Seismic Imaging: Past, Present and Future (Tentative)

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

[a]*Department of Physics, University of Toronto, Toronto, Ontario, M5S1A7, Canada*
[b]*Department of Physics, University of Alberta, Edmonton, Alberta, T6G2G7, Canada*

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 shown in this study, they are becoming increasingly resolvable with the fast expanding global and regional seismic networks and improving forward modeling and inversion techniques.

This review paper aims to provide an overview of traditional seismic imaging methods and key debates pertaining to the Earth's mantle, as well as to highlight recent theoretical and computational advances in the forward-modeling methods as well as calculations of accurate sensitivity kernels for the inverse problem. The first part of the paper reviews the traditional surface and body wave tomography based on path-average approximations, partitioned waveforms and free oscillations. The second part highlights inversion techniques that are presently popular due to the proliferation of seismic array data. Key topics include, and not limited to, seismic array imaging and inversion, noise-correlation tomography, and the adaptation of more accurate 'sensitivity kernels' and matrix adjoint in inverse problems. We explore the strengths and weaknesses of each method, and speculate on their possible future improvement based on the expansion of seismic data and computation resources. Particular attention will be given to the recent developments of full 3D numerical simulations and their accuracy, resolution, and potential challenges.

Keywords: seismic tomography, sensitivity kernels, inverse problem, array seismology, computational seismology, adjoint methods

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 thus are 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 though 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 rooted in the same principles as the Abel transform of linear paths in a radially concentric sphere (Herglotz, 1907; Wiechert, 1910). Mathematically the approach was identical to the construction of a spherically symmetrically quantum mechanical potential in 3D (Keller et al., 1956; Snieder and Trampert, 1999). More general forms of linear tomography was presented as the central slice theorem by Johan Radon (1917), which was later shown to reduce to Abel transform for Earth-like geometry (Deans, 1983; Vest, 1974; Novak, 1990). While the mathematics introduced by Radon (1917) are beyond the scope of this review, the simple concept of integrating and inverting for a projection function between two generic points x, y along an arbitrary line embedded in a plane with suitable regularity conditions (Radon, 1917) laid the foundation for future seismic and radiological tomography. The proofs of three important theorems pertaining to the path integration and the generalized form in n-dimensional Euclidian space Rn (see Appendix to Radon, 1917) were instrumental not only for seismic and radiological tomography, but also for effective data de-noising and reconstruction (Vest, 1974; Deans, 1983; Sacchi and Uli; Gu and Sacchi, 2010).

It is unclear whether the subsequent development of CT and seismic tomography are contemporaneous or that the former, inspired the early pioneering efforts of seismic data analysis. In 1971, the first practical CT brain scanner by Dr. G. N. 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" (original refs; Cunningham and Judy, 2000). This breakthrough in medical radiography 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 the 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 inversions of normal mode observations based on Rayleigh's Principle [Rayleigh, 1877], which were instrumental for the development of one-dimensional reference Earth [Dziewonski and Anderson, 1981] and 3D tomographic [e.g., Woodhouse, 1974; Woodhouse and Dziewonski, 1984] structures. In retrospect, the delay of 40+ years from Johan Radon's Central Slice Theorem to the Frechet derivatives (Jeffreys , 1961; Backus and Gilbert, 1967; Takeuchi and Saito, 1972; Woodhouse, 1974) is, to a certain degree, reminiscent of the progression from Alfred Wagner's original theory of 'continental drift' [Wagner, 1915] to the validating observations of magnetic reversals in the late 1950s and early 1960s. The key concepts of Backus of Gilbert [1967] were later successfully applied to regional seismic travel time observations [Aki and Lee, 1976; Lee and Aki, 1976] and free-oscillations [Dziewonski, 1978].

By the late 1970s and early 1980s, the era of seismic tomographic imaging has matured beyond its infancy. Not surprisingly, the driving force behind the impetus of the three-dimension (3D) tomography is, in fact, the milestone development of the first well-accepted average one-dimensional (1D), radially anisotropic model of the Earth – the Preliminary Reference Earth Model (PREM, Dziewonski and Anderson, 1981). This study combines normal mode, travel times, attenuation

observations with physical parameters such as Earth's mass and moment of inertia (Dziewonski and Anderson, 1981) to establish an accurate reference frame, with typical errors of less than 1% between predicted and observed travel times for teleseismic arrivals, for advanced 3D surveys. This seminal study eventually earned the two original authors well-deserved Crawford Award, the Nobel price equivalence for all of geosciences. Shortly after this study was published, a series of pioneering studies both in linear (Nakanishi and Anderson, 1982; Woodhouse and Dziewonski, 1984; Dziewonski, 1984; Dziewonski and Woodhouse, 1986) and non-linear (Tarantola and Valett, 1982; Ho-Liu et al., 1988, 1989) demonstrated increasing levels of syphistication (as well as artistry) and helped propel 'global 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 in the waveform analysis by the Harvard group had to resort to the extended use of 'equivalence' statement to circumvent computer memory limitations. It is fair to state that, after four decades from Backus and Gilbert's first study, the pursuit of higher resolution and accuracy of the tomographic images (e.g., Japan papers?) continues to push computer software and hardware capacities to their allowable limits.

Largely owing to the appealing presentation of the Earth's structure and dynamics, seismic tomography quickly established itself as a critical link between the geology, geochemistry, geodynamics, and Earth's magnetic field analyses (see later sections). Tomographic images are among the most convincing 'observations' in the increasingly multi-disciplinary approach to Earth sciences. This would be a safe assumption if the tomographic problem is solved under ideal datasets with dense, uniform coverage and sensitivity, and if the solution to the tomographic problem is unique. In reality, none of the above conditions are truly met, hence to varying degrees imperfect data coverage and simplifying assumptions have to be tolerated in extracting these images. Adding to the uncertainty is the nonuniqueness of underdetermined inverse problems [Tarantola and Valett, 1984; Menke, 1984]. That is, the same dataset could yield multiple 'models' despite the simple truth that 'there is only one Earth'. Even the interpretation of a data projection, the final stage of a tomographic problem, is not immune from potential pitfalls due to its highly 'visual' nature (i.e., dependency on color and choice of projection angle), as well as to routine references to past tectonic events in spite of the projection's inherent lack of 'memory' beyond the oldest record used in the tomographic inversion.

While a chronological account of some of the tomographic techniques and results may be tempting, a truly objective/exhaustive search of all relevant research results is impractical, if not impossible, due to space limitation. This emphasis of this review is on global seismic applications and the key equations, assumptions and characteristics associated with them. We refer readers to to many reviews written on the subject of seismic tomography, particularly regarding areas that are thinly explored in this study (e.g. Dziewonski amd Woodhouse 1987;Woodhouse and Dziewonski 1989; Masters 1989; Romanowicz 1991; Montagner 1994; Masters and Shearer 1995; Ritzwoller and Lavely 1995; Dziewonski 1996; Masters et al. 2000; Grand, 2000; Fukao et al. 2001, 2009; Romanowicz 1993, 2003; Trampert and van der Hilst 2005; Rawlinson et al., 2010).

2. The Global 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 solution of) the forward problem (Engl et al., 2000). For a forward operator G that maps model parameters to observations based on selected rules or principles of physics,

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

Then the problem of recovering the best model parameters $\mathbf{m}$ from a set of observations is the associated inverse problem (see Wiggins, 1972; Menke, 1984; Zhdanov; Nolet, 1993; Iyer and Hirahara, 1993). Similarly, the tomographic approach introduced by Backus and Gilbert (1967, 1968, 1970) takes advantage of integration (of the simplest kind) and differentiation, two rudimentary mathematical operations that exemplify the forward-inverse relationship. Simple solution of this inverse problem involves the following key steps (Menke, 1984; ??):

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, in most cases a known 1D model structure
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.

Based on Rayleigh's principle (Rayleigh, 1877, 1906; Pekeris and Jarosch, 1958) and path integrals involving differential kernels for the normal modes of Earths oscillations, early work by George Backus and Freeman Gilbert provides the formalism for inverting for Earths density and bulk and shear moduli. While the approach was fundamental for the development of waveform inversions (see section 2.X), the governing principles were equally applicable to the extraction of elastic parameters based on travel time observations. For simplicity, this study first reviews the latter approach despite its delayed debut in geophysical literature in the mid 1970s.

2.2. Travel time inversions

The forward problem involving the calculation of the travel time of a ray can be expressed as

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

where dl is an element of the ray path, v is the wave speed, and $\mathbf{x}_s$ and $\mathbf{x}_r$ are the spatial coordinates of the source and receiver, respectively. The above expression can be evaluated via ray tracing between the two end points as a boundary value problem involving three-dimensional structures. The most common approaches are the 'shooting' method (ref) based on a ray incidence angle and integration. For faster convergence, 'bending' methods have also been adopted at the expense of

more complex formulations at discontinuous boundaries within the Earth (see Julian and Gubbins, 1977). Assuming a slowness vector $\mathbf{s}(t)$, which is tangential to the ray path with a magnitude equal to the inverse of the local seismic wave speed, the rate of change of slowness $\mathbf{s}'(t)$ along ray path can be written as (Chernov, 1960)

$$\mathbf{s}' = -\frac{\Delta v}{v}, \tag{3}$$

causing a perturbation of arrival time Δt relative to the unperturbed model prediction t. The main objective of the inverse problem is to determine $\mathbf{s}'$ for all solvable segments along the ray path.

2.3. Linearization and Parameterzations

Combinations of a set of basis functions with desirable properties (e.g., orthogonality) are generally adopted to digitize (or 'parameterize') slowness perturbation as a set of linear equations with discrete unknown weights (or, 'model coefficients'). Early tomographic studies adopted a back-projection approach based on gridded 'blocks' (e.g., Aki and Lee, 1976; Chou and Booker, 1979; Hearn and Clayton, 1986). This parameterization was influential for global travel time inversions in the 1990s using large data sets (e.g., Grand 1997; van der Hilst and Karason, 1997; see Section ??) and, largely owing to its simplicity, remains popular choice in regional and exploration seismic tomographic applications. The formulation falls directly from the discrete form of the travel time

$$\delta t_i = \sum_{j=1}^{N_p} \delta s_j l_{ij} \tag{4}$$

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

Uniform grids based on rectangular cells or blocks are an obvious choice of parameterization due to 1) simplistic assumption of piece-wise straight-line path segments within the blocks, and 2) affinity to the discrete form of forward travel time computation. The shapes or the distribution of the 'blocks' vary broadly among studies that adopt this parameterization approach. In fact, to circumvent the lack of sensitivity of 'great-circle paths' (Backus, 1964) of surface waves to local heterogeneities, early surface wave inversions adopted a priori regionalization schemes that effectively divided the Earth s surface into a few irregular, continent-sized blocks (Kanamori, 1970; Mills, 1978; Dziewonski and Steim, 1982). Problems associated with a priori regionalization surfaced soon after (Kawakatsu, 1982).

The desirable properties of orthogonality and smoothness made spherical harmonics, which was first proposed as the solutions to the Laplace's equation, a popular choice of horizontal basis functions in early global tomographic studies using travel time (Dziewonski, 1984) and waveform

(Woodhouse and Dziewonski, 1984) information. For

$$\frac{\delta v(r,\theta,\phi)}{v(r)} = \sum_{k=0}^{K}\sum_{l=0}^{L}\sum_{m=0}^{l} f_k(r)p_l^m\left({}_kA_l^m\cos(m\phi) + {}_kB_l^m\sin(m\phi)\right) \tag{5}$$

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 plm is the associated Legendre function 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 vertical axis (Press et al., 199?; Su et al., 1992). This 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 angular order of 8). The Earth can be safely considered close to spherically symmetric, and departures of the true Earth from reference models like PREM ('lateral heterogeneities') can be considered to be minor. In the matrix form, the set of model coefficients can be determined via set of linear equations,

$$\mathbf{Gm} = \mathbf{d} \tag{6}$$

where $\mathbf{G}$ contains the wave sensitivity along the assumed path connecting source and receiver, $\mathbf{m}$ corresponds to the set of model coefficients and $\mathbf{d}$ represents the set of travel time or waveform observations. In fact, various gridding or tesselation schemes have been adopted,

2.4. Waveform inversions

2.4.1. One stage inversion technique

2.4.2. Path Average Approximation (PAVA)

Waveform and travel time tomography differs in two major aspects: The input data consist of the seismic waveforms themselves (as opposed to travel times, amplitudes or some other secondary attribute of the recorded data). The underlying numerical method is based on the full wave equation (as opposed to a ray approximation or Born approximation) (REF). Hence, waveform inversions are more accurate than its travel time counterpart, despite considerably more complex formulations and increased nonlinearity associated with waveform fitting.

Woodhouse and Dziewonski (1984) marks a milestone effort in early stages of global waveform tomography. Having recognized the merits of surface wave analyses (e.g., Brune and Dorman, 1963; Knopoff, 1972;) and its largely untapped potential on the global scale (Masters et al., 1982), this study abandoned the controversial a priori regionalization schemes (Toksoz and Anderson, 1966; Kanamori, 1970; Dziewonski, 1971; Jordon, 1981; Kawakatsu, 1983; Souriau and Souriau, 1983) and formally introduced the 'path integral approximation' (or 'path average approximation', for short PAVA; see also Panza, 1985; Woodhouse and Dziewonski, 1986; Mosca and Trampert, 2009) 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 average the source-receiver trajectory. Despite the well-documented pitfalls of this approximation, particularly the inaccurate assumptions of path sensitivities of body wave, triplications, and

frequency-dependent surface waves (Snieder, 1987, 1996; Li and Romanowicz, 1996) along the 'great-circle' arc (Backus, 1964), Woodhouse and Dziewonski (1984) is worthy of a brief review here not only for its ingenuity in combining inverse theory with spherical harmonics, normal mode oscillations and surface/body wave observations, but also for its contribution to global seismic inversions and near-real-time earthquake source parameter inversions (e.g., Centroid Moment Tensor (CMT)) (Dziewonski and Woodhouse, 1981; Chou, Woodhouse and Dziewonski,1989).

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

$$\delta\phi = \int_0^T \delta\omega_{\text{local}}\, dt + n \int_0^T \delta\omega_{\text{local}}\, dt, \tag{7}$$

where $\delta\omega_{\text{local}}$ local denotes the local mantle wave phase perturbations in response to lateral heterogeneities, and the integral is performed over the group travel time along the source-station great circle path T_x for either a major or minor arc. The second term is 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 phase shift $\delta\omega$ (???) where

$$\delta\omega = \frac{1}{T}\int_0^{T_x} \delta\omega_{\text{local}}\, dt = \text{path average}. \tag{8}$$

The essence of PAV is to rewrite the above phase perturbation in terms of this apparent frequency term and source-receiver distance perturbation $\delta\theta$. Given the angular order l of a normal mode and its simple relationship with wavenumber $k = (l + 1/2)/a$ (a is the radius of the Earth), the total phase perturbation becomes (Woodhouse and Dziewonski,1984),

$$\delta\Psi = \pm(l + \frac{1}{2})\delta\theta + \delta\omega(t_y + nt) \tag{9}$$

where t_y corresponding group travel time for either even (major arc, leading by a positive sign) or odd (minor arc, negative) orbits. Perturbation term can be calculated using a simple relationship based on difference between the fictitious phase perturbations of great-circle and minor arc (see Equation 11 of Dziewonski and Woodhouse, 1984). The apparent frequency and distance shifts are different for each mode, and the Frechet kernels (Gilbert, 1967) for the waveform inversion problem are partial derivatives of synthetic waveforms with respect to these two terms.

Assuming PREM (Dziewonski and Anderson, 1981), synthetic seismograms can be computed by normal mode superposition (Gilbert, 1971; Gilbert and Dziewonski, 1975; Tanimoto 1984; Romanowicz et al. 2008) and the PAV facilitates the formulation of linear inverse problem involving model perturbations. In theory, the PAV is sensitive both to phase and to amplitude information of surface wave waveforms; in practice, however, perturbations in waveforms in terms of reduction in data variance is dominated by phase shifts (e.g., Woodhouse and Dziewonski, 1984). Consequently, waveform inversions based on PAVA falls short of properly accounting for the amplitude information on surface waves (e.g., Li and Romanowicz, 1995; Zhao and Dahlen, 1995; Dahlen and Tromp, 1998; Panning et al., 2009). This method is later used in deriving a series of global models of mantle heterogeneities by the Harvard seismology group (e.g., Su et al., 1992, 1994; Liu

and Dziewonsk, 1997; Ekstrom and Dziewonski, 1998; Gu et al., 2001, 2003, 2005; Kustovski et al., 2008; Nettles et al., 2008) and the results are generally consistent with those derived from other waveform or travel-time based tomographic approaches (see Gu et al., 2001; Becker and Boschi, 2002; Laske and Masters, 2003? Mosca and Trampert, 2009 for model reviews).

2.4.3. PAVA in classical travel time tomography

The application of PAVA goes far beyond waveform tomography. While it has been earlier suggested (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 $v(r)$, ray theoretical travel time $t(p, r)$ of a seismic phase (Brune, 1966; Woodhouse, 1978) can be written as,

$$t(p,r) = 2\int_0^{r_b} \left[\frac{1}{v^2(r)} - \frac{p^2}{r^2}\right]^{1/2} dr \tag{10}$$

where r is the instantaneous position of a ray segment from the center of the Earth in kilometers and p represents the ray parameter. Hence, small local vertical travel time perturbations can be formulated as (Mosca and Trampert, 2009),

$$\delta t_{\text{local}} = -2\int_{r_b}^{6371} \frac{1}{v^2(r)}\left[\frac{1}{v^2(r)} - \frac{p^2}{r^2}\right]^{-1/2} \delta \ln v(r,\theta,\phi)\, dr \tag{11}$$

where the $\delta \ln v(r, \theta, \phi)$ represents the natural log perturbation of velocity due to lateral heterogeneity along the ray path. The key is to recognize the inherent connection between the horizontal ray parameter p and the travel time equivalence of the 'local' frequency-phase velocity relationship (Dahlen and tromp, 1998; Mosca and Trampert, 2009) where

$$\frac{\delta\omega_{\text{local}}}{\omega} = \frac{\Delta(p)/T(p)}{c}\frac{\delta c_{\text{local}}}{c} \tag{12}$$

where $\Delta(p)$ and $T(p)$ correspond the ray parameter dependent epicentral and travel times, respectively, and phase velocity c is inherently linked to ray parameter by the simple relationship of $c = p^{-1}$. Equation 12 above is the travel time equivalence of Equation 8, where the total travel time perturbation below is a simple integration of local perturbations in the exact form of Equation 7 (Mosca and Trampert, 2009),

$$\delta T(p) = \frac{1}{\Delta}\int_0^{\Delta} \delta t_{\text{local}}(p,\theta,\phi)\, d\Delta \tag{13}$$

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

2.4.4. Improved waveform approaches

PAVA detailed by Woodhouse and Dziewonski (1984) was initially applied to low-pass filtered (cutoff period: 135 s) long-period mantle wave (named in view of the dispersion of long-period surface waves and its broad depth-sensitivity to mantle structure) and body waves (cutoff period: 45 s). Low computation cost was a crucial advantage for PAVA-based waveform tomography in 1980 s due to the assumed dependence of synthetic seismogram on model perturbations averaged along source-receiver path. However, with the vastly improved computer memory and processor speed, tomographers placed much greater emphasis on accuracy of the inversion outcomes rather than savings on computer speed and memory, once a key advantage of PAVA. Even for a relatively 'smooth' Earth, a condition that PAVA is based upon, there are well documented caveats to this assumption. First, PAVA is inaccurate for body-wave phases, which are sensitive to the geometrical banana-shaped (Dahlen and Nolet, ??) paths rather than horizonally averaged model perturbations, in a laterally heterogeneous Earth. More accurate kernels for surface waves were developed by considering mode-branch coupling (Li and Romanowicz, 1995) while reducing finite-frequency sensitivity to planar, 2-D sensitivity function path. Scalar Exponent Approximations (SEA; Marquering and Snieder, 1995) were also introduced to account for mode coupling. In the same year, Zhao and Dahlen (1995) derived the asymptotic expressions for the Frechet kernels of modal eigenfunctions for computations of exact waveforms for a heterogeneous Earth. Differential waveforms (Passier and Snieder, 1995) have also been suggested as a partial solution to improve the accuracy of PAVA. The majority of the approaches were computationally expensive at the time though, as evidence by a wide range of signature global studies in recent literature (e.g., Megnin and Romanowicz, 200?; Gung et al., 2004; Panning et al., 200?; XX and Romawicz, ??; Yuan and Romaniwicz,??), their usage has became more routine with ever-improving computational capability.

Inaccurate body wave path assumption is only part of a problem associated with PAVA.Even with improved geometrical path integrals, PAVA would remain a gross assumption due to the lack of consideration finite frequency effects of body (Nolet and Dahlen, ?) and surface (Zhou et al., 2004) waves. This problem is highlighted by the use of 'wiry' ray theoretical kernel in travel times, the waveform equivalent of PAVA (see last section), and requires further discussions (see Section 3).

