 2004. Q tomography of the upper mantle using three-component long-period waveforms. Geophysical Journal International 157, 813–830.

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

Hearn, T.M., Clayton, R.W., 1986. Lateral velocity variations in southern California. I. Results for upper crust from Pg waves. Bulletin of the Seismological Society of America 76, 495–509.

Helffrich, G., 2000. Topography of the transition zone seismic discontinuities. Reviews of Geophysics 38, 141–158.

Herglotz, G., 1907. Über das Benndorfsche problem der fortplanzungsgeschwindigkeit der erdbeben-strahlen. Zeitschrift für Geophysik 8, 145–147.

Hermann, V., Käser, M., Castro, C.E., 2010. Non-conforming hybrid meshes for efficient 2-D wave propagation using the Discontinuous Galerkin Method. Geophysical Journal International 184, 746–758.

Hess, H.H., 1964. Seismic anisotropy of the uppermost mantle under oceans. Nature 203, 629–631.

Hingee, M., Tkalčić, H., Fichtner, A., Sambridge, M.S., 2011. Seismic moment tensor inversion using a 3-D structural model: applications for the Australian region. Geophysical Journal International 184, 949–964.

Hirn, A., 1977. Anisotropy in the continental upper mantle: possible evidence from explosion seismology. Geophysical Journal of the Royal Astronomical Society 49, 49–58.

Hjörleifsdóttir, V., 2007. Earthquake source characterization using 3D numerical modeling. Ph.D. thesis. Caltech.

Ho-Liu, P., Kanamori, H., Clayton, R.W., 1988. Application of attenuation tomography to Imperial Valley and Coso-Indian Wells region, southern California. Journal of Geophysical Research 93, 10,501–10,520.

Ho-Liu, P., Montagner, J.P., Kanamori, H., 1989. Comparison of iterative back-projection inversion and generalized inversion without blocks: case studies in attenuation tomography. Geophysical Journal 97, 19–29.

Houser, C., Masters, G., Shearer, P.M., Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms. Geophysical Journal International 174, 195–212.

Hull, D.G., Tapley, B.D., 1977. Square-root variable-metric methods for minimization. Journal of Optimization Theory and Applications 21, 251–259.

Hung, S.H., Nolet, G., Dahlen, F.A., 2000. Fréchet kernels for finite-frequency traveltimes — II. Examples. Geophysical Journal International 141, 175–203.

Hung, S.H., Shen, Y., Chiao, L.Y., 2004. Imaging seismic velocity structure beneath the Iceland hot spot: a finite frequency approach. Journal of Geophysical Research 109. http://dx.doi.org/10.1029/2003JB002889.

Hung, S.H., Chen, W.P., Chiao, L.Y., Tseng, T.L., 2010. First multi-scale, finite-frequency tomography illuminates 3-D anatomy of the Tibetan Plateau. Geophysical Research Letters 37. http://dx.doi.org/10.1029/2009GL041875.

Hung, S.H., Chen, W.P., Chiao, L.Y., 2011. A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet. Journal of Geophysical Research 116. http://dx.doi.org/10.1029/2010JB008190.

Igel, H., 1996. P–SV wave propagation in the Earth's mantle using lowermost mantle structure. Geophysical Research Letters 23, 415–418.

Igel, H., Djikpéssé, H., Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio. Geophysical Journal International 124, 363–371.

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

Ishii, M., 2004. Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes. Physics of the Earth and Planetary Interiors 146, 113–124.

Ishii, M., Shearer, P.M., Houston, H., Vidale, J.E., 2005. Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array. Nature 435, 933–936.

Ita, J., Stixrude, L., 1992. Petrology, elasticity, and composition of the mantle transition zone. Journal of Geophysical Research 97, 6849–6866.

Iyer, H.M., Hirahara, K., 1993. Seismic Tomography, Theory and Practice. Chapman and Hall, London.

Jordan, T.H., 1978. A procedure for estimating lateral variations from low-frequency eigenspectra data. Geophysical Journal of the Royal Astronomical Society 52, 441–455.

Jordan, T.H., 1981. Global tectonic regionalization for seismological data analysis. Bulletin of the Seismological Society of America 71, 1131–1141.

Julian, B.R., Gubbins, D., 1977. Three-dimensional seismic ray tracing. Journal of Geophysics 43, 95–113.

Jung, H., Karato, S., 2001. Water-induced fabric transitions in olivine. Science 293, 1460–1463.

Kanamori, H., 1970. Velocity and Q of mantle waves. Physics of the Earth and Planetary Interiors 2, 259–275.

Kanamori, H., Hauksson, E., Heaton, T.H., 1991. Terrascope and CUBE Project at Caltech. EOS Transactions AGU 72.

Kárason, H., van der Hilst, R.D., 2000. Constraints on mantle convection from seismic tomography. The history and dynamics of global plate motions: Geophysical monograph, 121. AGU, pp. 277–288.

Kárason, H., van der Hilst, R.D., 2001. Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, Pdiff). Journal of Geophysical Research 106, 6569–6587.

Karato, S., 1998. Seismic anisotropy in the deep mantle, boundary layers and the geometry of mantle convection. Pure and Applied Geophysics 151, 565–587.

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

Katsura, T., Ito, E., 1989. The system Mg_2SiO_4–Fe_2SiO_4 at high pressure and temperatures: precise determination of stabilities of olivine, modified spinel, and spinel. Journal of Geophysical Research 94, 15,633–15,670.

Kawakatsu, H., 1983. Can 'pure-path' models explain free oscillation data? Geophysical Research Letters 10, 186–189.

Keilis-Borok, V.I., Yanovskaja, T.B., 1967. Inverse problems of seismology (structural review). Geophysical Journal of the Royal Astronomical Society 13, 223–234.

Keller, H., 1968. Numerical Methods for Two-point Boundary-value Problems. Blaisdell, Waltham.

Keller, J.B., Kay, I., Shmqys, J., 1956. Determination of the potential from scattering data. Physical Review 102, 557–559.

Kellogg, L.H., Hager, B.H., van der Hilst, R.D., 1999. Compositional stratification in the deep mantle. Science 283, 1881–1884.

Kelly, K.R., Ward, R.W., Treitel, S., Alford, R.M., 1976. Synthetic seismograms: a finite-difference approach. Geophysics 41, 2–27.

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

Kennett, B.L.N., Sambridge, M.S., Williamson, P.R., 1988. Subspace methods for large inverse problems with multiple parameter classes. Geophysical Journal 94, 237–247.

Kennett, B.L.N., Widiyantoro, S., van der Hilst, R.D., 1998. Joint seismic tomography for bulk sound and shear wave speed. Journal of Geophysical Research 103, 469–493.

Khan, A., Lantz, E., 2002. Who really discovered Snell's law? Physics World 64–84 (April).

Khan, A., Connolly, J.A.D., Taylor, S.R., 2008. Inversion of seismic and geodetic data for the major element chemistry and temperature of the Earth's mantle. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2007JB005239.

Kim, Y., Liu, Q., Tromp, J., 2011. Adjoint centroid-moment tensor inversions. Geophysical Journal International 186, 264–278.

Kissling, E., Husen, S., Haslinger, F., 2001. Model parametrization in seismic tomography: a choice of consequence for the solution quality. Physics of the Earth and Planetary Interiors 123, 89–101.

Knopoff, L., 1972. Observation and inversion of surface-wave dispersion. Tectonophysics 13, 497–519.

