

A data-adaptive, multiscale approach of finite-frequency, traveltimes tomography with special reference to P and S wave data from central Tibet

Shu-Huei Hung,¹ Wang-Ping Chen,^{2,3} and Ling-Yun Chiao⁴

Received 31 December 2010; revised 17 March 2011; accepted 24 March 2011; published 21 June 2011.

[1] We discuss an innovation in traveltimes tomography that combines wavelet-based, multiscale parameterization and finite-frequency theory to solve two outstanding issues that inevitably arise from uneven source station distributions and from the three-dimensional (3-D) nature of wavefront healing: how to objectively address the intrinsically multiscale nature of data coverage while simultaneously maintain model resolution at each scale level. We apply the new, integrated methodology to investigate 3-D variations of P and S wave speeds ($\delta \ln V_P$ and $\delta \ln V_S$) beneath the Himalayan-Tibetan orogen. In particular, we are able to constrain variations in the Poisson's ratio via $\delta \ln(V_P/V_S)$. The formulation is naturally data adaptive, resolving features at each scale only if the required data converge is available. The very first, long-wavelength feature that emerges is a clear anomaly of high $\delta \ln V$ that extends over more than 500 km beyond the northern edge of the Lhasa terrane at places. Farther northward, a strong negative anomaly underlies the region where recent volcanism occurs in northern Tibet. Regions of negative $\delta \ln(V_P/V_S)$ delineate a slab-like, subhorizontal feature concentrated between depths of ~ 100 – 250 km. Such characteristics are consistent with the notion that chemically refractory, and therefore buoyant, mantle lithosphere of the Indian shield ("Greater India") has advanced subhorizontally northward far beyond the surficial Bangong-Nujiang suture. In the crust, two isolated regions of low $\delta \ln V$, each extending to depths near 100 km, occur along the Lunggar and the Yadong-Gulu active rifts in southern Tibet. Deep penetrating rifts imply that only a limited amount of horizontal displacement is being accommodated on subvertical structures.

Citation: Hung, S.-H., W.-P. Chen, and L.-Y. Chiao (2011), A data-adaptive, multiscale approach of finite-frequency, traveltimes tomography with special reference to P and S wave data from central Tibet, *J. Geophys. Res.*, 116, B06307, doi:10.1029/2010JB008190.

1. Introduction

[2] Since its inception in the mid-1970s [e.g., *Aki and Lee*, 1976], seismic traveltimes tomography has been a key method in probing the earth's interior at various scales, ranging from the whole mantle to local crustal structures [e.g., *Zhao and Kanamori*, 1993; *Masters et al.*, 1996; *Grand et al.*, 1997; *Zhang and Thurber*, 2003; *Montelli et al.*, 2004a; *Li et al.*, 2008]. Its ability to unravel three-dimensional (3-D) variations makes traveltimes tomography indispensable in providing vital insights to global tectonics and geodynamics. Over the past few decades, there has been significant progress in several fronts to improve the resolution of tomographic images.

[3] First, both the quality and the quantity of data have greatly improved. Globally, the coverage of and quality of data from the Global Seismographic Network (GSN) has been increasing steadily. Meanwhile, many deployments of broadband seismographs provided much needed coverage in key geological areas. Some recent experiments, such as during the Project Hi-CLIMB over the Himalayan-Tibetan orogen, made hundreds of deployments in spite of severe logistic constraints [*Chen et al.*, 2010; *Hung et al.*, 2010]. Such efforts are culminating as the USArray component of the EarthScope, a national initiative, is well underway across the North American continent, leading to unprecedented lateral coverage and high-resolution at depth [e.g., *Sigloch et al.*, 2008; *Tian et al.*, 2009; *Obrebski et al.*, 2010].

[4] Second, in the last decade, more sophisticated theory emerged for forward calculations, taking into account the frequency-dependent nature intrinsic to traveltimes [e.g., *Dahlen et al.*, 2000; *Hung et al.*, 2000; *Zhao et al.*, 2000]. Combining ray theory for body waves, Born approximation for forward scattering, and paraxial approximation for wavefronts away from the central ray, *Dahlen et al.* [2000] and *Hung et al.*

¹Department of Geosciences, National Taiwan University, Taipei, Taiwan.

²Department of Geology, University of Illinois-Urbana, Urbana, Illinois, USA.

³Department of Ocean Science and Engineering, Zhejiang University, Hangzhou, China.

⁴Institute of Oceanography, National Taiwan University, Taipei, Taiwan.

[2000] devised an efficient method for computing the Fréchet derivative (or data kernel) which relates traveltimes, precisely measured at different frequencies by cross correlation of waveforms, to 3-D perturbations in seismic wave speeds for the whole Earth.

[5] For each source-receiver pair, the resulting kernel occupies a banana-shaped region surrounding the geometric ray and extending to a width determined by the first Fresnel zone, which bounds the volume of maximum sensitivity associated with this particular measurement [e.g., *Hung et al.*, 2001]. At any finite frequency, the center of the banana-shaped region is hollow because the geometric ray itself represents the trajectory of maximum sensitivity when frequency approaches infinity. This formulation encapsulates the essence of diffractive wavefront healing whose effect is more pronounced at low frequencies. To this end, frequency-dependent traveltimes take full advantage of the broadband nature of signals from earthquake sources: With traveltimes measured at multiple frequency bands, each measurement, depending on its frequency band, is sensitive to heterogeneities at a different spatial range from the central ray; and measurements at different frequency bands complement each other [e.g., *Hung et al.*, 2004; *Montelli et al.*, 2004b; *Yang et al.*, 2006].

[6] Third, tomography is inherently an inverse problem. Appropriate parameterization of model and proper regularization in implementing the inversion are particularly critical in dealing with unevenly distributed data which is always the case for actual observations [e.g., *Trampert and Snieder*, 1996; *Chiao and Kuo*, 2001; *Sambridge and Faletić*, 2003; *Loris et al.*, 2007]. Localized parameterization, using either regular or variable sizes, has been a common practice in regional traveltime tomography. Unlike global tomography, which usually involves parameterization in terms of expansions in spherical harmonics or in basis functions of spherical splines (thus emphasizing spectral resolution), localized schemes preferentially emphasizes spatial resolution [e.g., *Chiao and Kuo*, 2001; *Romanowicz*, 2003]. Furthermore, spatially stationary regularization, with a priori assumptions regarding smoothness, is often imposed during the inversion in order to reduce nonuniqueness of the solution and to suppress small-scale, uncorrelated fluctuations in the solution which arise from ill-constrained model parameters, errors in the data, finite truncation intrinsic to discrete parameterization, null vectors, and approximations in the formulation of the forward problem.

[7] A more desirable parameterization should be data adaptive and capable of resolving spatially varying scales of heterogeneities. That is, a scheme favorably preserves the

long-wavelength amplitude spectra in sparse-sampled regions without penalizing high spatial resolution in places with dense coverage of data [e.g., *Trampert*, 1998]. One promising approach is a multiscale parameterization based on wavelet transforms [*Chiao and Kuo*, 2001; *Chiao and Liang*, 2003; *Chiao et al.*, 2006]. *Hung et al.* [2010] are the first to extend this technique to 3-D, and then combined the extended multiscale parameterization with finite-frequency tomography, leading to the most recent, 3-D distribution of both $\delta \ln V_P$ and $\delta \ln V_S$, or fractional changes in the speed of P and S waves, respectively, under the Himalayan-Tibetan orogen. Limited by space, *Hung et al.* [2010; hereinafter paper 1] emphasized results for V_P and V_S and their geologic interpretations, describing the methodology only briefly.

[8] In this study, we enunciate, in some detail, the procedure of extending multiscale parameterization to 3-D and how to combine this procedure with finite-frequency tomography. Since the Hi-CLIMB data set used in paper 1 is comprised of both a dense-spaced, north-northwest trending linear array and a broad, lateral (east-west) regional network (Figure 1a; data particularly well suited for multiscale analysis) we perform further analysis of this data set to obtain 3-D images of fractional changes in the ratio between V_P and V_S , $\delta \ln(V_P/V_S)$, a quantity directly related to variations in Poisson's ratio. We then compare the new results, vis-à-vis, with results of both $\delta \ln V_P$ and $\delta \ln V_S$ to place additional constraints on properties of converging lithosphere beneath the Himalayan-Tibetan orogen where both the greatly thickened crust and the mantle have been investigated by several independent, complementary methods.

[9] It is important to point out that although theoretically the relationship among $\delta \ln V_P$, $\delta \ln V_S$, and $\delta \ln(V_P/V_S)$ is simple [e.g., *Chou et al.*, 2009], reliable estimates of $\delta \ln(V_P/V_S)$ requires a careful selection of data to ensure that only source-receiver pairs common to both P and S wave are being used. Subsequently, the entire process of inversion, along with additional resolution tests, is carried out anew. While the images for $\delta \ln V_P$ and $\delta \ln V_S$ are similar to those shown in paper 1 (an expected outcome) most results presented in this study are entirely new.

2. Preliminary Analysis of Data

[10] The Hi-CLIMB seismic array covers a vast region over the Himalayas and the Tibetan plateau. The array is roughly L shaped, with apertures of about 800 km and 500 km along north-northwest—south-southeast (NNW-SSE) and east-west (E-W) trends, respectively (see Figure 1a).

Figure 1. Examples of frequency-dependent, traveltime residuals along the N-S trending, linear array of Hi-CLIMB. (a) The map on the left shows locations of seismic stations (triangles) and the region being investigated (box outlined by dashed lines). Gray solid lines indicate locations of profiles along which images of $\delta \ln V_P$, $\delta \ln V_S$ and $\delta \ln(V_P/V_S)$ are shown in Figures 5–9. Other features are major geologic boundaries (solid curves), including (from north to south) KF, the Kunlun fault; JRS, the Jinsha River Suture; BNS, the Bangong-Nujiang Suture; IYS, the Indus-Yarlung Suture; STD, the South Tibet Detachment System; MCT, the Main Central Thrust; and MBT, the Main Boundary Thrust. The map on the left shows locations of epicenters of earthquakes E1 and E2 that generated the P and S wave data, respectively, shown below. (b) Results from P waves. Notice that at a given station, the amplitude of residual from the high-frequency band (“high”) is generally larger than that from the low-frequency band (“low”), reflecting a greater degree of wavefront healing associated with diffraction in the latter case. In both cases, long- and short-wavelength fluctuations are both present along the profile, calling for a multiscale analysis. (c) Corresponding results for S waves. The overall patterns of traveltime residuals are similar to those in Figure 1b. Five-letter codes at the bottom are names of stations for internal bookkeeping.

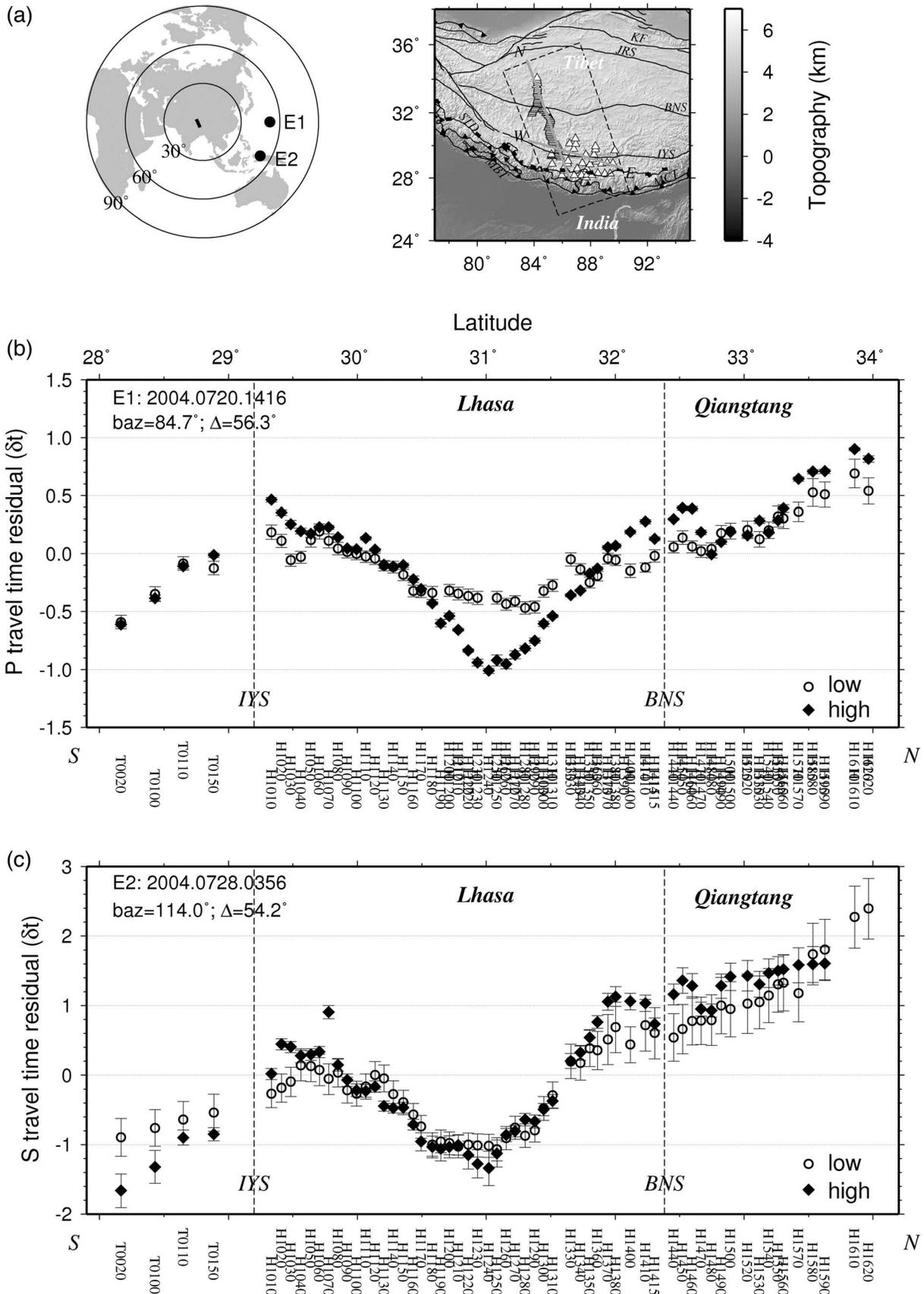


Figure 1

[11] A number of studies have shown that body wave arrivals are dispersive, a result of diffractive wavefront healing [e.g., *Hung et al.*, 2004; *Yang et al.*, 2006]. For the Hi-CLIMB data set, we apply a fourth-order, zero-phase Butterworth band-passed filter at two passbands, then measured traveltimes using the method of multichannel cross correlation (MCCC) [*VanDecar and Crosson*, 1990]. For P phases, measured from the vertical component of seismograms, the high and low range of frequencies are 0.3–2.0 Hz and 0.030–0.125 Hz, respectively. Corresponding values for S phases, measured from the tangential component of seismograms are 0.2–0.1 Hz and 0.03–0.10 Hz, respectively. For additional details of data processing and the resulting coverage of raypaths, see paper 1.

[12] Figures 1b and 1c show a representative example of how measured traveltimes for P and S arrivals, respectively, vary across the long, linear array. The first notable feature is that in general, the magnitude of traveltime anomalies at the low-frequency band is subdued when compared with their counterpart at higher frequencies. This trend is precisely the expected effect from wavefront healing: Heterogeneities, distributed in 3-D, cause the wavefront to become irregular as the wavefield passes through the medium; subsequent healing of the wavefront is more pronounced at low frequencies than high frequencies. In the former case, along a given geometric raypath, the propagating wave is sensitive to properties of the medium away from the path; whereas in the latter case, the wave is responsive to small-scale heterogeneities close to the path, leading to a larger traveltime anomaly.

[13] That is, under the framework of finite-frequency tomography, sensitivity kernels of traveltimes broaden with period, effectively not sensing small-scale heterogeneities in the hole of the “banana-doughnut” [*Hung et al.*, 2001; *Baig et al.*, 2003; *Yang and Hung*, 2005]. In contrast, classic ray theory based traveltime tomography assumes that the signal’s frequency approaches infinity and traveltime residuals of the same phase measured at different frequencies have exactly the same sensitivity confined to the raypath itself [e.g., *Montelli et al.*, 2004b; *Hung et al.*, 2004].

[14] The second important feature of the data is that at both passbands, both long- and short-wavelength variations are apparent (Figure 1), calling for an approach that can competently resolve heterogeneities at different length scales. To this end, because of varying subsurface structures and different configurations of receiver-source pairs, each data set would exhibit its own distinct characteristics. As such, a truly versatile multiscale approach also must be data adaptive.

[15] For our data set, measured traveltime residuals are strongly dependent on the back azimuths of earthquake sources. This feature is apparent when we group the selected earthquake sources into a succession of swaths, each spanning 20° in back azimuth. For each group, the average traveltime residual at each station is calculated by the mean of all measured traveltime residuals, each weighted by the inverse of estimated standard deviation in measurement error [*VanDecar and Crosson*, 1990].

[16] Figure 2 displays the average residuals of P and S arrivals measured at the high-frequency band from two event groups at opposite back azimuths. With some shift in baseline, as expected in part from differences in wave speeds, overall patterns of residuals for P and S arrivals are similar. Lateral, or roughly east-west trending, variations along the overall trend of the orogen is apparent from a single azimuth

of sources. For instance, for sources from the southwest (Figure 2, right), traveltime residuals are strongly positive (implying lower wave speed than that of the reference model) on the southeastern portion of the area of coverage, then become negative toward the west. Along the opposite azimuth, most anomalies along the NNW-SSE trending linear array shifted to more positive values; and many strong positive anomalies of the E-W trending regional array in southern Tibet become either smaller or even negative. Taken together, strong lateral variations across both NNW-SSE and E-W trends (Figures 1 and 2), and azimuthal dependence of residuals (Figure 2) point to the 3-D nature of subsurface structures. In short, characteristics of the traveltime data in hand require an inversion scheme that combines finite-frequency tomography with a data-adaptive, multiscale expansion in 3-D.

3. Theory and Methodology

[17] To fully illustrate the advantage of a multiscale approach, below we briefly review relevant aspects of the forward problem and regularization of the inverse problem, then discuss a wavelet-based, multiscale inversion.