2.4.5. Multi-stage inversion schemes

Based on Frechet kernels, the inversion approach outlined the preceding sections aimed to resolve perturbations in seismic velocities in reference to a starting 1D (or known 3D) model directly from travel time or waveform/phase perturbations. Woodhouse and Dziewonski (1984) suggested that a direct inversion for the coefficients are more reliable based on the consideration of the roughness of Rayleigh and Love waves at 200 s period. Still, this one-step inversion process is not a requisite but a choice. In fact, the earliest tomographic applications (especially in medical imaging) based the central slice theorem (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 proposed in the recovery of mantle velocity structures, most notably Partitioned Waveform Inversion (PWI) (Nolet, 1990) applied to data from the Network of Autonomously Recording Seismographs (NARS).

The PWI method introduced by Nolet (1990, 2008) contains two key steps: 1. produce model

1D waveforms of observations based on nonlinear optimation methods, and 2. solve for 3D structure via linear inversions of linear model constraints from path- averaged misfits (e.g., Nolet, 1990; Maquering et al., 1996; van der Lee and Nolet, 1997; Das and Nolet, 1997; Maggie and Priesley, 2005). The first step is the more conceptually important in the Nolet (1990) and computationally intensive if the number of model parameters is sizeable. Unlike equation (8), the author chose to recast the path integral in terms of a more confined model domain and define the path average shear velocity perturbation as (Nolet, 1990),

$$\delta v(r)^{(j)} = \frac{1}{\Delta_j} \int_{\text{source}}^{\text{receiver}} [v(r,\theta,\phi) - v^{(j)}]\, d\Delta \tag{14}$$

where $v^{(j)}(r)$ represents the radially symmetric reference speed, $v(r,\theta,]phi)$ corresponds to the 3D shear velocity variation 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,

$$\delta k_n(\omega)^{(j)} = \int_0^a \left[\frac{\partial k_n^{(j)}(\omega)}{\partial v^{(j)}(r)} \right] \delta v^{(j)}(r)\, dr \tag{15}$$

where a is the radius of the Earth and the dependent variable is the radially varying shear velocity along the path. By expressing the path-average wavenumber perturbation as a function of the depth-dependent velocity perturbation, the spectrum of the wavefield $S_j(\omega)$ at this station becomes an exponential function of the path integral (Nolet, 1990). For the WKBJ approximation, the velocity term can be further discretized using a set of radial basis function $h_n(z)$ where $\delta v(r)^{(j)} = \sum_{n=1}^{N} C_n h_n(z)$, z is a point along the depth axis, and $n = 1, \cdots, N$ corresponds to the total number of horizontal layers in the 1D path-averaged model.

To solve for the unknown coefficient C_n, Nolet (1990) defined an objective function containing the squared difference between the observed and synthetic time series after filtering, inverse Fourier transform and windowing. An optional Bayesian term was also introduced to optimize the weight of each datum (Nolet, 1990). The final operation during the partitioned inversion (step 1) is to seek model coefficients $C_n^{(j)}$ (for the n-th layer along j-th path) using a nonlinear Conjugate Gradient method (Press et al., 1986) that minimize the objective function. Since there is no specific dependence on the background model, multiple 1D models can be used in unison to characterize a single path.

To construct 3D variations in shear velocities $\delta v(r,\theta,\phi)$ from the model coefficients $C_n^{(j)}$ (i.e., step 2), Nolet (1990) defines a new data vector d based on the vector C containing the model coefficients,

$$\mathbf{d} = \mathbf{S}^T \mathbf{C} \tag{16}$$

where $\mathbf{S}^T$ is the transpose of a matrix containing eigenvectors of the Hessian matrix computed from the 2nd derivative of the objective function with respect to model coefficients. The main advantage of eigenvalue decomposition is to remove the linear dependencies in the original Hessian matrix (Nolet, 1990). Fewer linearly independent equations involving partitioned model coefficients are

further expanded using an orthogonal set of basis functions,

$$\delta v(r)^{(j)} = \sum_{i=1}^{M}\sum_{j=1}^{M} S_{ij} d_j h_i(r) = \sum_{i=1}^{M}\sum_{j=1}^{M} d_j g_j(r) \tag{17}$$

where M is the row rank of the eigenmatrix and $h_i(r)$ is the radial basis function as a function of radius. The main equation that links the newly defined datum to the 3D velocity perturbation $v(r, \theta, \phi)$ relative to partitioned1D velocity model $v^{(j)}(r)$ for path j is

$$\frac{1}{\Delta_j}\int_0^{\pi}\int_{\text{path j}} \left[v(r,\theta,\phi) - v^{(j)}(r)\right] g_i^{-1}(r)\, dr d\Delta = d_i \pm \delta d_i, \tag{18}$$

where $g_i^{-1}(r)$ represents the estimated inverse of $g_i(r)$ for the i-th path and Δ_j corresponds to the source-receiver distance for the j-th path. This equation is in the form $\mathbf{Gm} = \mathbf{d}$ with $g_i^{-1}(r)$ as the sensitivity kernel and $[v(r, \theta, \phi)$ as the unknown model vector. Hence, the 3D velocity perturbations can be solved directly via a linearized inversion approach (Nolet, 1990).

A key advantage of PWI is flexibility, since multiple starting models can be simultaneous introduced during the 1D waveform fitting procedure depending on the path properties of individual source-receiver pairs. Various content-scale studies have benefited from PWI in the investigation of crustal thickenss (Das and Nolet, 1998; Manaman et al., 2011) and upper mantle structure (van der Lee and Nolet, 1997; Maggi and Priesley, 2005; van der Lee and Frederiksen, 2005; Maggie and Priesley, 2005; Fletcher, 2011) Similar to PAVA (Woodhouse and Dziewonski, 1984), however, the original form of PWI (Nolet, 1990) approximates the path sensitivity of body waves based on the WKBJ assumption associated with surface wave modes. For this reason, path inaccuracy, finite frequency effects, and weak amplitude sensitivity remained problematic for PWI, although the former issue was largely resolved in the subsequent modifications that incorporated mode coupling (e,g., Marquering and Snieder., 1995; Marquering et al., 1996) and amplitude information (Panning et al., 2009).

A different kind of two-step inversion approach has also been frequently adopted in surface wave analyses. These studies first seek solutions to surface wave phase velocity (Nakanishi and Anderson, 1984; Nafaf et al., 1984; Tanimoto and Anderson, 1985; Montagner, 1986; Trampert and Woodhouse, Ekstrom and Dziewonski, 1998; Ritswoller, 2002; Yang and Ritswoller, 2006), group velocity (Ekstrom et al., 1998;) or arrival angle information(Larson et al., 1998) for a range of frequencies, then construct the 3D structures by inverting the resulting local dispersion curves or phase/group velocity maps (e.g., Ekstrom and Dziewonski, 1998; Shapiro and Ritswoller, 2002). Stutzmann and Montagner (1993) proposed a two-step waveform inversion scheme for surface wave fundamental and higher modes that incorporated amplitude information. Instead of targeting full waveforms, as did in 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

'The Earth is usually assumed to be isotropic to the propagation of seismic waves. It should be stressed that this assumptions made for mathematical convenience. The fact that a large body

of seismic data can be satisfactorily modeled with this assumption does not prove that the Earth is isotropic. There is often a direct trade-off between anisotropy and heterogeneity.' —– Don Anderson in Theory of the Earth (1989)

Since the early 1960's, the dependence of seismic wave speed variations on particle motion or direction of propagation (generally known as 'seismic anisotropy') have been documented through the discrepant observations of Pn (Hess, 1964; Raitt et al., 1971) and surface (McEvilly, 1976; Aki, 1968; Forsyth, 1975; Yu and Mitchell, 1979; Schule and Knopoff, 1977) waves, as well as anomalous particle motions associated with body waves (Anderson, 1966; Hirn, 1977; Ando et al., 1980). A range of causes on the Earth's crust and mantle long been recognized as tradeoffs between heterogeneity and anisotropy have long been recognized in tomographic modeling (e.g., Smith and Dahlon, 1973; Dziewonski and Anderson, 1981; Woodhouse and Dziewonski , despite the often-adopted isotropic assumption for mathematical or computational convenience (Anderson, 1989).

Inversions of general anisotropic bodies require the recovery of 21 independent elastic parameters according to Hooke2019s 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., Ekstrom and Dziewonski, 1998), a back-of-envelope calculation suggests a total number of unknowns to be $(20+1)2(13+1)21 \approx 130,000$. For a one-step, non-gradient based method such as Cholesky decomposition (Su et al., 1994; Press et al., 1993; Ekstrom and Dziewonski, 1998; Gu et al., 2001), this would require 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 inversions for a fully anisotropic mantle, tomographic inversions targeting the anisotropic wave speeds garnered considerable attention in the early 1980s (Dziewonski and Anderson, 1981; Woodhouse, 1981; Anderson and Dziewonski, 1983; Nataf, 1984) and the trend continued in the mid 19902019s following the increased affordability and accessibility of supercomputers. Simplifying geometrical considerations were introduced based on symmetry axis of anisotropic mineral (Postma, 1955; Anderson, 1961, Backus, 1962; Christensen and Crosson, 1968; Christensen and Salisbury, 1979; Dziewonski and Anderson, 1981). By assuming an axis of symmetry, generic anisotropy is reduced to transversely isotropy (or radial anisotropy) (e.g., Anderson and Dziewonski, 1982) that adopted by five independent elastic parameters A, C, N. L, and F (Love, 1927). The first four elastic constants can be determined by measurements of velocities in two orthogonal directions and F is a function of the velocities at intermediate directions (Dziewonski and Anderson, 1981). Transverse isotropy is the most physical model of anisotropic solids due to its wide-ranging applicability to laminated or finely layered solids, oriented cracks or melt zones and, more importantly, preferentially oriented mantle minerals such as olivine and pyroxine due to shear deformation (Anderson, 1989; Silver, 1995).

The inversion for Linearized anisotropic inversion methods pertaining to transverse isotropic solids were provided by Dziewonski and Anderson (1981) and the addendum to this study provided by Woodhouse (1981) for waveform and travel time data, respectively. Based on theory initially

developed by Takeuchi and Saito (1972) and the definition of Frechet derivatives (e.g., Backus and Gilbert, 1967), a relative change in the squared eigenfrequency of a normal mode can be written as (Dziewonski and Anderson, 1981; Anderson and Dziewonski, 1982), ????

$$\frac{\delta\omega^2}{\omega^2} = \int_0^l r^2 dr(A\delta A + C\delta C + N\delta N + L\delta L + F\delta F + R\delta\rho) \tag{19}$$

where the symbols with tilde signs represent the differential kernels (Takeuchi and Saito, 1972 for alternative expressions) expressible as functions of the eigenfunctions (or the derivatives of eigenfunctions) and the angular degree of spherical harmonics (see Dziewonski and Anderson, 1981 for details). For practical reasons, this expression was recast into Frechet derivatives of more measurable quantities such as wave speeds, the 'horizontal' (V_{XH} where 'X' can be either P or S wave speed), and the last term corresponds to perturbations to the density of the anisotropic structure. Simple scaling relationships between P, S and density became widely accepted (Ekstrom and Dziewonski, 1998; Gu et al., 2005; Yuan and Romanowicz, 2008) to reduce the number of free parameters in the inversion process.

Formulations pertaining to anisotropic inversions based on travel time observations were introduced by Woodhouse (1981) as an addendum to Dziewonski and Anderson (1981). For simplicity we briefly outline the key results on travel-time integral of SH waves and refer the readers to the seminal study by Woodhouse (1981) for SH-wave distance expression and more complex formulations for P-SV waves. Using the same notation for a transversely isotropic solid (A, C, N, L and F), a ray tracing equations is expressible for an SH-polarized wave as an eigenvalue problem

$$\frac{dy}{dr} = \begin{bmatrix} 0 & \omega L^{-1} \\ \omega r^{-2} p^2 N - \omega\rho & 0 \end{bmatrix} y(r) \tag{20}$$

where y is the stress displacement vector, r is the radius, and ω, ρ and p are angular frequency, density and ray parameter, respectively. An eigenvalue of the eigenvalue problem is iq_r, where q_r is an integrand for one-way travel time path integral based on asymptotic ray theory (Brune, 1964; Brodskii, 1975 1978; Woodhouse, 1978, 1981; Aki and Richards, ??),

$$\tau(p) = \int_0^{r_b} q_r(p, r)\, dr \tag{21}$$

where r_b corresponds to the bottoming point of a seismic phase and

$$q_\tau = \left(\frac{\rho}{L} - \frac{N}{L}\frac{p^2}{r^2} \right)^{1/2} \tag{22}$$

This generic expression is the key for computing partial derivatives of SH-wave travel time with respect to all five elastic parameters (Woodhouse, 1981). To obtain explicit expressions practical for seismic velocities from anisotropic travel time inversions, one must map the Frechet derivatives of q_τ with the five elastic constants to horizontal and vertical velocities (omitted by Woodhouse, 1981 but useful for actual implementations) through changes of variable based on simple relationships

$N = \rho V_{SH}^2$ and $L = \rho V_{SV}^2$. The resulting Frechet derivative becomes,

$$\frac{\partial q_\tau}{\partial V_{SH}}\Big|_p = \frac{\partial q_\tau}{\partial N}\frac{\partial N}{\partial V_{SH}} = -\frac{\partial V_{SH}}{\partial q_\tau V_{SV}^2}\frac{p^2}{r^2} \tag{23}$$

$$\frac{\partial q_\tau}{\partial V_{SV}}\Big|_p = \frac{\partial q_\tau}{\partial L}\frac{\partial L}{\partial V_{SV}} = -\frac{q_\tau}{V_{SV}} \tag{24}$$

Substituting these expressions into the path integral over radius, the anisotropic travel time perturbation of a given seismic phase can be expressed as,

$$\delta t = 2\int_0^{r_b} \frac{\partial q}{\partial V_{SH}}\delta V_{SH}\, dr + 2\int_0^{r_b} \frac{\partial q_r}{\partial V_{SV}}\delta V_{SV}\, dr \tag{25}$$

where δV_{SH} and δV_{SV} are the unknown perturbations in horizontal and vertical velocities, respectively, along ray path. Assuming the same level of velocity perturbations in all directions $\delta V \approx \delta V_{SH} \approx \delta V_{SV}$, the above expression simplifies to,

$$\delta t = 2\int_0^{r_b} \left(\frac{\partial q_\tau}{\partial V_{SH}} + \frac{\partial q_\tau}{\partial V_{SV}}\right)\delta V\, dr \tag{26}$$

where the Frechet derivative relative to an apparent isotropic speed V is simply the sum of the Frechet derivatives of the two anisotropic wave speeds. Based on this linear relationship, perturbations to the travel time of an SH-polarized wave due to propagations within an anisotropic Earth can be properly established by 1) normalizing these velocity perturbations by the respective reference wave speeds, and 2) expanding the fractional velocity via a desired choice of 3D basis functions similar to procedures outlined in Sections 2.2 and 2.3.

Perturbations to V_{SV} and V_{SH} are often determined from separate inversions of Rayleigh and Love wave data sets. However, higher-mode surface waves at low wave numbers and frequencies are sensitive to both vertical and horizontal velocities. Hence a simultaneous inversion for both speeds is more accurate than separate inversions of them using polarized kernels. For a given toroidal or spheroidal mode, the sensitivities to V_{SV} and V_{SH} speeds are nonzero and vary in a complex manner as a function of depth (see Fig. ??). The sample ray-theoretical travel time kernels are sensitive to wave polarization and propagation directions (Fig. ??). This is especially important for SH-polarized waves near the source and the receiver, where the wave vectors are close to vertical, and the turning points where the wave vectors are perpendicular to the symmetry axis. SV-polarized waves are only sensitive to vertical shear speed but the sensitivity varies as a function of ray propagation angle (see Fig. ??)

Finally, tomography based on radially isotropic assumption assumes minimal lateral variations in wave speeds in the plane of symmetry. This is a gross approximation considering the abundance of evidence, particularly from Pn wave travel times (Hess, 1964; Raitt et al., 1971) and shear-wave bierfringence (see Long and Silver, 2009 for review). Smith and Dahlen (1973) provided following formulations for changes in wave speed with the azimuth of the ray path with respect to a fixed

direction (also known as azimuthal anisotropy),

$$\frac{\delta c(\mathbf{r},\phi)}{c} = a_0(r) + a_1(r)\cos 2\phi + a_2(r)\sin 2\phi + a_3(r)\cos 4\phi + a_4(r(\sin 4\phi \qquad (27)$$

where $\mathbf{r}$ is a 2D positional vector (2D), ϕ is the relative azimuth of the ray path and coefficients $a_i(i = 1, \cdots, 4)$ represent the unknown coefficients to the major (2ϕ terms) and secondary (4ϕ terms) (e.g., Tanimoto and Anderson, 1984, 1985; Boschi and Woodhouse, 2006). This is a convenient linear equation where the unknown coefficients are easily invertible based on observed phase velocity perturbations $\delta c/c$. A number of global azimuthally anisotropic maps of shear velocity have been produced (Tanimoto and Anerson, 1984, 1985; Montagner and Tanimoto, 1996; Laske and Masters, 1998; Larson et al., 1998; Ekstrom, 200l; Trampert and Woodhouse, 2003). These studies undoubtedly offer critical constraints on the horizontally oriented fabric/flow at crust and mantle depths.

2.6. Inversion with other geophysical or geochemical constraints

The accuracy of tomographic imaging strongly hinges on the data density and sensitivity to the targeted structure. Despite a fast growing number of seismic stations worldwide and the increased diversity of seismic data types, global seismic tomography remains a under-determined inverse problem. A reliable appraisal of the mantle could benefit from additional data constraints on geometry, density and composition during the inversion procedure. For example, boundary sensitive seismic waves such as SS precursors (Gu et al., 2003; Houser et al., 2008) and receiver functions (Chang et al., 2010) can effectively resolve the tradeoff between seismic speeds and depth variations of a seismic interface.

A joint inversion represents a significant improvement over a priori crustal corrections and assumptions of flat interfaces below crust. Simultaneous inversions utilizing traditional body/surface wave phases and added structure-sensitive data generally take the form,

$$\begin{bmatrix} G_0 \\ G_1 \end{bmatrix} \delta\mathbf{v} = \begin{bmatrix} d_0 \\ d_1 \end{bmatrix} \qquad (28)$$

where d_0 adn d_1 and represent conventional seismic phases and other supporting data, respectively, and matrices and denote the corresponding sensitivity matrices. The model vector $\delta\mathbf{v}$ is seismic velocity perturbation, which is modified by the additional weighted equations of constraint during regularization and minimization of the data errors. Additional unknowns could be appended to the velocity vector, for instance, the depths of Moho (Chang et al., 2010; Xu and Song, 2010), core-mantle boundary (Morreli and Dziewonski, 1987? Song,) and transition zone discontinuities (Gu et al., 2003; Lawrence and Shearer, 2006; Reif et al., 2008). Other contributing variables pertaining to magnetotelluric observations (Moorkamp et al., 2007, 2010; Roux et al., 2011), convection-related geodynamical constraints (Forte et al., 1994; Forte, 2000; Forte and Mitrovica, 2001; Trampert et al., 2004; Simmons et al., 2006, 2007), geodesy (Combat and Vallette, 2001; Khan et al., 2008) could all be introduced to the simultaneous inversion procedure.

Data integration, which is a driving force behind some of the recent tomographic efforts (e.g., Yuan and Romanowicz, 2010; Chang et al.,2010; Forte ??), is vital for the pursuit of a self-consistent Earth model (see later sections). The compatibility of the different data sets and the

relative weights to assign the vastly different datasets in the inverse problem, however, remain keys issues of additional, more quantitative investigations.

2.7. 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, 2005; Snieder et al., 2009). The data and application 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 (Clarebout, 1968), and 2) to locate and characterize the source of dominant noise signals (e.g., Sabra et al., 2005; Gu et al., 2007; Yang and Ritswoller, 2008). When the noise sources are uniformly distributed (or, 'ambient'), it can be mathematically proven (see below) that the cross-correlation (or its derivative) of noise records at two nearby receivers is symmetrical about the zero lag time and equivalent to the Green2̌019s 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., Shapiro et al., 2005; Sabra et al., 2005), southern China (Yao et al., 2006) and Europe (Yang et al., 2006; Brzak et al., 2009).

The theory behind noise-correlations underwent stages of development since the equivalence of reflection seismograms. Autocorrelation functions was first documented over 40 years ago in studies of plane waves in layered medium [e.g., Kunetz, 1962; Claerbout, 1968]. Following an expansion of the original hypothesis to what is presently referred to as 'Clarebout conjecture' [Rickett and Claerbout, 1999], the presence of the seismic response in the time-domain cross-correlation function was validated both in theory [Lobkis and Weaver, 2001] and in observations of unfiltered seismic coda waves [Campillo and Paul, 2003]. Subsequent studies using reciprocity and scattered waves [Wapenaar et al., 2004; Snieder, 2004] led to increased confidence in tomographic maps under the assumption of ambient noise sources [Sabra et al., 2005b; Shapiro et al., 2005; Yao et al., 2006; Cho et al., 2007; Gudmundsson et al., 2007; Lin et al., 2007; Moschetti et al., 2007; Lin et al., 2008]. Devoid of dominant, non-uniform noise sources, Greens functions can be sufficiently extracted from both surface [e.g., Sabra et al., 2005a; Snieder, 2004] and body [Roux et al., 2005] waves after averaging months to years of continuous seismic records.