Komatitsch, D., Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation. Geophysical Journal International 139, 806–822.

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

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

Komatitsch, D., Liu, Q., Tromp, J., Süss, M.P., Stidham, C., Shaw, J.H., 2004. Simulations of ground motion in the Los Angeles Basin based upon the spectral-element method. Bulletin of the Seismological Society of America 94, 187–206.

Komatitsch, D., Michéa, D., Erlebacher, G., 2009. Porting a high-order finite-element earthquake modeling application to NVIDIA graphics cards using CUDA. Journal of Parallel and Distributed Computing 69, 451–460.

Komatitsch, D., Erlebacher, G., Göddeke, D., Michéa, D., 2010. High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. Journal of Computational Physics 229, 7692–7714.

Koper, K.D., Wysession, M.E., Wiens, D.A., 1999. Multimodal function optimization with a niching genetic algorithm: A seismological example. Bulletin of the Seismological Society of America 89, 978–988.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Kremers, S., Fichtner, A., Brietzke, G.B., Igel, H., Larmat, C., Huang, L., Käser, M., 2011. Exploring the potentials and limitations of the time-reversal imaging of finite seismic sources. Solid and Earth Discussions 3, 217–248.

Kustowski, B., Ekström, G., Dziewonski, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2007JB005169.

Lailly, P., 1984. Migration methods: partial but efficient solutions to the seismic inverse problem. In: Santosa, F. (Ed.), Inverse Problem of Acoustic and Elastic Waves. Society for Industrial and Applied Mathematics, pp. 182–214.

Langston, C.A., 1977. Corvallis, Oregon, crustal and upper mantle receiver structure from teleseismic P and S waves. Bulletin of the Seismological Society of America 67, 713–724.

Langston, C.A., 1989. Scattering of teleseismic body waves under Pasadena, California. Journal of Geophysical Research 94, 1935–1951.

Langston, C.A., Hammer, J.K., 2001. The vertical component P-wave receiver function. Bulletin of the Seismological Society of America 91, 1805–1819.

Larmat, C., Liu, Q., Tromp, J., Montagner, J.P., 2008. Time reversal location of glacial earthquakes. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2008JB005607.

Larmat, C., Guyer, R.A., Johnson, P.A., 2010. Time-reversal methods in geophysics. Physics Today 63, 31–35.

Larson, E.W.F., Tromp, J., Ekström, G., 1998. Effects of slight anisotropy on surface waves. Geophysical Journal International 132, 654–666.

Laske, G., Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data. Journal of Geophysical Research 101, 16,059–16,075.

Lawrence, J.F., Prieto, G.A., 2011. Attenuation tomography of the western United States from ambient seismic noise. Journal of Geophysical Research 116. http://dx.doi.org/10.1029/2010JB007836.

Lawrence, J.F., Shearer, P.M., 2006. A global study of transition zone thickness using receiver functions. Journal of Geophysical Research 111. http://dx.doi.org/10.1029/2005JB003973.

Lee, S.J., Chen, H.W., Liu, Q., Komatitsch, D., Tromp, J., Huang, B.S., 2008. Three-dimensional simulations of seismic-wave propagation in the Taipei basin with realistic topography based upon the Spectral-element method. Bulletin of the Seismological Society of America 98, 253–264.

Lee, E.J., Chen, P., Jordan, T.H., Wang, L.Q., 2011. Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. Geophysical Journal International 186, 311–330.

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

Lekić, V., Panning, M., Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography. Geophysical Journal International 182, 264–278.

Levander, A., Nolet, G. (Eds.), 2005. Seismic earth: array analysis of broadband seismograms. : Geophysical Monograph Series, 157. AGU.

Levander, A., Humphreys, E.D., Ekström, G., Meltzer, A.S., Shearer, P.M., 1999. Proposed project would give unprecedented look under North America. EOS Transactions AGU 80, 245–251.

Levander, A., Niu, F.L., Symes, W.W., 2005. Imaging teleseismic P to S scattered waves using the Kirchhoff integral. In: Levander, A., Guust, N. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 149–169.

Leveque, J.J., Debayle, E., Maupin, V., 1998. Anisotropy in the Indian Ocean upper mantle from Rayleigh- and Love-waveform inversion. Geophysical Journal International 133, 529–540.

Levin, V., Roecker, S., Graham, P., Hosseini, A., 2008. Seismic anisotropy indicators in Western Tibet: shear wave splitting and receiver function analysis. Tectonophysics 462, 99–108.

Levshin, A.L., Yang, X.N., Barmin, M.P., Ritzwoller, M.H., 2010. Midperiod Rayleigh wave attenuation model for Asia. Geochemistry, Geophysics, Geosystems 11. http://dx.doi.org/10.1029/2010GC003164.

Li, X.Q., Nfibelek, J.L., 1999. Deconvolution of teleseismic body waves for enhancing structure beneath a seismometer array. Bulletin of the Seismological Society of America 89, 190–201.

Li, X.D., Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling. Geophysical Journal International 121, 695–709.

Li, X.D., Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. Journal of Geophysical Research 101, 22,245–22,272.

Li, X.D., Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical earth model. Geophysical Journal International 112, 92–102.

Lin, F.C., Moschetti, M.P., Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. Geophysical Journal International 173, 281–298.

Lin, F.C., Ritzwoller, M.H., Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. Geophysical Journal International 177, 1091–1910.

Lin, F.C., Ritzwoller, M.H., Yang, Y., Moschetti, M.P., Fouch, M.J., 2010. Complex and variable crustal and uppermost mantle seismic anisotropy in the western United States. Nature Geoscience 4, 55–61.

Lin, F.C., Ritzwoller, M.H., Shen, W.S., 2011. On the reliability of attenuation measurements from ambient noise cross-correlations. Geophysical Research Letters 38. http://dx.doi.org/10.1029/2011GL047366.

Liu, Q., 2009. A database of global finite-frequency sensitivity kernels based upon spectral-element simulations. Abstract S33B-1761: Eos Trans. AGU, 90.

Liu, Q., Chen, C.W., 2011. High-resolution array imaging using teleseismic converted waves based on adjoint methods, abstract S13C-03. 2011 Fall Meeting. AGU.

Liu, X.F., Dziewonski, A.M., 1998. Global analysis of shear wave velocity anomalies in the lowermost mantle. In: Gurnis, M., Wysession, N., Knittle, E., Buffett, B. (Eds.), The Core–Mantle Boundary Region. AGU, pp. 21–36.

Liu, Q., Tromp, J., 2006. Finite-frequency kernels based on adjoint methods. Bulletin of the Seismological Society of America 96, 2283–2397.

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

Liu, Q., Tromp, J., Polet, J., Komatitsch, D., 2004. Spectral-element moment tensor inversions for earthquakes in southern California. Bulletin of the Seismological Society of America 94, 1748–1761.

Lobkis, O., Weaver, R.L., 2002. On the emergence of the Green's function in the correlations of a diffuse field: pulse-echo using thermal phonons. Ultrasonics 40, 435–439.

Long, M.D., Silver, P.G., 2009. Shear wave splitting and mantle anisotropy: measurements, interpretations, and new directions. Surveys in Geophysics 30, 407–461.

Love, A.E.H., 1927. A Treatise on the Mathematical theory of Elasticity. Cambridge University Press.

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

Luo, Y., Zhu, H.J., Nissen-Meyer, T., Morency, C., Tromp, J., 2009. Seismic modeling and imaging based upon spectral-element and adjoint methods. The Leading Edge 28, 568–574.