3.1. The Forward Problem: Finite-Frequency Formulation Versus Ray Theory

[18] The starting point of linearized formulation for traveltime tomography is

$$\delta t = \int \int \int_{\oplus} K(\mathbf{x}) \delta s(\mathbf{x}) d^3 \mathbf{x}, \quad (1)$$

where δt is an observed traveltime residual relative to that predicted by a radially symmetric, reference Earth model, and $K(\mathbf{x})$ the Fréchet derivative or sensitivity kernel that relates δt to perturbations in P or S wave slowness (s , the reciprocal of wave speed) at every point \mathbf{x} throughout the region, \oplus , to be imaged. To date, there are two main theories for constructing the sensitivity kernel: conventional ray theory and finite-frequency theory that was first developed in the early 2000s and is continually being refined [e.g., *Tromp et al.*, 2005]. The differences between these two theories and their applications to seismic tomography are subjects of extensive investigation in recent literature [e.g., *Montelli et al.*, 2004b; *Hung et al.*, 2004; *Zhou et al.*, 2005].

[19] For seismic phases relevant to our study, sensitivity kernels from the finite-frequency theory all have the so-called “banana-doughnut” characteristics, with minimal sensitivity surrounding the vicinity of geometrical raypaths [*Hung et al.*, 2000]. When the lateral length scale of heterogeneity in wave speeds is considerably greater than the diameter of the 3-D kernel, or equivalently the frequency of the wave approaches infinity, equation (1) simply reduces to linearized ray theory. In this case, the Fréchet derivative essentially approaches zero everywhere except along the geometrical ray; and equation (1) reduces to a line integral of δs at each incremental arc length, evaluated along the geometrical raypath which is determined solely by the 1-D reference model. Based on synthetic traveltime residuals, a number of studies have shown that the two different approaches produce comparable results in general terms, but images from finite-frequency formulation are superior in recovering both the amplitude and the geometrical configuration of anomalies [*Montelli et al.*, 2004b; *Hung et al.*, 2004].

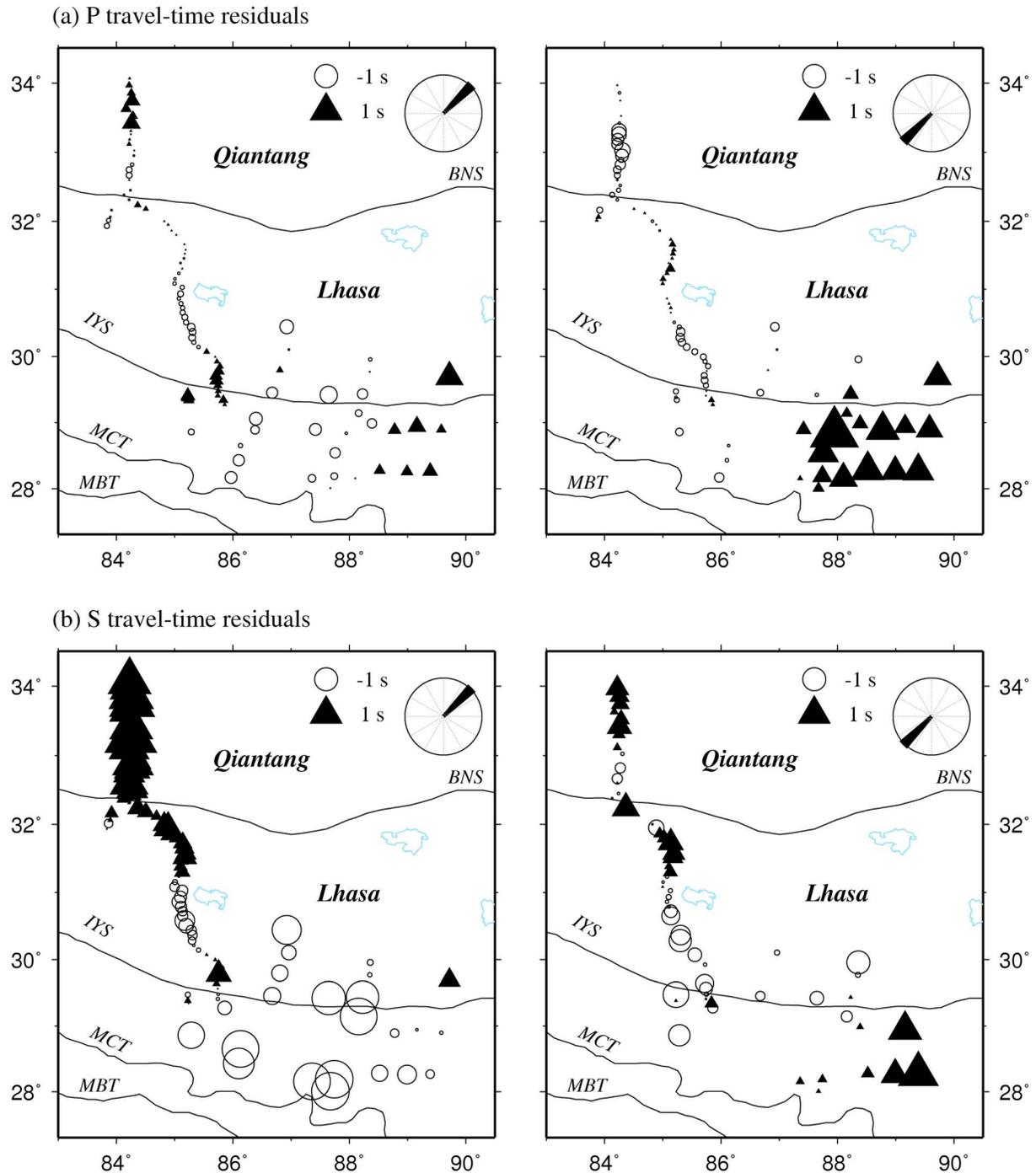


Figure 2. Map views of how traveltime residuals of (a) P and (b) S waves vary with back azimuths. For each map, the data are from the high-frequency band and averaged over a swath of 20° in back azimuths (darkened sectors in inset on the top right). For each image, we show how the data vary from opposing back azimuths. Such variations, when combined with profiles shown in Figure 1, indicate that subsurface anomalies are 3-D in nature.

3.2. Grid Parameterization and Regularization of the Inverse Problem

[20] The region of interest under the L -shaped Hi-CLIMB arrays is a spherical cap that laterally covers an area of 9.4° by 4.7° (approximately 1,000 km by 500 km) with a vertical

extent between the surface and a depth of 900 km (Figure 1). Using a parameterization for regional tomography similar to that of *Hung et al.* [2004], perturbations to the slowness of P and S waves within the volume of interest are determined at a total of $65(2^6 + 1)$ by $33(2^5 + 1)$ by $65(2^6 + 1)$ ($= 139,425$)

nodes, resulting in approximately equal spacing of about 15 km between grids in all three dimensions.

[21] For finite-frequency tomography, numerical integration of the sensitivity kernel within a single voxel, $K(\mathbf{x})d\mathbf{x}^3$ (see equation (1)), is evaluated by weighted sum of 27 (or 3^3) sampling points using the Gaussian quadrature [Zienkiewicz and Taylor, 1989]. (In comparison, when only ray theory is invoked, explicit line integral along the geometrical raypath is directly determined from the arc length of raypath within the voxel.) Summing up contribution of numerically computed kernels from all voxels surrounding each individual node leads to a concise data equation that relates data and model parameters of the form

$$\delta t_i = G_{ij}m_j = \langle \mathbf{g}_i, \mathbf{m} \rangle, \quad (2)$$

or in a matrix form

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

where δt_i is the i th measurement out of a total of N data points that form the data vector \mathbf{d} . It is expressed in equation (2) as the discrete inner product between the coefficient vector of the data kernel, \mathbf{g}_i or the transpose of each row of \mathbf{G} , and the model vector \mathbf{m} , comprised of M (139,425) unknown values of slowness perturbations at nodes of the study volume, $[m_1, m_2, \dots, m_M]^T$. In matrix form, the matrix \mathbf{G} comprises of N row vectors of the data kernel \mathbf{g}_i^T , that is, $\mathbf{G}^T = [\mathbf{g}_1 \ \mathbf{g}_2 \ \dots \ \mathbf{g}_N]$; and G_{ij} , or the j th component of the row vector \mathbf{g}_i^T , represents the sensitivity of the i th data point to slowness perturbation at the j th node. The unknown model parameters are obtained by solving all N data equations (equation (2)) simultaneously.

[22] Since nonuniform distributions of stations and sources inevitably leave some model parameters unresolved due to very sparse or no constraints from data, regularizations, such as minimum-norm and/or maximum smoothness, are often imposed on underdetermined inverse problems in order to obtain solutions in which large amplitudes are damped out so fluctuations are smooth enough to reveal long-wavelength features. One standard choice is the weighted, damped least squares (DLS) solution [e.g., Menke, 1984]:

$$\hat{\mathbf{m}} = (\mathbf{G}^T \mathbf{C}^{-1} \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T \mathbf{C}^{-1} \mathbf{d}, \quad (3)$$

where $\hat{\mathbf{m}}$ represents a particular estimate of the model vector under a fixed-scale parameterization of nodes, λ is a non-negative damping factor assigned a priori to regulate the degree of minimization in model norms, \mathbf{I} an $M \times M$ identity matrix, and \mathbf{C} the $N \times N$ data covariance matrix defined as $\mathbf{C} = \sigma_i^2 \mathbf{I}$, where σ_i is the estimated error for the i th data, assuming that each measure is independent.

[23] A competing consideration is the call for over-parameterization in order to alleviate spectral leakage or aliasing arising from errors in discretization and truncated expansion of continuous model parameters [Trampert and Snieder, 1996]. Heavy-handed norm damping is then necessary to diminish noisy fluctuations due to a significant increase in model parameters, many of which are poorly constrained by data. Consequently, such a practice degenerates the local coherence of resulting images and suppresses, en masse, amplitude of heterogeneities. Additionally, in order to preserve the spectral resolution of long-wavelength fea-

tures which are usually adequately constrained by data, imposing smoothness constraints on the model becomes indispensable, but at the expense of penalizing the spatial resolution in regions of dense data coverage.

[24] An alternative regularization scheme is to enforce convolution quelling, which broadens the cross-path width of rays to a prescribed value in an ad hoc manner, and penalizes model roughness by imposing an a priori correlation length to the model [e.g., Hung et al., 2004; Chiao et al., 2006]:

$$\hat{\mathbf{m}} = \mathbf{Q}(\mathbf{Q}^T \mathbf{G}^T \mathbf{C}^{-1} \mathbf{G} \mathbf{Q} + \lambda \mathbf{I})^{-1} \mathbf{Q}^T \mathbf{G}^T \mathbf{C}^{-1} \mathbf{d}, \quad (4)$$

where \mathbf{Q} is the convolution operator that prescribes a stationary correlation length to the model [Meyerholtz et al., 1989]. Notice that if \mathbf{Q} is an identity matrix, then equation (4) reduces to the weighted DLS solution (equation (3)). The convolution-quelling scheme operates similar to the ‘‘fat ray’’ approach [Hung et al., 2004; Chiao et al., 2010] which accounts for the frequency-dependent sensitivity of broadband signals but not for the effect of wavefront healing.

3.3. Multiscale Parameterization

[25] To achieve both spectral resolution for long-wavelength features in sparse-sampled regions and spatial resolution in regions of dense data coverage, we introduce a method of multiscale parameterization to naturally and automatically account for regional variability in data coverage and to yield models with spatially varying resolutions. In cross-section view, Figure 3 illustrates qualitatively salient principles behind the multiscale parameterization. In regions where the density of raypaths is high and where crisscrossing paths are abundant, model grids are closely spaced. The reverse is true in regions that are poorly sampled by observations. The exact configuration and subsequent refinement of model grids are data-dependent.

[26] To this end, a flexible, data-adaptive scheme of non-stationary regularization is the hierarchical wavelet decomposition [e.g., Chiao and Kuo, 2001; Chiao et al., 2006]. Rather than directly solving for model parameters specified at nodes as implied in equations (3) and (4), we expand the model vector \mathbf{m} in terms of wavelet basis functions in 3-D [e.g., Mallat, 1998]:

$$\hat{\mathbf{m}} = \mathbf{W}\mathbf{m}, \quad (5)$$

where \mathbf{W} represents a 3-D primary wavelet transform operator that expands the model vector of slowness perturbations at grid nodes in terms of a set of primary bases in the wavelet space. In essence, the inverse problem is now recast as seeking coefficients of the wavelet basis functions. Finally, the corresponding inverse wavelet transform \mathbf{W}^{-1} that satisfies the relationship $\mathbf{W}^{-1} \mathbf{W} = \mathbf{I}$ reconstitutes the estimate of \mathbf{m} .

[27] This algorithm is similar to that of Chiao et al. [2006] who applied it to a 2-D, spherical surface to investigate crustal magnetization of Mars. We extend the method to 3-D, invoking a discrete wavelet transform based on linear, biorthogonal bases, and apply it to determine slowness perturbations, δs , across an entire study volume. There are two steps in this procedure. First, we perform a 2-D wavelet transform in the horizontal directions and a 1-D wavelet transform in the vertical direction. Then tensor product of these two transforms yields the final 3-D wavelet transform. Details of the

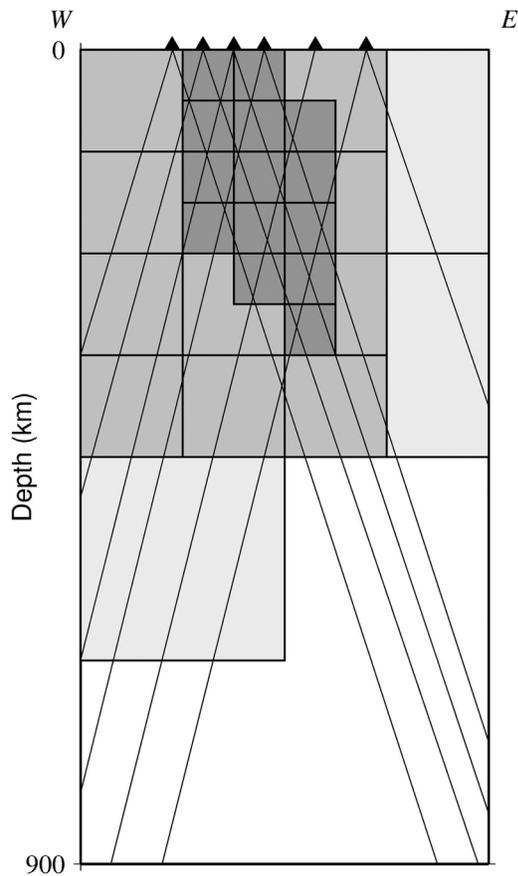


Figure 3. A schematic cross section illustrating the basic concept of a data-adaptive, multiscale parameterization scheme. Slant lines represent intrinsically uneven-distributed raypaths to seismic stations on the surface (solid triangles). In regions of sparse raypaths or few crossing rays, the parameterization is coarse, intended to capture only long-wavelength variations. Meanwhile, where sampling is dense, multiscale expansion is carried out to progressively finer scales (shown as increasing levels of gray) in order to resolvable short-wavelength features. See also Figure A1 for specifics.

multiscale parameterization, including an illustrative example, appear in Appendix A.

[28] To reformulate the data equation (2) in terms of $\tilde{\mathbf{m}}$ (instead of \mathbf{m}), we also introduce two dual-wavelet transforms associated with both the forward and the inverse primary transforms (\mathbf{W} and \mathbf{W}^{-1} , respectively). Each dual transform uses a set of dual bases that are biorthogonal to the primary bases. Biorthogonality between the primary and dual bases allows us to recast equation (2) as (see Appendix A for further details)

$$\delta t_i = \langle (\mathbf{W}^{-1})^T \mathbf{g}_i, \tilde{\mathbf{m}} \rangle. \quad (6)$$

In other words, the transpose of the inverse wavelet transform acting on each individual row vector, \mathbf{g}_i^T in the original Gram matrix, \mathbf{G} , of equation (2) would directly give the transformed Gram matrix to solve for the wavelet representation of the model vector.

[29] To impose minimum-norm regularization, the choice of the optimum damping factor is determined by analyzing the tradeoff between model variance and data misfit. As discussed further in the next paragraph, instead of exerting a uniform damping directly on all model parameters specified at each node, such as the case in the DSL solution (equation (3)), damping in the multiscale parameterization acts on wavelet coefficients, resulting in approximately equal model variances among different scales. We then construct the final image through the inverse wavelet transform of the multiresolution wavelet coefficients $\tilde{\mathbf{m}}$:

$$\begin{aligned} \hat{\mathbf{m}} &= \mathbf{W}^{-1} \tilde{\mathbf{m}} \\ &= \mathbf{W}^{-1} \left(\left((\mathbf{W}^{-1})^T \mathbf{G}^T \mathbf{C}^{-1} \mathbf{G} \mathbf{W}^{-1} + \lambda \mathbf{I} \right)^{-1} (\mathbf{W}^{-1})^T \mathbf{G}^T \mathbf{C}^{-1} \mathbf{d} \right). \end{aligned} \quad (7)$$

Even though equations (3), (4) and (7) have similar symbolic forms, the concepts they represent are miles apart. Among the differences, that in damping is particularly noteworthy. In general terms, damping sets up a threshold for the smallest permissible singular values of the Gram matrix during inversion. However, unlike conventional, grid-based parameterization in which a single damping value acts uniformly on the entire model, damping in equation (7) actually exerts spatially different effects on the model according to data coverage. This highly desirable characteristics naturally arises from the multiscale parameterization as it decomposes the model into a hierarchy of wavelet coefficients at various scales. Usually components of the decomposition that represent large scales would naturally have good constraints from the data, resulting in larger singular values than those associated with smaller scales and thus becoming well resolved.

[30] Similarly, for regions effectively covered by dense sensitivity kernels, a greater number of singular values associated with small-scale wavelet coefficients are retained above the threshold value. In contrast, for regions with sparse sampling, most singular values are damped out. Consequently, the cutoff value of the wavelet coefficients, as determined by the imposed damping to annihilate ill-constrained, small singular values, permeates through the hierarchy of spatial scales so that the finest scales being resolved from the data are spatially varying, being naturally determined by variations in the magnitude of coefficients.

[31] In essence, as we construct the final image through the inverse wavelet transform of $\tilde{\mathbf{m}}$; a process through which the data-adaptive algorithm builds up the result starting from robustly resolvable, long-wavelength components and successively adds more and more resolvable details at shorter wavelengths. This is a key distinction from spatially uniform regularization schemes. In section 4, we shall show specific examples of this progression based on actual observations.