Technically, for a spatially uniform broadband noise distribution and uniform medium with velocity v, the time derivative of the noise correlation function (NCF) C 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 wavefield (Snieder, 2004; Roux et al., 2005; Sabra et al., 2005). Mathematically,this is expressible as (Weaver and Lobkis, 2003; Sabra et al., 2005),

$$\frac{dC_{ij}}{dt} = -G_{ij}(r_1; r_2, t) + G_{ji}(r_2; r_1, t) \tag{29}$$

where C_{ij} corresponds to the NCF for the i-th station and j-th component. The time-domain Greens functions $G_{ij}(r_1; r_2, t)$ relates a unit concentrated impulse displacement in direction i at spatial location to the displacement response in direction j at location (Sabra et al., 2005). The above approximation presents an effective means to extract the Green s function based on derivative of

the correlation function, which can be further approximated by the correlation function itself (e.g., Campillo and Paul, 2003; Shapiro and Campillo, 2004; Bensen et al., 2007), between two station records. The lag times of the correlation peaks offer reliable constraints on regional shear/Rayleigh wave speeds (Shapiro et al., 2005; Yang et al., 2006; Bensen et al., 2007) and anisotropy (e.g., Yao et al., 2006) down to 100 km depth.

Noise correlation tomography is, in principle, no different from earthquake tomography beyond the ambient source assumption. For this reason, all the tomographic approaches from earthquake tomography pertaining to surface and body waves are equally applicable to noise tomography. That being said, the extraction of the Green s function from noise is not as straight forward or clean as conventional earthquake-based analyses, however. Contamination during the retrieval of Green s functions may arise from accidental inclusions of large-amplitude earthquake signals, or from unevenly distributed noise sources near the region of interest. While the influence of earthquakes can be sufficiently suppressed by removing known events based on published catalogues, one-bit normalization, and spectral whitening [Bensen et al., 2007], existence of dominant/persistent noise sources remains a possible origin of highly asymmetric correlation peaks [Stehly et al., 2006; Gu et al., 2007; Yang and Ritzwoller, 2008; Brzak et al., 2009]. Other sources of detectable and potentially contaminating sources for noise-correlation tomography mainly arise from ocean swells and shoreline-wave interactions [Gutenberg, 1931; Longuet-Higgins, 1950], particularly in the frequency range $0.05 - 0.2$ Hz [Schulte-Pelkum et al., 2004; Rhie and Romanowicz, 2006; Gerstoft and Tanimoto, 2007; Webb, 2007; Yao et al., 2009]. These microseismic signals often exhibit strong directivity and, in some areas, seasonal or daily fluctuations [e.g., Schulte-Pelkum et al., 2004; Gu et al., 2007; Yang and Ritzwoller, 2008; Brzak et al., 2009]. Distinctive amplitude asymmetries are often present on the cross-correlations of the surface waves [Paul et al., 2005; Gu et al., 2007; Yang and Ritzwoller, 2008] as direct consequences of source distribution, geometry, strength, and mechanism [Stehly et al., 2006; Brzak et al., 2009]. For the above reasons, the effects of dominant and/or persistent noise sources need to be carefully evaluated both before and after tomographic imaging.

2.8. Resolution of existing global models

Over the years there have been an abundance of . Much of the discussion relied on somewhat qualitative assessment of certain structure and anomalies whose strengths and shapes are, at times, too reliant on the the viewing angles and color scales.

The structures on information provided The resolution of global models

========== Qinya's part is below ==========

2.9. Ray vs Finite-frequency Kernels

One of seismologists' ultimate goals is to model every wiggle on recorded seismograms, and in the process of fitting seismic data, retrieve as much detailed information on velocity structures of the earth's interior as data permit. This is a very challenging task. First of all, 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, most seismologists still rely on approximation theories (e.g., high-frequency ray approximation for body-wave modeling or asymptotical methods for surface-wave modeling) or assume simple 1D

earth structure (e.g., normal-mode theory) to compute Greens functions. The accuracy of these solutions direct affects that of the sensitivity kernels for measurements (also known as 'Fréchet derivatives', referring to functional derivatives of measurements with respect to model parameters) used in seismic imaging problems. Geometrical ray theory for high-frequency waves is simply a zeroth order approximation to seismic wave equations and deals primarily with arrival times of high-frequency seismic waves (onset of phases). Under geometrical, sensitivity kernels of high-frequency arrival times reduce to simple geometrical ray paths connecting sources and receivers. The applicability of ray theory relies on the assumption that lateral heterogeneity in the studied regions are weak and large compared to the dominant wavelength of the seismic waves passing through. For anomaly sizes of the same order or smaller than the dominant wavelength of these waves, high-frequency approximation is no longer valid, and finite-frequency effects need to be accounted for to compute sensitivities for body waves (examined usually at $10 - 50$ seconds) and surface waves (at 50 seconds and longer) at global scales.

Decades of theoretical and computational development have aimed to go beyond simple ray theory by combining solutions to seismic wave equations with first-order Born approximation, and computed sensitivities of broadband seismic phases by mode coupling (e.g., Li and Tanimoto, 1993; Li and Romanowicz, 1995; 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 et al., 2004; Zhou, 2009), normal-mode summation (e.g., Zhao and Jordan, 1998; Zhao et al., 2006; Zhao and Chevrot, 2011a; Capdeville, 2005) and full 3D numerical simulatios (Liu and Tromp, 2008). 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 on infinitesimal rays. In particular, Marquering et al. (1999) and Dahlen et al. (2000) showed that counter-intuitively sensitivity kernels for finite-frequency cross-correlation traveltime measurements have exactly zero values right on the geometrical ray paths, therefore acquire the shape of 3D 'hollow bananas' (namely banana-doughnut). The use of paraxial ray approximation simplified the calculations of traveltimes and geometrical spreading factors of body waves, and allowed sensitivity kernels to be constructed fairly efficiently, enabling them to be use for long period body waves in both global (e.g., Montelli et al., 2004, 2006b; Tian et al., 2009) and regional tomography problems (e.g., Hung et al., 2004; Sigloch et al., 2008; Hung et al., 2010). The counter-intuitive 'banana-doughnut' shape of these sensitivity kernels (and the novelty of its name), as well as some starking 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 (Dahlen and Nolet, 2005; de Hoop and van Der Hilst, 2004; van Der Hilst and de Hoop, 2005, 2006; Montelli et al., 2006a). de Hoop and van Der Hilst (2004) pointed out that the zero sensitivity of these banana-doughnut kernels on the ray paths is mainly the result of a combination of factors (such as the use of ray-theory, choice of measurement, assumption of background media, and model dimensionality), and does not hold in face of realistic media (where caustics are present), different choices of measurements, and model parameterization. In general, these sensitivity kernels can be considered as a regularization of infinitely thin rays, therefore will have similar effects as ray-based tomography once model parameterization and damping are applied. In principle, the use of appropriate sensitivity kernels for broadband measurements of seismic data takes into account

the finite-frequency nature of observations, allows different dataset to be consistently combined in tomography inversions, and makes it possible to image heterogeneities of sizes similar to the first Fresnel zone (e.g., Hung et al., 2004; Chevrot and Zhao, 2007). However, as approximate 2D and 3D sensitivity kernels have been used previously in global tomography studies (e.g., Li and Romanowicz, 1995; Karason and van Der Hilst, 2000) and no significantly discernible improvement in image resolution has been identified as the result of theoretical advancement over simple ray-theory. Also as the global tomography problem is under-determined or at least mix-determined, the main issue hindering the resolution of these output images is the uneven distribution of sources and receivers. As regularization is inevitable for stable tomography inversions and data errors associated with measurements are largely unknown, any 'fair' comparison of finite-frequency-kernel-based and ray-based tomography images is likely impossible (Panning and Romanowicz, 2006; Montelli et al., 2006b).

However, this does not diminish the importance of using proper sensitivity kernels for mapping lateral heterogeneities in global and regional scales, particularly with increasing number of measurements made on broadband seismometers available from global seismic network and regional seismic arrays (Romanowicz, 2008). Indeed, using the consistent sensitivity of measurements based on improved theory gets us one step closer in correctly dealing with the inverse problem, and should be done whenever computational resources allow.

3. Seismic tomography based on numerical simulations of seismic wave equation and adjoint methods

With the advancement of numerical modeling techniques and expansion of computational resources, numerical simulations of seismic wave propagation for complicated and realistic 3D heterogeneous media at both global and regional scales have now become routine practices (e.g., Komatitsch and Tromp, 2002a; Moczo et al., 2007; Castro and Käser, 2010; Tromp et al., 2010). Factors such as gravity forcing, rotation, attenuation, anisotropy, topography of the free surface and internal discontinuities, important for global and regional seismology can now be accounted for accurately. This naturally renders the possibility of calculating finite-frequency sensitivity kernels based on more accurate forward modeling methods (Zhao et al., 2005; Liu and Tromp, 2006, 2008). However, numerically calculating kernels for individual source-receiver pairs in realistic regional and global tomography problems is still prohibitively expensive computationally.

Enpowerd by the numerical tools, and inspired by adjoint methods prevalent in the field of meteorology (e.g., Talagrand and Courtier, 1987; Courtier and Talagrand, 1987) and recently appeared in geodynamics (e.g., Bunge et al., 2003; Liu and Gurnis, 2008), Tromp et al. (2005) made connections between 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 et al., 1989; Fink, 1992, 1997) where the recorded data residual at receivers are physically time-reversed and transmitted to focus back at either original source location or possible locations of model inadequacy, data residuals have been time-reversed to obtain updates on earthquake source parameters based on 3D background models (Tape et al., 2007; Hjörleifsdóttir, 2007), the first step of which is simply a more sophisticated version of the back-projection technique used to image earthquake rupture processes (e.g., Ishii et al., 2005). 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 elastic tensor) 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 the receivers as simultaneous fictitious sources, i.e. two simulations of seismic wave equations. Applications of adjoint methods to seismological problems have been similarly derived by Fichtner et al. (2006a,b).

Seismic inverse problems may be treated as optimization problems, and the 'best' solutions can be sought to minimize some misfit criteria that quantify the difference between data and synthetics. Again due to considerations of computational efficiency, when numerical solvers are used for the forward problem, the optimization/inverse problems are usually solved based solely on local gradients (i.e.,Frechét derivatives) (Epanomeritakis et al., 2008; Tape et al., 2010; Fichtner et al., 2009b) without forming and accessing the Hessian matrix (the second order derivatives of the misfit function). Depending on the technical details of computational practice, this type of optimization method has appeared under many different names, causing some confusion in the literature. We hope to clarify this in later sections, but for the consistency of this paper, we will call it 'adjoint tomography', as a short name for 'tomographic techniques based on numerical simulations of seismic waves and local gradients computed by adjoint methods'. In the following sections, we will first review classical literature on this subject introduced in the exploration setting, focusing on important lessons learned from both synthetic tests and applications to field data that may potentially benefit the earthquake seismology community. We will then provide a straightforward and intuitive derivation of the Fréchet derivatives (a key ingredient of adjoint tomography) based on Born approximation. Applications of adjoint tomography in both local and regional studies will be reviewed and future applications in tomography problems will be discussed.

3.1. Literature on full-wave inversions in exploration seismology

There is natural overlap between 'adjoint tomography' that recently becomes popular in earthquake seismology and the 'full-wave inversion' (FWI) techniques that have been around 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 related to full-wave inversions (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 acousticimaging problem as an optimization problem which seeks the 'best' model parameters that minimizes the generalized least-square sum of weighted data residual (difference between observed seismograms and predicted synthetic seismograms based on a background velocity model). Reflected wavefield is treated as scattered wavefield generated by secondary sources related to perturbations to the background velocity model, and based on the single-scatter approximation, the gradient of the nonlinear misfit function (or the model update direction) 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 residual 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 was shown to be equivalent to pre-stack migration techniques based on imaging principles (Claerbout, 1971). However, instead of reflectivity images with no direct physical interpretation from pre-stack migration, nonlinear waveform inversions allow retrieval of model parameters with physical

meaning (e.g., velocities and density) through successive model updates over iterations. Diffractions, refractions and multiple reflections are also automatically accounted for through successive iterations when forward problems are calculated based on numerical simulations. A priori model may be introduced to improve the well-poseness of the nonlinear inversion. 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 velocity while short-period waveforms are mostly sensitivity to impedance contrast (Tarantola, 1986), corroborating scattering properties of elastic waves observed by Wu and Aki (1985).

Gauthier et al. (1986) and Mora (1987, 1988) proved the viability of full-wave inversion methods based on finite-difference scheme. Monte Carlo inversion techniques were discarded as computational costly and inefficient, while local gradient techniques are much more efficient, and only one forward and one backward simulation are needed to compute the gradient direction and one (or few) more simulation(s) are required to estimate the step length for model update in the gradient direction. Serious concerns were also raised on the nonlinearity of the waveform misfit function, where the final model may be trapped in local minimum. Initial solutions need to close enough to the true model to keep the misfit function in the quasi-linear regime with respect to model perturbations. Numerical tests showed that surface reflected data, due to its nonlinearity with respect to large wavelength structure, recovers mostly the high frequency content of the original model in a few iterations, while transmission data (VSP) 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 full waveform inversion problem under matrix formulation, and pointed out that the best preconditioner is the inverse Hessian matrix, in which case the optimization problem can be solved by one quasi-Newton step.

Moderate success has been achieved when applying full waveform inversion techniques to realistic industry data (Crase et al., 1990; Igel et al., 1996), and serious challenges regarding the nonlinearity of waveform misfit function still remain. Luo and Schuster (1991) remedy this problem by introducing misfit functions that quantifies the traveltime difference between observed and synthetic seismograms instead, and noted that no phase picking is needed when numerical modeling is used. Traveltime misfit functions are much more linear compared to full-waveform misfit, however with reduced model resolution. Another approach proved successful is the frequency-domain iterative waveform inversion, where based on numerical simulations to the frequency-domain seismic wave equations, model were inverted starting from low-frequency where the initial guess model is sufficiently close to the true model to guarantee the linearity of the waveform function, and successively move towards higher frequency using inversion results from the previous frequency as initial models for the current iteration (e.g., Pratt and Worthington, 1990a,b; Pratt, 1999; Sirgue and Pratt, 2004; Brenders and Pratt, 2007; Plessix, 2009). As traveltime tomography (based on full-wave equations) resolves velocity anomalies of the size of the first Fresnel zone, reflection 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 (Luo and Schuster, 1991;

Pratt and Shipp, 1999). Equivalently, multi-scale inversions in the time-domain also proved to be promising based on similar principles (e.g., Virieux and Operto, 2009).

3.2. Application 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, and abundant recordings of local/regional earthquakes are available in a few selected regions. Secondly, 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) which could be complicated by effects such as attenuation or surface-wave focusing and defocusing. Measurements are usually made for individual body-wave phases or as 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, inversions may only be performed in the time-domain. Lastly, due to the computation cost of a single numerical simulation at global or regional scales, ranging 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 as much as possible.

From inversion setup point of view, individual sensitivity kernels of source-receiver pairs form row vectors of the design matrix $\mathbf{G}$ for tomography inversions (section 3.4.3) and are often computed to demonstrate the effect of underlying approximations and finite-frequency effect in global seismology. However, in the exploration literature, as full waveforms are inverted, computing sensitivities for individual data points of the scattered waves (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, computed in two simulations, gives the weighted sum of sensitivity kernels for scattered waves between all source-receiver pairs.

3.3. Forward solver 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 in seismology to solve seismic wave equations for 3D heterogeneous models before continuing our discussion on adjoint tomography. This section will by no means include all references in the related fields, and further references can be found within the given ones.

With the development of compute technology and numerical techniques, significant advances have been made in solving seismic wave equations numerically. Finite difference (FD) methods gained popularity since the early 80s due to its relative simple formulation (e.g., Kelly et al., 1976; Olsen et al., 1995; Igel, 1996; Igel et al., 2002; Moczo et al., 2007). Pseudo-spectral methods (e.g., Tessmer et al., 1992; Carcione, 1994; Carcione et al., 2005) have also been used, although mostly restricted to smooth models. Both methods have difficulty with implementing free-surface boundary conditions for the accurate propagation of surface waves. On the other hand, finite-element type of methods (FEM), although more involved programmingly, implicitly incorporates natural boundary conditions (e.g., Bao et al., 1998; Akcelik et al., 2002). In particular, a specific type of finite-element method, spectral-element method (SEM), was introduced into seismology from

computational fluid dynamics (Patera, 1984) both for global (Komatitsch and Tromp, 2002a,b; Capdeville et al., 2003; Chaljub and Valette, 2004; Fichtner et al., 2009a) and regional (Komatitsch et al., 2004; Peter et al., 2011) seismic wave propagation in 3D heterogeneous media (see Chaljub et al. (2007) for a review). It combines the geometrical flexibility of finite-element methods with the 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 fact that velocity increases with depth to minimize computational cost. Unlike classical finite-element method wherein large linear systems need to be solved, the mass matrix of the SEM method is diagonal due to the choice of Gauss-Lobbatto-Legrendre points as both interpolation and integration points within spectral elements. This makes SEM extremely efficient and scalable on parallel computers, in particular when ported to take advantage of recent Graphics Processing Unit (GPU) hardwares (Komatitsch et al., 2009, 2010). Spectral element methods have been applied to automatically simulate seismic wave propagation for $M_w \geq 3.5$ eevents 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., 2010, http://www.globalshakemovie.princeton.edu). Despite the computational efficiency, it still remains a big challenge to design proper conforming unstructured hexhedral SEM mesh for regions with strong velocity variations that honor predefined internal surfaces and material discontinuities and maintain consistent number of points-per-wavelength throughout the entire model which dictates the accuracy of simulations (Casarotti et al., 2008; Peter et al., 2011). To efficiently simulate wave propagation on such a complicated mesh in parallel also requires considerations on load balancing between individual processors (Peter et al., 2011).

One variant of the spectral-element method is proposed by Capdeville et al. (2003), wherein wave propagation is calculated 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 methodology significant speed up the simulation without loss of accuracy, and has been used in various tomography problems (Lekić and Romanowicz, 2011).

Discontinuous Gerlekin methods also gained popularity in seismology in recent years (e.g., Käser and Dumbser, 2006; Dumbser and Käser, 2006; Hermann et al., 2010; Castro and Käser, 2010). Being more general in formulation than the spectral-element method, it can handle material discontinuities across element boundaries and uses both tetrahedral and hexhedral meshes. In particular, it has proved to be very useful for dynamic rupture simulations where high frequency waves and super-shear near the fault may be captured by denser mesh and better time stepping schemes (e.g., de la Puente et al., 2009). Although 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 regional seismic wave simulations.

In the context of seismic imaging and inverse problem, most material interfaces beneath the surface are not necessarily precisely known. To first order, most geological regions are composed of large blocks with relatively smooth velocity variations within, and spectral-element method may be a more efficient approach, striking a good balance between computational cost and accuracy due to mesh design. Certainly at the scale of the globe, where velocity variations are believed to be long-wavelength ($100 \sim 1000$ km) and smooth ($< 10\%$ in V_s) in most regions, spectral-element

method is still one of the only viable solutions to simulate wave propagation in 3D heterogeneous models.

3.4. Theory and implementations

The basic formulation of adjoint tomography has been derived by numerous researchers, either based on Born approximation and reciprocity of the Greens function (Tromp et al., 2005), or a 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 Greens functions and reciprocity is required (Fichtner et al., 2006a,b). The Lagrange multiplier approach also enabled Liu and Tromp (2008) to derive expressions of Fréchet derivatives for the more complicated global seismic wave equation including factors such as ellipticity, anelasticity and gravity. Here we only symbolically outline expressions for local gradients based on the frequency-domain wave equation, focusing on the difference between adjoint and tradition tomography (either ray or finite-frequency) techniques, from the choice of misfit function to the computation and storage challenges associated with utilizing numerical forward solvers.

3.4.1. Definition of misfit function

The goal of tomographic inversions is to utilize as much information of recorded seismograms as possible to map the internal structures of the earth. It can also be casted as an optimization problem, where given a background velocity model $\mathbf{m}(\mathbf{x})$, we seek to minimize a misfit functional that quantifies the difference between observed seismograms and synthetics calculated for the background model $\mathbf{m}$. For example, a least-square misfit that quantifies the weighted residual between a specific type of measurement a^0_{erp} made to the the p'th phase (or wave group) of the r'th receiver for the e'th event and the corresponding prediction $a_{erp}(\mathbf{m})$ from synthetics is given by

$$\Phi = \frac{1}{2}\sum_e W_e \sum_{rp} W_{rp}[a_{erp}(\mathbf{m}) - a^o_{erp}]^2 = \sum_e W_e \Phi_e, \tag{30}$$

where W_{rp} is the weight assigned based upon measurement error and relative importance of this particular phase and receiver in the inversion and W_e is the weight assigned to the e'th earthquake. We also define the misfit functional associated with only the e'th event as

$$\Phi_e = \frac{1}{2}\sum_{rp} W_{rp}[a_{erp}(\mathbf{m}) - a^o_{erp}]^2. \tag{31}$$

For an minimization problem such as (30), global optimization methods such as the Monte Carlo and genetic algorithms can be used to achieve the global minimum (e.g., Tarantola, 2005; Sambridge, 1999; **?**; Meier et al., 2007). However, these methods often involve many number of evaluations of the forward problem, and are impractical due to the current computational cost associated with solving seismic wave equations fully numerically. Instead local gradient methods which linearize the misfit function with respect to a reference model are favored for realistic data inversions at global and regional scales, and special care should be taken to reduce the nonlinearity of the misfit function and the non-uniqueness of the inverse problem (**?**).