Maggi, A., Priestley, K., 2005. Surface waveform tomography of the Turkish-Iranian plateau. Geophysical Journal International 160, 1068–1080.

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

Manaman, N.S., Shomali, H., Koyi, H., 2011. New constraints on upper-mantle S-velocity structure and crustal thickness of the Iranian plateau using partitioned waveform inversion. Geophysical Journal International 184, 247–267.

Manners, U.J., 2008. Investigating the structure of the core–mantle boundary region using S and P diffracted waves. Ph.D. thesis. University of California, San Diego.

Maraschini, M., Foti, S., 2010. A Monte Carlo multimodal inversion of surface waves. Geophysical Journal International 182, 1557–1566.

Marone, F., Romanowicz, B., 2007. The depth distribution of azimuthal anisotropy in the continental upper mantle. Nature 447, 198–201.

Marone, F., Gung, Y., Romanowicz, B., 2007. Three-dimensional radial anisotropic structure of the North American upper mantle from inversion of surface waveform data. Geophysical Journal International 171, 206–222.

Marquering, H., Snieder, R., 1995. Surface-wave mode coupling for efficient forward modelling and inversion of body-wave phases. Geophysical Journal International 120, 186–208.

Marquering, H., Snieder, R., Nolet, G., 1996. Waveform inversions and the significance of surface-wave mode coupling. Geophysical Journal International 124, 258–278.

Marquering, H., Nolet, G., Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels. Geophysical Journal International 132, 521–534.

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

Masters, G., 1989. Low-frequency seismology and the three-dimensional structure of the Earth. Philosophical Transactions of the Royal Society of London, Series A 328, 329–349.

Masters, G., Shearer, P.M., 1995. Seismic models of the Earth: elastic and anelastic. In: Ahrens, T.J. (Ed.), Global Earth Physics, A Handbook of Physical Constants. AGU, pp. 88–103.

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

Masters, G., Laske, G., Bolton, H., Dziewonski, A.M., 2000. The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure. Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale: Geophysical monograph, 117. AGU, pp. 63–87.

Mégnin, C., Romanowicz, B., 2000. A comparison between tomographic and geodynamic models of the Earth's mantle. History and Dynamics of plate Motions. : Geophysical Monograph, 121. AGU, pp. 257–276.

Meier, U., Curtis, A., Trampert, J., 2007a. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. Geophysical Research Letters 34. http://dx.doi.org/10.1029/2007GL030989.

Meier, U., Curtis, A., Trampert, J., 2007b. Global crustal thickness from neural network inversion of surface wave data. Geophysical Journal International 169, 706–722.

Menke, W., 1984. Geophysical Data Analysis: Discrete Inverse Theory. Academic Press.

Menke, W., Schaff, D., 2004. Absolute earthquake locations with differential data. Bulletin of the Seismological Society of America 94, 2254–2264.

Mercier, J.P., Bostock, M.G., Cassidy, J.F., Dueker, K., Gaherty, J.B., Garnero, E.J., Revenaugh, J., Zandt, G., 2009. Body-wave tomography of western Canada. Tectonophysics 475, 480–492.

Metropolis, N., Ulam, S., 1949. The Monte Carlo method. Journal of the American Statistical Association 44, 335–341.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculation by fast computing machines. The Journal of Chemical Physics 21, 1087–1092.