4. Results

4.1. Comparison Among Regularization Schemes

[32] To illustrate the utility of the multiscale approach, we also conduct finite-frequency tomography using more conventional regularization schemes. Two commonly employed schemes are damped least squares and convolution-quelling solutions, both directly solve for model parameters using

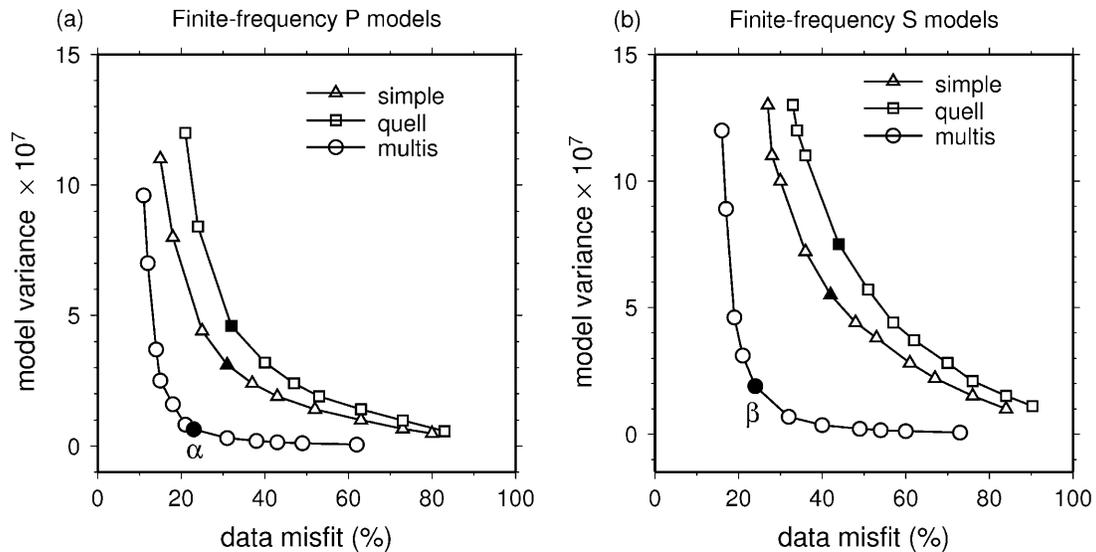


Figure 4. Tradeoff between model variance and data misfit for three different schemes of regularization. Results for (a) P and (b) S waves, respectively. The labels are: triangles (“simple”), minimum-norm, damped least squares; squares (“quell”), convolution quelling over a distance of ~ 30 km in all directions; and circles (“multi”), multiscale parameterization. Solid symbols denote preferred solutions whose images are shown in Figure 5 for further comparison. It is evident that the multiscale approach achieves the best tradeoff between model variance and data misfit. Labels α and β mark optimal solutions under the multiscale approach for P and S waves, respectively. Perturbations to these two solutions are investigated further in Appendix A (Figure A4).

grids of regular spacing. In the former case, we apply only the minimum-norm criterion; while in the latter case, we impose an additional smoothness constraint by a prescribed Gaussian function with a correlation length or standard deviation of twice the grid spacing (~ 30 km) in all three directions. It is worth noting again that conventional schemes apply regularization and smoothness constraints uniformly throughout the grids, thus unsuitable for retrieving features whose spatial and spectral distributions are innately nonstationary. In contrast, multiscale parameterization decomposes the nonstationary, final image to be solved into a natural hierarchy of image components on various length scales. In essence, the multiscale parameterization exploits simultaneously spectrally resolvable and spatially localized characteristics of the data over a range of scales, and then synthesizes results from each scale to achieve resolutions that are inherently data adaptive.

[33] Figure 4 shows comparisons among three different schemes of regularization in terms of tradeoffs between model variance and misfit of data: the former is a measure of uncertainty in model parameters resulting from errors in the data, while the latter is estimated from the sum of the squared differences between observed and predicted data, after being normalized by the sum of the squares of observed data. To obtain a preferred model, we follow the established practice in inverse theory of staying near the point of maximum curvature along the tradeoff curves (denoted by solid symbols in Figure 4 [Parker, 1994]). Such a solution gives good fits of the data but avoids amplifying errors in the data which often lead to superfluous features with extremely large uncertainties. In addition, we tested different values of damping which produce images that are highly compatible with each other, with some differences only in the amplitude of resulting anomalies, so long as one stays near the point of maximum

curvature along the tradeoff curve (see further discussions related to Figure 7 below and in Appendix A).

[34] In Figure 5, we compare images, using the same three schemes as those in Figure 4, along a cross section following the NNW-SSE trending linear array. Overall, there are similarities among the three sets of images in that except for the shallow anomaly near the Indus-Yarlung Suture (IYS, the surficial boundary between the Lhasa and the Tethyan Himalaya terranes), negative anomalies mostly occur near the left end of the profile, while positive anomalies dominate the center. Meanwhile, there are also important differences. For instance, the two conventional schemes lead to much lower amplitudes of anomalies than the multiscale approach. We have adjusted the range of scales by over a factor of two in Figures 5b and 5c, such that these images show approximately the same range of color as their counterparts in Figure 5a.

[35] Images obtained from simple minimum-norm constraints exhibit many fragmented features with low amplitudes (Figure 5b), an expected shortcoming [e.g., Chiao and Liang, 2003; Chiao et al., 2006]. To suppress such poorly constrained, short-wavelength variations, it is a common practice to impose a priori smoothness constraints that minimize the first derivatives and/or second derivatives of model parameters [e.g., Montelli et al., 2004b; Wolfe et al., 2009]. We did not apply smoothness constraints so Figure 5b can be directly compared with Figure 5a based on the multiscale approach where there is no smoothing. It is evident that for both P and S waves, the multiscale parameterization results in the lowest model variance, by far, under the same level of data misfit (Figure 4). Similarly, for a given value of model variance, the multiscale parameterization clearly yields the best fit to the data.

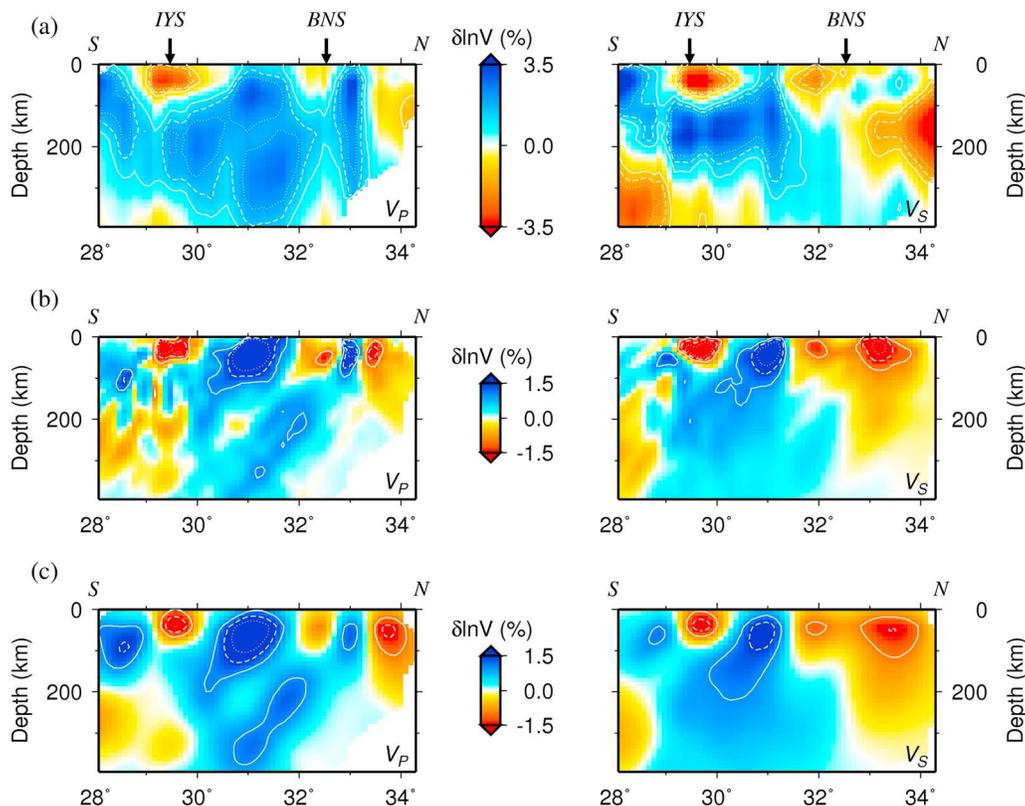


Figure 5. Comparison of images of $\delta \ln V_P$ and $\delta \ln V_S$ along a N-S trending profile using the three different schemes of regularization as in Figure 4: (a) multiscale, (b) simple minimum-norm, and (c) convolution-quelling. While all three methods place a large-scale anomaly of positive $\delta \ln V$ in the upper mantle near the center of the profile, the bottom of this subhorizontal anomaly (Figure 5a) is smeared out along raypaths in the latter two cases. This artifact is most obvious for $\delta \ln V_P$ near the bottom of Figure 5b (cross-hatched patterns between about 29° to 32°N). In addition, the latter two schemes of conventional regularization underestimate the amplitude of anomalies (The range of color scale for Figure 5a is $\pm 3\%$, twice as large as that for Figures 5b and 5c ($\pm 1.5\%$)).

[36] Instead of introducing another damping factor to reduce model roughness (often defined by the first spatial derivatives or the Laplacian of model parameters), we undertake an alternative approach of applying convolutional quelling. This procedure solves for model parameters that are subject to a smoothing operator in the form of a Gaussian filter, blurring out details of the image and removing noisy undulations in $\delta \ln V$ [Meyerholtz *et al.*, 1989]. In one particular aspect, this approach shares a property of the preferred method of multiscale inversion in that the \mathbf{Q} operator in equation (4) is simply the inverse of the symmetric Gaussian operator, equivalent to the transpose of the inverse wavelet transform in equation (7). At comparable levels of data misfit, images from convolutional quelling achieve better spectral resolution in long-wavelength features in the upper mantle than damped least squares (see Figures 5b and 5c).

[37] However, in both cases where regular-spaced, grids and spatially stationary regularization are used, the strong positive anomaly in the upper mantle, extending over horizontal distances of several hundreds of kilometers in Figure 5a, is smeared obliquely along sparse-crossing raypaths (Figures 5b and 5c). Meanwhile, short-wavelength (~ 100 km), negative anomalies in the crust, well-resolved by the multiscale approach

(Figure 5a), are also less apparent (Figures 5b and 5c). Overall, quelling resulted in somewhat worse performance than simple damped least squares because a prescribed, nonadaptive correlation length among anomalies cannot fit all short-wavelength variations in the data [e.g., Chiao and Kuo, 2001].

[38] To further illustrate the multiscale approach, Figure 6 shows how the final solution is built up, adding one hierarchical level at a time, for $\delta \ln V_P$ and $\delta \ln V_S$ along the same profile as shown in Figure 5. Notice that long-wavelength features emerge on the coarsest scale (level = 1), with the demarcation between positive and negative anomalies farther to the north of the Bangong-Nujiang suture (BNS, the surficial boundary between the Lhasa and the Qiangtang terranes in southern and central Tibet, respectively [Yin and Harrison, 2000]). Fitting of data progresses steadily as finer-scale features appear. The fitting of P wave data improved rapidly after level 4 is incorporated; resulting in a well-resolved negative anomaly throughout the thick crust near the IYS.

4.2. Anomalies in V_P , V_S , and V_P/V_S

[39] In Figure 7, we present a summary of our results. Three maps, taken at depths of 56, 155, and 295 km, and two cross sections, trending roughly N-S and E-W, illustrate 3-D var-

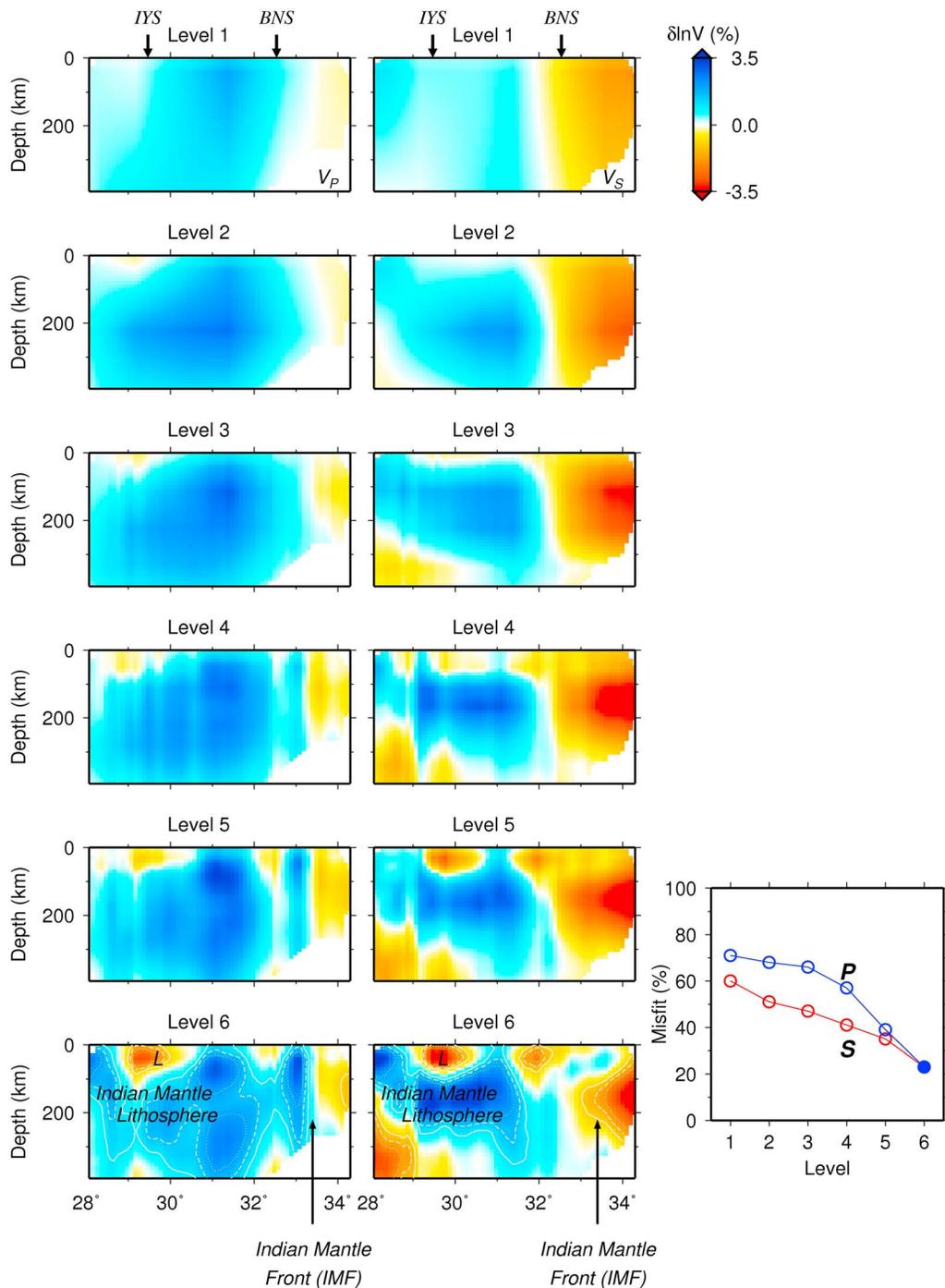


Figure 6. Hierarchical built up of the multiscale images for $\delta \ln V_P$ and $\delta \ln V_S$ along the same N-S trending profile as in Figure 5. The images are formed by successive superposition of images from the coarsest (“Level 1”: level 1 only) to the finest scales (“Level 6”: sum of levels 1 to 6). The right panel shows how data fits improve as each smaller-scale features are incorporated in the model. Solid, dashed, and dotted contours correspond to $\pm 1\%$, $\pm 1.5\%$, and $\pm 2\%$ of anomalies, respectively. A subhorizontal anomaly of high V_P and V_S , and the Indian mantle front (IMF), are the most obvious features imaged in the upper mantle. Notice that in the crust, regions of low V_P and V_S near the IYS (the active Lunggar rift, L) and the BNS become prominent only when levels 5 and above are included.

iations in V_P , V_S , and corresponding changes in V_P/V_S under a large portion of the Himalayan-Tibetan orogen. All results for the fractional change in V_P/V_S , $\delta \ln(V_P/V_S)$, are new from this study, while the map views for $\delta \ln V_P$ and $\delta \ln V_S$ complement

the recent report in paper 1 in which horizontal slices at depths of 42, 112, 225 and 338 km were shown.