3.4.2. Gradient of misfit function

We will first study the gradient of misfit function for a single event Φ_e. An alternative approach takes advantage of the sparsity of inidividual sensitivity kernels and uses a data reduction scheme that stacks multiple event recordings for a common station (Capdeville et al., 2005). For brevity, we drop the index dependence on event e until later sections. Let us assume that the variation of the measurement $\delta a(\mathbf{m})$ is related to the variation of synthetics $\delta\mathbf{s}$ through a 'connective function' (Luo and Schuster, 1991):

$$\delta a_{rp}(\mathbf{m}) = \int \mathbf{g}_{rp}(\tau)\delta\mathbf{s}(\mathbf{x}_r, \tau; \mathbf{m})\, d\tau. \tag{32}$$

The applicability of adjoint tomography on a particular type of measurement dataset hinges on the ability to derive the above analytical expression $\mathbf{g}(t)$ (Fichtner et al., 2006a; Tape et al., 2010), which may be mathematically quite involved at times (Chen et al., 2010b).

We now substitute expression (32) into the variation of the misfit function (31) to obtain

$$\delta\Phi_e = \int \sum_{rp} W_{rp}[a_{rp}(\mathbf{m}) - a^o_{rp}]\mathbf{g}_{rp}(\tau)\delta\mathbf{s}(\mathbf{x}_r, \tau; \mathbf{m})\, d\tau. \tag{33}$$

This expression can be also rewritten as integration over space and frequency based on Parseval's theorem:

$$\delta\Phi_e = \int\int \sum_{rp} W_{rp}[a_{rp}(\mathbf{m}) - a^o_{rp}]\delta(\mathbf{x} - \mathbf{x}_r)\mathbf{g}^*_{rp}(\omega)\delta\mathbf{s}(\mathbf{x}, \omega; \mathbf{m})\, d\omega dV, \tag{34}$$

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

Given the seismic wave equation in the frequency domain (for simplicity, we ignore the effects of ellipticity, rotation and gravity)

$$-\rho\omega^2\mathbf{s} = \nabla \cdot (C : \nabla\mathbf{s}) + \mathbf{f}, \tag{35}$$

where $\rho(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$ are the background density distribution and elastic tensor relating stress to strain[1] and $\mathbf{f}$ is the force term associated with the event itself. We linearize the dependence of synthetics $\mathbf{s}(\mathbf{x}, \mathbf{m})$ on model $\mathbf{m}$ based on Born approximation (i.e., single-scatter approximation, Nolet, 2008, Chapter 4):

$$-\rho\omega^2\delta\mathbf{s} = \nabla \cdot (C : \nabla\delta\mathbf{s}) + [\omega^2\mathbf{s}\,\delta\rho + \nabla \cdot (\delta C : \nabla\mathbf{s})], \tag{36}$$

i.e. the perturbation to the current wavefield $\delta\mathbf{s}$ is generated by secondary sources from interactions of density and elastic tensor perturbations $(\delta\rho, \delta C)$ with the current wavefield $\mathbf{s}$. This is only a first-order approximation excluding contributions from multiple-scattering, and the perturbed wave field

[1]Note here C can be generalized to a complex number $C = C(\omega)$, representing a visco-elastic medium where stress is related to strain by $T(\omega) = C(\omega) : \nabla\mathbf{s}(\omega)$.

calculated based on (36) will always arrive later than the unperturbed field (Fichtner et al., 2006b).

Let us assume that we have the ability to obtain the Greens function $G(\mathbf{x}, \mathbf{x}_s, \omega)$ to the forward problem (35). This may involve either approximate or semi-analytical methods for 1D or 3D reference models or numerical simulations for 3D reference models. We can then express the perturbed (scattered) wavefield as

$$\delta s(\mathbf{x}, \omega) = \int G(\mathbf{x}, \mathbf{x}', \omega) \left[\omega^2 \mathbf{s}(\mathbf{x}') \delta\rho(\mathbf{x}') + \nabla \cdot (\delta C(\mathbf{x}') : \nabla \mathbf{s}(\mathbf{x}'))\right] \, d\mathbf{x}', \tag{37}$$

which can be substitute into (34) to obtain the relation between event misfit and material properties:

$$\delta\Phi_e = \int \int \sum_{rp} W_{rp}[a_{rp}(\mathbf{m}) - a^o_{rp}] \mathbf{g}^*_{rp}(\omega) G(\mathbf{x}_r, \mathbf{x}', \omega) \left[\omega^2 \mathbf{s}(\mathbf{x}') \delta\rho(\mathbf{x}') + \nabla \cdot (\delta C(\mathbf{x}') : \nabla \mathbf{s}(\mathbf{x}'))\right] d\mathbf{x}' \, d\omega. \tag{38}$$

Based on the source-receiver reciprocity of Greens function (Aki and Richards, 2002, Chapter 2), we define the adjoint wavefield generated by injecting time-reversed weighted residuals at receivers as virtual adjoint sources:

$$\mathbf{s}^\dagger(\mathbf{x}', \omega) = \sum_{rp} W_{rp}[a_{rp}(\mathbf{m}) - a^o_{rp}] \mathbf{g}^*_{rp}(\omega) G(\mathbf{x}', \mathbf{x}_r, \omega), \tag{39}$$

and the variation of the misfit function becomes

$$\delta\Phi_e = \int \int \mathbf{s}^\dagger(\mathbf{x}', \omega) \left[\omega^2 \mathbf{s}(\mathbf{x}', \omega) \delta\rho(\mathbf{x}') + \nabla \cdot (\delta C(\mathbf{x}') : \nabla \mathbf{s}(\mathbf{x}', \omega))\right] d\mathbf{x}' d\omega. \tag{40}$$

The expressions for Fréchet derivatives (functional derivatives) of Φ_e with respect to $\mathbf{m}$, also known as the 'event kernels' (Tape et al., 2007), can be extracted from above expression:

$$\begin{aligned} K^e_\rho(\mathbf{x}) &= \rho(\mathbf{x}) \int \omega^2 \mathbf{s}(\mathbf{x}, \omega) \cdot \mathbf{s}^\dagger(\mathbf{x}, \omega) \, d\omega \\ K^e_{\mathbf{c}}(\mathbf{x}) &= -\mathbf{c}(\mathbf{x}) \int \nabla \mathbf{s}(\mathbf{x}, \omega) \nabla \mathbf{s}^\dagger(\mathbf{x}, \omega) \, d\omega, \end{aligned} \tag{41}$$

involving 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); Sieminski et al. (2007b,a). Then the variation of the misfit function can be written as

$$\delta\Phi_e = \int \left[K^e_\rho(\mathbf{x}) \frac{\delta\rho}{\rho} + K^e_{\mathbf{c}}(\mathbf{x}) \frac{\delta \mathbf{c}}{\mathbf{c}}\right] d\mathbf{x}. \tag{42}$$

Note the adjoint sources used to generate the adjoint wavefield in (39) stem entirely from the definition of the misfit function (31) and its variation (33). The choices of measurement and misfit function affect the gradient of the misfit function (42) through the sensitivity kernels involving the adjoint field.

3.4.3. Hessian Matrix

What has been computed so far is only the gradient of the misfit function. To solve the minimization problem, it may be desirable to obtain second-order derivatives of the misfit function. For brevity, we assume we have casted the spatial function of model parameters $\mathbf{m}(\mathbf{x})$ into discrete model parameters $\underline{m}$ through a set of basis functions. The misfit function can be linearized with respect to a reference model $\underline{m}$ as

$$\Phi(\underline{m} + \Delta\underline{m}) = \Phi(\underline{m}) + \mathbf{g}^T \delta\underline{m} + \frac{1}{2}\Delta\underline{m}^T \mathbf{H}\, \Delta\underline{m} + o(|\Delta\underline{m}|^2), \tag{43}$$

where $\mathbf{g} = \frac{\partial\Phi}{\partial\underline{m}}$ is the gradient vector obtained from representing (41) by the basis functions and $\mathbf{H} = \frac{\partial^2\Phi}{\partial\underline{m}^2}$ is the Hessian matrix. It follows from the definition of of event misfit (31) that

$$\frac{\partial^2\Phi_e}{\partial\underline{m}^2} = \sum_{rp} W_{rp} \left\{ [a_{rp}(\underline{m}) - a^o_{rp}] \frac{\partial^2 a_{rp}}{\partial\underline{m}^2} + \frac{\partial a_{rp}}{\partial\underline{m}} \frac{\partial a_{rp}}{\partial\underline{m}} \right\}. \tag{44}$$

Under the assumption that the measurement function $a_{rp}(\mathbf{m})$ is quasi-linear with respect to $\mathbf{m}$, or the reference model is sufficiently close to the true model such that the residual $[a_{rp}(\underline{m}) - a^o_{rp}]$ is sufficiently small, the first term in the (44) can be ignored, which gives us the approximate Hessian matrix

$$\mathbf{H} = \frac{\partial^2\Phi_e}{\partial\underline{m}^2} \sim \sum_{rp} W_{rp} \frac{\partial a_{rp}}{\partial\underline{m}} \frac{\partial a_{rp}}{\partial\underline{m}} \tag{45}$$

Note that the computation of approximate Hessian requires access to derivatives of individual measurements $\frac{\partial a_{rp}}{\partial\underline{m}}$. If we define matrix $G_{ij} = \frac{\partial a_i}{\partial m_j}$, where we have used a composite index i for measurement of phase p at receiver r, then with weights ignored,

$$\mathbf{H} = G^T G. \tag{46}$$

Note that neither $\mathbf{G}$ nor $\mathbf{H}$ involves actual data a^0_{rp}, and are completely determined by source-receiver geometry and a given background model. In particular, $\mathbf{G}$ is refered to as the 'design matrix'. The optimal model parameters that minimize the quadratic function (43) can be obtained by solving the linear system

$$\mathbf{H}\Delta m = -\mathbf{g}. \tag{47}$$

In general, misfit function (31) can be approximated only locally by quadratic functions (43), and successive iterations in which both H and g are recomputed based on updated velocity models are necessary for the final convergence. If the full Hessian matrix is used as in (44), the inverse method is referred to as the Newton's method, while if the approximate Hessian matrix is used as in eqn (45), it becomes Gauss-Newton's method.

If we set $\Phi_e = a_{rp}(\mathbf{m})$ in section 3.4.2, i.e., an individual measurement, then its gradient is also given by (42), with the adjoint field now generated by the time-reversed connectivity function $g^*_{rp}(\omega)$ as adjoint source. Therefore, in addition to one forward calculation for $\mathbf{s}$, one adjoint calculation is required to compute $\frac{\partial a_{rp}}{\partial\mathbf{m}}$, and the total number of calculations to assemble row

vectors of $\mathbf{G}$ relating to event e is $\sim 2N_R N_P$ 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 (42) requires just one additional adjoint simulation where residuals for all phases and receivers are injected back as adjoint sources simultaneously. When Greens functions are obtained through numerical methods and therefore computationally expensive, assembling $\mathbf{H}$ becomes extremely costly, and is avoided in general. Optimization methods such as nonlinear conjugate-gradient techniques based only on gradient functions (Fletcher, 1987) are used instead (Tape et al., 2007; Fichtner et al., 2009b). These methods effectively approximate the operation of $\mathbf{H}^{-1}$ by applying polynomial functions of $\mathbf{H}$ through successive iterations.

3.4.4. 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 simulation is carried out for the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{x}, t)$ by injecting time-reversed weighted residuals as adjoint sources. However, notice that when expression (41) are transformed into the time-domain,

$$K_\rho^e(\mathbf{x}) = -\rho(\mathbf{x}) \int_0^T \partial_t^2 \mathbf{s}(\mathbf{x}, T-t) \cdot \mathbf{s}^\dagger(\mathbf{x}, t) dt, \tag{48}$$

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 T. This requires access to time t slice of $\mathbf{s}^\dagger$ and time $T-t$ slice of $\mathbf{s}$, which is not generally satisfied when forward and adjoint simulation 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. Next launch the adjoint simulation, and at given time t, access the $T-t$ slice of the $\mathbf{s}$ stored in disk. This poses serious storage requirement on the computation system, since typical global and regional wave propagation simulations may have tens-to-hundreds of millions of grid points, and tens of thousands of time steps. Realizing that the computation of the sensitivity kernels require many fewer number of grid points than that is required to accurately simulate seismic waves, this issue can be remedied by subsample both in time and space based on the necessary spatial and temporal resolutions (e.g., Fichtner et al., 2009b). Another solution, which is computationally more expensive, but avoids the storage requirement comes from recognizing that the elastic seismic wave equation (where $C(\omega) = C$) is invariant under time reflection, therefore can be run backward in time with the last frame of the wavefield as end condition. In this case, when the forward field is reconstructed backward in time, the adjoint simulation is performed forward in time simultaneously, and at any given time t, the $T-t$ slice of $\mathbf{s}$ and the t slice of $\mathbf{s}^\dagger$ will be both update in computer memory, which can be then combined to obtain the contribution of time slice t to the sensitivity kernel as in (41) (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 to reconstruct the forward field backward in time, which also doubles memory requirement for kernel calculations. Figure 3.4.4 is a 2D synthetic example showing the implementation of this technique to calculate sensitivity kernels of a single measurement. Also note 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 check-pointing scheme needs to be implemented to periodically reset the reconstructed wavefield, adding extra storage

requirement (Liu and Tromp, 2008).

As pointed out in section 3.4.3, the gradient of individual measurements may be expensive to compute. On the other hand, the overall gradient of the event misfit function (so-called 'event kernels', or $\mathbf{g}$ as in (47)) can be simply computed based on the interaction of the forward and adjoint wavefield. The gradient of the misfit function of all events (30) is then simply a weighted sum of all event kernels. Figure 2 shows examples of event kernels generated for a synthetic check-board pattern based on a 2D uniform background velocity model. Each event kernel can be interpreted as the velocity anomalies 'seen' by the measurements of this particular event, 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 can be updated in one iteration step simply as $\Delta m = H^{-1}\mathbf{g}$. However, when Hessian is not available, multiple (typically $6-7$) nonlinear conjugate-gradient iterations are necessary to resemble the effect of one Hessian iterations (Akcelik et al., 2002; Chen et al., 2007c; Tape et al., 2007).

3.4.5. Comparison with traditional tomography methods

Also note that although derived under the framework of optimization methods, eqn (47) is also the key equation used in traditional tomography inversions. In this sense, classical ray-based and finite-frequency-based tomography only differ in how H and g are computed. In the classical ray tomography, the design matrix $\mathbf{G}$ is simply filled up by the effective length of rays associated with each model parameter for given source-receiver geometries, while in the case of finite-frequency tomography, different approximate or semi-analytical methods can be used to calculate the Greens function $G(\mathbf{x}, \mathbf{x}_r, t)$, and corresponding sensitivity kernels used in $\mathbf{G}$ and $\mathbf{H}$ matrices. Of course with these approximations and semi-analytical methods, Greens functions can be computed relatively quickly and $\mathbf{H}$ matrix can be assembled efficiently to obtain a model update. However, approximate nature of these techniques restrict them to the application of specific wave types or simple 1D starting models. Even though the misfit function is nonlinear with respect to model parameters, especially for regions with strong velocity heterogeneities, such as the crust, only one Hessian-based inversion step is taken. Any successive iteration which requires recomputing $\mathbf{H}$ matrix under the updated model will be infeasible or computationally expensive, therefore, is rarely attempted (Bijwaard and Spakman, 2000).

3.4.6. Sensitivity kernels for 1D reference velocity models

At global scale, 3D velocity anomalies are relatively small with respect to 1D reference models, except for the upper 200 km and parts of the D$''$ layer. Sensitivity kernels for long-period body and surface wave phases do not differ much for 1D reference models and 3D tomographic models (Liu and Tromp, 2008; Zhou et al., 2011), and traditional tomography inversions based on 1D reference models may still be an efficient and sufficient way to provide velocity model update at long wavelength scales (Capdeville et al., 2003). Therefore there has always been great interest in computing sensitivity kernels (or $\mathbf{G}$) for 1D velocity models based on accurate forward modeling tools such as normal-mode theory (Zhao and Chevrot, 2011a,b), efficient frequency-domain algorithms (Friederich and Dalkolmo, 1995), algorithms that reduce 3D to 2D spectral-element calculations (Nissen-Meyer et al., 2007, 2008) or full numerical simulations (Liu, 2009). This is especially beneficiary for phases that can not be handled accurated by ray theory or asymptotically

methods, such as diffracted phases, important for mapping structures at the core-mantle boundary (Manners, 2008, Chapter 5).

3.4.7. Scattered-integral methods

Another flavor of tomography methods based on full numerical simulations, so-called scattering-integral methods (Chen et al., 2005, 2007b; Chen, 2011), differs from adjoint tomography by its approach to obtain the adjoint wavefield in eqn (39). When only limited number of receivers is involved, Greens functions $G(\mathbf{x}, \mathbf{x}_r, t)$ can be computed numerically for all receivers and stored on temporal and spatial grid points (a total of $3N_r$ number of simulations Eisner and Clayton, 2001). With an additional N_s number of simulations to compute the forward wavefield $\mathbf{s}(\mathbf{x}, \mathbf{x}_s, t)$, sensitivity kernels for a single measurement made on a source-receiver pair can be constructed based on (41). This provides extreme flexibility to compute both the gradient vector g and the building blocks for the design matrix $\mathbf{G}$ and Hessian matrix H, and the inverse problem falls back into the traditional tomography case of solving (47). It also enables experimentation with inversion schemes, such as different weighting schemes (W_e, W_{rp}), exclusion and inclusion of certain sources, receivers or phases, without any extra simulation for either the gradient vector g or the Hessian matrix H. The biggest gain is in terms of a formal resolution analysis. Checker-board tests can be now performed to understand the errors and resolution associated with the final inverted model.

This can be quite an appealing approach when the number of receivers are not excessive (i.e., $3N_r$ forward solutions can be done in reasonable amount of simulation time) and the simulation domain is of relative small size (so that $G(\mathbf{x}, \mathbf{x}_r, t)$ as a function of $(\mathbf{x}, t)$ can be stored on hard disks for efficient access in later computations). Even when interpolation schemes are introduced to reduce the storage from simulation grids onto a coarse grid suitable for the frequency content of the seismic waves (Zhao et al., 2006; Chen et al., 2007b), storage can still be a daunting issue, if the simulation domain is expanded, or shorter period of seismic waves are inverted. The other disadvantage is that the gradient of single measurements are computed based on the reference model, hence the Hessian matrix is only valid for this background model. When the reference model differs significantly from the true model, such as up to $10 - 30\%$ as is the case in some local (Tape et al., 2009) and regional (Fichtner et al., 2009b) tomography problems, one Hessian iteration is not enough to fully eliminate the residuals between observations and synthetic predictions, and successive iterations with updated Greens functions and Hessian matrix based on updated velocity models are necessary. For example, structural inversions for Los Angeles basin in Chen et al. (2007b) showed a maximal 10% of shear-wave speed variations in one quasi-Newton step, while adjoint tomography inversions in Tape et al. (2010) displayed up to 30% shear velocity perturbation for a similar region. Since one Hessian iteration is equivalent to approximately $6 - 7$ gradient-only iterations, for regions with very strong velocity heterogeneities, the computation and storage requirement for the scattering-integral methods may over-weigh the flexibility of Hessian inversions, hence deeming it impractical. Model iterations may be less serious an issue for global tomography inversions, where the P and S velocity anomalies are well within 10% for most parts of the Earth's mantle (Masters and Laske, 2000; Romanowicz, 2003), and sensitivity kernels for 1D background models may not differ that much from a weakly heterogeneous 3D mantle model (Liu and Tromp, 2008; Lekić and Romanowicz, 2011; Zhou et al., 2011). However, the storage of global seismic

wavefield based on 3D reference models well exceeds the current typical storage capabilities, and will be infeasible without significant advance of storage technology.

3.5. Adjoint tomography

In this section, we will discuss several aspects of the adjoint tomography inversion schemes, in particular adapted to the intense computation and storage requirements of numerical simulations. Theoretically, tomography problem can be treated as an unconstrained optimization problem, with vast amount of available literature from applied and compuational math (Nocedal and Wright, 2006). This 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 problem of crust and mantle imaging require more practical considerations in terms of choices of misfit functions, weighting schemes, preconditioners used for nonlinear conjugate techniques, parametrization and regularization. Careful considerations of these subjects offer huge pay-off in decreasing the ill-poseness of the inverse problem and speeding up the ultimate convergence of velocity models.