Mewes, A., Kulessa, B., McKinley, J.D., Binley, A.M., 2010. Anisotropic seismic inversion using a multigrid Monte Carlo approach. Geophysical Journal International 183, 267–276.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Mitchell, B.J., Romanowicz, B. (Eds.), 1998. Q of the Earth: Global, Regional, and Laboratory Studies: PAGEOPH, vol. 153. Birkhauser Verlag, Basel.
Moczo, P., Bystricky, E., Kristek, J., Carcione, J.M., Bouchon, M., 1997. Hybrid modeling of P–SV seismic motion at inhomogeneous viscoelastic topographic structures. Bulletin of the Seismological Society of America 87, 1305–1323.
Moczo, P., Robertsson, J.O.A., Eisner, L., 2007. The finite-difference time-domain method for modeling of seismic wave propagation. Advances in Geophysics 48. http://dx.doi.org/10.1016/S0065-2687(06)48008-0.
Montagner, J.P., 1994. Can seismology tell us anything about convection in the mantle. Reviews of Geophysics 32, 115–137.
Montagner, J.P., Nataf, H.C., 1986. A simple method for inverting the azimuthal anisotropy of surface waves. Journal of Geophysical Research 91, 511–520.
Montagner, J.P., Tanimoto, T., 1990. Global anisotropy in the upper mantle inferred from the regionalization of phase velocities. Journal of Geophysical Research 95, 4797–4819.
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.
Montelli, R., Nolet, G., Dahlen, F.A., 2006a. Comment on 'Banana-doughnut kernels and mantle tomography' by van der Hilst and de Hoop. Geophysical Journal International 167, 1204–1210.
Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., 2006b. A catalogue of deep mantle plumes: new results from finite-frequency tomography. Geochemistry, Geophysics, Geosystems 7. http://dx.doi.org/10.1029/2006GC001248.
Moorkamp, M., Jones, A.G., Eaton, D.W., 2007. Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: are seismic velocities and electrical conductivities compatible? Geophysical Research Letters 34. http://dx.doi.org/10.1029/2007GL030519.
Moorkamp, M., Jones, A.G., Fishwick, S., 2010. Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data. Journal of Geophysical Research 115. http://dx.doi.org/10.1029/2009JB006369.
Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics 52, 1211–1228.
Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data. Geophysics 53, 750–759.
Morelli, A., Dziewonski, A.M., 1987. Topography of the core–mantle boundary and lateral homogeneity of the liquid core. Nature 325, 678–683.
Mosca, I., 2010. Probabilistic tomography using body wave, normal-mode and surface wave data. Ph.D. thesis. University Utrechet.
Mosca, I., Trampert, J., 2009. Path-average kernels for long wavelength traveltime tomography. Geophysical Journal International 177, 639–650.
Moschetti, M.P., Ritzwoller, M.H., Shapiro, N.M., 2007. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps. Geochemistry, Geophysics, Geosystems 8. http://dx.doi.org/10.1029/2007GC001655.
Moschetti, M.P., Ritzwoller, M.H., Lin, F., Yang, Y., 2010. Seismic evidence for widespread western-US deep-crustal deformation caused by extension. Nature 464, 885–889.
Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse Problems 14, 405–426.
Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problem. Journal of Geophysical Research 100, 12,431–12,447.
Mosegaard, K., Tarantola, A., 2002. Probabilistic approach to inverse problems. International handbook of earthquake and engineering seismology, vol. 81A, pp. 237–265.
Nábelek, J., Hetényi, G., Vergne, J., Sapkota, S., Kafle, B., Jiang, M., Su, H.P., Chen, J., Huang, B.S., 2009. Underplating in the Himalaya-Tibet collision zone revealed by the Hi-CLIMB experiment. Science 325, 1371–1374.
Nakanishi, I., Anderson, D.L., 1982. Worldwide distribution of group velocity of mantle Rayleigh wave as determined by spherical harmonic inversion. Bulletin of the Seismological Society of America 72, 1185–1194.
Nakanishi, I., Anderson, D.L., 1984. Measurements of mantle wave velocities and inversion for lateral heterogeneity and anisotropy — II. analysis by the single-station method. Geophysical Journal of the Royal Astronomical Society 78, 573–617.
Nataf, H.C., Nakanishi, I., Anderson, D.L., 1984. Anisotropy and shear-velocity heterogeneities in the upper mantle. Geophysical Research Letters 11, 109–112.
Neele, F., de Regt, H., Vandecar, J., 1997. Gross errors in upper-mantle discontinuity topography from underside reflection data. Geophysical Journal International 129, 194–204.
Nettles, M., Dziewonski, A.M., 2008. Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2006JB004819.
Nishida, K., Kawakatsu, H., Obara, K., 2008. Three-dimensional crustal S wave velocity structure in Japan using microseismic data recorded by Hi-net tiltmeters. Journal of Geophysical Research 113. http://dx.doi.org/10.1029/2007JB005395.
Nishimura, C.E., Forsyth, D.W., 1989. The anisotropic structure of the upper mantle in the Pacific. Geophysical Journal 96, 203–229.
Nissen-Meyer, T., Dahlen, F.A., Fournier, A., 2007a. Spherical-earth Fréchet sensitivity kernels. Geophysical Journal International 168, 1051–1066.
Nissen-Meyer, T., Fournier, A., Dahlen, F.A., 2007b. A two-dimensional spectral-element method for computing spherical-earth seismograms — I. Moment-tensor source. Geophysical Journal International 168, 1067–1092.
Nissen-Meyer, T., Fournier, A., Dahlen, F.A., 2008. A 2-D spectral-element method for computing spherical-earth seismograms — II. Waves in solid–fluid media. Geophysical Journal International 174, 873–888.
Niu, F.L., James, D.E., 2002. Fine structure of the lowermost crust beneath the Kaapvaal craton and its implications for crustal formation and evolution. Earth and Planetary Science Letters 200, 121–130.
Nocedal, J., Wright, S.J., 2006. Numerical Optimization, 2nd edition. Springer.
Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs. Journal of Geophysical Research 95, 8499–8512.
Nolet, G., 1993. Solving large linearized tomographic problems. In: Iyer, H.M., Hirahara, K. (Eds.), Seismic Tomography — Theory and Practice. Chapman & Hall, London, pp. 227–247.
Nolet, G., 2008. A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun. Cambridge University Press.
Nolet, G., Dahlen, F.A., Montelli, R., 2005. Traveltimes and amplitudes of seismic waves: a re-assessment. In: Levander, A., Nolet, G. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 37–47.
Nowack, R.L., 1990. Tomography and the Herglotz–Wiechert inverse formulation. Pure and Applied Geophysics 133, 305–315.
Obayashi, M., Fukao, Y., 1997. P and PcP travel time tomography for the core–mantle boundary. Journal of Geophysical Research 102, 17,825–17,841.
Obrebski, M., Allen, R.M., Pollitz, F., Hung, S.H., 2011. Lithosphere–asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities. Geophysical Journal International 185, 1003–1021.
Obrebski, M., Allen, R.M., Zhang, F.X., Pan, J.T., Wu, Q.W., Hung, S.H., 2012. Shear wave tomography of China using joint inversion of body and surface wave constraints. Journal of Geophysical Research 117. http://dx.doi.org/10.1029/2011JB008349.
Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H., Yamamoto, A., 2004. Recent progress of seismic observation networks in Japan. Earth, Planets and Space 56, xv–xxviii.
Olsen, K.B., Pechmann, J.C., Schuster, G.T., 1995. Simulation of 3D elastic wave propagation in the Salt Lake basin. Bulletin of the Seismological Society of America 85, 1688–1710.
Opršal, I., Matyska, C., Irikura, K., 2009. The source-box wave propagation hybrid methods: general formulation and implementation. Geophysical Journal International 176, 555–564.
Panning, M., Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle. Geophysical Journal International 167, 361–379.
Panning, M., Capdeville, Y., Romanowicz, B., 2009. Seismic waveform modelling in a 3-D Earth using the Born approximation: potential shortcomings and a remedy. Geophysical Journal International 177, 161–178.
Panza, G.F., 1985. Synthetic seismograms: The Rayleigh waves modal summation. Journal of Geophysics 58, 125–145.
Park, J., Levin, V., 2002. Seismic anisotropy: tracing plate dynamics in the mantle. Science 296, 485–489.
Park, J., Lindberg, C.R., Thomson, D.J., 1987. Multiple-taper spectral analysis of terrestrial free oscillations: part I. Geophysical Journal of the Royal Astronomical Society 91, 755–794.
Parker, R.L., 1994. Geophysical Inverse Theory. Princeton University Press.
Passier, M.L., Snieder, R., 1995. Using differential waveform data to retrieve local S velocity structure or path-averaged S velocity gradients. Journal of Geophysical Research 100, 24,061–24,078.
Patera, A., 1984. A spectral element method for fluid dynamics: laminar flow in a channel expansion. Journal of Computational Physics 54, 468–488.
Pavlis, G.L., 2005. Direct imaging of the coda of teleseismic P waves. In: Levander, A., Guust, N. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 171–185.
Pekeris, C.L., Jarosch, H., 1958. The free oscillations of the earth. Contributions in Geophysics in Honor of Beno Gutenberg. Pergamon, New York, pp. 171–192.
Peter, D., Komatitsch, D., Luo, Y., Martin, R., Goff, L., Casarotti, E., Loher, P.L., Magnoni, F., Liu, Q., Blitz, C., Nissen-meyer, T., Basini, P., Tromp, J., 2011a. Forward and adjoint simulations of seismic wave propagation on unstructured hexahedral meshes. Geophysical Journal International 186, 721–739.
Peter, D., Savage, B., Rodgers, A., Morency, C., Tromp, J., 2011b. Adjoint tomography of the Middle East, abstract S44A-04. 2011 Fall Meeting. AGU.
Pilz, M., Parolai, S., Stupazzini, M., Paolucci, R., Zschau, J., 2011. Modelling basin effects on earthquake ground motion in the Santiago de Chile basin by a spectral element code. Geophysical Journal International 187, 929–945.
Plesch, A., Tape, C.H., Graves, J.R., Small, P., Ely, G., Shaw, J.H., 2011. Updates for the CVM-H including new representations of the offshore Santa Maria and San Bernardino basin and a new Moho surface. Proceedings and Abstracts, vol. 21.
Plessix, R.E., 2009. Three-dimensional frequency-domain full-waveform inversion with an iterative solver. Geophysics 74, WCC149–WCC157.
Pollitz, F., 1999. Regional velocity structure in northern California from inversion of scattered seismic surface waves. Journal of Geophysical Research 104, 15,043–15,072.
Postma, G.W., 1955. Wave propagation in a stratified medium. Geophysics 20, 780–806.
Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model. Geophysics 64, 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, 902–914.
Pratt, R.G., Worthington, M.H., 1990a. Inverse theory applied to multi-source cross-hole tomography. Part 2: elastic wave-equation method. Geophysical Prospecting 38, 311–329.
Pratt, R.G., Worthington, M.H., 1990b. Inverse theory applied to multi-source cross-hole tomography. Part 1: acoustic wave-equation method. Geophysical Prospecting 287–310.
Pratt, R.G., Shin, C., Hicks, G.J., 1998. Gauss–Newton and full Newton methods in frequency-space seismic waveform inversion. Geophysical Journal International 133, 341–362.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes: The Art of Scientific Computing, 3rd edition. Cambride University Press.