[40] Theoretically, $\delta \ln(V_P/V_S)$ is simply the difference between $\delta \ln V_P$ and $\delta \ln V_S$ [e.g., *Chou et al.*, 2009]. Although

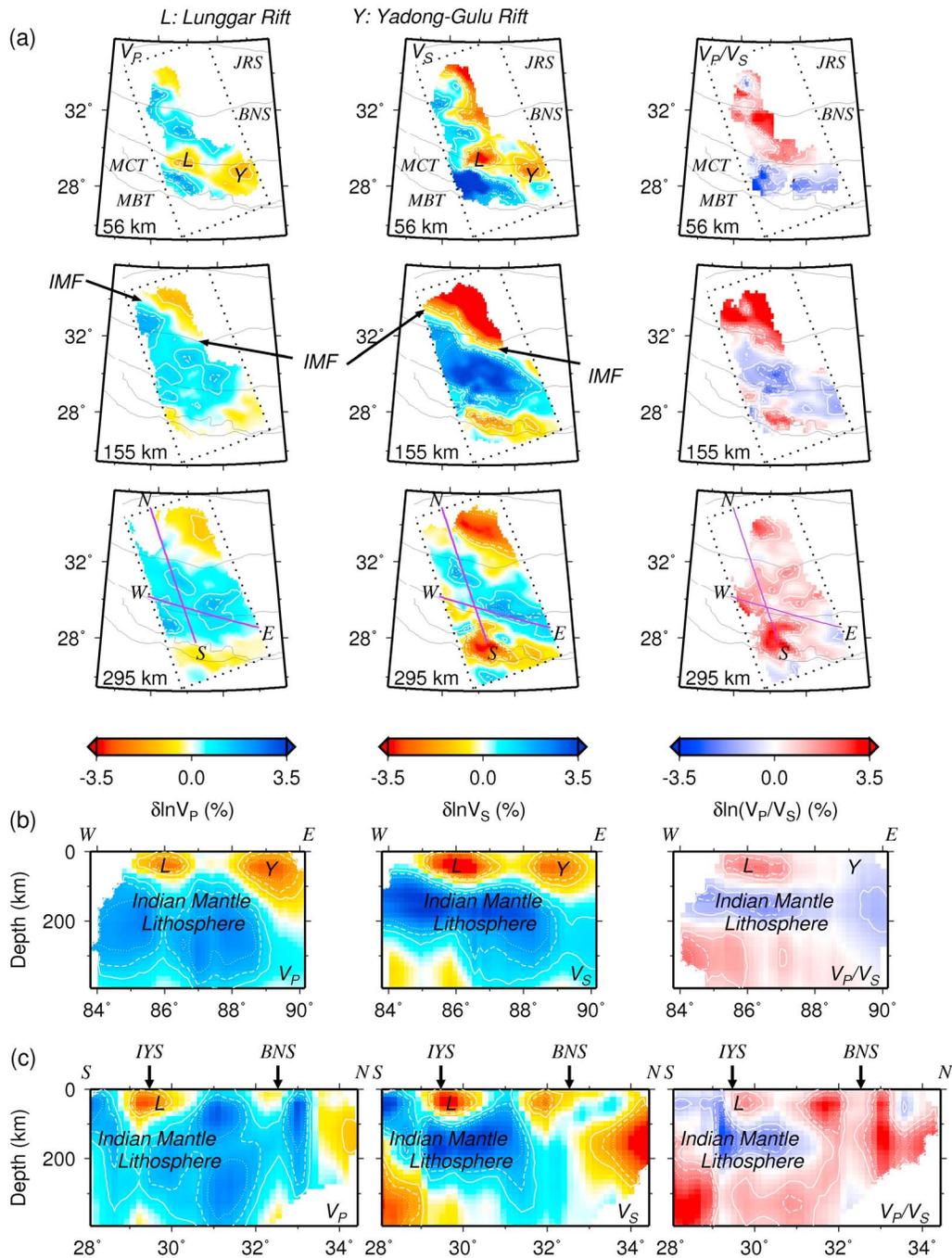


Figure 7. Images showing 3-D variations of V_p , V_s , and V_p/V_s beneath central Tibet from multiscale, finite-frequency travelt ime tomography. (a) Map views of $\delta \ln V_p$, $\delta \ln V_s$, and $\delta \ln(V_p/V_s)$ at three selected depths. Solid curves show major geologic boundaries as those in Figure 1. IMF marks the Indian mantle front. (b and c) Cross sections of $\delta \ln V_p$, $\delta \ln V_s$, and $\delta \ln(V_p/V_s)$, trending approximately east-west and north-south (E-W and N-S, respectively) and extending down to a depth of 400 km. In general, results for long wavelength variations in V_p and V_s are in good agreement. More important, the subhorizontal configuration of positive anomalies of V_p and V_s in the upper mantle stands out as a negative anomaly in V_p/V_s (whose color scheme is the reverse of those for V_p and V_s). This feature is interpreted as underthrust Indian mantle lithosphere of Archean age, thus being chemically depleted and buoyant. In all images, we masked out portions of the models where sampling is sparse and resolution is poor [Hung et al., 2010]. Solid, dashed, and dotted contours correspond to $\pm 1\%$, $\pm 1.5\%$, and $\pm 2\%$ of anomalies, respectively.

the configuration of sensitivity kernels are roughly comparable between P and S waves at each frequency band used here, differences in coverage of source-receiver pairs at places can lead to discordant resolutions between $\delta\ln V_P$ and $\delta\ln V_S$ which, in turn, may produce fictitious anomalies in V_P/V_S . Consequently, we select a subset of events for which traveltimes residuals of both P and S waves are available simultaneously. This subset amounts to 78% and 82% of the original data set for P and S waves, respectively. An unavoidable side effect is a slight reduction in the spatial coverage of images when compared with those in paper 1, but the overall distribution of $\delta\ln V_P$ and $\delta\ln V_S$ remain unaffected. To visualize regions reliably sampled by both P and S waves simultaneously, we plot squared root values of the diagonal elements of the matrix $\mathbf{G}^T\mathbf{G}$ in Figure A2, where \mathbf{G}^T is the transpose of the Gram matrix \mathbf{G} in equation (2). $\mathbf{G}^T\mathbf{G}$ is akin to the model resolution matrix, indicating how well model parameters are being constrained by the combined high- and low-frequency traveltimes residuals. For comparison, we present corresponding values for the full data set in Figure A3 and there are no significant differences between Figures A2 and A3. For $\delta\ln V_P$ and $\delta\ln V_S$, resolution tests in paper 1 indicate that robustly constrained regions roughly correspond to nodes where their indicators of robust observational control have values exceeding 0.5% of the overall medium at all nodes. Thus at each node, we use this value as the threshold below which the result of inversion is deemed not reliable enough for display. The corresponding threshold for $\delta\ln(V_P/V_S)$ is then 0.7%, reflecting the propagation of independent errors when the difference between $\delta\ln V_P$ and $\delta\ln V_S$ is taken. Imposing these thresholds on the results of inversion leads to slight differences in the spatial coverage of images among $\delta\ln V_P$, $\delta\ln V_S$, and $\delta\ln(V_P/V_S)$ (Figure 7).

[41] For $\delta\ln V_P$ and $\delta\ln V_S$, there are two important features as pointed in paper 1: First, in the upper mantle, a positive anomaly, concentrated between depths of about 100 to 250 km, extends subhorizontally over a distance of about 600 km to the north of the BNS (near 33°N). This feature is most obvious in $\delta\ln V_P$ (Figure 7c, left) and the corresponding feature in $\delta\ln V_S$ is where its value rapidly changes from near naught to strongly negative. (Regional traveltimes tomography based on teleseismic sources is sensitive only to gradients in $\delta\ln V$, as its baseline depends on the choice of reference model (in this case model AK135 [Kennett et al., 1995]). Second, in the crust, there are two isolated negative anomalies straddling the IYS, each extending down to a depth close to 100 km (Figure 7b, left and middle).

[42] For $\delta\ln(V_P/V_S)$, a slab-like, subhorizontal negative anomaly in the upper mantle is concentrated between depths of 100 and 200 km. This feature is obvious along the E-W profile and extends over approximately two thirds of the N-S trending cross section (Figure 7, right). The result that the anomaly in $\delta\ln(V_P/V_S)$ does not extend as deep as either that of $\delta\ln V_P$ or $\delta\ln V_S$ is a direct consequence of high $\delta\ln V_S$ being most pronounced above 200 km (Figure 7, middle and left).

[43] In the crust, while the two isolated anomalies of negative $\delta\ln V_P$ are comparable in amplitude, the anomaly on the west has greater amplitude in $\delta\ln V_S$ than the other anomaly about 200 km farther to the east (Figure 7). As such, the former anomaly shows up with a high ratio of V_P/V_S but the latter anomaly becomes null in terms of V_P/V_S (Figure 7).

[44] In both the upper mantle and the crust, a comparison of $\delta\ln V_P$ and $\delta\ln V_S$, vis-à-vis, with new results for $\delta\ln(V_P/V_S)$

yields additional information which is useful in constraining geological processes responsible for these anomalies. In section 5, we first discuss resolution tests in addition to those already performed in paper 1. These new tests then form the basis for further geological interpretations.

5. Discussions and Tectonic Significance

5.1. Resolution Tests

[45] In paper 1, we have conducted a suite of resolution tests, demonstrating the capability of multiscale, finite-frequency tomography in restoring true configuration of heterogeneities in both V_P and V_S . Specifically, we used a variety of different distribution of anomalies including variable-sized checkerboards in the entire volume being investigated, a channel-like layer in the lower crust with a thickness equal to the smallest grid spacing (~15 km), and isolated, crustal volumes of low V_P and low V_S which extend to a depth of about 100 km.

[46] Here we emphasize tests designed to assess how robust features in $\delta\ln(V_P/V_S)$ are resolved. First, we explore effects of norm damping in restoring anomalies in V_P/V_S . For both P and S waves, the tradeoff between model variance and data misfit has a well-defined point of maximum curvature (Figure 4) and we try several different combinations of norm damping between P and S waves in the vicinity of optimal solutions. In general, salient features in $\delta\ln(V_P/V_S)$ are insensitive to perturbations in norm damping, even though the amplitude of anomalies do show some fluctuations as one would expect. We discuss details of this procedure in Appendix A (Figure A4), and show examples of resulting images in Figures A5–A7.

[47] Second, we generate synthetic traveltimes residuals for specific features of $\delta\ln(V_P/V_S)$ that appear in our results (Figure 7) and perform inversion to ensure that such features are well resolved. In doing so, we calculate synthetic residuals using exactly the same source-receiver configurations, the multiscale expansion, and other relevant parameters as those used in the analysis of actual data. Moreover, prior to the inversion, we add random errors whose mean amplitude is about 5% of the average size of signals. In sections 5.2 and 5.3, we summarize results of such tests in conjunction with geological interpretations of features being tested (Figures 8 and 9).

5.2. Large-Scale Anomalies in the Upper Mantle

[48] To verify the presence of a slab-like, subhorizontal region of low V_P/V_S in the upper mantle, we place a tabular layer of limited lateral extent beneath the northern Himalayas and the Lhasa terrane. This isolated feature has a uniform amplitude of -2% in $\delta\ln(V_P/V_S)$, with corresponding anomalies of $+2\%$ and $+4\%$ in $\delta\ln V_P$ and $\delta\ln V_S$, respectively. Figure 8 shows the result of this test. While there are some inevitable smearing, configurations of all anomalies in V_P , V_S , and V_P/V_S are well resolved, as outlined by regions where the increase in amplitude is rapid. Likewise, peak amplitudes of anomalies are slightly underestimated by about only 10–13%. Clearly a difference of about 2% or less in $\delta\ln(V_P/V_S)$ is easily resolvable. This estimated minimum resolution is only about half the overall contrast between the slab-like anomaly and the region farther to the north, leaving little doubt that there is a subhorizontal region of low V_P/V_S in the upper mantle under the northern Himalayas and

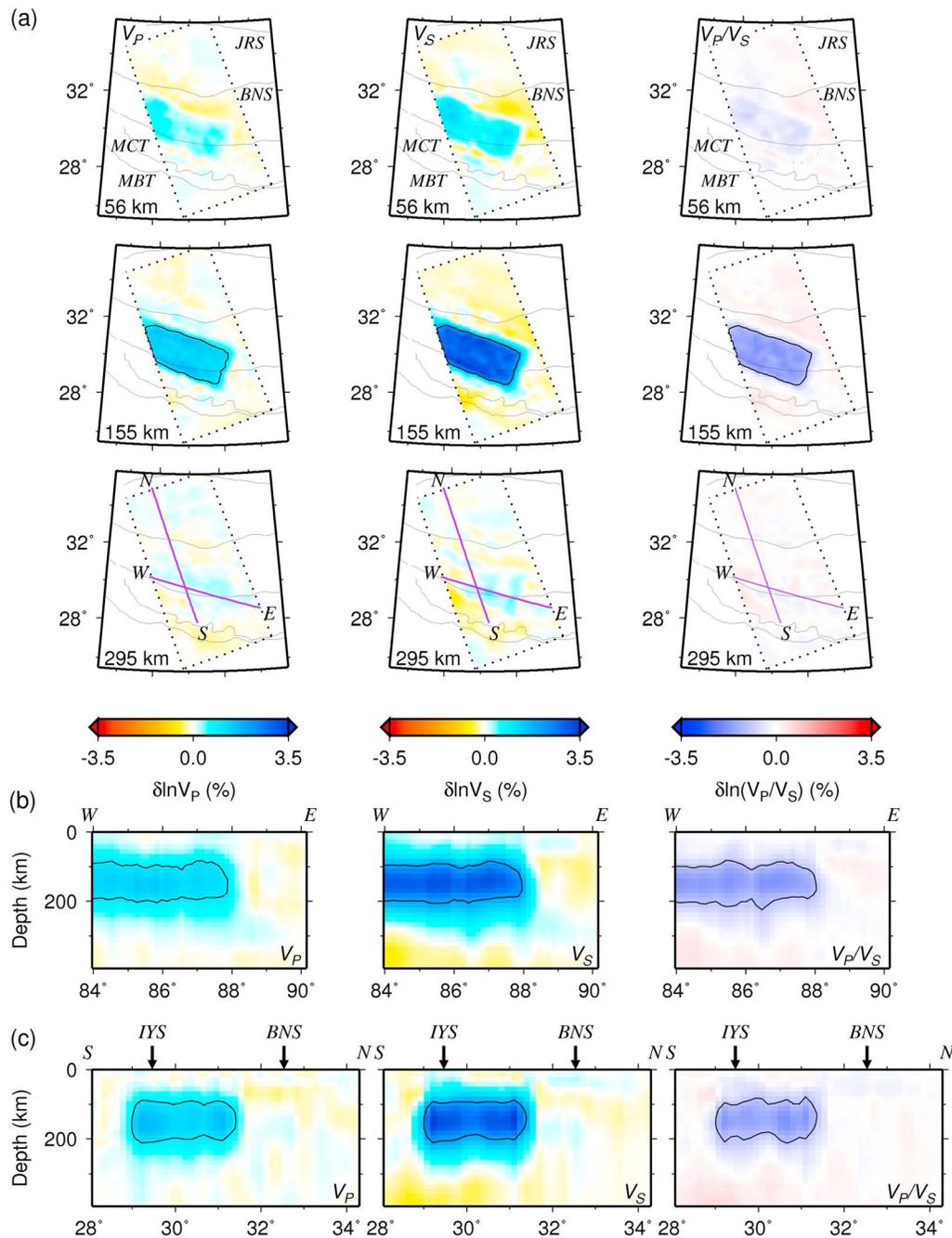


Figure 8. A resolution test designed to corroborate the subhorizontal, slab-like anomaly of high V_P and V_S but low (V_P/V_S) . The input is comprised of a horizontal layer, embedded between depths of 100 and 200 km beneath the Lhasa terrane. Its lateral extent is set to about 500 km by 200 km, with +2% and +4% increases in V_P and V_S , respectively (or a -2% reduction in V_P/V_S). Contours mark 1%, 2%, and -1% in $\delta \ln V_P$, $\delta \ln V_S$, and $\delta \ln(V_P/V_S)$ where the amplitude of anomalies increase rapidly. Exactly the same data coverage and algorithms of analysis apply to both synthetic test and real observations, with a random error of 5% in amplitude added to the former case. The layout is the same as that of Figure 7.

the Lhasa terrane. In contrast, anomalously high values of V_P/V_S stand out north of the BNS near 33° N (Figure 7, right).

[49] Based on a suite of complementary evidence, including previous work on modeling of plate flexure using Bouguer gravity anomalies, *Chen et al.* [2010] surmised that positive anomalies of $\delta \ln V_P$ and $\delta \ln V_S$ are most likely associated with

subhorizontally advancing mantle portion of underthrust Indian lithosphere (so-called “Greater India”). Along the western side of the area of our study, the northern edge of the Greater India, or the Indian mantle front (IMF), terminates about 100 km north of the BNS (Figure 7) [*Chen and Tseng, 2007*]. In particular, the IMF is marked by rapid decrease in

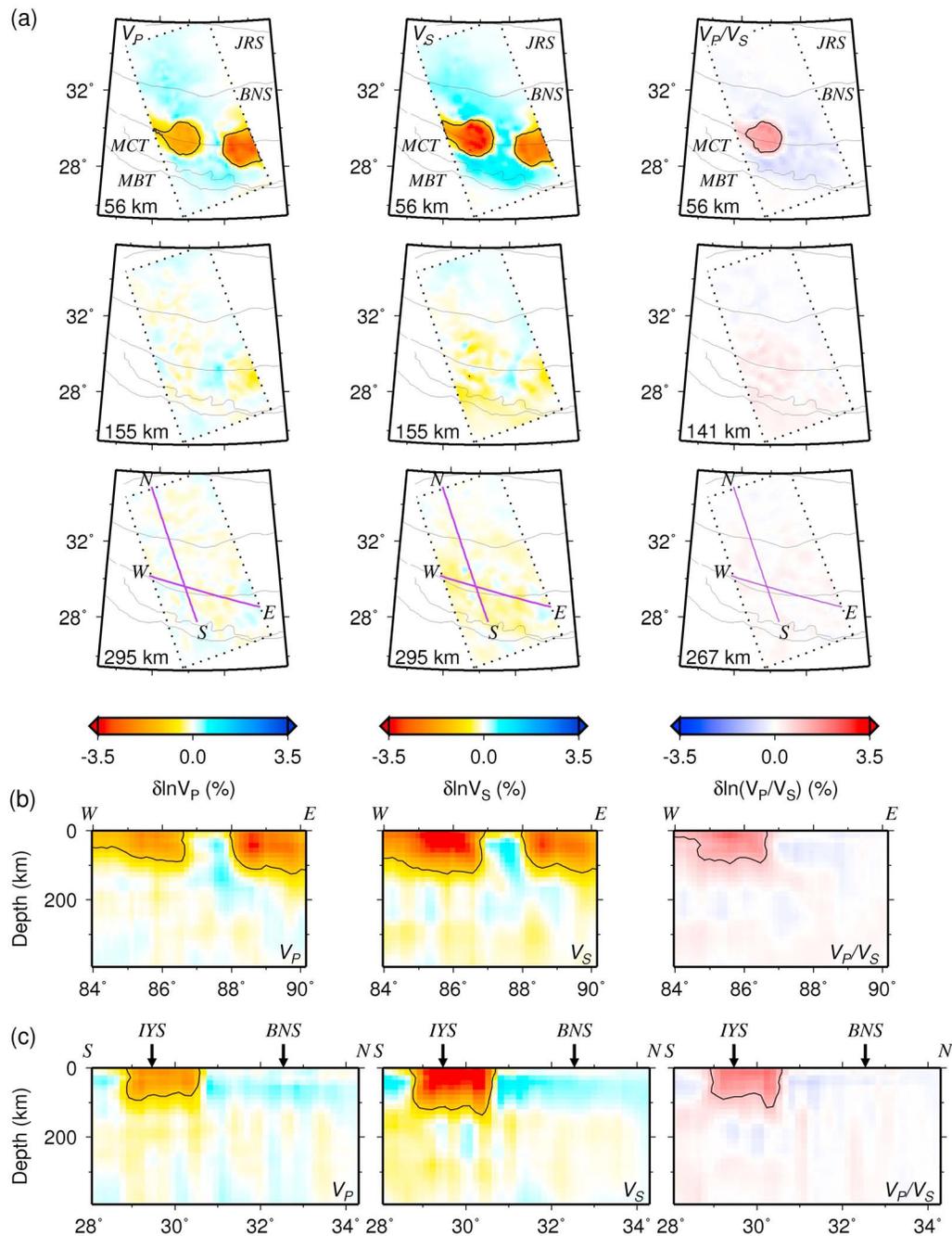


Figure 9. A resolution test aimed to verify the isolated nature of anomalies in the crust. (Layout is the same as that in Figure 8.) The input consists of two isolated cylinders, each with a radius of ~ 100 km and extends down to a depth of 100 km. For the cylinder on the west, reductions in V_P and V_S , are -2% and -4% , respectively (corresponding to a 2% increase in V_P/V_S). Corresponding values for the other cylinder are reductions of -3% in both V_P and V_S , resulting in a null anomaly in V_P/V_S . Synthetic traveltimes are generated under the same conditions as those of Figure 8.

both $\delta \ln V_P$ and $\delta \ln V_S$, and a corresponding increase in the birefringence of vertical incident shear waves [Chen *et al.*, 2010].