3.5.1. Misfit function and Measurements

The choice of measurements made to data and synthetics may vary from typical cross-correlation traveltime and amplitude measurements for body waves (Luo and Schuster, 1991; Hung et al., 2000; Sigloch and Nolet, 2006), frequency-dependent phase and amplitude measurements for surface waves (Zhou et al., 2004; Tape et al., 2010) based on multi-taper techniques (Thomson, 1982; Park et al., 1987; Laske and Masters, 1996), to other frequency-dependent functionals such as those based on short-time Fourier transform (Fichtner and Igel, 2008), or generalized seismological density functional (GSDF) (Chen, 2011). Waveform difference between data and synthetics prevalent in exploration problems (section 3.1) has also been utilized in global tomography 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 requires the initial velocity model to be very close to the true model. In comparison, the Rytov approximation relating traveltime delay ΔT to model perturbation is quasi-linear, valid for traveltime delays up to a quarter of the dominant period (Chen et al., 2007c). Although attempts have been made recently to calculate derivatives $\frac{\partial \mathbf{s}}{\partial \mathbf{m}}$ more accurately by including ad-hoc path average phase delays for global surface waves (Romanowicz, 2008; Panning et al., 2009), most global and regional tomography inversion applications choose to invert traveltime or frequency-dependent phase measurements, while using amplitude or waveform information (which may also be affected by uncertainty in earthquake source parameters) to verify final inverted models (Chen et al., 2007b; Tape et al., 2010; Fichtner et al., 2009b).

Selecting proper phases and wave-groups with reasonable similarity in waveform and relatively small traveltime delays is crucial for the successful application of adjoint tomography. However, except in relatively small problems, where manual selection and screening is possible (e.g., Chen et al., 2007b; Fichtner et al., 2009b), some automatic or semi-automatic scheme is needed to mine through the ever-growing broadband data sets. Automatic phase picking algorithms have been applied for body-wave phase selection for decades (e.g., Earle and Shearer, 1994; Bai and Kennett,

2000; Sigloch and Nolet, 2006; Houser et al., 2008). The use of 3D simulation made it possible to select windows based solely on waveform similarity and phase delays without any a priori knowledge on 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. It has been successful applied to both global (Bozda et al., 2010) and regional dataset (Zhu et al., 2010; Kim et al., 2011) and is a key component of adjoint tomography for southern California crust (Tape et al., 2010).

Weights used in the misfit function (30) 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., 2007b)). Alternatively they can be associated with the remaining differences between data and synthetics after the measurement correction is applied (Tape et al., 2010), which basically further emphasizes the requirement on window similarity. Weight functions can also be related to wave amplitudes within windows in order to balance the contribution from different wave types (Tape et al., 2010; Fichtner et al., 2009b).

3.5.2. 3D initial Model and parameterization

One of the main advantages of adjoint tomography compared to traditional tomography methods is that it allows the use of 3D reference models. Numerical simulations are of the same computational cost for either 1D or 3D reference models. Therefore 3D models, derived either from existing tomography studies (Fichtner et al., 2009b) or integrated regional models built on a combination of tomography, local geology and seismic reflection and refraction profiles (Komatitsch et al., 2004; Plesch et al., 2008), can be used as a starting point of adjoint tomography inversions. Adjoint tomography based on 3D reference models is also sometimes dubbed *3D-3D* tomography inversion, where the first and second *3D* refer to 3D background velocity model and 3D physical simulation domain respectively, while the traditional tomography studies are mostly *1D-3D* inversions (Tape et al., 2007). The use of 3D initial models usually results in synthetics that more closely resemble the data, allows more windows or phases to be selected for iterative inversions, and further ensures the applicability of Born or Rytov approximation on the nonlinear misfit functions.

For global and continental scale applications, surface-wave modeling is strongly affected by the 3D heterogenous crust (Bozda and Trampert, 2008; Lekić et al., 2010), and accuracy of surface-wave propagation simulation is closely related to the ability to mesh the crust accurately, even at relatively long period waves (Capdeville and Marigo, 2007). However honoring the Moho such as those provided by CRUST2.0 (*Laske et al.*, http://mahi.ucsd.edu/Gabi/rem.dir/crust/crust2.html), in particular, adapting to the thin oceanic crust significantly decreases mesh sizes, and increases the computation cost for forward simulations. Possible remedies are proposed by introducing a smooth and effective medium for the crust that is equivalent to the original crustal model either by applying corrections based on homogenization of varying coefficients (Capdeville and Marigo, 2007), or matching observations such as local dispersion curves (Fichtner and Igel, 2008), or inverting a smooth crustal model based on short period group velocity dispersion maps (Lekić and Romanowicz, 2011).

Depending on the dataset and measurements used (Section 3.5.1), either isotropic (Chen et al., 2007b) or anisotropic (Fichtner et al., 2010) material properties can be inverted. Tarantola (2005,

section 6.3.3) pointed out that the the bulk-sound and shear-wave speeds are independent, therefore compose a better set of parameters than the commonly used compressional and shear velocities. Travel time or phase delay measurements are less sensitive to density variations for inversions of relatively long period waves (Tarantola, 1986). However density kernels have proved to be instructive in identifying internal sharp velocity contrasts at shorter periods (Liu and Tromp, 2006; Zhu et al., 2009; Luo et al., 2009). Attenuation structure has been inverted together with elastic structures for 2D synthetic cases based on waveform misfit functions (Askan et al., 2007), however, it is much less well-determined as in the case of traditional tomography, due to its trade-off with focusing and de-focusing effect of 3D elastic structures on seismic amplitudes. Therefore in the following sections, we mainly focus on inversions of elastic velocity structure.

3.5.3. Regularization

Due to the non-uniform coverage of sources and receivers and the uneven spatial distribution of sensitivities of phases or wavegroups used, the inverse problem of solving the linear system (47) in earthquake seismology, formally or effectively, becomes under-determined or at least mixed-determined, and therefore ill-posed and non-unique. Some forms of regularization need to be applied 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 chosen from L-curves that show a trade-off between model norm and reduction in misfit function value (Tape et al., 2007).

In principle, basis functions of inversion parameters can be built on the grid points used in the mesh for 3D numerical simulations (Akcelik et al., 2002; Tape et al., 2010). However, grid spacings used in simulations (dictated by the computation accuracy requirements) are generally much smaller than the actual model resolvability based on the bandwidth of given dataset, and using all mesh points makes the inverse problem more ill-posed. Therefore regularization need to be explicitly introduced into the misfit function, providing either a priori model information or some constraints on model variations (such as norm or roughness) (Tarantola, 2005). Although L_2 norms are ubiquitously used in regularization of tomography 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). On the other hand, adjoint tomography relies mainly on the gradient vector of the misfit function for model update at each iteration, and no formal inverse of the Hessian matrix or solution to a linear system with large condition number is sought, therefore, spurious model updates in regions of inadequate source-receiver coverage are less likely to occur. However, implicit regularization should still be introduced by simple local smoothing of the gradient vectors, which may also get rid of excessively large amplitudes of sensitivity kernels at sources and receivers due to the singularities introduced by point moment-tensors or forces in the forward and adjoint simulations (Liu and Tromp, 2008; Tape et al., 2007; Fichtner et al., 2009b).

3.5.4. Inversion schemes: step size and preconditioner

As full 3D forward calculations of the seismic wave equation are numerically expensive, it is desirable to design inversion schemes that reduce the number of numerical simulations as much as possible. When only the gradient vector of the misfit is available to minimize (30), nonlinear conjugate gradient (CG) methods are the most natural solutions (Fletcher, 1987). In nonlinear CG methods, successive search (or so-called conjugate-gradient) directions are defined as linear

combinations of current gradient directions and previous search directions. Model update in a search direction requires the computation of a step length that minimize the misfit function in this direction, i.e. a line search (Nocedal and Wright, 2006). Cubic interpolation which requires two more simulations (one forward and one adjoint simulation) or quadratic interpolation which only requires one more forward simulation can be performed to obtain the optimal step length (Tromp et al., 2005). In practice, for a quasi-linear misfit function, where the misfit function can be approximated locally by a quadratic function of step length in the search direction, either scheme offers similar convergence rate (Tape et al., 2007), therefore quadratic interpolations are favored in line searches of most adjoint-type of inversions (Gauthier et al., 1986; Vigh et al., 2009; Fichtner et al., 2009b). Other 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 in the search direction (Akcelik et al., 2002). Other practical constraints can be placed to restrict the step length. For example, Fichtner et al. (2009b) used envelope misfit reduction as a secondary criterion for selecting a proper step length.

Weights defined in the misfit function (30) emphasize some measurements while down-weighing others, and are effectively preconditioners applied to the Hessian matrix. More generally, a preconditioner P can be applied to (47) to give an equivalent linear system

$$P\,\mathbf{H}\Delta m = P\,\mathbf{g}. \tag{49}$$

P is usually a positive definite matrix that help reduce the condition number of this linear system and speed up the convergence. In particular, the most ideal preconditioner is the inverse Hessian matrix H^{-1} (Pratt et al., 1998), which may reduce $6-7$ steps of nonlinear conjugate-gradient iterations to a simple one-step inversion based on the current model (Tape et al., 2007; Chen et al., 2007c). As mentioned in Section 3.4.3, the H^{-1} preconditioner acts as a 'lens' that scales and focuses the image of a gradient vector into an actual model update. The diagonal terms of 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, the Hessian matrix is not formulated and applying H^{-1} as preconditioner deems impractical. However, when sensitivity kernels of measurements can be relatively cheaply calculated by approximate methods (Li and Tanimoto, 1993; Hung et al., 2000; Zhou et al., 2004) or based on 1D background velocity model (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). Realizing that the Hessian matrix basically captures information on the source-receiver geometry and measurement bandwidth, Igel et al. (1996) and Fichtner et al. (2009b) used less elaborate preconditioner based on simple geometrical arguments to reduce the excessively large sensitivity at the source and receiver, which is also closely tied to the diagonal of the Hessian matrix (Virieux and Operto, 2009). This also has a similar effect as regularization demonstrated by Tape et al. (2007, Figure 10). Alternatively, Tape et al. (2009, 2010) computed a 'best' model update in each iteration that is a linear combination of event kernels based on a source subspace method (Sambridge et al., 1991), which exploits the common features of different event kernels and effectively preconditions the gradient vector. Note in this case no extra test solution needs to be evaluated for a quadratic interpolation in the gradient direction, and the number of equivalent iterations to one Hessian-based step in the nonlinear inverse problem may be further reduced from

6-7, speeding up the convergence.

On the other hand, Hessian matrix and the local curvature of the misfit function can be also approximated based on previous iterations (Akcelik et al., 2003). An extra adjoint simulation used in cubic interpolation for step length also provides local curvature information that can be incorporated for updates to the approximate H^{-1}, as those used in BFGS-type of quasi-Newton optimization algorithms (Nocedal and Wright, 2006; Epanomeritakis et al., 2008), also known as variable metric methods (Tarantola, 2005, section 3.4.3). Applications of these approximate H^{-1} as preconditioners to gradient vectors speed up the convergence and are considered to work better 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 H^{-1} matrices, adaptable to the procedures of adjoint tomography. Most importantly, the final approximate 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) and Akcelik et al. (2003) still approached the nonlinear optimization problem with Gauss-Newton methods. H is not formulated explicitly (for the same reasons discussed in Section 3.4.3), 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 (47) 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. (2007c). However, the total number of outer iterations drastically exceeds those in Tape et al. (2009) and Fichtner et al. (2009b), partly due to the significantly larger velocity perturbation ($\sim 100\%$) used in their synthetic examples, but also partly due to the choice of simple unweighted waveform misfit function. In practical tomography problems, a reasonable 1D or 3D model from other studies is always sought as an initial model, therefore velocity perturbations will be much less (e.g., $< 30\%$ in regional studies and $< 10\%$ in continental scale studies), and convergence will be achieved in many fewer, typically less than 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, the initial solution needs to be sufficiently close to the 'true' solution. Dataset are first filtered at longer periods and inverted for longer wavelength structure or model parameters on a coarser mesh, and its final solution is then used as the initial model in the subsequent inversion of shorter periods data for finer scale of structure, 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 finer and finer scale structures in the close vicinity of the 'true' model. Practical variations of this technique include successively decreasing the upper cut-off period over iterations (Fichtner et al., 2009b) or combining measurements from different period bands, where the longer period data dominate the first few iterations, and successively more measurements from shorter periods are incorporated in later iterations (Tape et al., 2010).

Figure 3 shows the application of adjoint tomography to the inversion of southern California crustal structure based on an automatic windowing scheme (Maggi et al., 2009). Note the large

velocity variations resolved by 16 iterations shown in the final model update (last column of Figure 3a) and detailed hetergeneous structures depicted by the final shear velocity model (2nd column of Figure 3a). Waveform fits are significantly improved in successive iterations such that usable windows are automatically selected from the transverse component seismogram in the final iteration.

3.5.5. Model verification, model error 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 for their final model, and showed significant reduction in both phase delay measurements and waveform misfits. For inversions based on traveltime or phase delay measurements, amplitude measurements also provide independent datasets to validate the final velocity model (Chen et al., 2007b; Tape et al., 2010; Fichtner et al., 2009b). 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.

The diagonal terms of the H matrix provides information on how well a domain is sampled based on given source-receiver geometry and selected phases and wavegroups. Similar to the ray density maps common in traditional tomography, volumetric coverage kernels can be constructed by omitting measurement residuals in the adjoint sources defined in (39). It provides an ad hoc way to understand 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 traditional tomography 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

$$R = (G^T G + \Gamma)^{-1}(G^T G), \tag{50}$$

where Γ is a matrix representing the applied regularization, the lack of individual sensitivity kernels (i.e., row vectors of $\mathbf{G}$) in adjoint tomography deems it infeasible to compute the full resolution matrix. Even checker-board tests common in traditional tomography require the full adjoint inversion machinery when only gradient vectors are available, i.e. one checker-board test is numerically as costly as one full 3D adjoint inversion. Nevertheless, attempts were made to design checker-board tests in order to evaluate model resolutions at both long and short wavelengths. Figure 4 illustrates checker-board tests performed by Fichtner et al. (2009b) for adjoint tomography inversions of the Australasian region. The input model ('true' model) is a checker-board pattern ($3^\circ \times 3^\circ$) onto a high-velocity patch ($\sim 25^\circ \times 25^\circ$) mimicking positive anomalies in the central and western part of the region. When velocity perturbations of the checker-boards are set to be 3%, 6% and even as large as 9%, both the long and short wavelength structures are very well recovered except in regions with limited ray coverage. Note that in traditional tomography, checker-board test is simply a machinery to test the resolution of design matrix ($\mathbf{G}$) or Hessian matrix (H), where the linearization of measurements with respect to model parameters has already be made. In contrast, these check-board tests compute 3D synthetic data based on full 3D numerical simulations, therefore provide a means to understand not only the effect of source-receiver and kernel coverage, but also

the effect of the nonlinearity of the misfit function, particularly in the presence of strong velocity variations.

Since the inverse Hessian matrix H^{-1} provides an approximation to the covariance matrix of inverted model parameters (provided individual data are uncorrelated, Tarantola, 2005), Stadler (2010) made efforts to estimate significant eigenvalues and corresponding eigenvectors of the H matrix, through numerical methods such as Lancos algorithm based on only matrix-vector product of H with a given directional vector q (Saad, 2011). Hq can be achieved by an extra forward and adjoint simulation (Akcelik et al., 2002). Hundreds of such operations may be required to obtain the eigenvalues of H, which deems it still impractical for realistic problems. Alternatively if a cubic interpolation scheme is used for step-length selection (Tromp et al., 2005), and BFGS-type of inversion schemes are selected (refer to section 3.5.4), the approximate H^{-1} matrix saved over iterations can be used as an approximate posterior model covariance matrix. The diagonal terms of the approximate H^{-1} give error-bars associated with model parameters, while the off-diagonal terms are pertinent to correlations between different model parameters. Only one extra adjoint simulation is required each iteration to construct the approximate H^{-1} matrix, and considering the better convergence maybe brought by BFGS-type of inversion algorithms, and potentially more accurate line-search based on cubic interpolation, the ultimate computation cost may even be less than those of nonlinear CG methods. Therefore BFGS-type of inversion schemes 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 H^{-1} instead of H^{-1} is required (Tarantola, 2007), prompting the use of square-root variable metric inversion techniques (Hull and Tapley, 1977).

An evident advantage of the scattering-integral method over the adjoint method is the ability to perform formal resolution analysis, thanks to the availability of the entire design matrix $\mathbf{G}$ (Chen et al., 2007b). The effect of nonlinearity of the misfit functions will also be reflected in checker-board tests as synthetics are calculated numerically for 3D synthetic 'true' model. However, as pointed out in section 3.4.7, for regions with strong velocity heterogeneities, more than one scatter-integral iterations may be required to achieve further misfit reduction and ultimate convergence, making it computationally onerous.

3.5.6. Source Inversions

As 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 traditional tomography inversions (Manners, 2008, Chapter 2), adjoint tomography inversions may explicitly include source parameters as part of the inversion parameter set, i.e. joint inversions as in Chen et al. (2007b). However, for most cases, alternate source and structure inversions are performed, where updates in one help reduce the bias in the other (Tape et al., 2009; Fichtner et al., 2009b). As model updates are moderate between conjugate gradient iterations, source inversions for location and/or mechanisms only need to be performed at selected iterations. Even though Akcelik et al. (2003) inverted both a simple finite-fault source model (with fault plane given) and structural parameters from displacement recordings of a 2D synthetic example, source-time history of events are not inverted in most practical tomography problems due to the severe nonlinearity of related misfit

functions. Events are carefully selected based on magnitudes and event durations are generally much smaller than main period bands of the data used for structural inversions (Tape et al., 2010; Fichtner et al., 2009b). In practice, finite-fault inversions based on adjoint techniques (Kremers et al., 2011) can be well decoupled from structural inversions, and are simple extensions of the popular back-projection techniques for fault rupture imaging (e.g., Ishii et al., 2005).

3.6. Discussions

Here we summarize major advantages of adjoint tomography methods. The use of accurate Greens function through 3D full numerical simulations of seismic wave propagation gives rise to 'true' sensitivity kernels of selected time windows on seismograms. No prior knowledge of phases related to these time windows is assumed, allowing all possible time windows (either as impulses or wave groups) to be included in adjoint inversions. An important consequence is that 3D models, inverted based on traditional tomography techniques, either ray-based or finite-frequency based, can now be exploited as initial starting solutions for adjoint tomography methods, as 3D models cost similar amount of computation time as 1D models to simulate numerically. The ability to recalculate gradients of misfit functions at every iteration is critical and successive linearization of the inverse problem at updated models are necessary for the ultimate convergence of the solution, particularly for regions of large velocity variations. Synthetics for final models can be computed again by 3D numerical simulations, enabling true validation of the quality of final inversion results.

At global scale where velocity variations are moderate, Lekić and Romanowicz (2011) proposed a hybrid quasi-Newton method where misfit functions are evaluated by accurate 3D numerical simulations to minimize the effect of heterogeneous crustal structure on surface waves of 60 seconds and longer, while an approximate partial derivatives $\mathbf{G}$ is computed based on asymptotic methods. The accurate 3D Greens function reduces systematic errors associated with measurements, while the approximate $\mathbf{G}$ matrix emancipates the inversion from onerous gradient calculations as in the adjoint tomography. However, it is not clear how the ultimate convergence should be gauged, whether further iterations are needed, and how they should be performed. Although computationally expensive, adjoint tomography may be the ultimate tool to produce new-generation high-resolution images for laterally highly 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 (Chen et al., 2007a), hot-spots (Montelli et al., 2004) and continental collision zones (Chen et al., 2010a).

Adjoint tomography, besides being computationally intensive, also has its own share of challenges. It suffers from common issues associated with inverse problems. Localized gradient methods require the initial solution to be sufficiently close to the true medium so that the final solution will not be trapped in local minima. This in turns requires either a very good initial guess of the medium, or misfit functions to be designed cleverly so that they stay in the linear regime. Convergence can be extremely slow if the gradient scheme is not properly pre-conditioned. Designing the appropriate preconditioner for seismic inverse problems involving millions of model parameters can be extremely difficult, although its benefit in terms of reduction in number of iterations can be also vast. On the other hand, gradient techniques avoids the problems related to formally inverting the Hessian matrix, which may be quite poorly conditioned with uneven source-receiver distributions. Empirical damping is replaced by implicit regularization of the gradient vector as residual

measurements and their composite sensitivities designate regions of the model that requires updated, rendering the inversion process stable. However, without the Hessian matrix (or its inverse), formal resolution analysis becomes difficult, and standard check-board tests which provides the resolvability of anomalies of various sizes require actually running the full optimization problem over again, 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 feasible remedy at this stage. Formal check-board test and resolution analysis for adjoint tomography are stuill areas of very active research (Fichtner and Trampert, 2011a,b).