Radon, J., 1917. On the determination of functions from their integrals along certain manifolds. Berichte über die Verhandlungen der Sächsische Akademie der Wissenschaften 69, 262–277.

Raitt, R., Shorjr, G., Morris, G., Kirk, H., 1971. Mantle anisotropy in the Pacific Ocean. Tectonophysics 12, 173–186.

Rawlinson, N., Pozgay, S., Fishwick, S., 2010. Seismic Tomography: A Window into Deep Earth. Physics of the Earth and Planetary Interiors 178, 101–135.

Rayleigh, L., 1877. The Theory of Sound. Macmillan, London.

Rayleigh, L., 1906. On the dilatational stability of the Earth. Proceedings of the Royal Society of London Series A 77, 486–499.

Resovsky, J., Trampert, J., 2003. Using probabilistic seismic tomography to test mantle velocity–density relationships. Earth and Planetary Science Letters 215, 121–134.

Revenaugh, J., Jordan, T.H., 1989. A study of mantle layering beneath the Western Pacific. Journal of Geophysical Research 94, 5787–5813.

Rickett, J., Claerbout, J.F., 1999. Acoustic daylight imaging via spectral factorization: helioseismology and reservoir monitoring. The Leading Edge 18, 957–960.

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

Ritsema, J., van Heijst, H.J., 2000. Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow. Science Progress 83, 243–259.

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

Ritzwoller, M.H., Lavely, E.M., 1995. Three-dimensional seismic models of the Earth's mantle. Reviews of Geophysics 33, 1–66.

Ritzwoller, M.H., Shapiro, N.M., Barmin, M.P., Levshin, A.L., 2002. Global surface wave diffraction tomography. Journal of Geophysical Research 107. http://dx.doi.org/10.1029/2002JB001777.

Ritzwoller, M.H., Lin, F.C., Shen, W.S., 2011. Ambient noise tomography with a large seismic array. Comptes Rendus Geoscience 343, 558–570.

Robertsson, J.O.A., Chapman, C.H., 2000. An efficient method for calculating finite-difference seismograms after model alterations. Geophysics 65, 907–918.

Roecker, S., Thurber, C.H., Roberts, K., Powell, L., 2006. Refining the image of the San Andreas Fault near Parkfield, California using a finite difference travel time computation technique. Tectonophysics 426, 189–205.

Romanowicz, B., 1990. The upper mantle degree 2: constraints and inferences. Journal of Geophysical Research 95, 11,051–11,071.

Romanowicz, B., 1991. Seismic tomography of the Earth's mantle. Annual Review of Earth and Planetary Sciences 19, 77–99.

Romanowicz, B., 1994. Anelastic tomography: a new perspective on upper mantle thermal structure. Earth and Planetary Science Letters 128, 113–121.

Romanowicz, B., 1995. A global tomographic model of shear attenuation in the upper mantle. Journal of Geophysical Research 100, 12,375–12,394.

Romanowicz, B., 1998. Attenuation tomography of the Earth's mantle: a review of current status. Pure and Applied Geophysics 153, 257–272.

Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years. Annual Review of Earth and Planetary Sciences 31, 303–328.

Romanowicz, B., 2008. Using seismic waves to image Earth's internal structure. Nature 451, 266–268.

Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves. Surveys in Geophysics 30, 377–405.

Rondenay, S., Bostock, M.G., Fischer, K.M., 2005. Multichannel inversion of scattered teleseismic body waves: practical considerations and applicability. In: Levander, A., Guust, N. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 187–203.

Rost, S., Thomas, C., 2002. Array seismology: methods and applications. Reviews of Geophysics 40. http://dx.doi.org/10.1029/2000RG000100.

Rost, S., Thomas, C., 2009. Improving seismic resolution through array processing techniques. Surveys in Geophysics 30, 271–299.

Roth, E.G., Wiens, D.A., 2000. An empirical relationship between seismic attenuation and velocity anomalies in the upper mantle. Geophysical Research Letters 27, 601–604.

Roux, P., Sabra, K.G., Gerstoft, P., Kuperman, W.A., 2005. P-waves from cross-correlation of seismic noise. Geophysical Research Letters 32. http://dx.doi.org/10.1029/2005GL023803.

Roux, E., Moorkamp, M., Jones, A.G., Bischoff, M., Endrun, B., Lebedev, S., Meier, T., 2011. Joint inversion of long-period magnetotelluric data and surface-wave dispersion curves for anisotropic structure: application to data from Central Germany. Geophysical Research Letters 38. http://dx.doi.org/10.1029/2010GL046358.

Rychert, C.A., Shearer, P.M., 2011. Imaging the lithosphere–asthenosphere boundary beneath the Pacific using SS waveform modeling. Journal of Geophysical Research 116. http://dx.doi.org/10.1029/2010JB008070.

Rychert, C.A., Fischer, K.M., Rondenay, S., 2005. A sharp lithosphere–asthenosphere boundary imaged beneath eastern North America. Nature 436, 542–545.

Saad, Y., 2011. Numerical Methods for Large Eigenvaue Problems, 2nd edition. SIAM.

Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A., Fehler, M.C., 2005. Extracting time-domain Green's function estimates from ambient seismic noise. Geophysical Research Letters 32. http://dx.doi.org/10.1029/2004GL021862.

Saltzer, R.L., van der Hilst, R.D., Kárason, H., 2001. Comparing P and S wave heterogeneity in the mantle. Geophysical Research Letters 28, 1335–1338.

Sambridge, M.S., 1990. Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D. Geophysical Journal International 102, 653–677.

Sambridge, M.S., Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion. Geophysical Journal International 109, 323–342.

Sambridge, M.S., Kennett, B.L.N., 1986. A novel method of hypocentre location. Geophysical Journal of the Royal Astronomical Society 87, 679–697.

Sambridge, M.S., Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems. Reviews of Geophysics 40. http://dx.doi.org/10.1029/2000RG000089.

Sambridge, M.S., Rawlinson, N., 2005. Seismic tomography with irregular meshes. In: Lavender, A., Nolet, G. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 49–65.

Sambridge, M.S., Tarantola, A., Kennett, B.L.N., 1991. An alternative strategy for non-linear inversion of seismic waveforms. Geophysical Prospecting 39, 723–736.

Saygin, E., Kennett, B.L.N., 2012. Crustal structure of Australia from ambient seismic noise tomography. Journal of Geophysical Research 117. http://dx.doi.org/10.1029/2011JB008403.

Scales, J.A., Snieder, R., 1997. To Bayes or not to Bayes? Geophysics 62, 1045–1046.

Scales, J.A., Snieder, R., 2000. The anatomy of inverse problems. Geophysics 65, 1708–1710.

Scales, J.A., Smith, M.L., Fischer, T.L., 1992. Global optimization for multimodal inverse problems. Journal of Computational Physics 103, 258–268.

Schlue, J.W., Knopoff, L., 1977. Shear-wave polarization anisotropy in the Pacific Basin. Geophysical Journal of the Royal Astronomical Society 49, 145–165.

Shapiro, N.M., Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. Geophysical Research Letters 31. http://dx.doi.org/10.1029/2004GL019491.

Shapiro, N.M., Ritzwoller, M.H., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle. Geophysical Journal International 151, 88–105.

Shapiro, N.M., Campillo, M., Stehly, L., Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise. Science 307, 1615–1618.

Shaw, P.R., 1992. Quantitative comparison of seismic data sets with waveform inversion: testing the age-dependent evolution of crustal structure. Journal of Geophysical Research 97, 19,981–19,991.