[50] The results that most of this slab-like anomaly of high V_P and V_S is associated with low values of V_P/V_S (Figure 7) further reinforce this interpretation: The Indian shield, like most other shield regions, is likely to have a chemically

refractory mantle lithosphere which provides the necessary buoyancy to counteract thermal contraction caused by more than 2.5 Ga of cooling since the Archean. Meanwhile, Lee [2003] suggested that there is a negative correlation between the magnesium number (Mg#) and V_P/V_S ; so one expects high Mg#, indicating depleted mantle materials whose V_P/V_S is low. Furthermore, the negative anomaly of $\delta \ln(V_P/V_S)$ is apparently

restricted to depths no deeper than 200 km, somewhat shallower than the corresponding positive anomalies of V_P and V_S . This result is consistent with the isopycnic hypothesis [e.g., Jordan, 1979, 1988] which predicts that extraction of basalt in the Archean would have left the mantle lithosphere to be most depleted near its top. However, some recent studies questioned if the relationship proposed by Lee [2003], an extrapolation based on data measured at standard temperature and pressure, may be too strong [Faul and Jackson, 2005; Wagner et al., 2008]. So a precise, quantitative relationship awaits more measurements under high pressures and temperatures.

[51] At any rate, it is curious to note that strong negative anomaly of $\delta\ln(V_P/V_S)$ does not extend as far north as positive anomalies of $\delta\ln V_P$ and $\delta\ln V_S$ (a direct consequence of a gradual termination of positive $\delta\ln V_S$ between about 31.5°N and 33°N) even though the strongest gradients all occur near 33°N (Figure 7c). At the moment, the cause for a gradual termination of positive $\delta\ln V_S$ is not clear. We suspect that the transitional region corresponds to the northernmost border of underthrust Greater India, comprised of gradually refertilized Indian continental lithosphere as it was intruded by mafic dykes and thinned during the opening of the Neo-Tethyn ocean: a well-documented process during major orogeny in the Phanerozoic (such as the Calidonides [e.g., Gee et al., 2008; Milnes, 1998]). Indeed, some previous work attributed highly modified Indian crust that used to lie outboard of stable, Greater India, as the source material making up the Greater Himalayas or the overthrust nappe above the Main Central Thrust of the Himalayas [e.g., DeCelles et al., 1998], but a competing hypothesis favors southward transport and exhumation of midcrust to lower crust materials beneath southernmost Tibet [e.g., Searle et al., 2003].

[52] North of 33°N, the upper mantle beneath the Qiangtang terrane is characterized by high V_P/V_S ratios. This is approximately the southern portion of a region in northern Tibet where a number of previous studies have argued for an unusually warm upper mantle, including an unusually thin crust over a regional scale (implying thermal isostasy), and perhaps lateral (E-W trending) flows in the upper mantle, as indicated by very large values (~ 2 s) of shear wave birefringence, or even partial melts [Chen et al., 2010; Chen and Özalaybey, 1998; Jiménez-Munt et al., 2008; McNamara et al., 1997; Molnar and Chen, 1984; Ni and Barazangi, 1983; Owens and Zandt, 1997; Tseng et al., 2009].

[53] The most direct evidence for partial melts, which would easily account for very high V_P/V_S ratios, comes from the observation of volcanism on the surface. The most recent episode of magmatism in Tibet, ranging in age from about 15 Ma to recent, is largely comprised of lavas with mafic to intermediate composition; and this volcanism is restricted exclusively in regions north of 34°N [e.g., Chung et al., 2005, and references therein]. Chen and Tseng [2007] pursued the tectonic context of this recent episode of igneous activity, attributing it to convective removal of thickened, hydrous subcontinental mantle lithosphere of the Qiangtang terrane, whose remnant is now resting directly over the top of the lower mantle or the 660 km discontinuity [Tseng and Chen, 2008].

5.3. Rifting of the Thickened Crust

[54] As pointed out in paper 1, the two negative anomalies in the crust of southern Tibet are most likely associated with (from west to east) the Lunggar and the Yadong-Gulu active

riffs, respectively: The center of each anomaly coincides with the rift axis; and the longitudinal extent of each anomaly corresponds to zones of low electric resistivity deduced from magnetotelluric (MT) profiles along the rift valley [e.g., Unsworth et al., 2005].

[55] Obvious from a bimodal pattern of positive traveltime residuals along E-W trending profiles and from resolution tests reported in paper 1, the two crustal anomalies are separated from each other. To ensure the reliability of results for V_P/V_S , here we use the same configuration of synthetic anomalies as in paper 1, with two isolated cylinders of ~ 100 km in radius and in depth (Figure 9). For the purpose at hand, we specify anomalies of -2% and -4% in V_P and V_S , respectively, near the Lunggar rift, so the expected value of $\delta\ln(V_P/V_S)$ is $+2\%$. The corresponding values for the anomaly near the Yadong-Gulu rift are -3% in both V_P and V_S , so $\delta\ln(V_P/V_S)$ should be close to null.

[56] Figure 9 shows the resulting images which recover the overall shape and amplitude of both anomalies well. The range of $\delta\ln(V_P/V_S)$ being confidently resolved in this test is 2% , comparable to the contrast appeared in the crust (Figure 7). Once again, damping during the inversion caused an unavoidable but only moderate amount of vertical smearing and decrease in peak amplitude. Nonetheless, lateral smearing is sufficiently small that there is no ambiguity in identifying each anomaly as an isolated feature. To a large extent, the high resolution is also a direct consequence of dense-spaced stations, about 5–8 km apart at most, along the N-S trending linear array (Figure 1).

[57] More interesting is the difference between the two anomalies which is discernable from $\delta\ln V_S$ but becomes obvious when viewed as $\delta\ln(V_P/V_S)$. For the Yadong-Gulu rift, which apparently reaches somewhat below the base of thickened crust (~ 70 km in thickness), very small $\delta\ln(V_P/V_S)$ implies concordant changes between $\delta\ln V_P$ and $\delta\ln V_S$, consistent with a predominantly thermal origin of this anomalies [Jackson et al., 2000; Sinogeikin et al., 2003]. If fluids, such as partial melt or volatiles, are present deep in the crust, one would expect highly elevated values of $\delta\ln(V_P/V_S)$ [e.g., Hammond and Humphreys, 2000; Zhang et al., 2004], which are not observed. In fact, it has been known for almost 30 years that frequent seismicity, including several moderate-sized events, occurs in the uppermost mantle near the Yadong-Gulu rift [Chen et al., 1981; Chen and Molnar, 1983; Chen and Yang, 2004; Zhu and Helmberger, 1998]. Thus the temperature in the uppermost mantle must be well below the solidus of peridotite, probably in the range of only 600–800°C, so enough elastic strain can accumulate and cause earthquakes [Chen and Molnar, 1983; Wiens and Stein, 1983].

[58] In contrast, high $\delta\ln(V_P/V_S)$ along the Lunggar rift indicates the likelihood of volatiles or even partial melt in the crust. The latter, if present, must be in very small quantities because recent igneous extrusion is all but absent [Taylor and Yin, 2009]. Since subcrustal seismicity is prevalent only near the Yadong-Gulu rift, one of the longest in the entire Himalayan-Tibetan orogen, it appears to have a different origin or is at a different stage of evolution from numerous other rifts which share an approximately N-S trend. Equally important, deep penetration of the two anomalies implies that active extension concentrates on steep-dipping zones which, unlike subhorizontal detachment faults, can accommodate only very limited amount of horizontal displacement. Further-

more, the predominately N-S trend of rifts, spaced several hundreds of kilometers apart but occurring throughout Tibet [e.g., *Taylor and Yin*, 2009], would lead to crustal blocks with quasi-linear edges: a feature that may be a useful clue in identifying collapsed plateaus from ancient orogens.

6. Conclusions

[59] Using high-quality, broadband data recorded by the Hi-CLIMB array, we illustrate how a multiscale parameterization of tomographic inversion, coupled with finite-frequency theory to handle frequency dependence of traveltimes and then in a data-adaptive manner, can resolve spatial-varying anomalies of seismic wave speeds at different length scales. Unlike models built up on fixed-scale grids, multiresolution parameterization preserves both spectral and spatial resolutions which are automatically regulated by the quality and the coverage of data.

[60] Our results corroborate that a subhorizontal anomaly of high $\delta \ln V_P$ and $\delta \ln V_S$ underlies much of the Himalayan and Lhasa terranes. Broad aperture of data coverage constrains the minimal lateral (east-west) span of the anomaly to be between approximately 84° and 89°E in longitude. This feature extends northward up to 33°N at places and is concentrated between depths of about 100 to 200 km. Much of this anomaly shows low V_P/V_S , providing additional evidence for the interpretation that such an anomaly is associated with subhorizontal underthrusting of chemically depleted, buoyant mantle lithosphere of the Indian shield [e.g., *Chen et al.*, 2010; *Hung et al.*, 2010].

[61] In the crust, two anomalies of low V_P and V_S occur along two N-S trending, geothermal active rift zones that are about 300 km apart. Each anomaly is cylindrical in shape, with a diameter of about 100 km and extending throughout the thickened crust near the IYS. Locations of these anomalies correlate well with regions of low electrical resistivity in the crust, but the two anomalies of low V_P and V_S are disconnected [*Hung et al.*, 2010]. Moreover, due to particularly low values of V_S the western anomaly along the Luggar rift is characterized by high V_P/V_S ratio, an attribute not shared by the Yadong-Gulu rift to the east. In the latter case, a null anomaly in $\delta \ln(V_P/V_S)$ is consistent with frequent occurrence of moderate-sized earthquakes in the uppermost mantle [e.g., *Chen et al.*, 1981; *Chen and Molnar*, 1983] where the temperature must be far below the solidus of peridotite to allow elastic strain to accumulate: a prerequisite for generating earthquakes. Taken together, all available geophysical evidence points to the conclusion that pervasive, southward channel flow of the lower crust, if ever active, was limited in both space and duration [e.g., *Searle et al.*, 2003; *Hung et al.*, 2010].

Appendix A

[62] Here we present additional discussions on three topics that have not been explored fully in the main text: the wavelet transform, multiscale parameterization, and regularization.

A1. Wavelet Transforms

[63] Following *Chiao et al.* [2010], we briefly review the basic properties of and relationships between the primary and the inverse wavelet transforms and their associated dual transforms.

[64] At original grid nodes, we expand the model vector, \mathbf{m} , in terms of a set of basis functions, Φ_j ($j = 1, \dots, M$), associated with a 3-D, primary wavelet transform operator, \mathbf{W} . We then define an associated dual-wavelet transform \mathbf{W}^* and apply it to the vector of data kernel, \mathbf{g}_i , in order to obtain the Gram matrix in a new form that relates the model vector, now expressed in wavelet coefficients, to the data vector. This dual wavelet transform involves the expansion in terms of a second set of dual bases, Φ_j^* , which is biorthogonal to Φ_j . (Notice that in general, the bases for an invertible wavelet transform are not necessarily mutually orthogonal to each other.) Here biorthogonality is defined by the inner product between these two sets of bases, $\langle \Phi_j, \Phi_{j'}^* \rangle = \delta_{jj'}$. Specifically,

$$\mathbf{W}\mathbf{m} = \sum_{j=1}^M \langle \mathbf{m}, \Phi_j^* \rangle \Phi_j, \quad (\text{A1})$$

$$\mathbf{W}^* \mathbf{g}_i = \sum_{j=1}^M \langle \mathbf{g}_i, \Phi_j \rangle \Phi_j^*. \quad (\text{A2})$$

Taking the inner product of equation (A1) and the data kernel, \mathbf{g}_i , we rewrite our formulation in terms of dual wavelets using relations expressed by (A1) and (A2):

$$\begin{aligned} \langle \mathbf{g}_i, \mathbf{W}\mathbf{m} \rangle &= \left\langle \mathbf{g}_i, \sum_{j=1}^M \langle \mathbf{m}, \Phi_j^* \rangle \Phi_j \right\rangle = \sum_{j=1}^M \langle \mathbf{m}, \Phi_j^* \rangle \langle \mathbf{g}_i, \Phi_j \rangle \\ &= \left\langle \sum_{j=1}^M \langle \mathbf{g}_i, \Phi_j \rangle \Phi_j^*, \mathbf{m} \right\rangle = \langle \mathbf{W}^* \mathbf{g}_i, \mathbf{m} \rangle. \end{aligned} \quad (\text{A3})$$

With the commutative property of inner product, equation (A3) can be carried forward as matrix multiplication and rewritten as

$$\langle \mathbf{g}_i, \mathbf{W}\mathbf{m} \rangle = (\mathbf{W}\mathbf{m})^T \mathbf{g}_i = \mathbf{m}^T (\mathbf{W}^T \mathbf{g}_i) = \langle \mathbf{W}^T \mathbf{g}_i, \mathbf{m} \rangle. \quad (\text{A4})$$

In Hilbert space, each linear operator, \mathbf{W} , has a corresponding adjoint operator, denoted by \mathbf{W}^\dagger , that satisfies the following relationship:

$$\langle \mathbf{g}_i, \mathbf{W}\mathbf{m} \rangle = \langle \mathbf{W}^\dagger \mathbf{g}_i, \mathbf{m} \rangle. \quad (\text{A5})$$

Comparing the three equivalent expressions, equations (A3)–(A5), it follows that the dual transform \mathbf{W}^* is simply the adjoint (conjugate transpose) of \mathbf{W} . In Hilbert space, instead of simply summing up the product of each component from two vectors, their inner product is evaluated by first taking the complex conjugate of the second of the two vectors. In traveltime tomography, all quantities are real and therefore so is \mathbf{W} , or equivalently $\mathbf{W}^T = \mathbf{W}^\dagger$ (see equations (A4) and (A5)).

[65] Likewise, we introduce the inverse transform \mathbf{W}^{-1} and its corresponding dual transform $(\mathbf{W}^{-1})^*$ and apply them to the model vector and the vector of data kernel (now both in wavelet space), respectively. Following the same operations as in equations (A3)–(A5) and using the identity $\mathbf{W}^{-1} \mathbf{W} = \mathbf{I}$, equation (2) is recast as

$$\delta t_i = \langle \mathbf{g}_i, \mathbf{W}^{-1}(\mathbf{W}\mathbf{m}) \rangle = \langle (\mathbf{W}^{-1})^* \mathbf{g}_i, \tilde{\mathbf{m}} \rangle = \langle (\mathbf{W}^{-1})^T \mathbf{g}_i, \tilde{\mathbf{m}} \rangle. \quad (\text{A6})$$

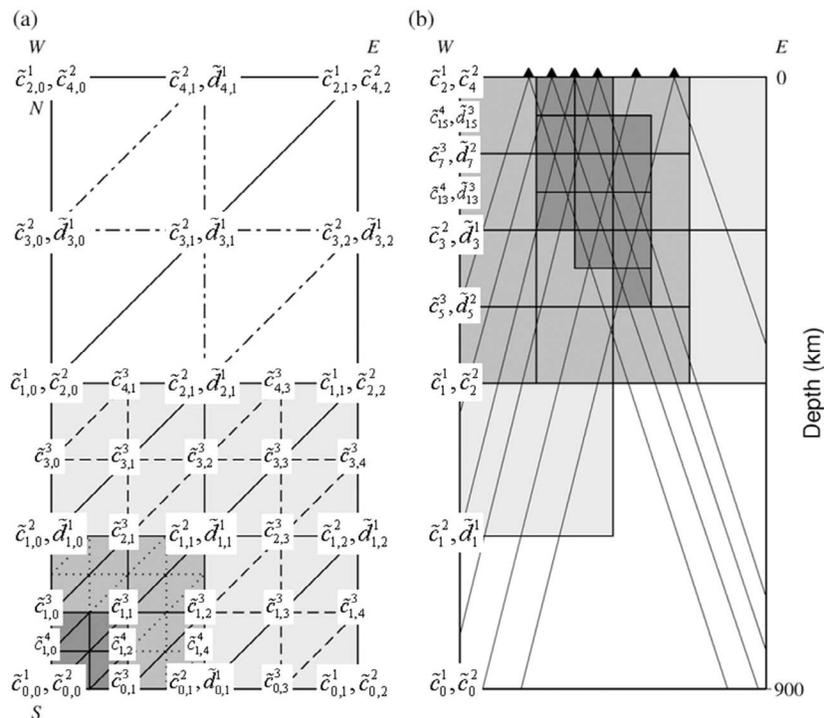


Figure A1. Schematic diagrams illustrating how to build a multiscale parameterization, from levels 1 to 4, for the model of slowness perturbations. (a) Map view of a 2-D spherical surface which is successively divided into triangular meshes. At each node, the two integers in subscript specify node indices in the N-S and the E-W direction, respectively. The integer in superscript indicates the scale level in a 2-D wavelet transform. When more than one parameter is shown at a given node, their relationships are defined by the lifting schemes described in section A3. (b) Cross-sectional view of how a thick sector of shell is being progressively bisected into finer layers. At each depth, the integer in subscript indicates the node index in the vertical direction while the integer in superscript indicates the scale level in a 1-D wavelet transform. The lifting schemes in section A3 describe relationships among parameters that occupy the same node position. Slant, solid lines represent intrinsically uneven-distributed raypaths. In regions of sparse raypaths or few crossing rays, the parameterization is coarse, intended to capture only long-wavelength variations. Meanwhile, where sampling is dense, multiscale expansion is carried out to fine scales in order to resolve short-wavelength features.