3.7. Future directions

With ever-expanding computation power, it becomes more and more feasible to apply adjoint tomography inversions based on full 3D numerical simulations and generate robust and high-resolution images of various regions of the earth's interior, at local (Chen et al., 2007b; Tape et al., 2010), regional (Fichtner et al., 2009b, 2010; Zhu et al., 2010) and global (Bozda et al., 2010) scales. Applications to other geologically and tectonically interesting regions are also on the horizon, especially with the improvement of numerical techniques, computational infrastructure and inversion schemes adapted to local-gradient based inversions. Combined with geodynamical modelling, interpretation of these improved high-resolution images will further our understanding of the temperature, composition and internal dynamics of the earth's interior.

Incorporating amplitude measurements, or performing full waveform inversions as those in exploration seismology (Virieux and Operto, 2009) may be a further step in improving the quality of tomography images. Full waveform inversions may help resolve fine scale features as small as the wavelength of the dominant seismic waves, and may provide information to structures that are less well probed by inversions based solely on travel time information, such as low-velocity zones (Brenders and Pratt, 2007). With improved 3D elastic models, attenuation structures can be also jointly inverted from amplitudes and full waveforms, providing clues to the temperature and composition of studied regions (Askan et al., 2007).

Regional-scale high-resolution imaging has become popular in the past decade thanks to the densely spaced seismic arrays deployed around the world. However, due to the limited spatial distributions of earthquake sources globally, few earthquakes of proper magnitudes may occur in specific regions covered by seismic arrays, and only tele-seismic waveforms can be utilized. For instance, receiver function techniques examines the converted tele-seismic P waves at the Moho to map out the thickness of the crust (e.g., Langston, 1977; Zhu and Kanamori, 2000; Niu and James, 2002). Scattered imaging techniques standard in exploration seismology, such as Kirchhoff migration (Levander, 2003) and generalized Radon transform (Bostock et al., 2001; Wang et al., 2008) are also used to map local anomalies based on diffracted or scattered waves. Adjoint tomography can be very well adapted to robustly image localized structure based on converted or scattered waves of main phases, where sensitivity kernels are constructed based on the interaction of forward plane waves (corresponding to the main phases) and adjoint wavefield generated by converted or scattered waves (Liu, 2009). Forward wavefield can be also generated accurately based on hybrid methods (Moczo et al., 1997; Bielak et al., 2003; Opršal et al., 2009). This provides great potential for full waveform inversions of scattered tele-seismic waves based on localized 3D numerical

simulations, with all the advantages of adjoint tomography, such as model iterations and validation, and will help gain new insights into finer-scale structures of tectonically and dynamically interesting regions.

Acknowledgement

References

Akcelik, V., Bielak, J., Epanomeritakis, I., Biros, G., Ghattas, O., Fernandez, A., Kim, E.J., Lopez, J., OHallaron, D., Tu, T., Urbanic, J., 2003. High resolution forward and inverse earthquake modeling on terascale computers, in: IEEE Conference on Supercomputing.

Akcelik, V., Biros, G., Ghattas, O., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in: Proceedings of IEEE/ACM SC2002 Conference, pp. 1–15.

Aki, K., Richards, P.G., 2002. Quantitative seismology. University Science Books. 2nd edition.

Askan, A., Akcelik, V., Bielak, J., Ghattas, O., 2007. Full Waveform Inversion for Seismic Velocity and Anelastic Losses in Heterogeneous Structures. Bull. Seism. Soc. Am. 97, 1990–2008.

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. Comput. Methods Appl. Mech. Engrg. 162, 85–102.

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

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, 30771–30782.

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

Bozda, E., Zhu, H., Peter, D., Tromp, J., 2010. Towards global adjoint tomography, in: 2010 Fall Meeting, AGU, San Francisco, Calif., pp. S31A–2012.

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

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

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

Capdeville, Y., 2005. An efficient Born normal mode method to compute sensitivity kernels and synthetic seismograms in the Earth. Geophys. J. Int. 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., Romanowicz, B., Gung, Y., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking. Geophys. J. Int. 162, 541–554.

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. Geophys. J. Int. 152, 34–67.

Cara, M., Leveque, J.J., 1987. Waveform inversion using secondary observables. GRL 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.

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, in: 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. Geophys. J. Int. 182, 250–264.

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.

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.

Chen, M., Huang, H., Yao, H., van Der Hilst, R.D., 2010a. Adjoint tomography using Green's functions from ambient noise, in: 2010 Fall Meeting, AGU, San Francisco, Calif.,, pp. S31B–07.

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

Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances. Pure appl. geophys. , submitted.

Chen, P., Jordan, T.H., Lee, E.J., 2010b. Perturbation kernels for generalized seismological data functionals (GSDF). Geophys. J. Int. 183, 869–883.

Chen, P., Zhao, L., Jordan, T.H., 2005. Finite-moment tensor of the 3 September 2002 Yorba Linda earthquake. Bull. Seism. Soc. Am. 95, 1170–1180.

Chen, P., Zhao, L., Jordan, T.H., 2007b. Full 3D tomography for the crustal structure of the Los Angeles region. Bull. Seism. Soc. Am. 97, 1094–1120.

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

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

Claerbout, J.F., 1971. Toward a unified theory of reflector mapping. Geophysics 36, 467–481.

Courtier, P., Talagrand, O., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II : Numerical results. Q. J. R. Meteorol. Soc. 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.

Dahlen, F.A., Nolet, G., 2005. Comment on 'On sensitivity kernels for 'wave-equation' transmission tomography' by de Hoop and van der Hilst. Geophys. J. Int. 163, 949–951.

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

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.

Earle, P.S., Shearer, P., 1994. Characterization of global seismograms using an automatic-picking algorithm. Bull. Seism. Soc. Am. 84, 366–376.

Eisner, L., Clayton, R.W., 2001. A reciprocity method for multiple-source simulations. Bull. Seism. Soc. Am. 91, 553.

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, doi:10.1088/0266–5611/24/3/034015.

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 seismologyII. Applications: traveltimes and sensitivity functionals. Physics of The Earth and Planetary Interiors 157, 105–123.

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). Geophys. J. Int. 173, 519–533.

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 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. Geophys. J. Int. 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.

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. Geophys. J. Int , submitted.

Fink, M., 1992. Time reversal of ultrasonic fields. I. Basic principles. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 39, 555–66.

Fink, M., 1997. Time-reversed Acoustics. Physics Today , 34–40.

Fink, M., Prada, C., Wu, F., Cassereau, D., 1989. Self focusing in inhomogeneous media with time reversal acousticmirrors. IEEE Ultrasonics Symposium , 681–686.

Fletcher, R., 1987. Practical methods of optimization. Wiley-Interscience publication. 2nd edition.

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

Gauthier, O., Virieux, J., Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. Geophysics 51, 1387–1403.

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.

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

de Hoop, M.V., van Der Hilst, R.D., 2004. On sensitivity kernels for wave-equation transmission tomography. Geophys. J. Int. 160, 621–633.

Houser, C., Masters, G., Shearer, P., Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms. Geophys. J. Int. 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., 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, 1–5.

Hung, S.H., Nolet, G., Dahlen, F.A., 2000. Fréchet kernels for finite-frequency traveltimes-II. Examples. Geophys. J. Int. 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, 1–16.

Igel, H., 1996. P-SV wave propagation in the Earth's mantle using lowermost mantle structure. Geophys. Res. Lett 23, 415–418.

Igel, H., Djikpéssé, H., Tarantola, A., 1996. Waveform inverstion 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., 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, doi:10.1038/nature03675 LETTERS.

Karason, H., van Der Hilst, R.D., 2000. Constraints on mantle convection from seismic tomography, in: The History and Dynamics of Global Plate Motions, Geophysica edition. pp. 277–288.

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.

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

Kim, Y., Liu, Q., Tromp, J., 2011. Adjoint centroid-moment tensor inversions. Geophysical Journal International , doi: 10.1111/j.1365–246X.2011.05027.x.

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.

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. Bull. Seism. Soc. Am. 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., Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation - I. Validation. Geophys. J. Int. 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. Geophys. J. Int. 150, 303–318.

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 Earth Discussions 3, 217–248.

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. The effect of planar dipping structure on source and receiver responses for constant ray parameter. BSSA 67, 1029–1050.

Laske, G., Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data. Journal of Geophysical Research 101, 16059–16075.

Lekić, V., Panning, M., Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography. Geophys. J. Int. , no–no.

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

Levander, A., 2003. USArray design implications for wavefield imaging in the lithosphere and upper mantle. The leading Edge , 250–255.

Li, X.D., Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling. Geophys. J. Int. 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, 22245–22272.

Li, X.D., Tanimoto, T., 1993. Waveforms of long-period body waves In a slightly aspherical earth model. Geophys. J. Int. 112, 92–102.

Liu, L.J., Gurnis, M., 2008. Simultaneous inversion of mantle properties and initial conditions using an adjoint of mantle convection. Journal of Geophysical Research 113, 1–17.

Liu, Q., 2009. A database of global finite-frequency sensitivity kernels based upon spectral-element simulations, in: Eos Trans. AGU, 90(52), Fall Meet. Suppl.,, pp. Abstract S33B–1761.

Liu, Q., Tromp, J., 2006. Finite-frequency kernels based on Adjoint Methods. Bull. Seism. Soc. Am. 96, 2283–2397.

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

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 , 568–574.

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

Manners, U.J., 2008. Investigating the structure of the core-mantle boundary region using S and P diffracted. Ph.D. thesis.

Marquering, H., Nolet, G., Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels. Geophys. J. Int. 132, 521–534.

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

Masters, G., Laske, G., 2000. The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in: Earth's deep interior: Mineral physics and tomography from the atomic to the global scale. American Geophysical Union, Geophysica edition. pp. 63–87.

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

Moczo, P., Bystricky, E., Kristek, J., Carcione, J.M., Bouchon, M., 1997. Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures. Bull. Seism. Soc. Am. 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, DOI 10.1016/S0065–2687(06)48008–0.

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

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., 2006b. A catalogue of deep mantle plumes: new results from finite-frequency tomography: Geochemistry Geophysics Geosystems. Geochemistry Geophysics Geosystems 7, Q11007.

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

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.

Nissen-Meyer, T., Fournier, A., Dahlen, F.A., 2007. A two-dimensional spectral-element method for computing spherical-earth seismograms - I. Moment-tensor source. Geophysical Journal International 168, doi: 10.1111/j.1365–246X.2006.03121.x.

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. Springer. 2nd edition.

Nolet, G., 2008. A breviary of seismic tomography: imaging the interior of the Earth and Sun. Cambridge University Press.

Olsen, K.B., Pechmann, J.C., Schuster, G.T., 1995. Simulation of 3D elastic wave propagation in the Salt Lake basin. Bull. Seism. Soc. Am. 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., Capdeville, Y., Romanowicz, B., 2009. Seismic waveform modelling in a 3-D Earth using the Born approximation: potential shortcomings and a remedy. Geophys. J. Int. 177, 161–178.

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

Park, J., Lindberg, C.R., Thomson, D.J., 1987. Multiple-taper spectral analysis of terrestrial free oscillations: part I. Geophysical Journal International 91, 755–794.

Patera, A., 1984. A spectral element method for fluid dynamics: Laminar flow in a channel expansion. Journal of Computational Physics 54, 468–488.

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., 2011. Forward and adjoint simulations of seismic wave propagation on unstructured hexahedral meshes. GJI , doi: 10.1111/j.1365–246X.2011.05044.x.

Plesch, A., Shaw, J.H., Hauksson, E., Tanimoto, T., 2008. SCEC Community Velocity Model (CVM-H 5.5), in: Southern California Earthquake Center Annual Meeting, Proceedings and Abstracts,, p. 142.

Plessix, R.E., 2009. Three-dimensional frequency-domain full-waveform inversion with an iterative solver. Geophysics 74, WCC149–157.

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

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 1: acoustic wave-equation method. Geophysical Prospecting , 287–310.

Pratt, R.G., Worthington, M.H., 1990b. Inverse Theory Applied To Multi-Source Cross-Hole Tomography. Part 2: elastic wave-equation method. Geophysical Prospecting 38, 311–329.

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, doi:10.1029/2008JB006271.

Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years. Annu. Rev. Earth Planet. Sci. 31, 303–328.

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

Saad, Y., 2011. Numerical methods for large eigenvaue problems. SIAM. 2nd edition.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space. Geophysical Journal International 138, 479–494.

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.

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

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., McQuarrie, N., Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography. Nature Geoscience 1, 458–462.

Sigloch, K., Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes. Geophys. J. Int. 167, doi: 10.1111/j.1365–246X.2006.03116.x.

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

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

Talagrand, O., Courtier, P., 1987. Variatonal assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. Q. J. R. Meteorol. Soc. 113, 1311–1328.

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

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

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

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. PAGEOPH 128, 365–399.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. volume 128. SIAM.

Tarantola, A., 2007. Course on inverse problems: square root variable metric algorithm for least-squares problems.

Tarantola, A., Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion. Reviews of Geophysics 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.

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

Tromp, J., Komatitsch, D., Hjörleifsdóttir, V., Liu, Q., Zhu, H.J., 2010. Near realtime simulations of global CMT earthquakes. Geophys. J. Int. 183, 381–389.

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

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

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

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

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

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. Geophys. J. Int. 175, 175103–115.

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.

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., Jordan, T.H., 1998. Sensitivity of frequency-dependent traveltimes to laterally heterogeneous, anisotropic Earth structure. Geophys. J. Int. 133, 683–704.

Zhao, L., Jordan, T.H., Chapman, C.H., 2000. Three-dimensional Frechet differential kernels for seismic delay times. Geophys. J. Int. 141, 558–576.

Zhao, L., Jordan, T.H., Chen, P., 2006. Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. Bull. Seism. Soc. Am. 96, 1753–1763.

Zhao, L., Jordan, T.H., Olsen, K.B., Chen, P., 2005. Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. Bull. Seism. Soc. Am. 95, 2066–2080.

Zhou, Y., 2009. Surface-wave sensitivity to 3-D anelasticity. Geophysical Journal International 178, 1403–1410.

Zhou, Y., Liu, Q., Tromp, J., 2011. Surface-wave sensitivity : mode summation versus adjoint SEM. Geophys J. Int , submitted.

Zhou, Y., Nolet, G., Dahlen, F.A., 2004. Three-dimensional sensitivity kernels for surface wave observables. Geophys. J. Int. 158, 142–168.

Zhu, H.J., Bozda, E., Peter, D., Tromp, J., 2010. Adjoint tomography of Europe, in: 2010 Fall Meeting, AGU, San Francisco, Calif., pp. S31A–2011.

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, L., Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions. Journal of Geophysical Research 105, 2969–2980.

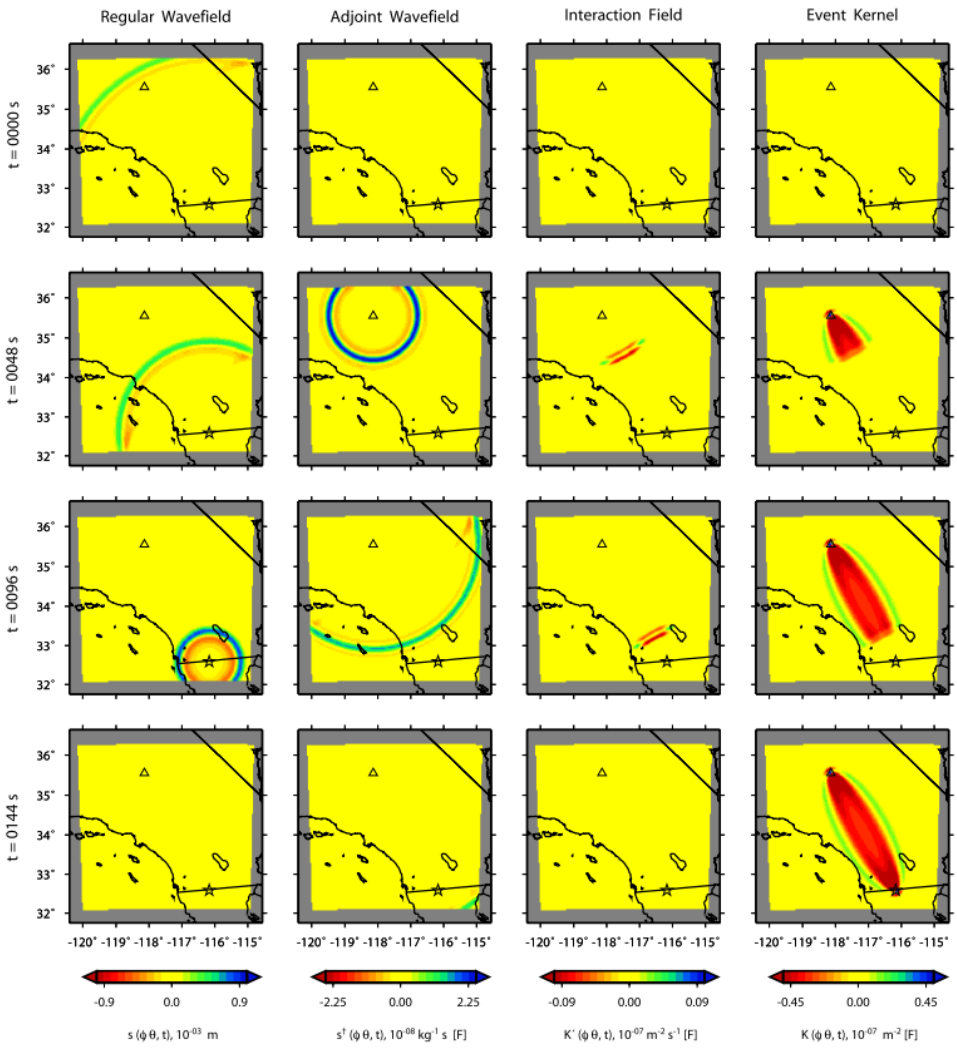


Figure 1: Computation of sensitivity kernels by the interaction of the forward wavefield (first column) and the adjoint wavefield (second column) at four different time slices in a simple 2D homogeneous model. The contribution to sensitivity kernels by one particular time slice is shown in the third column, and the total integration 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)).

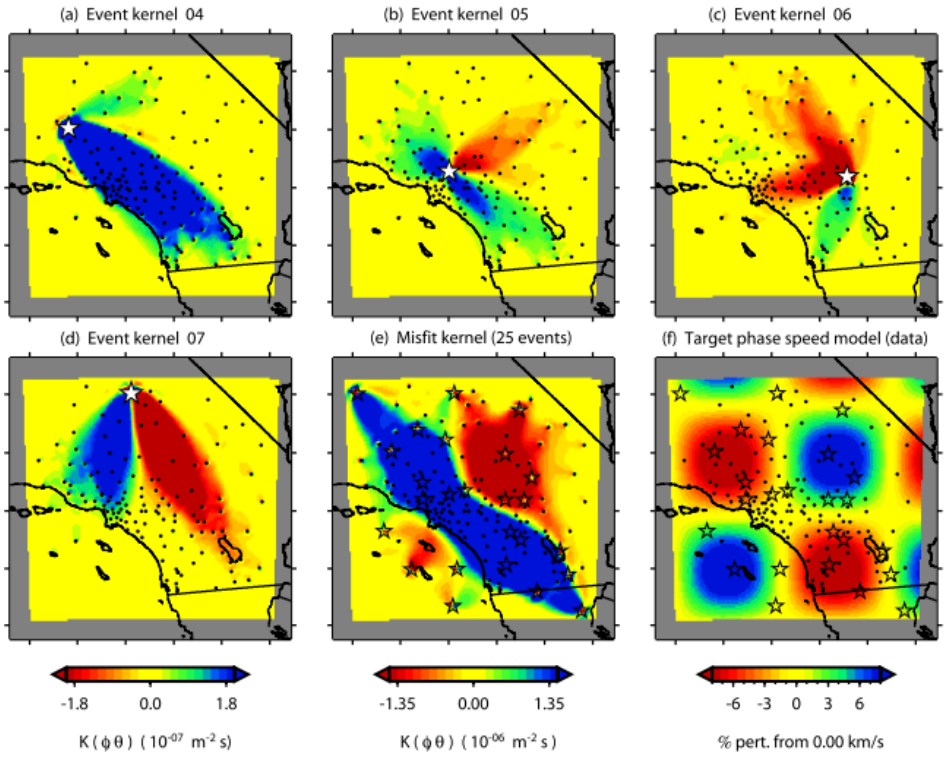


Figure 2: Event kernels generated for a 2D synthetic example. The 'true' model (shown in the lower-right graph) is a checker-board pattern in terms of phase velocity perturbations from a uniform background model. The first 4 graphs (a-d) show event kernels for four different events, illustrating their ability to 'see' the true model. The sum of all event kernels (i.e. $\mathbf{g}$) is shown in graph (e), indicating the direction of model update in the first iteration (Courtesy of Tape et al. (2007)).