Shearer, P.M., 1991. Imaging global body wave phases by stacking long-period seismograms. Journal of Geophysical Research 96, 20,353–20,364.

Shearer, P.M., 1993. Global mapping of upper mantle reflectors from long-period SS precursors. Geophysical Journal International 115, 878–904.

Shearer, P.M., 2000. Upper mantle seismic discontinuities. Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale. : Geophysical Monograph, 117. AGU, pp. 115–131.

Shearer, P.M., Masters, T.G., 1992. Global mapping of topography on the 660-km discontinuity. Nature 355, 791–796.

Shin, C., Jang, S., Min, D.J., 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory. Geophysical Prospecting 49, 592–606.

Shragge, J., Bostock, M.G., Rondenay, S., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves, 2. Numerical examples. Journal of Geophysical Research 10630 (30), 783–793.

Sieminski, A., Liu, Q., Trampert, J., Tromp, J., 2007a. Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods. Geophysical Journal International 171, 368–389.

Sieminski, A., Liu, Q., Trampert, J., Tromp, J., 2007b. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods. Geophysical Journal International 168, 1153–1174.

Sigloch, K., Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes. Geophysical Journal International 167, 271–287.

Sigloch, K., McQuarrie, N., Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography. Nature Geoscience 1. http://dx.doi.org/10.1038/ngeo231.

Silver, P.G., 1996. Seismic anisotropy beneath the continents: probing the depths of geology. Annual Review of Earth and Planetary Sciences 24, 385–432.

Simmons, N.A., Forte, A.M., Grand, S.P., 2006. Constraining mantle flow with seismic and geodynamic data: a joint approach. Earth and Planetary Science Letters 246, 109–124.

Simmons, N.A., Forte, A.M., Grand, S.P., 2007. Thermochemical structure and dynamics of the African superplume. Geophysical Research Letters 34. http://dx.doi.org/10.1029/2006GL028009.

Simons, F.J., van der Hilst, R.D., 2003. Seismic and mechanical anisotropy and the past and present deformation of the Australian lithosphere. Earth and Planetary Science Letters 211, 271–286.

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

Smith, M.L., Dahlen, F.A., 1973. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. Journal of Geophysical Research 78, 3321–3333.

Snellius, W., 1621. Cyclometricus. Reprinted by Kessinger Publishing, 2010.

Snieder, R., 1996. Surface wave inversions on a regional scale. In: Boschi, E., Ekstrom, G., Morelli, A. (Eds.), Seismic Modelling of Earth Structure: Editrice Compositori, vol. 25, pp. 149–173.

Snieder, R., 1998. The role of nonlinearity in inverse problems. Inverse Problems 14, 387–404.

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase. Physical Review E 69. http://dx.doi.org/10.1103/PhysRevE.69.046610.

Snieder, R., Nolet, G., 1987. Linearized scattering of surface waves on a spherical Earth. Journal of Geophysics 61, 55–63.

Snieder, R., Romanowicz, B., 1988. A new formalism for the effect of lateral heterogeneity on normal modes and surface waves — I: isotropic perturbations, perturbations of interfaces and gravitational perturbations. Geophysical Journal 92, 207–222.

Snieder, R., Trampert, J., 1999. Inverse problems in geophysics. In: Wirgin, A. (Ed.), Wavefield Inversion. Springer-Verlag, New York, pp. 119–190.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Snieder, R., Miyazawa, M., Slob, E., Vasconcelos, I., Wapenaar, K., 2009. A comparison of strategies for seismic interferometry. Surveys in Geophysics 30, 503–523.

Souriau, A., Souriau, M., 1983. Test of tectonic models by great circle Rayleigh waves. Geophysical Journal of the Royal Astronomical Society 73, 533–551.

Spakman, W., Nolet, G., 1988. Imaging algorithms, accuracy and resolution in delay time tomography. In: Vlaar, N.J., Nolet, G., Wortel, M.J.R., Cloetingh, S. (Eds.), Mathematical Geophysics: A Survey of Recent Developments in Seismology and Geodynamics. D. Reidel, pp. 155–188.

Stadler, G., 2010. Spectral DG for wave propagation and inversion in coupled acoustic-elastic media. 1st QUEST Workshop, Sardinia, Italy.

Stehly, L., Campillo, M., Shapiro, N.M., 2006. A study of the seismic noise from its long-range correlation properties. Journal of Geophysical Research 111. http://dx.doi.org/10.1029/2005JB004237.

Stupazzini, M., 2006. 3D ground motion simulation of the Grenoble Valley by GeoELSE. Proceedings of the 3rd International Symposium on the Effects of Surface Geology on Seismic Motion (ESG), Grenoble, France.

Stupazzini, M., Paolucci, R., Igel, H., 2009. Near-fault earthquake ground-motion simulation in the Grenoble Valley by a high-performance spectral element code. Bulletin of the Seismological Society of America 99, 286–301.

Stutzmann, E., Montagner, J.P., 1993. An inverse technique for retrieving higher mode phase velocity and mantle structure. Geophysical Journal International 113, 669–683.

Su, W.J., Dziewonski, A.M., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. Physics of the Earth and Planetary Interiors 100, 135–156.

Su, W.J., Woodward, R.L., Dziewonski, A.M., 1992. Deep origin of mid-ocean-ridge seismic velocity anomalies. Nature 360, 149–152.

Su, W.J., Woodward, R.L., Dziewonski, A.M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle. Journal of Geophysical Research 99, 6945–6980.

Suda, N., Shibata, N., Fukao, Y., 1991. Degree-2 pattern of attenuation structure in the upper mantle from apparent complex frequency measurements of fundamental spheroidal modes. Geophysical Research Letters 18, 1119–1122.

Takeuchi, H., Saito, M., 1966. Seismic surface waves. Methods in Computational Physics, 11. Academic Press, New York, pp. 217–295.

Talagrand, O., Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. Quarterly Journal of the Royal Meteorological Society 113, 1311–1328.

Tanimoto, T., 1984. A simple derivation of the formula to calculate synthetic long-period seismograms in a heterogeneous earth by normal mode summation. Geophysical Journal of the Royal Astronomical Society 77, 275–278.

Tanimoto, T., Anderson, D.L., 1985. Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s. Journal of Geophysical Research 90, 1842–1858.

Tape, C.H., Liu, Q., Tromp, J., 2007. Finite-frequency tomography using adjoint methods — methodology and examples using membrane surface waves. Geophysical Journal International 168, 1105–1129.

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

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

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

Tarantola, A., 1984b. Linearized inversion of seismic reflection data. Geophysical Prospecting 32, 998–1015.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data. Geophysics 51, 1893–1903.

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

Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM.

Tarantola, A., 2007. The Square-root Variable Metric Algorithm.

Tarantola, A., Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion. Reviews of Geophysics and Space Physics 20, 219–232.

Tessmer, E., Kosloff, D., Behle, A., 1992. Elastic wave propagation simulation in the presence of surface topography. Geophysical Journal International 108, 621–632.

Thomson, D.J., 1982. Spectrum estimation and harmonic analysis. Proceedings of the IEEE 70, 1055–1096.

Thybo, H., Perchuc, E., 1997. The seismic 8° discontinuity and partial melting in continental mantle. Science 275, 1626–1629.

Tian, Y., Sigloch, K., Nolet, G., 2009. Multiple-frequency SH-wave tomography of the western US upper mantle. Geophysical Journal International 178, 1384–1402.