Therefore, coefficient vectors of the data kernel, which relates the model vector in wavelet space, $\tilde{\mathbf{m}}$, to the i th traveltimes data, can be found by applying the dual-wavelet transform associated with the inverse wavelet transform to the original data kernel in the node-based parameterization, or equivalently calculated by the matrix multiplication between the conjugate transpose matrix of \mathbf{W}^{-1} and column vector, \mathbf{g}_i . In practice, this inverse wavelet transformation is the very starting point of all operations that eventually lead to the solution for \mathbf{m} .

A2. Hierarchical Refinement of the Multiscale Model

[66] Starting with the grid system described in section 3.2, the continuous 3-D model of slowness perturbation is discretized into M regular-spaced nodes. Three indices, in subscript, specify a given grid node in the elongated N-S, the narrow E-W, and the vertical direction, respectively. The other index l , in superscript, denotes the wavelet decomposition of \mathbf{m} at the l th hierarchical scale level (Figure A1a). (For simplicity, we shall drop the index for the vertical

direction in Figure A1a and in subsequent notations that pertain only to map views.)

[67] In Figure A1, we show how the hierarchical refinements build up to a highest scale level of $l = 4$. As the total length along the E-W direction is about half of those along the other two dimensions, we first divide the elongated study region into two approximately square cells. Similar to the spherical triangular meshes adopted by *Chiao and Kuo* [2001], the mesh refinement begins from the root level $l = 1$, with four ($= 2^2$) isosceles triangles occupying the two cells; and each triangle has two congruent sides about 500 km in length (Figure A1a).

[68] While the mesh is in ordinary space, eventually two types of parameters attach to each node of the mesh: the value of a component in the model vector \mathbf{m} , and a related wavelet coefficient, \tilde{c} , which, in turn, is associated with $\tilde{\mathbf{m}}$. To avoid confusion, we denote parameters in wavelet space and in ordinary space by symbols with and without a tilde, respectively. This is necessary as the lifting scheme described in the next section involves both types of parameters. As an example, the vector $\tilde{\mathbf{c}}^1$, representing slowness perturbations at the root level (in wavelet space) and being specified at the

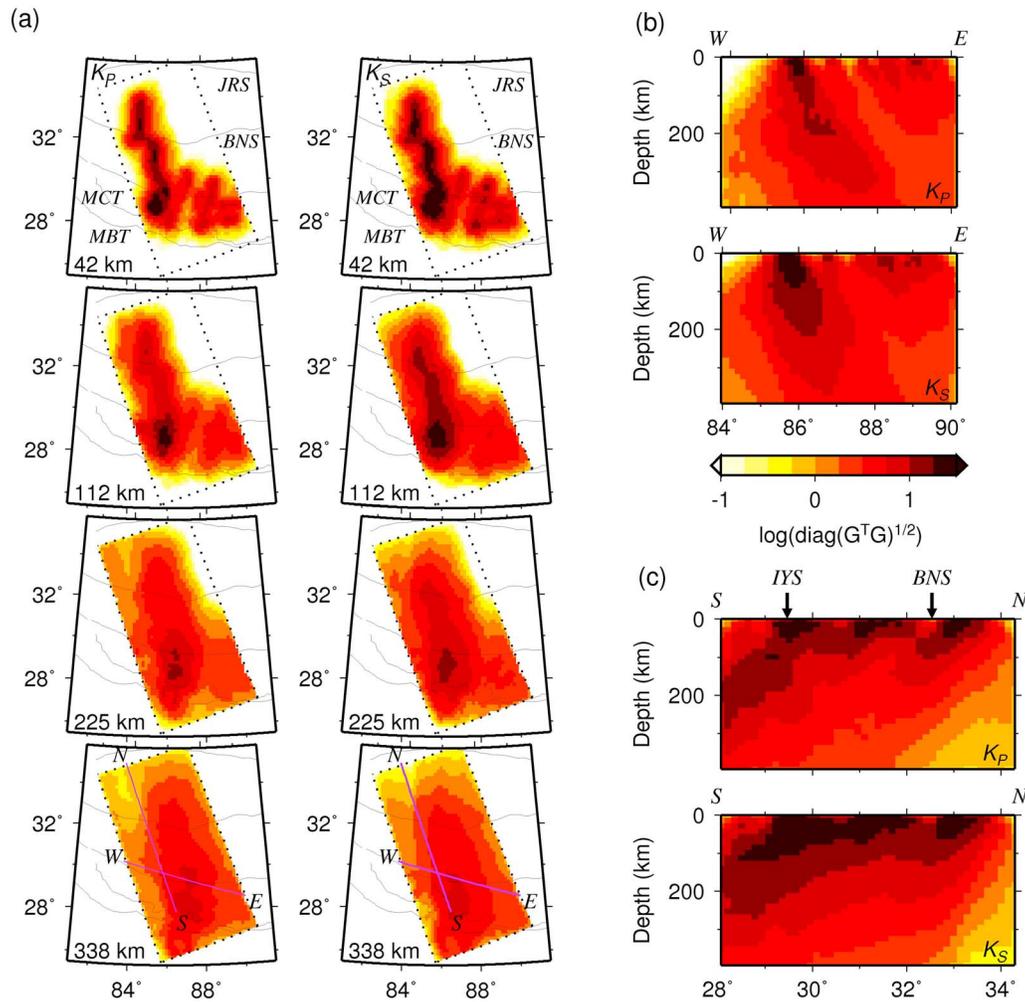


Figure A2. Square root values for diagonal elements of the matrix $\mathbf{G}^T \mathbf{G}$, where \mathbf{G} is the Gram matrix of the inverse problem (equation (2)). Each value corresponds to the overall sensitivity of combined high- and low-frequency traveltimes data to the variation in slowness at each node. The plots are presented in logarithmic scale, with high values indicating regions where constraints are particularly strong. Here we use only a subset of data: those with simultaneous recorded P and S waves in order to estimate $\delta \ln(V_P/V_S)$. The layouts are directly comparable to Figure 7.

vertexes of the four triangles, has six degrees of freedom (DOF's): three ($2^1 + 1$) in the long (N-S) and two ($2^0 + 1$) in the short (E-W) dimension, respectively (Figure A1a),

$$\tilde{\mathbf{c}}^1 = [\tilde{c}_{0,0}^1, \tilde{c}_{1,0}^1, \tilde{c}_{2,0}^1, \tilde{c}_{0,1}^1, \tilde{c}_{1,1}^1, \tilde{c}_{2,1}^1]^T.$$

Notice that in general, at a given scale level l , there are a total of 2^{2l} children or subtriangles with $(2^l + 1) \times (2^{l-1} + 1)$ DOF's for the nodes at a particular depth:

$$\tilde{\mathbf{c}}^l = [\tilde{c}_{0,0}^l, \tilde{c}_{1,0}^l, \dots, \tilde{c}_{2^l-1,0}^l, \tilde{c}_{2^l,0}^l, \dots, \tilde{c}_{0,2^l-1}^l, \tilde{c}_{1,2^l-1}^l, \dots, \tilde{c}_{2^l-1,2^l-1}^l, \tilde{c}_{2^l,2^l-1}^l]^T.$$

For the next level of refinement, $l = 2$, we connect the mid-points on each side of the four triangular meshes to yield 16 ($= 2^4$) children triangles with a total of 15 ($(2^2 + 1) \times (2^1 + 1)$) DOF's (vertexes). The wavelet coefficients, i.e., components of the transformed model vector $\tilde{\mathbf{m}}$ are closely

related to $\tilde{\mathbf{c}}$ at different levels. The determination of $\tilde{\mathbf{c}}$ is tied to the lifting scheme we use for both the forward and the inverse wavelet transforms. In the next section, we will demonstrate how the lifting scheme connects $\tilde{\mathbf{c}}$ to slowness perturbations in ordinary space at different scale lengths as shown in Figure 6.

[69] At any fixed horizontal position, the slowness perturbations across 2^l layers or segments (connecting $2^l + 1$ nodes) in the vertical direction are also represented by hierarchical variations that begin with a linear trend across each of two segments of equal length on the longest length scale (whose value is approximately equal to that in the E-W direction) with just three DOF's (Figure 3). With the addition of two DOF's to express the needed details at the next scale level, we end up with four segments of equal length, each at half of the previous length. As the refinement goes on with additional 2^l DOF's for the $(l + 1)$ th scale level, the

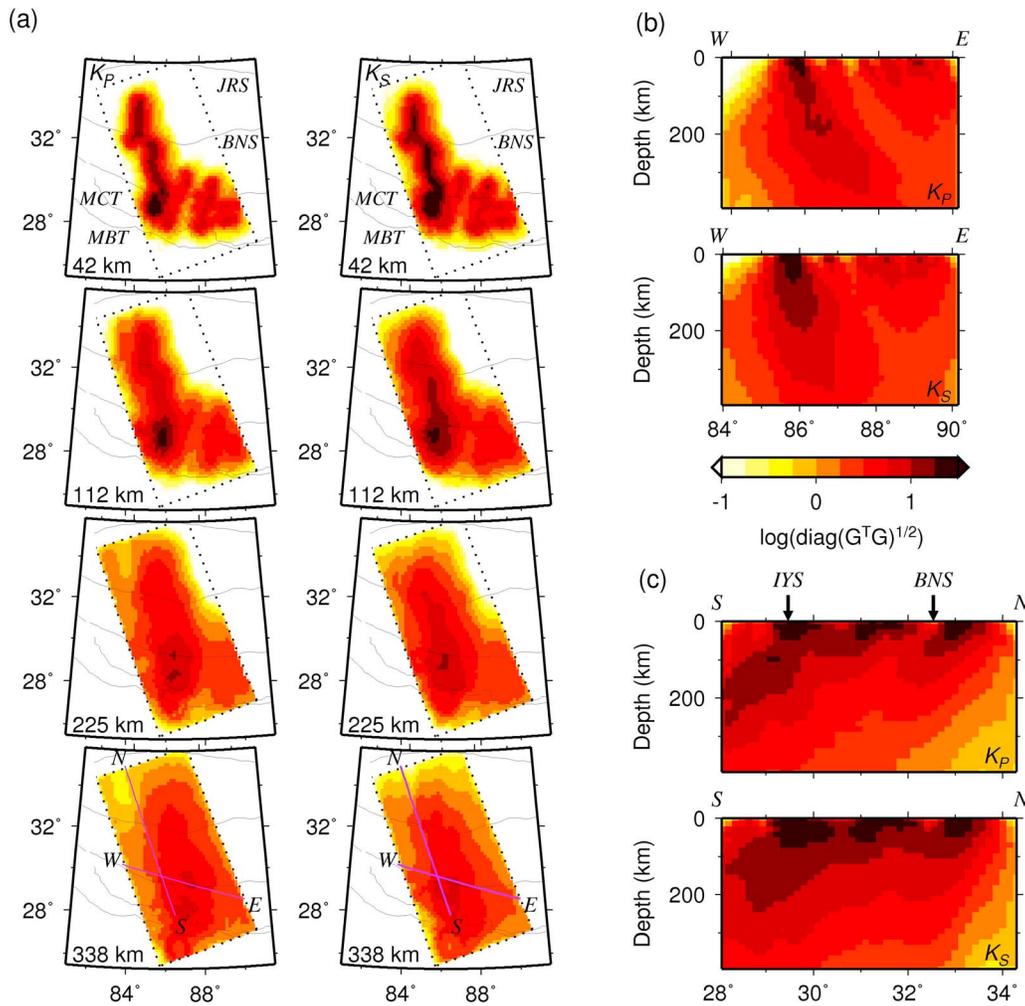


Figure A3. Same plots as in Figure A1, except that we include all available data for P and S waves. Figure A3 supersedes Figure S4 of Hung *et al.* [2010].

2^L segments are grouped into L scale levels and the total DOF's, $2^L + 1 = 3 + \sum_{l=2}^L 2^{l-1}$, remain unchanged.

A3. Lifting Scheme for Wavelet Transform

[70] For fast and straightforward implementation of both 2-D and 1-D wavelet transforms, we use the hierarchical streaming lifting scheme first developed by Sweldens [1996]. This scheme has complementary low-pass and high-pass wavelet filters that capture both smooth and detailed variations, respectively; and is well suited for our purpose. Unlike classical wavelet transforms, the lifting scheme can easily transform discrete, irregularly sampled data or model parameters in curvilinear coordinate systems. In particular, because all operations are directly carried out in the spatial domain, there is no need to make ad hoc assumptions such as periodicity for a data set of finite length, or bounded model space as required in Fourier-based wavelet transforms. While our main objective, as described in equation (6), involves the inverse wavelet transform, we take a heuristic approach and shall begin with a general discussion of the lifting scheme for the forward transform.

[71] The forward transform employs a hierarchical, tree-like algorithm which iteratively halves the number of variables at each reduced level until reaching the final, root level. Each iteration involves three steps and we shall illustrate this procedure by the 1-D wavelet transform of slowness perturbations specified at grid nodes in the vertical direction. To this end, take a set of arbitrary model parameters as the input. First, split the discrete model parameters into two subsets according to even and odd index of nodes in the original model grids. In the case of Figure A1b, there are a total of 17 nodes, indexed from 0 to 16 (in subscript) at the highest scale level (i.e., 4).

[72] The second step is to calculate expected values of the model parameters in the odd subset based on a local criterion defined by the average of values at adjacent nodes in the even subset. Differences between the original values at odd index nodes and corresponding predictions from averaging values at adjacent even index nodes represent details of the original input at the smallest scale. This procedure is then carried out at each successive level until the root level is reached. So in general the resulting difference has the following form:

$$\tilde{d}_{2j+1}^l = c_{2j+1}^{l+1} - \frac{1}{2} (c_{2j}^{l+1} + c_{2j+2}^{l+1}), \quad (\text{A7})$$

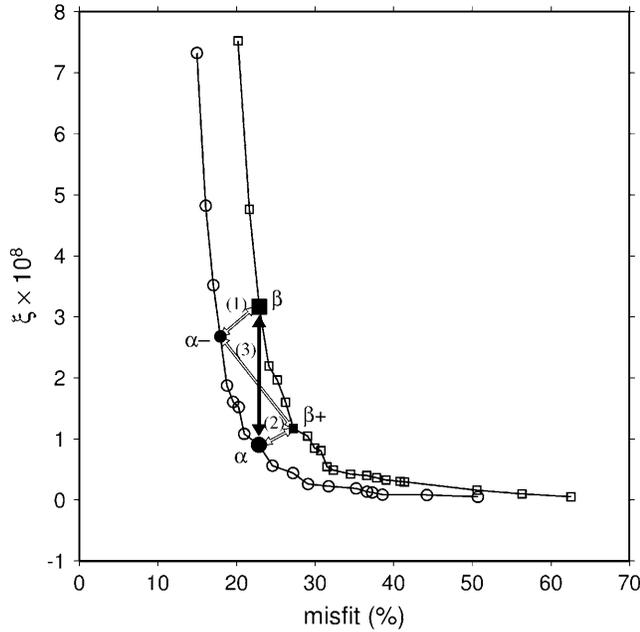


Figure A4. A plot of how the derivative of model variance with respect to reduction in the variance of data misfit (i.e., 1– data misfit), ξ , or the slopes of a curve in Figure 4 after a sign change, varies as a function of data misfit. Here α and β mark preferred solutions for $\delta \ln V_P$ and $\delta \ln V_S$, respectively, and α^- and β^+ indicate alternative choices of an underdamped result for $\delta \ln V_P$ and an overdamped result for $\delta \ln V_S$, respectively. Pairing these four preferred or trial solutions provides a test for the reliability of estimated $\delta \ln(V_P/V_S)$. The best pairing is between the preferred solutions for both $\delta \ln V_P$ and $\delta \ln V_S$ as shown in Figure 7. Results for the other three possible pairings (marked as (1), (2), and (3)) are shown in Figures A5–A7.

where the superscript, $l = 3, 2, 1$ indicates the scale level, and the subscript, $j = 0, \dots, 2^l - 1$, represents the node index in the vertical dimension at a given horizontal position. The resulting difference, \tilde{d}_{2j+1}^l , calculated from values specified at level $l + 1$, is now assigned as the $(j + 1)$ th wavelet coefficient at the reduced level l .

[73] Third, update the values at the even index nodes to maintain certain global properties, such as preserving the mean of the original parameters at all scale levels:

$$\tilde{c}_j^l = \tilde{c}_{2j}^{l+1} + \frac{1}{4}(\tilde{d}_{2j-1}^l + \tilde{d}_{2j+1}^l). \quad (\text{A8})$$

(The fact that the mean of model parameters remains unchanged is readily seen by plugging equations (A7)–(A8)

and gathering even and odd terms of \tilde{c}_j^{l+1} , leading to $\sum_{j=0}^{2^{l+1}-1} \tilde{c}_j^{l+1} = \sum_{j=0}^{2^l-1} \tilde{c}_j^l$.) It is important to note that except for the first iteration (going from $l = 4$ to 3), where variables \tilde{c}_{2j+1}^{l+1} , \tilde{c}_{2j}^{l+1} , and \tilde{c}_{2j+1}^{l+1} on the left-hand side of equation (A7) are the same as the original parameters at grid nodes (in ordinary space), all three variables are updated values at even index nodes calculated from the previous iteration using equation (A8).

As such, all variables are turned into wavelet coefficients after applying equation (A8).

[74] Notice that the subset of coefficients with odd indices, calculated from equation (A7), represent detailed variations of the input, while those in subset with even indices correspond to smooth variations. The latter also form a new sequence of values which serves as the initial values to calculate the wavelet coefficients at the next, coarser scale. After applying the 1-D wavelet transform in the vertical direction through this lifting scheme, the resulting wavelet coefficients retain the same DOF's as the original parameters in ordinary space. However, the wavelet coefficients now explicitly express the longest wavelength variations in a subset with both even and odd indices at the root level, and detailed variations as subsets with only odd indices at all leaf levels:

$$\mathbf{W}_{1D}([c_0, c_1, \dots, c_{15}, c_{16}]^T) = [\tilde{c}_0^1, \tilde{d}_1^3, \tilde{d}_1^2, \tilde{d}_3^3, \tilde{d}_1^1, \tilde{d}_5^3, \tilde{d}_3^2, \tilde{d}_7^3, \tilde{c}_1^1, \tilde{d}_9^3, \tilde{d}_5^2, \tilde{d}_{11}^3, \tilde{d}_3^1, \tilde{d}_{13}^3, \tilde{d}_7^2, \tilde{d}_{15}^3, \tilde{c}_2^1]^T.$$

The inverse transform simply reverses both the order and the sign of operations (see equations (A8) and (A7)):

$$\tilde{c}_{2j}^{l+1} = \tilde{c}_j^l - \frac{1}{4}(\tilde{d}_{2j-1}^l + \tilde{d}_{2j+1}^l), \quad (\text{A9})$$

$$\tilde{d}_{2j+1}^{l+1} = \tilde{d}_{2j+1}^l + \frac{1}{2}(\tilde{c}_{2j}^{l+1} + \tilde{c}_{2j+2}^{l+1}). \quad (\text{A10})$$

This lifting scheme, involving only linear interpolations, is equivalent to the forward and inverse transforms using the CDF (2,2) wavelet (named after Cohen-Daubechies-Feauveau).