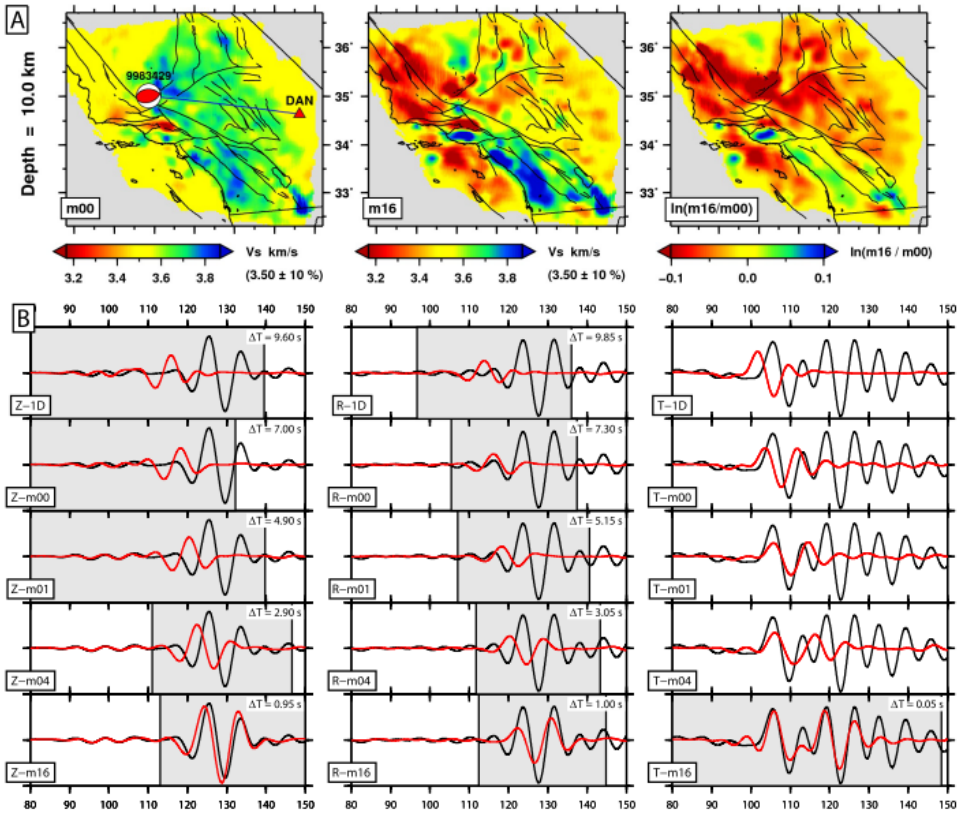


Figure 3: Adjoint tomography inversion of southern California crust. Shear velocity models are shown. (a) Comparisons of initial model, final model after 16 iterations and model update at the depth of 10 km. The final model exhibits velocity perturbations up to $\pm 30\%$ for this model and depicts geological features such as sedimentary basins, exhumed batholiths and lithological contrast across faults. Event 9983429 is denoted by a beachball 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 $6 - 30$ seconds. Windows automatically picked by FLEXWIN algorithm (Maggi et al., 2009) are indicated by grey 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).

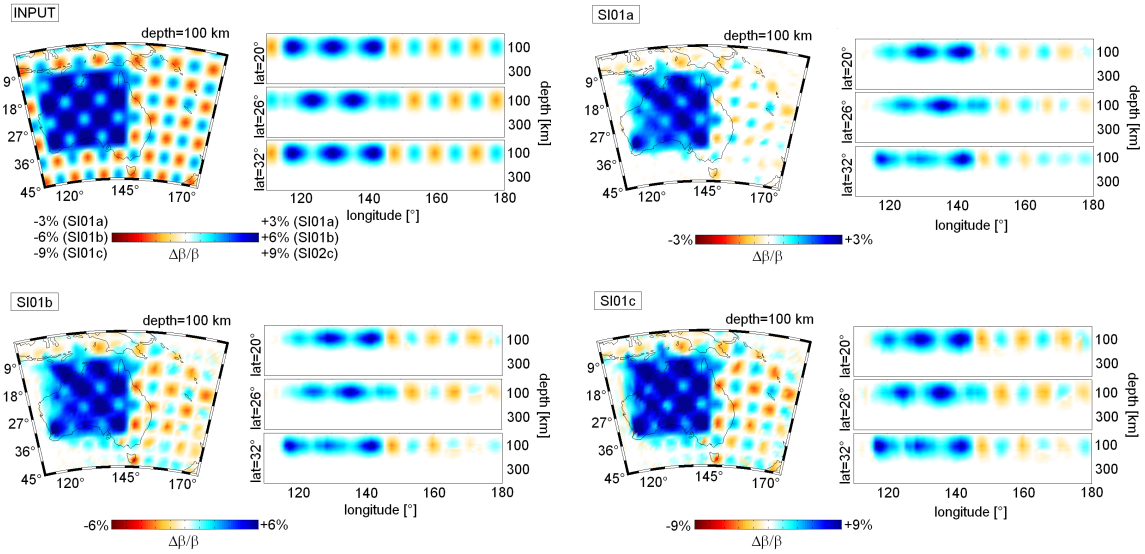
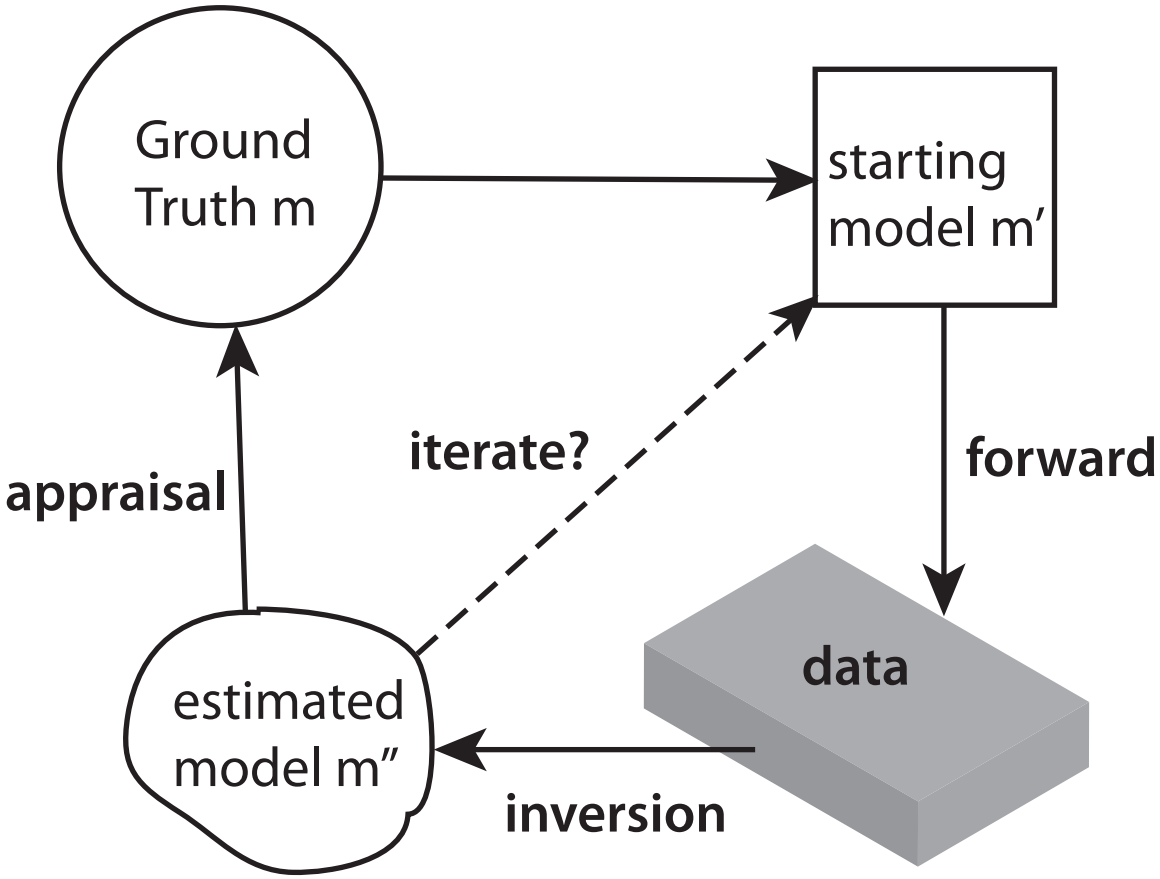
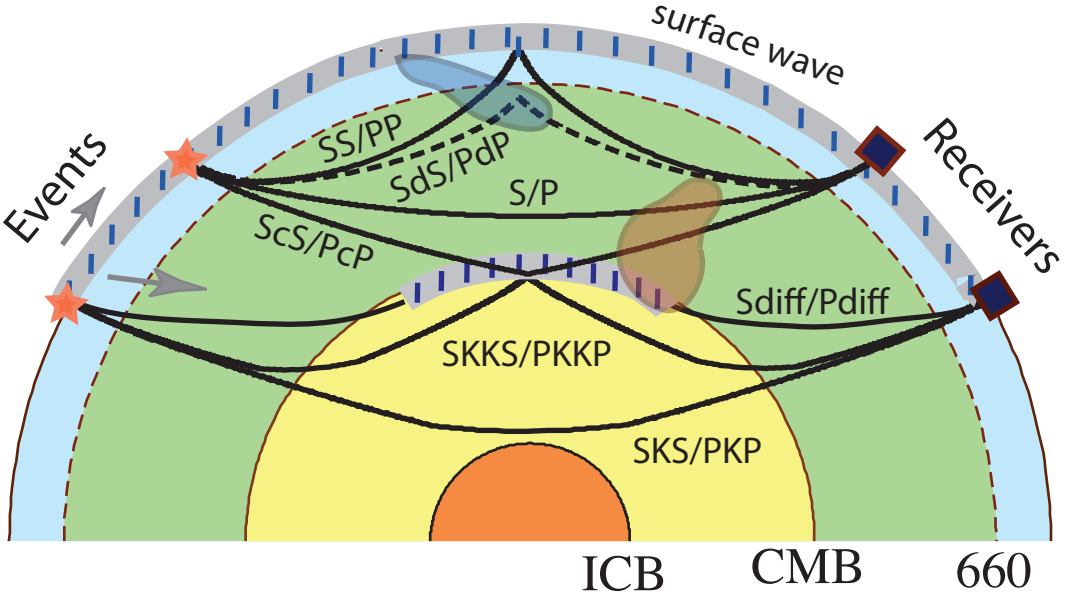
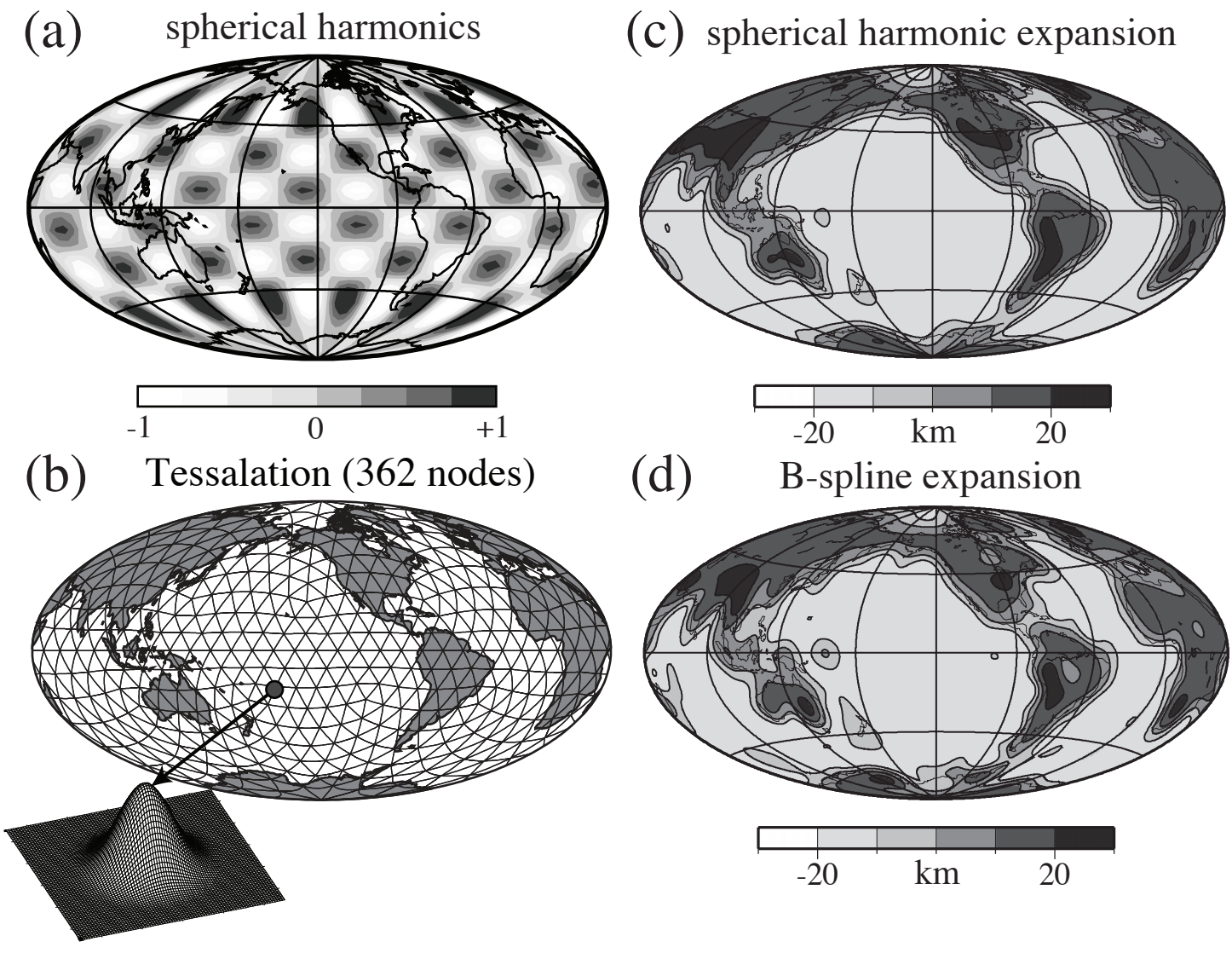


Figure 4: Resolution tests for adjoint tomography inversions of the Australasian region. Initial models are checker-board patterns ($3^\circ \times 3^\circ$) of various velocity perturbations ($3\%, 6\%, 9\%$) superposed onto a high-velocity patch ($\sim 25^\circ \times 25^\circ$). The 1D background velocity model suitable for regional structure is given in Fichtner et al. (2009b, Figure 7). Resolution tests go through the same adjoint tomography procedures as the actual 3D inversions and recover both the high velocity patch and checker-board patterns in regions with sufficient source-receiver coverage.

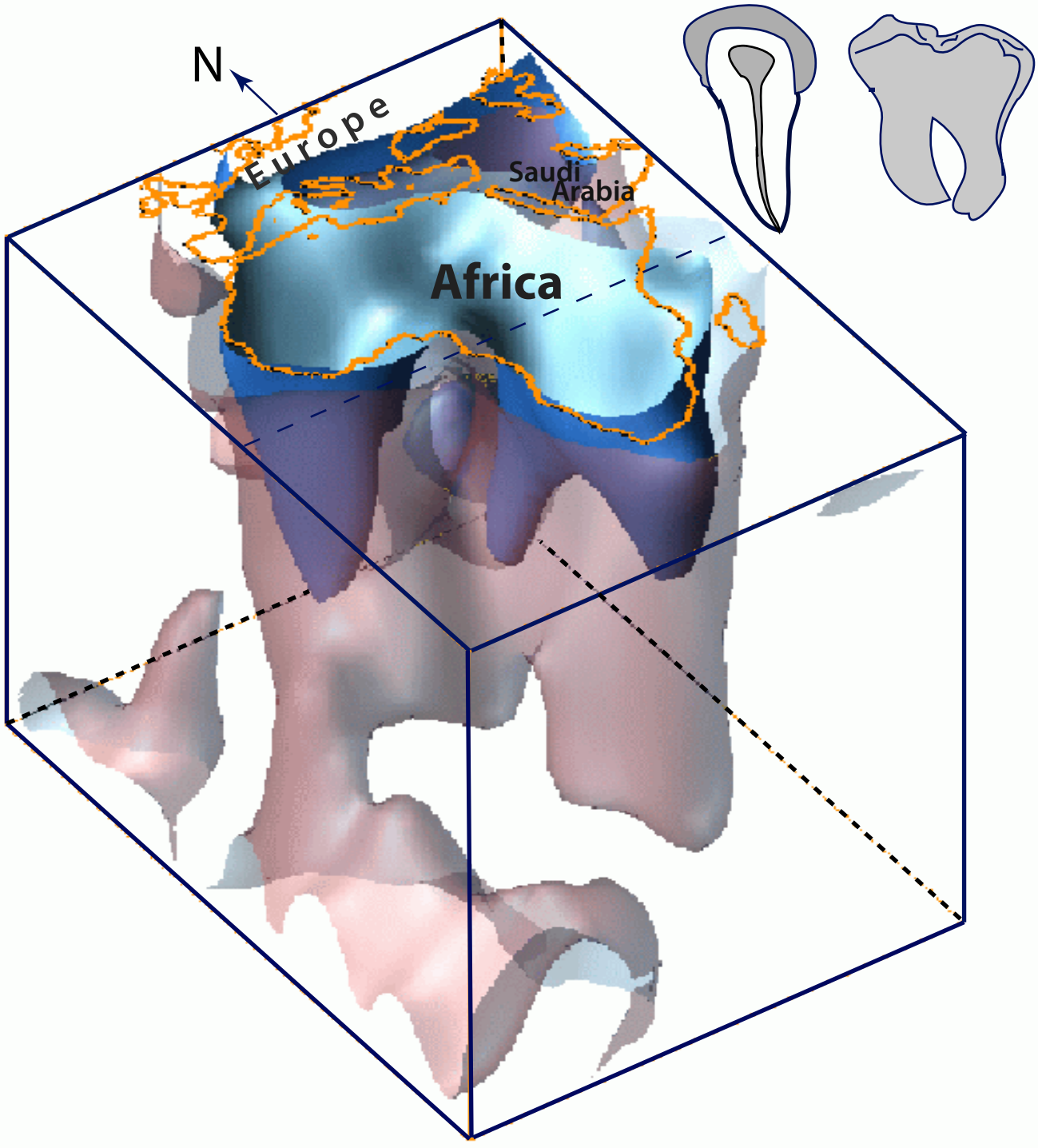

Ground
Truth m
starting
model m'
iterate?
forward
appraisal
data
estimated
model m"
inversion