Tkalčić, H., Romanowicz, B., Houy, N., 2002. Constraints on D″ structure using PKP(AB-DF), PKP(BC-DF) and PcP–P traveltime data from broad-band records. Geophysical Journal International 148, 599–616.

Toksoz, M.N., Anderson, D.L., 1966. Phase velocities of long-period surface waves and structure of the upper mantle. Journal of Geophysical Research 71, 1649–1658.

Trampert, J., van der Hilst, R.D., 2005. Towards a quantitative interpretation of global seismic tomography. Earth's deep mantle: Structure, composition, and evolution. AGU, pp. 47–62.

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

Trampert, J., Woodhouse, J.H., 1996. High resolution global phase velocity distributions. Geophysical Research Letters 23, 21–24.

Trampert, J., Woodhouse, J.H., 2003. Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s. Geophysical Journal International 154, 154–165.

Trampert, J., Deschamps, F., Resovsky, J., Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. Science 306, 853–856.

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

Tromp, J., Komatitsch, D., Hjörleifsdóttir, V., Liu, Q., Zhu, H.J., 2010a. Near real-time simulations of global CMT earthquakes. Geophysical Journal International 183, 381–389.

Tromp, J., Luo, Y., Hanasoge, S., Peter, D., 2010b. Noise cross-correlation sensitivity kernels. Geophysical Journal International 183, 791–819.

Valentine, A.P., Trampert, J., 2011. Reducing the dimension of seismic waveforms using Autoencoders: a step towards fully non-linear tomography, and a practical tool, Abstract S12C-05. 2011 Fall Meeting. AGU.

van der Hilst, R.D., 1999. Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: toward a hybrid convection model. Science 283, 1885–1888.

van der Hilst, R.D., de Hoop, M.V., 2005. Banana-doughnut kernels and mantle tomography. Geophysical Journal International 163, 956–961.

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

van der Hilst, R.D., Engdahl, E.R., Spakman, W., Nolet, G., 1991. Tomographic imaging of subducted lithosphere below northwest Pacific island arcs. Nature 353, 37–43.

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

van der Hilst, R.D., de Hoop, M.V., Wang, P., Shim, S.H., Ma, P., Tenorio, L., 2007. Structure of Earth's core–mantle boundary region. Science 315, 1813–1817.

van der Lee, S., Frederiksen, A., 2005. Surface wave tomography applied to the North American upper mantle. In: Levander, A., Guust, N. (Eds.), Seismic Earth: Array Analysis of Broadband Seismograms. AGU, pp. 67–80.

van der Lee, S., Nolet, G., 1997. Upper mantle S velocity structure of North America. Journal of Geophysical Research 102, 22,815–22,838.

Vandecar, J.C., Crosson, R.S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares. Bulletin of the Seismological Society of America 80, 150–169.

Vasco, D.W., Johnson, L.R., 1998. Whole earth structure estimated from seismic arrival times. Journal of Geophysical Research 103, 2633–2671.

Velasco, A.A., Gee, V.L., Rowe, C., Grujic, D., Hollister, L.S., Hernandez, D., Miller, K.C., Tobgay, T., Fort, M., Harder, S., 2007. Using small, temporary seismic networks for investigating tectonic deformation: brittle deformation and evidence for strike-slip faulting in Bhutan. Seismological Research Letters 78, 446–453.

Vest, C.M., 1974. Formation of images from projections: Radon and Abel transforms. Journal of the Optical Society of America 64, 1215–1218.

Vigh, D., Starr, E.W., Kapoor, J., 2009. Developing earth models with full waveform inversion. The Leading Edge 28, 432–435.

Villaseñor, A., Yang, Y., Ritzwoller, M.H., Gallart, J., 2007. Ambient noise surface wave tomography of the Iberian Peninsula: implications for shallow seismic structure. Geophysical Research Letters 34. http://dx.doi.org/10.1029/2007GL030164.

Vinnik, L.P., 1977. Detection of waves converted from P to SV in the mantle. Physics of the Earth and Planetary Interiors 15, 39–45.

Vinnik, L., Farra, V., 2002. Subcratonic low-velocity layer and flood basalts. Geophysical Research Letters 29. http://dx.doi.org/10.1029/2001GL014064.

Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics 74, WCC1–WCC26.

Waldhauser, F., Tolstoy, M., 2011. Seismogenic structure and processes associated with magma inflation and hydrothermal circulation beneath the East Pacific Rise at 9°50′N. Geochemistry, Geophysics, Geosystems 12. http://dx.doi.org/10.1029/2011GC003568.

Wang, P., de Hoop, M.V., van der Hilst, R.D., 2008. Imaging of the lowermost mantle (D″) and the core–mantle boundary with SKKS coda waves. Geophysical Journal International 175, 103–115.

Wapenaar, K., 2004. Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation. Physical Review Letters 93. http://dx.doi.org/10.1103/PhysRevLett.93.254301.

Weaver, R.L., Lobkis, O., 2003. Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons. The Journal of the Acoustical Society of America 113, 2611–2621.

Weaver, R.L., Lobkis, O., 2004. Diffuse fields in open systems and the emergence of the Green's function (L). The Journal of the Acoustical Society of America 116, 2731–2734.

Weber, Z., 2000. Seismic traveltime tomography: a simulated annealing approach. Physics of the Earth and Planetary Interiors 119, 149–159.

Wessel, P., Smith, W.H., 1991. Free software helps map and display data. EOS Transactions AGU 72, 441.

Widiyantoro, S., van der Hilst, R.D., 1997. Mantle structure beneath Indonesia inferred from high-resolution tomographic imaging. Geophysical Journal International 130, 167–182.

Wiechert, E., 1910. Bestimmung des weges der erdbebenwellen im erdinneren, I. theoretisches. Physikalische Zeitschrift 11, 294–304.

Wiggins, R.A., 1972. The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. Reviews of Geophysics and Space Physics 10, 251–285.

Wilson, D.C., Aster, R.C., 2003. Imaging crust and upper mantle seismic structure in the southwestern United States using teleseismic receiver functions. The Leading Edge 22, 232–235.

Wilson, D.C., Angus, D.A., Ni, J.F., Grand, S.P., 2006. Constraints on the interpretation of S-to-P receiver functions. Geophysical Journal International 165, 969–980.

Woodhouse, J.H., 1974. Surface waves in a laterally varying layered structure. Geophysical Journal of the Royal Astronomical Society 37, 461–490.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006