[75] As mentioned in section A2, at a given depth, the 2-D multiscale parameterization is built up through hierarchical refinements of triangular meshes. Vertices in the mesh at the final, finest level of refinement coincide with grid nodes of the model for which a solution is sought. Like the 1-D case, applying the 2-D forward transform begins at the top (finest level) where we divide the parameters into subsets with even and odd indices. The former comprises those with even indices in both dimensions, which would coincide with vertices of quadruple-sized triangular meshes at the next, coarser level. The latter includes the remaining vertices with at least one odd index, which are at the midpoints of each side of the enlarged triangles at the next level.

[76] Next, we replace the values at the vertices in the odd subset through a step analogous to equation (A7) in the 1-D transform:

$$\begin{aligned} \tilde{d}_{2p+1,2q}^l &= \tilde{c}_{2p+1,2q}^{l+1} - \frac{1}{2}(\tilde{c}_{2p,2q}^{l+1} + \tilde{c}_{2p+2,2q}^{l+1}), \\ \tilde{d}_{2p,2q+1}^l &= \tilde{c}_{2p,2q+1}^{l+1} - \frac{1}{2}(\tilde{c}_{2p,2q}^{l+1} + \tilde{c}_{2p,2q+2}^{l+1}), \\ \tilde{d}_{2p+1,2q+1}^l &= \tilde{c}_{2p+1,2q+1}^{l+1} - \frac{1}{2}(\tilde{c}_{2p,2q}^{l+1} + \tilde{c}_{2p+2,2q+2}^{l+1}). \end{aligned} \quad (\text{A11})$$

where $p = 0, \dots, 2^{l-1}$, $q = 0, \dots, 2^{l-2}$, and $\tilde{d}_{2p+1,2q}^l$, $\tilde{d}_{2p,2q+1}^l$, and $\tilde{d}_{2p+1,2q+1}^l$ form wavelet coefficients at level l , being specified at the midpoint of each side of the triangular meshes parallel to the N-S, E-W and diagonal directions, respectively. Last, the vertices in the even subset, located

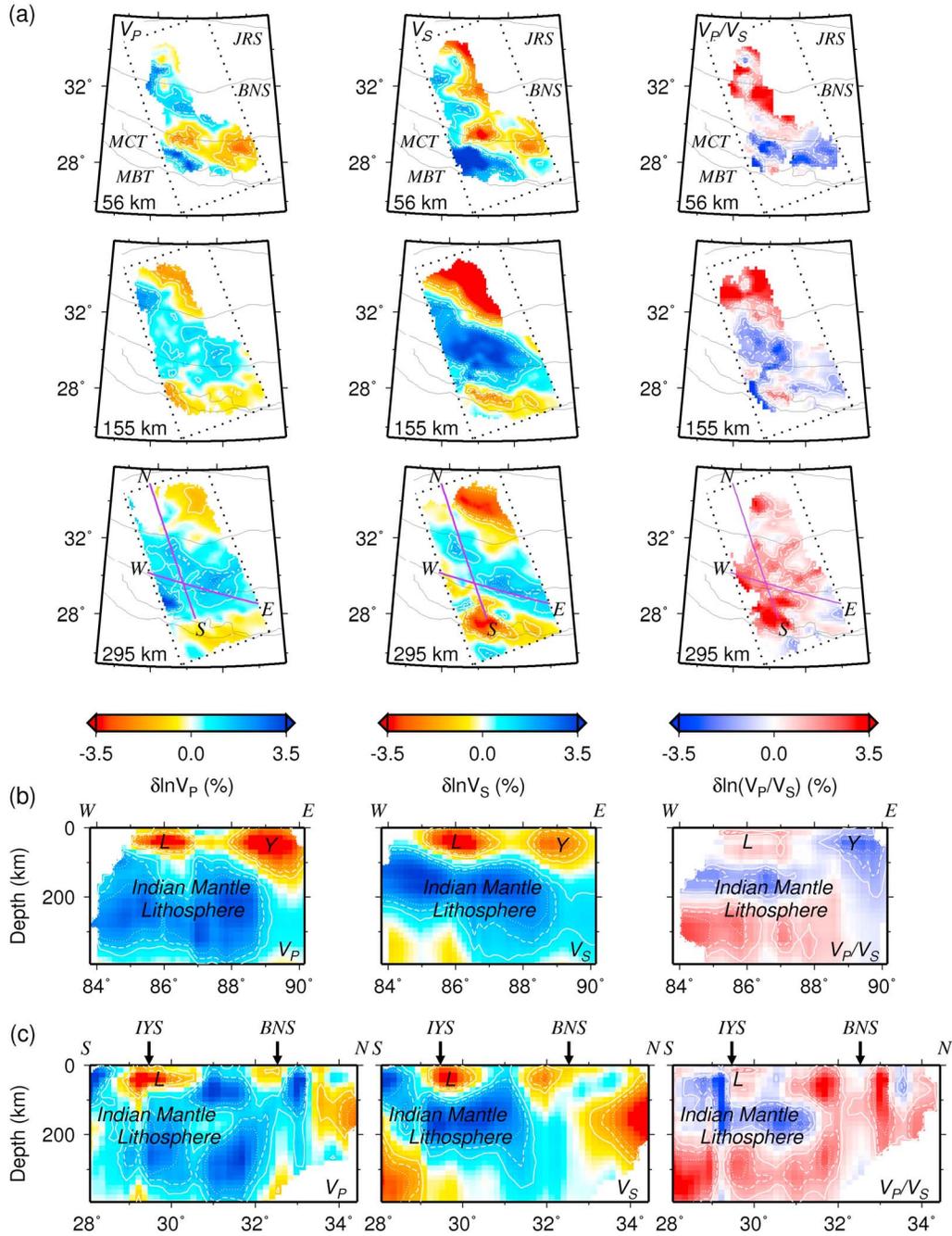


Figure A5. Resulting images for pairing 1: an underdamped solution for $\delta \ln V_P$ (α -) and the preferred solution for $\delta \ln V_S$ (β). There are some minor differences when compared with Figure 7, but only in short-wavelength features. This is an expected result in that the solution for $\delta \ln V_P$ is underdamped.

on three corresponding sides of the meshes are respectively updated as follows (see equation (A8)):

$$\begin{aligned} \tilde{c}_{p,q}^l &= \tilde{c}_{2p,2q}^{l+1} + \frac{1}{4} (\tilde{d}_{2p-1,2q}^l + \tilde{d}_{2p+1,2q}^l), \\ \tilde{c}_{p,q}^l &= \tilde{c}_{2p,2q}^{l+1} + \frac{1}{4} (\tilde{d}_{2p,2q-1}^l + \tilde{d}_{2p,2q+1}^l), \\ \tilde{c}_{p,q}^l &= \tilde{c}_{2p,2q}^{l+1} + \frac{1}{4} (\tilde{d}_{2p-1,2q-1}^l + \tilde{d}_{2p+1,2q+1}^l). \end{aligned} \quad (\text{A12})$$

Similar to the 1-D case (equations (A9) and (A10)), the inverse wavelet transform in 2-D is carried out by reversing the above lifting scheme, beginning at the root level ($l = 1$) where there are six DOF's in $\tilde{\mathbf{c}}^1$, whose components are associated with the vertexes of the four largest triangular meshes (Figure A1a). The vector $\tilde{\mathbf{c}}^{l+1}$, whose components are wavelet coefficients specified at the vertexes of subdivided triangular meshes at each finer level, is recursively built up by reversing the order and sign of the arithmetic operations of $\tilde{\mathbf{c}}^l$ and the wavelet coefficients $\tilde{d}_{p,q}^l$ at the

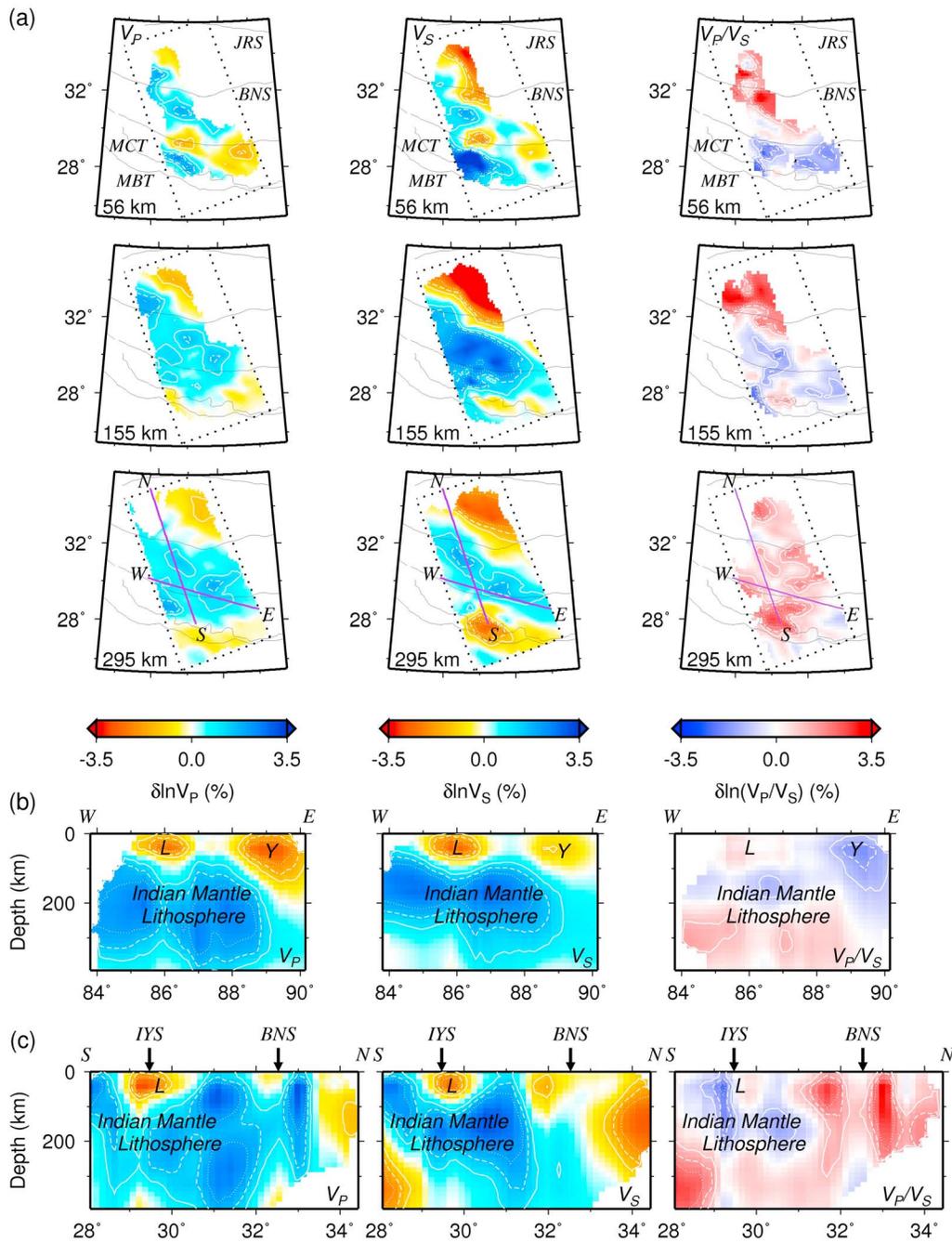


Figure A6. Resulting images for pairing 2: the preferred solution for $\delta \ln V_P$ (α) and the an overdamped solution for $\delta \ln V_S$ ($\beta+$). Overall characteristics of the images are similar to those shown in Figure 7, but appear at smaller amplitudes because the solution for $\delta \ln V_S$ is overdamped.

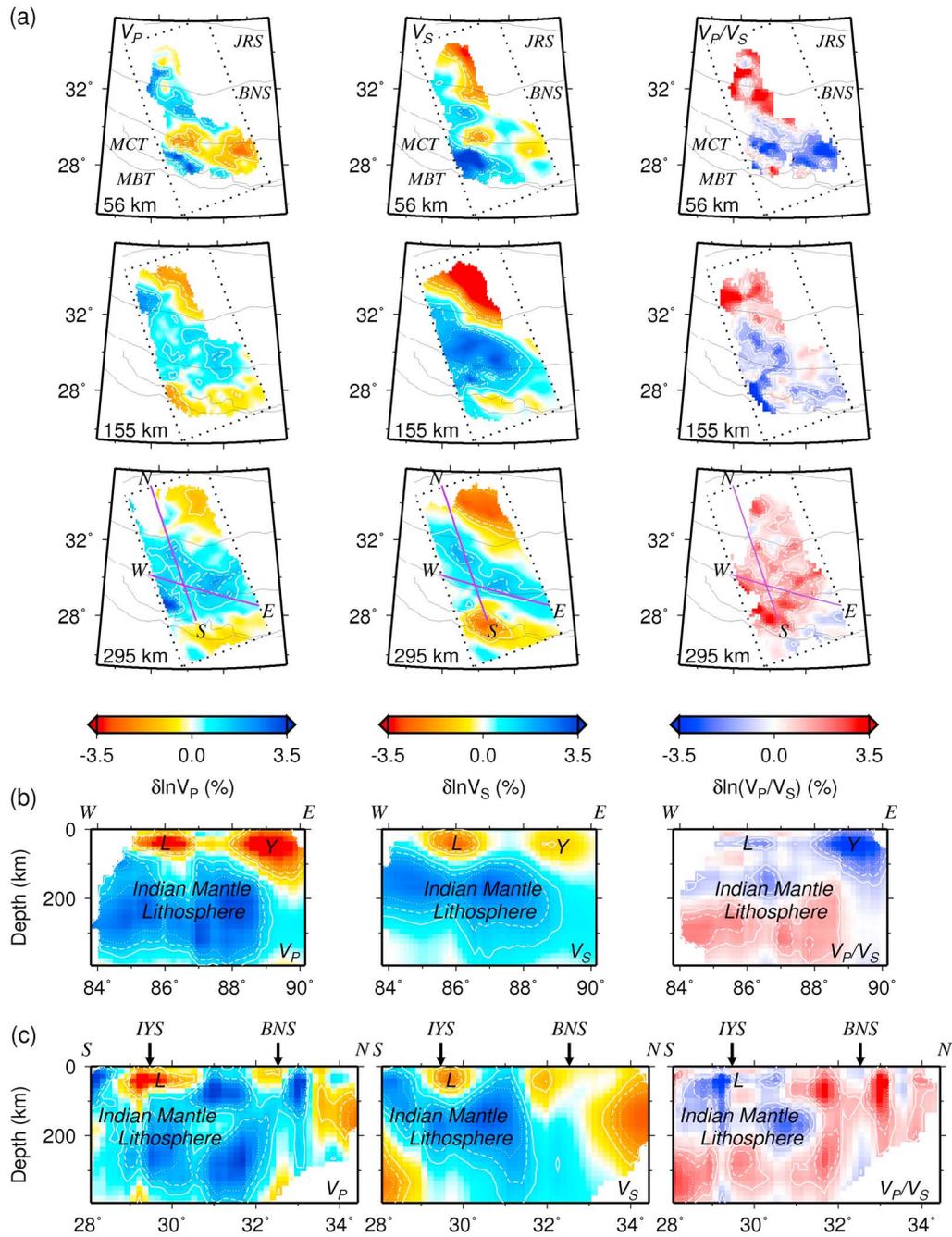


Figure A7. Resulting images for pairing 3: an underdamped solution for $\delta \ln V_P$ (α^-) and an overdamped solution for $\delta \ln V_S$ (β^+). This is the worst case scenario and not surprisingly, the images for $\delta \ln(V_P/V_S)$ becomes difficult to interpret.

current level l . (For our application at hand, the wavelet coefficients result from solving the inversion problem according to equation (7).)

[77] Specifically, at level 2, we have the following expressions:

$$\begin{aligned}\tilde{c}_{0,0}^2 &= \tilde{c}_{0,0}^1 - \frac{\tilde{d}_{1,1}^1}{4}, \tilde{c}_{2,2}^2 = \tilde{c}_{1,1}^1 - \frac{\tilde{d}_{1,1}^1}{4}; \\ \tilde{c}_{0,0}^2 &= \tilde{c}_{0,0}^1 - \frac{\tilde{d}_{0,1}^1}{4}, \tilde{c}_{0,2}^2 = \tilde{c}_{0,1}^1 - \frac{\tilde{d}_{0,1}^1}{4}; \\ \tilde{c}_{0,0}^2 &= \tilde{c}_{0,0}^1 - \frac{\tilde{d}_{1,0}^1}{4}, \tilde{c}_{2,0}^2 = \tilde{c}_{1,0}^1 - \frac{\tilde{d}_{1,0}^1 + \tilde{d}_{3,0}^1}{4}, \tilde{c}_{4,0}^2 = \tilde{c}_{2,0}^1 - \frac{\tilde{d}_{3,0}^1}{4}; \\ \tilde{c}_{1,1}^2 &= \tilde{d}_{1,1}^1 + \frac{\tilde{c}_{0,0}^2 + \tilde{c}_{2,2}^2}{2}, \tilde{c}_{3,1}^2 = \tilde{d}_{3,1}^1 + \frac{\tilde{c}_{2,0}^2 + \tilde{c}_{4,2}^2}{2}; \\ \tilde{c}_{0,1}^2 &= \tilde{d}_{0,1}^1 + \frac{\tilde{c}_{0,0}^2 + \tilde{c}_{2,2}^2}{2}, \tilde{c}_{2,1}^2 = \tilde{d}_{2,1}^1 + \frac{\tilde{c}_{2,0}^2 + \tilde{c}_{4,2}^2}{2}, \tilde{c}_{4,1}^2 = \tilde{d}_{4,1}^1 + \frac{\tilde{c}_{0,0}^2 + \tilde{c}_{4,2}^2}{2}; \\ \tilde{c}_{1,0}^2 &= \tilde{d}_{1,0}^1 + \frac{\tilde{c}_{0,0}^2 + \tilde{c}_{2,2}^2}{2}, \tilde{c}_{3,0}^2 = \tilde{d}_{3,0}^1 + \frac{\tilde{c}_{2,0}^2 + \tilde{c}_{4,2}^2}{2}, \tilde{c}_{1,2}^2 = \tilde{d}_{1,2}^1 + \frac{\tilde{c}_{0,2}^2 + \tilde{c}_{2,2}^2}{2}, \\ \tilde{c}_{3,2}^2 &= \tilde{d}_{3,2}^1 + \frac{\tilde{c}_{2,0}^2 + \tilde{c}_{4,2}^2}{2}.\end{aligned}$$

In actual practice, first we calculate $(\mathbf{W}^{-1})^T \mathbf{g}_i$ which involves the inverse waveform transform \mathbf{W}^{-1} and the data kernel \mathbf{g}_i (equation (6)). We then estimate \mathbf{m} , via $\tilde{\mathbf{m}}$, through equation (7).