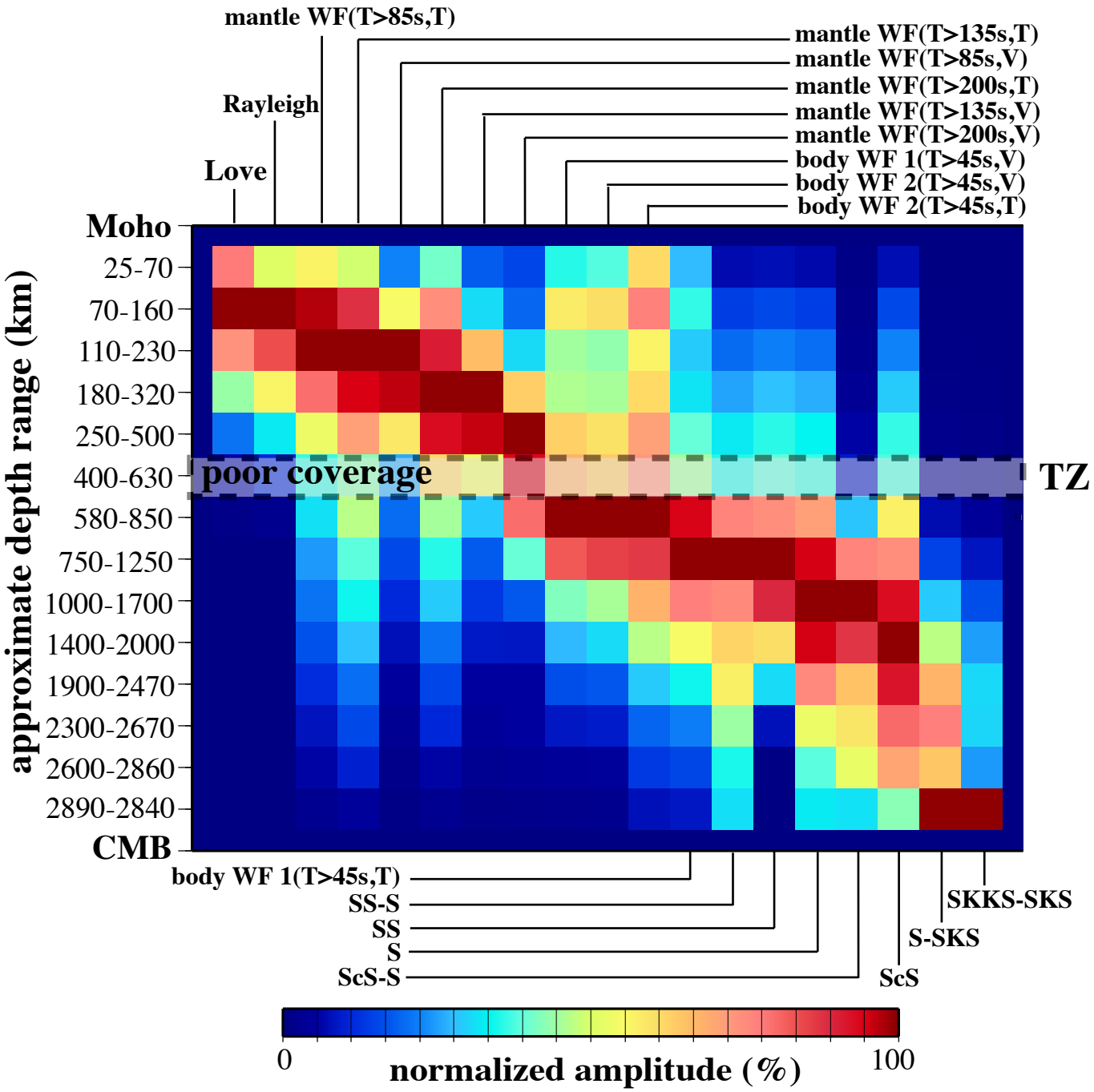
Global Data Constraints

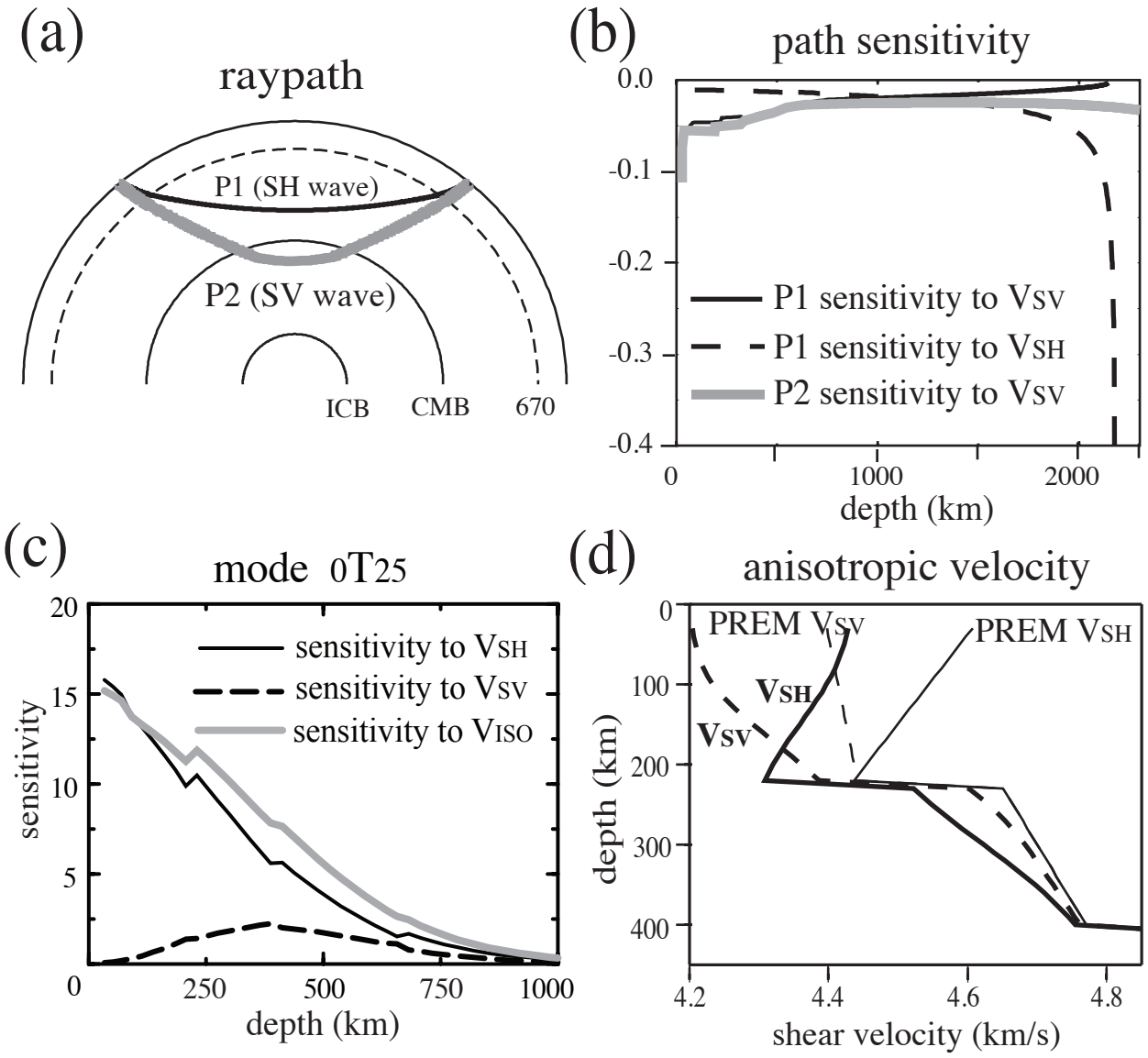


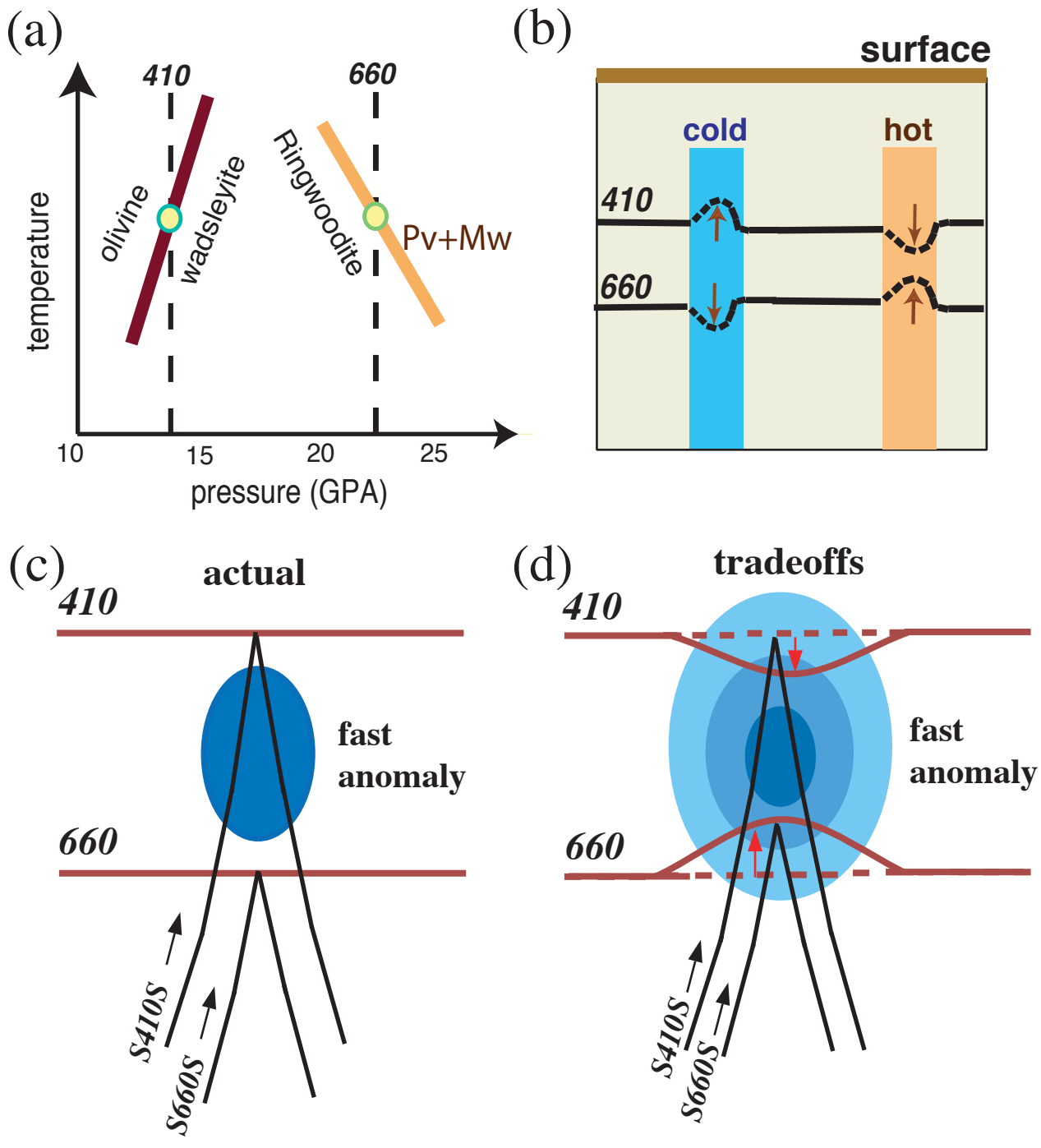

(a) spherical harmonics
-1
0
+1
(c) spherical harmonic expansion
-20
km
20
(b) Tessalation (362 nodes)
(d) B-spline expansion
-20
km
20

‘Teeth’ Beneath Our Feet

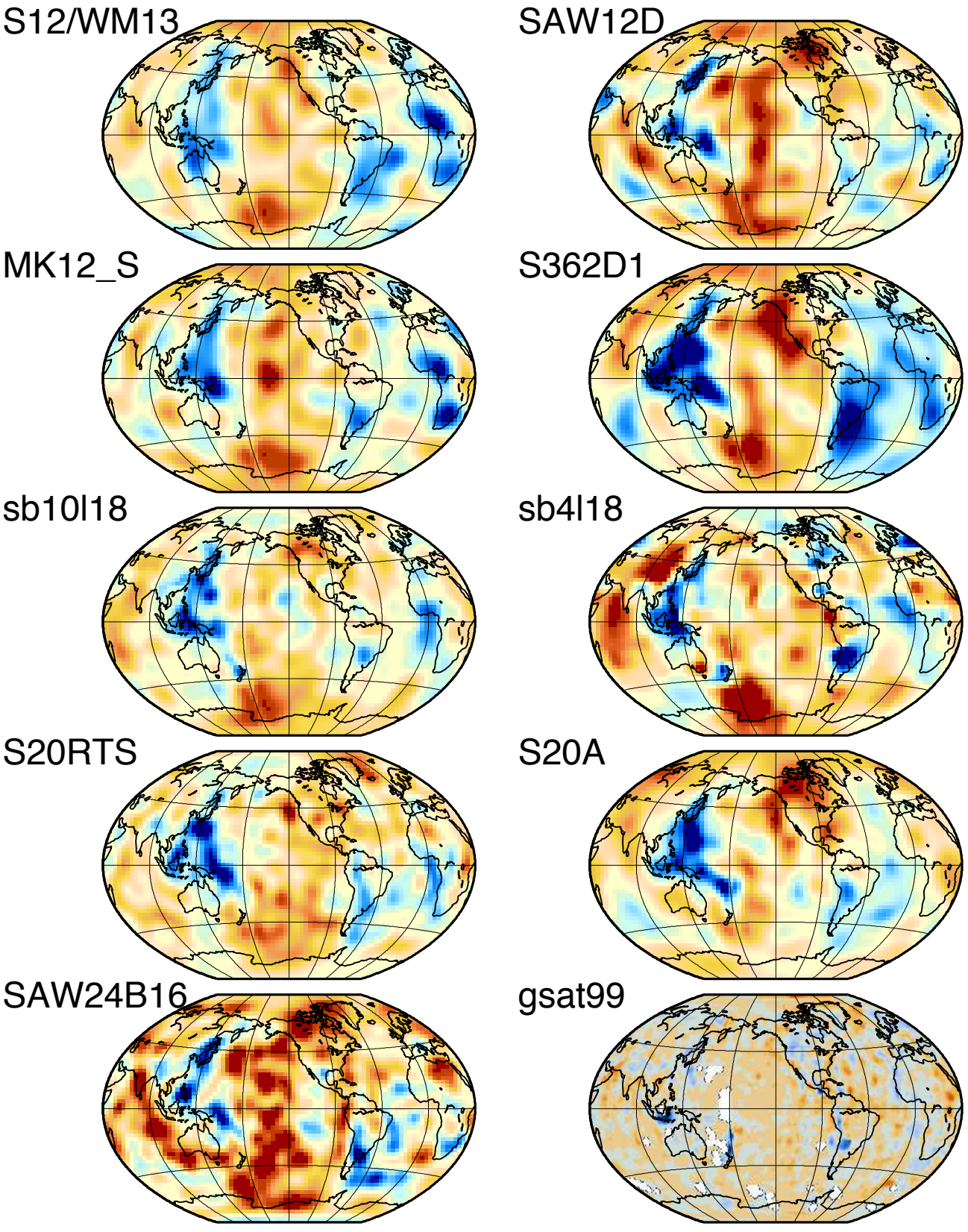


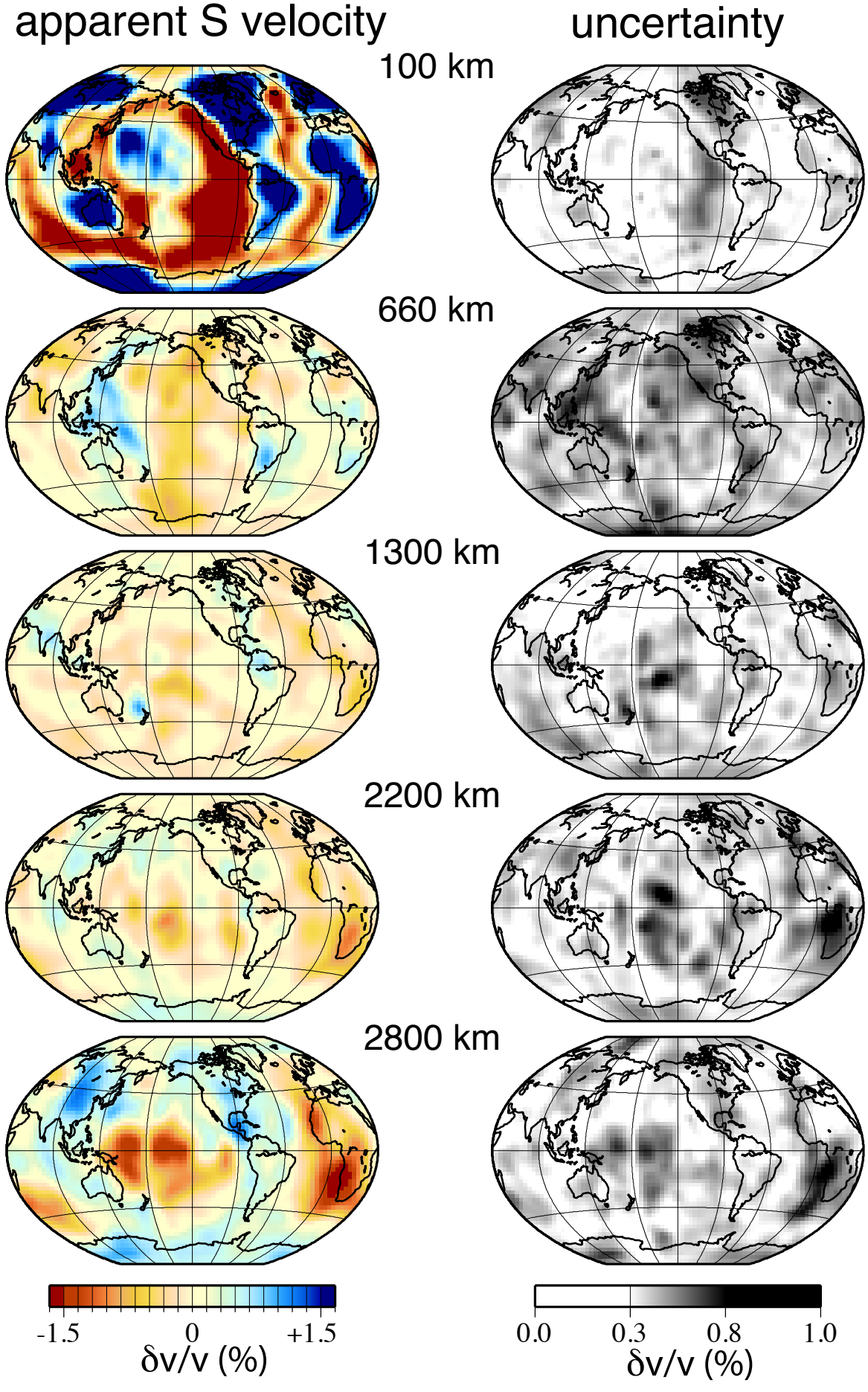

mantle WF(T>85s,T)
mantle WF(T>135s,T)
mantle WF(T>85s,V)
mantle WF(T>200s,T)
mantle WF(T>135s,V)
mantle WF(T>200s,V)
body WF 1(T>45s,V)
body WF 2(T>45s,V)
body WF 2(T>45s,T)
Rayleigh
Love
Moho
25-70
70-160
110-230
180-320
250-500
400-630
580-850
750-1250
1000-1700
1400-2000
1900-2470
2300-2670
2600-2860
2890-2840
CMB
poor coverage
TZ
approximate depth range (km)
body WF 1(T>45s,T)
SS-S
SS
S
ScS-S
ScS
S-SKS
SKKS-SKS
0
normalized amplitude (%)
100

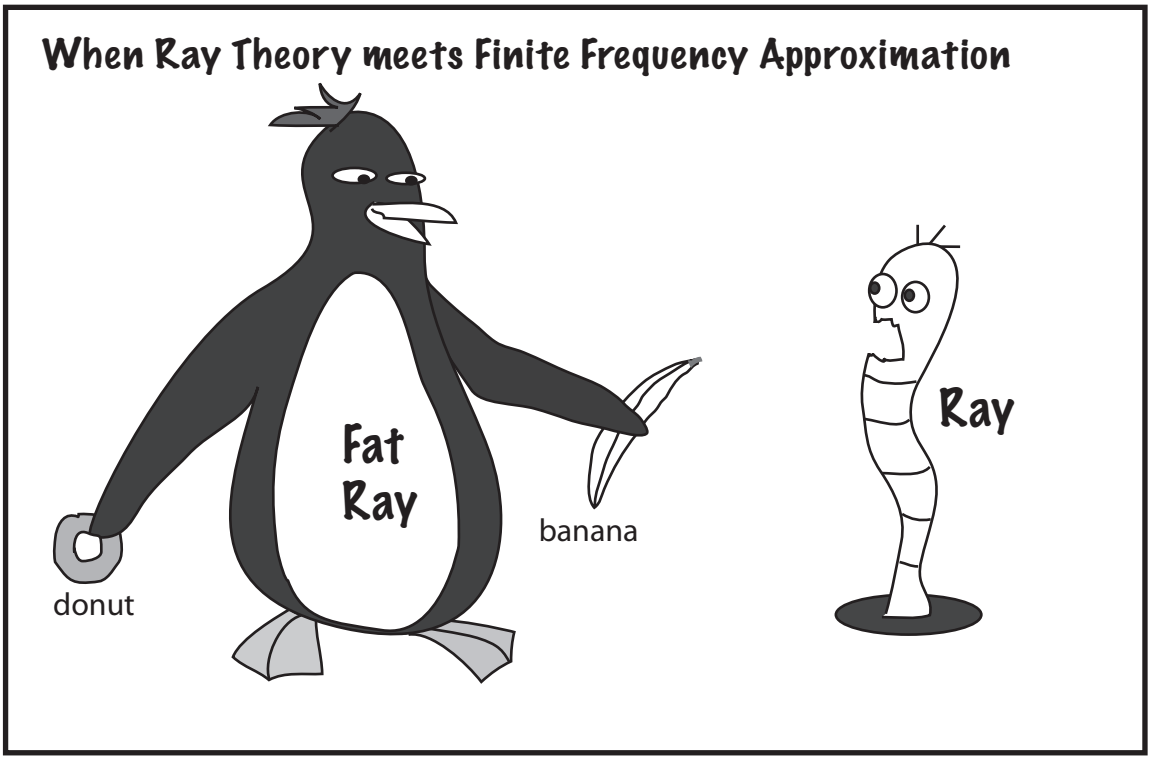

(a)
raypath
P1 (SH wave)
P2 (SV wave)
ICB
CMB
670
(b)
path sensitivity
0.0
-0.1
-0.2
-0.3
-0.4
0
1000
2000
depth (km)
P1 sensitivity to VSV
P1 sensitivity to VSH
P2 sensitivity to VSV
(c)
mode 0T25
20
15
10
5
0
250
500
750
1000
depth (km)
sensitivity
sensitivity to VSH
sensitivity to VSV
sensitivity to VISO
(d)
anisotropic velocity
0
100
200
300
400
depth (km)
PREM VSV
PREM VSH
VSH
VSV
4.2
4.4
4.6
4.8
shear velocity (km/s)


(a)
410
660
olivine
wadsleyite
Ringwoodite
Pv+Mw
temperature
10
15
20
25
pressure (GPA)
(b)
surface
cold
hot
410
660
(c)
actual
410
fast
anomaly
660
S410S
S660S
(d)
tradeoffs
410
fast
anomaly
660
S410S
S660S

Shear Velocity Perturbations (540 km)




apparent S velocity
uncertainty
100 km
660 km
1300 km
2200 km
2800 km
-1.5
0
+1.5
δv/v (%)
0.0
0.3
0.8
1.0
δv/v (%)


When Ray Theory meets Finite Frequency Approximation
Fat Ray
donut
banana
Ray