Woodhouse, J.H., 1976. On Rayleigh's Principle. Geophysical Journal of the Royal Astronomical Society 46, 11–22.
Woodhouse, J.H., 1978. Asymptotic results for elastodynamic propagator matrices in plane-stratified and spherically-stratified earth models. Geophysical Journal of the Royal Astronomical Society 54, 263–280.
Woodhouse, J.H., 1981. A note on the calculation of travel times in a transversely isotropic Earth model. Physics of the Earth and Planetary Interiors 25, 357–359.
Woodhouse, J.H., Dziewonski, A.M., 1984. Mapping the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms. Journal of Geophysical Research 89, 5953–5986.
Woodhouse, J.H., Dziewonski, A.M., 1986. Three-dimensional mantle models based on mantle wave and long period body wave data. Eos Transactions AGU 67, 307.
Woodhouse, J.H., Dziewonski, A.M., 1989. Seismic modelling of the Earth's large-Scale three-dimensional structure. Philosophical Transactions of the Royal Society of London, Series A 328, 291–308.
Wookey, J., Stackhouse, S., Kendall, J.M., Brodholt, J., Price, G.D., 2005. Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties. Nature 438, 1004–1007.
Wu, R.S., Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity. Geophysics 50, 582–595.
Wu, R.S., Toksoz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging. Geophysics 52, 11–25.
Wysession, M.E., 1996. Large-scale structure at the core–mantle boundary from diffracted waves. Nature 382, 244–248.
Xu, Z.J., Song, X.D., 2010. Joint inversion for crustal and Pn velocities and Moho depth in eastern margin of the Tibetan Plateau. Tectonophysics 491, 185–193.
Yan, Z., Clayton, R.W., 2007. Regional mapping of the crustal structure in southern California from receiver functions. Journal of Geophysical Research 112. http://dx.doi.org/10.1029/2006JB004622.
Yang, Y., Forsyth, D.W., 2006. Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California. Journal of Geophysical Research 111. http://dx.doi.org/10.1029/2005JB004180.
Yang, Y., Ritzwoller, M.H., 2008. Characteristics of ambient seismic noise as a source for surface wave tomography. Geochemistry, Geophysics, Geosystems 9. http://dx.doi.org/10.1029/2007GC001814.
Yang, Y., Ritzwoller, M.H., Levshin, A.L., Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe. Geophysical Journal International 168, 259–274.
Yang, Y., Li, A., Ritzwoller, M.H., 2008. Crustal and uppermost mantle structure in southern Africa revealed from ambient noise and teleseismic tomography. Geophysical Journal International 174, 235–248.
Yang, Y., Ritzwoller, M.H., Jones, C.H., 2011. Crustal structure determined from ambient noise tomography near the magmatic centers of the Coso region, southeastern California. Geochemistry, Geophysics, Geosystems 12. http://dx.doi.org/10.1029/2010GC003362.
Yao, H., van der Hilst, R.D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet. Geophysical Journal International 179, 1113–1132.
Yao, H., van der Hilst, R.D., de Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis: I. Phase velocity maps. Geophysical Journal International 166, 732–744.
Yao, H., van der Hilst, R.D., Montagner, J.P., 2010. Heterogeneity and anisotropy of the lithosphere of SE Tibet from surface wave array tomography. Journal of Geophysical Research 115. http://dx.doi.org/10.1029/2009JB007142.
Yu, G.K., Mitchell, B.J., 1979. Regionalized shear velocity models of the Pacific upper mantle from observed Love and Rayleigh wave dispersion. Geophysical Journal of the Royal Astronomical Society 57, 311–341.
Yuan, H.Y., Romanowicz, B., 2010. Lithospheric layering in the North American craton. Nature 466, 1063–1068.
Yuan, X.H., Kind, R., Li, X.Q., Wang, R.J., 2006. The S receiver functions: synthetics and data example. Geophysical Journal International 165, 555–564.
Yuan, H.Y., Romanowicz, B., Fischer, K.M., Abt, D.L., 2011. 3-D shear wave radially and azimuthally anisotropic velocity model of the North American upper mantle. Geophysical Journal International 184, 1237–1260.
Zhang, Y.S., Tanimoto, T., 1992. Ridges, hotspots and their interaction as observed in seismic velocity maps. Nature 355, 45–49.
Zhao, D.P., 2001a. New advances of seismic tomography and its applications to subduction zones and earthquake fault zones: a review. Island Arc 10, 68–84.
Zhao, D.P., 2001b. Seismic structure and origin of hotspots and mantle plumes. Earth and Planetary Science Letters 192, 251–265.
Zhao, D.P., 2009. Multiscale seismic tomography and mantle dynamics. Gondwana Research 15, 297–323.
Zhao, L., Chevrot, S., 2011a. An efficient and flexible approach to the calculation of three-dimensional full-wave Fréchet kernels for seismic tomography — I. Theory. Geophysical Journal International 185, 922–938.
Zhao, L., Chevrot, S., 2011b. An efficient and flexible approach to the calculation of three-dimensional full-wave Fréchet kernels for seismic tomography — II. Numerical results. Geophysical Journal International 185, 939–954.
Zhao, L., Dahlen, F.A., 1995. Asymptotic normal modes of the Earth — III. Frechet kernel and group velocity. Geophysical Journal International 122, 229–325.
Zhao, L., Jordan, T.H., 1998. Sensitivity of frequency-dependent traveltimes to laterally heterogeneous, anisotropic Earth structure. Geophysical Journal International 133, 683–704.
Zhao, D.P., Hasegawa, A., Kanamori, H., 1994. Deep structure of Japan subduction zone as derived from local, regional and teleseismic events. Journal of Geophysical Research 99, 22,313–22,329.
Zhao, L., Jordan, T.H., Chapman, C.H., 2000. Three-dimensional Frechet differential kernels for seismic delay times. Geophysical Journal International 141, 558–576.
Zhao, L., Jordan, T.H., Olsen, K.B., Chen, P., 2005. Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. Bulletin of the Seismological Society of America 95, 2066–2080.
Zhao, L., Jordan, T.H., Chen, P., 2006. Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. Bulletin of the Seismological Society of America 96, 1753–1763.
Zheng, S.H., Sun, X.L., Song, X.D., Yang, Y., Ritzwoller, M.H., 2008. Surface wave tomography of China from ambient seismic noise correlation. Geochemistry, Geophysics, Geosystems 9. http://dx.doi.org/10.1029/2008GC001981.
Zheng, Y., Shen, W.S., Zhou, L.Q., Yang, Y., Xie, Z.J., Ritzwoller, M.H., 2011. Crust and uppermost mantle beneath the North China Craton, northeastern China, and the Sea of Japan from ambient noise tomography. Journal of Geophysical Research 116. http://dx.doi.org/10.1029/2011JB008637.
Zhou, Y., 2009. Surface-wave sensitivity to 3-D anelasticity. Geophysical Journal International 178, 1403–1410.
Zhou, H.W., Clayton, R.W., 1990. P and S wave travel time inversions for subducted slab under the island arcs of the Northwest Pacific. Journal of Geophysical Research 95, 6829–6851.
Zhou, Y., Nolet, G., Dahlen, F.A., 2004. Three-dimensional sensitivity kernels for surface wave observables. Geophysical Journal International 158, 142–168.
Zhou, Y., Nolet, G., Dahlen, F.A., Laske, G., 2005. Finite-frequency effects in global surface-wave tomography. Geophysical Journal International 163, 1087–1111.
Zhou, Y., Liu, Q., Tromp, J., 2011. Surface-wave sensitivity: mode summation versus adjoint SEM. Geophysical Journal International 187, 1570–1576.
Zhu, L., Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions. Journal of Geophysical Research 105, 2969–2980.
Zhu, L., Rivera, L.A., 2002. A note on the dynamic and static displacements from a point source in multilayered media. Geophysical Journal International 148, 619–627.
Zhu, H.J., Luo, Y., Nissen-Meyer, T., Morency, C., Tromp, J., 2009. Elastic imaging and time-lapse migration based on adjoint methods. Geophysics 74, WCA167–WCA177.
Zhu, H.J., Bozda , E., Peter, D., Tromp, J., 2012. Adjoint tomography reveals structure of the European upper mantle. Nature Geoscience 5, 493–498.

Please cite this article as: Liu, Q., Gu, Y.J., Seismic imaging: From classical to adjoint tomography, Tectonophysics (2012), http://dx.doi.org/10.1016/j.tecto.2012.07.006