A4. Data Coverage and Effects of Norm Damping

[78] Figure A2 shows the square root values for diagonal elements of the matrix $\mathbf{G}^T \mathbf{G}$, a measure of the overall sensitivity of combined high- and low-frequency traveltime data from simultaneously recorded P and S waves. This data set lead to results shown in Figure 7 in which images of $\delta \ln(V_P/V_S)$ are shown in addition to those of $\delta \ln V_P$ and $\delta \ln V_S$. In Figure A3, based on the entire data sets for both P and S waves, we present the corresponding plots to those in Figure A2, applying exactly the same input values as those used by *Hung et al.* [2010] where an erroneous set of plots were inadvertently displayed. Save for a few minor details, Figures A2 and A3 are similar overall. In a very general sense, Figures A2 and A3 are analogous to plots of raypaths when only geometrical rays are used in tomographic inversion.

[79] Next we investigate how perturbations in norm damping affect $\delta \ln V_P$ and $\delta \ln V_S$ which, in turn, influence results for $\delta \ln(V_P/V_S)$. As illustrated in Figure 4, we choose the preferred values of damping around the point of maximum curvature along tradeoff curves between model variance and data misfit. To the left of preferred values in Figure 4, model variance rises precipitously without much improvement in data misfit. Figure A4 emphasizes this point in terms of the derivatives of model variance with respect to reduction in the variance of data misfit (i.e., 1- data misfit), ξ , or the slopes of curves in Figure 4 after a sign change, as a function of data misfit.

[80] On face value, the preferred model for $\delta \ln V_S$ appears underdamped relative to that for $\delta \ln V_P$ (labeled as β and α , respectively, in Figures 4 and A4). However, because the Gram matrices of P and S wave data are different, direct comparison of numerical values of damping offers no insight into these solutions. Theoretically [*Lanczos*, 1961; *Backus and Gilbert*, 1967], if singular values of the Gram

matrix, \mathbf{G} , are known, the solution for \mathbf{m} from generalized inverse is

$$\hat{\mathbf{m}}_P = \mathbf{V}_P \left(\frac{\mathbf{\Lambda}_P}{\mathbf{\Lambda}_P^2 + \lambda \mathbf{I}} \right) \mathbf{U}_P^T \mathbf{d}. \quad (\text{A13})$$

where $\mathbf{\Lambda}_P$ is a $P \times P$ diagonal matrix containing a total of P nonzero singular values of the Gram matrix; \mathbf{U}_P is a $N \times P$ unitary matrix of singular vectors corresponding to nonzero singular values that span the data space; \mathbf{V}_P is a $M \times P$ unitary matrix of singular vectors that span the model space; \mathbf{I} is a $P \times P$ identity matrix; and λ is the damping factor. So the true effect of damping depends on the ratio between λ , which damps out small singular values, and the maximum singular value. In practice, a typical Gram matrix, on the order of about five giga elements in our case, is simply too large to lend itself to singular value decomposition.

[81] As a pragmatic solution, we investigate underdamped solutions of $\delta \ln V_P$ at values of ξ close to that of the optimal solution for $\delta \ln V_S$ (such as the one marked as “ α –” in Figure A4), and overdamped solutions of $\delta \ln V_S$ at values of ξ comparable to that of the optimal solution for $\delta \ln V_P$ (e.g., the one marked as “ β +” in Figure A4). When compared with the optimal solutions (Figure 7), the perturbed solutions yield nearly identical patterns in the overall distribution of $\delta \ln V_P$ and $\delta \ln V_S$, with only slight differences in the amplitude of anomalies (left panels Figures A5–A7).

[82] Figures 7 and A5–A7 also illustrate how different combinations of the four trial solutions affect $\delta \ln(V_P/V_S)$. Overall, results for $\delta \ln(V_P/V_S)$ remain stable. Noticeable changes occur only in the worst case scenario where non-optimal solutions for both $\delta \ln V_P$ and $\delta \ln V_S$ are combined to estimate $\delta \ln(V_P/V_S)$.

[83] **Acknowledgments.** We thank C.-T. A. Lee, M. Searle, and S. Marshak for stimulating discussions. We thank Editor R. Nowack and two anonymous referees for their comments that help improve the manuscript. This work was supported by National Science Council of Taiwan grants 97-2745-M-002-011, 98-2116-M-002-025, and 99-2116-M-002-028 (S.-H.H.) and U.S. National Science Foundation grants EAR99-09362 (“Hi-CLIMB”), EAR05-51995, and EAR06-35419 (W.-P.C.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF.

References

- Aki, K., and W. H. K. Lee (1976), Determination of three-dimensional anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model, *J. Geophys. Res.*, *81*, 4381–4399, doi:10.1029/JB081i023p04381.
- Backus, G., and F. Gilbert (1967), Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc.*, *13*, 247–276.
- Baig, A. M., F. A. Dahlen, and S.-H. Hung (2003), Traveltimes of waves in three-dimensional random media, *Geophys. J. Int.*, *153*, 467–482, doi:10.1046/j.1365-246X.2003.01905.x.
- Chen, W.-P., and P. Molnar (1983), Focal depths of intracontinental and intraplate earthquakes and their implications for the thermal and mechanical properties of the lithosphere, *J. Geophys. Res.*, *88*, 4183–4214, doi:10.1029/JB088iB05p04183.
- Chen, W.-P., and S. Özalaybey (1998), Correlation between seismic anisotropy and Bouguer gravity anomalies in Tibet and its implications for lithospheric structures, *Geophys. J. Int.*, *135*, 93–101, doi:10.1046/j.1365-246X.1998.00611.x.
- Chen, W.-P., and T.-L. Tseng (2007), Small 660-km seismic discontinuity beneath Tibet implies resting ground for detached lithosphere, *J. Geophys. Res.*, *112*, B05309, doi:10.1029/2006JB004607.

- Chen, W.-P., and Z. Yang (2004), Earthquakes beneath the Himalayas and Tibet: Evidence for strong lithospheric mantle, *Science*, *304*, 1949–1952, doi:10.1126/science.1097324.
- Chen, W.-P., J. L. Nábelek, T. J. Fitch, and P. Molnar (1981), An intermediate depth earthquake beneath Tibet: Source characteristics of the event of September 14, 1976, *J. Geophys. Res.*, *86*, 2863–2876.
- Chen, W.-P., M. Martin, T.-L. Tseng, R. L. Nowack, S.-H. Hung, and B.-S. Huang (2010), Shear-wave birefringence and current configuration of converging lithosphere under Tibet, *Earth Planet. Sci. Lett.*, *295*, 297–304, doi:10.1016/j.epsl.2010.04.017.
- Chiao, L.-Y., and B.-Y. Kuo (2001), Multiscale seismic tomography, *Geophys. J. Int.*, *145*, 517–527, doi:10.1046/j.0956-540x.2001.01403.x.
- Chiao, L.-Y., and W.-Z. Liang (2003), Multiresolution parameterization for geophysical inverse problems, *Geophysics*, *68*, 199–209, doi:10.1190/1.1543207.
- Chiao, L.-Y., J.-R. Lin, and Y.-C. Gung (2006), Crustal magnetization equivalent source model of Mars constructed from a hierarchical multi-resolution inversion of the Mars Global Surveyor data, *J. Geophys. Res.*, *111*, E12010, doi:10.1029/2006JE002725.
- Chiao, L.-Y., H.-Y. Fang, Y.-C. Gung, Y.-H. Chang, and S.-H. Hung (2010), Comparative appraisal of linear inverse models constructed via distinctive parameterizations, *J. Geophys. Res.*, *115*, B07305, doi:10.1029/2009JB006867.
- Chou, H.-C., B.-Y. Kuo, L.-Y. Chiao, D. Zhao, and S.-H. Hung (2009), Tomography of the westernmost Ryukyu subduction zone and the serpentinization of the fore-arc mantle, *J. Geophys. Res.*, *114*, B12301, doi:10.1029/2008JB006192.
- Chung, S.-L., M.-F. Chu, Y. Zhang, Y. Xie, C.-H. Lo, T.-Y. Lee, C.-Y. Lan, X. Li, Q. Zhang, and Y. Wang (2005), Tibetan tectonic evolution inferred from spatial and temporal variations in post-collisional magmatism, *Earth Sci. Rev.*, *68*(3–4), 173–196, doi:10.1016/j.earscirev.2004.05.001.
- Dahlen, F. A., S.-H. Hung, and G. Nolet (2000), Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, *141*, 157–174, doi:10.1046/j.1365-246X.2000.00070.x.
- DeCelles, P. G., G. E. Gehrels, J. Quade, T. P. Ojha, P. A. Kapp, and B. N. Upreti (1998), Neogene foreland basin deposits, erosional unroofing, and the kinematic history of the Himalayan fold-thrust belt, western Nepal, *Geol. Soc. Am. Bull.*, *110*, 2–21, doi:10.1130/0016-7606(1998)110<0002:NFBDEU>2.3.CO;2.
- Faul, U. H., and I. Jackson (2005), The seismological signature of temperature and grain size variations in the upper mantle, *Earth Planet. Sci. Lett.*, *234*, 119–134, doi:10.1016/j.epsl.2005.02.008.
- Gee, D. G., F. Haakon, N. Henriksen, and A. K. Higgins (2008), From the early Paleozoic platforms of Baltica and Laurentia to the Caledonide orogen of Scandinavia and Greenland, *Episodes*, *31*(1), 44–51.
- Grand, S. P., R. D. van der Hilst, and S. Widiyantoro (1997), Global seismic tomography: A snapshot of convection in the Earth, *GSA Today*, *7*(4), 1–7.
- Hammond, W. C., and E. D. Humphreys (2000), Upper mantle seismic wave velocity: Effects of realistic partial melt geometries, *J. Geophys. Res.*, *105*, 10,975–10,986, doi:10.1029/2000JB900041.
- Hung, S.-H., F. A. Dahlen, and G. Nolet (2000), Fréchet kernels for finite-frequency traveltimes—II. Examples, *Geophys. J. Int.*, *141*, 175–203, doi:10.1046/j.1365-246X.2000.00072.x.
- Hung, S.-H., F. A. Dahlen, and G. Nolet (2001), Wavefront healing: A banana-doughnut perspective, *Geophys. J. Int.*, *146*, 289–312, doi:10.1046/j.1365-246x.2001.01466.x.
- Hung, S.-H., Y. Shen, and L.-Y. Chiao (2004), Imaging seismic velocity structure beneath the Iceland hot spot: A finite frequency approach, *J. Geophys. Res.*, *109*, B08305, doi:10.1029/2003JB002889.
- Hung, S.-H., W.-P. Chen, L.-Y. Chiao, and T.-L. Tseng (2010), First multi-scale, finite-frequency tomography illuminates 3-D anatomy of the Tibetan plateau, *Geophys. Res. Lett.*, *37*, L06304, doi:10.1029/2009GL041875.
- Jackson, J. M., S. V. Sinogeikin, and J. D. Bass (2000), Sound velocities and elastic properties of γ -Mg₂SiO₄ to 873 K by Brillouin spectroscopy, *Am. Mineral.*, *85*(2), 296–303.
- Jiménez-Munt, I., M. Fernández, J. Vergés, and J. P. Platt (2008), Lithosphere structure underneath the Tibetan plateau inferred from elevation, gravity and geoid anomalies, *Earth Planet. Sci. Lett.*, *267*, 276–289, doi:10.1016/j.epsl.2007.11.045.
- Jordan, T. H. (1979), Mineralogies, densities and seismic velocities of garnet lherzolites and their geophysical implications, in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics*, edited by F. R. Boyd and H. O. A. Meyer, pp. 1–14, AGU, Washington, D. C.
- Jordan, T. H. (1988), Structure and formation of the continental tectosphere, *J. Petrol.*, *1988*, 11–37.
- Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, *122*, 108–124, doi:10.1111/j.1365-246X.1995.tb03540.x.
- Lanczos, C. (1961), *Linear Differential Operators*, Van Nostrand Reinhold, London.
- Lee, C. T. A. (2003), Compositional variation of density and seismic velocities in natural peridotites at STP conditions: Implications for seismic imaging of compositional heterogeneities in the upper mantle, *J. Geophys. Res.*, *108*(B9), 2441, doi:10.1029/2003JB002413.
- Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for *P* wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, *9*, Q05018, doi:10.1029/2007GC001806.
- Loris, I., G. Nolet, I. Daubechies, and F. A. Dahlen (2007), Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients, *Geophys. J. Int.*, *170*, 359–370, doi:10.1111/j.1365-246X.2007.03409.x.
- Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Elsevier, New York.
- Masters, G., S. Johnson, G. Laske, H. Bolton, and J. H. Davies (1996), A shear-velocity model of the mantle, *Philos. Trans. R. Soc. London A*, *354*, 1385–1411, doi:10.1098/rsta.1996.0054.
- McNamara, D. E., W. R. Walter, T. J. Owens, and C. J. Ammon (1997), Upper mantle velocity structure beneath the Tibetan plateau from *Pn* travel time tomography, *J. Geophys. Res.*, *102*, 493–505, doi:10.1029/96JB02112.
- Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif.
- Meyerholtz, K. A., G. L. Pavlis, and S. A. Szpakowski (1989), Convolutional quelling in seismic tomography, *Geophysics*, *54*, 570–580, doi:10.1190/1.1442684.
- Milnes, A. G. (1998), Alpine and Caledonide tectonics—A brief comparative study, *GFF*, *120*, 237–247, doi:10.1080/11035899801202237.
- Molnar, P., and W.-P. Chen (1984), *S-P* wave travel time residuals and lateral inhomogeneity in the mantle beneath Tibet and the Himalaya, *J. Geophys. Res.*, *89*, 6911–6917, doi:10.1029/JB089iB08p06911.
- Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung (2004a), Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343, doi:10.1126/science.1092485.
- Montelli, R., G. Nolet, G. Masters, F. A. Dahlen, and S.-H. Hung (2004b), Global *P* and *PP* traveltimes tomography: Rays versus waves, *Geophys. J. Int.*, *158*, 637–654, doi:10.1111/j.1365-246X.2004.02346.x.
- Ni, J., and M. Barazangi (1983), High-frequency seismic wave propagation beneath the Indian Shield, Himalayan Arc, Tibetan plateau and surrounding regions: High uppermost mantle velocities and efficient *Sn* propagation beneath Tibet, *Geophys. J. R. Astron. Soc.*, *72*, 665–689.
- Obrebski, M., R. M. Allen, M. Xue, and S.-H. Hung (2010), Slab-plume interaction beneath the Pacific Northwest, *Geophys. Res. Lett.*, *37*, L14305, doi:10.1029/2010GL043489.
- Owens, T. J., and G. Zandt (1997), Implications of crustal property variations for models of Tibetan plateau evolution, *Nature*, *387*, 37–43, doi:10.1038/387037a0.
- Parker, R. L. (1994), *Geophysical Inverse Theory*, 386 pp., Princeton Univ. Press, Princeton, N. J.
- Romanowicz, B. (2003), Global mantle tomography: Progress status in the past 10 years, *Annu. Rev. Earth Planet. Sci.*, *31*, 303–328, doi:10.1146/annurev.earth.31.091602.113555.
- Sambridge, M., and R. Faletič (2003), Adaptive whole Earth tomography, *Geochem. Geophys. Geosyst.*, *4*(3), 1022, doi:10.1029/2001GC000213.
- Searle, M. P., R. L. Simpson, R. D. Law, R. R. Parrish, and D. J. Waters (2003), The structural geometry, metamorphic and magmatic evolution of the Everest massif, High Himalaya of Nepal-South Tibet, *J. Geol. Soc.*, *160*, 345–366, doi:10.1144/0016-764902-126.
- Sigloch, K., N. McQuarrie, and G. Nolet (2008), Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.*, *1*(7), 458–462, doi:10.1038/ngeo231.
- Sinogeikin, S. V., J. D. Bass, and T. Katsura (2003), Single-crystal elasticity of ringwoodite to high pressures and high temperatures: Implications for 520 km seismic discontinuity, *Phys. Earth Planet. Inter.*, *136*(1–2), 41–66, doi:10.1016/S0031-9201(03)00022-0.
- Sweldens, W. (1996), The lifting scheme: A custom-design construction of biorthogonal wavelets, *Appl. Comput. Harmonic Anal.*, *3*, 186–200, doi:10.1006/acha.1996.0015.
- Taylor, M., and A. Yin (2009), Active structures of the Himalayan-Tibetan orogen and their relationships to earthquake distribution, contemporary strain field, and Cenozoic volcanism, *Geosphere*, *5*, 199–214, doi:10.1130/GES00217.1.
- Tian, Y., K. Sigloch, and G. Nolet (2009), Multiple-frequency *SH*-wave tomography of the western US upper mantle, *Geophys. J. Int.*, *178*, 1384–1402, doi:10.1111/j.1365-246X.2009.04225.x.
- Trampert, J. (1998), Global seismic tomography: The inverse problem and beyond, *Inverse Probl.*, *14*(3), 371, doi:10.1088/0266-5611/14/3/002.
- Trampert, J., and R. Snieder (1996), Model estimations biased by truncated expansions: Possible artifacts in seismic tomography, *Science*, *271*, 1257–1260, doi:10.1126/science.271.5253.1257.

- Tromp, J., C. Tape, and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216, doi:10.1111/j.1365-246X.2004.02453.x.
- Tseng, T.-L., and W.-P. Chen (2008), Discordant contrasts of *P*- and *S*-wave speeds across the 660-km discontinuity beneath Tibet: A case for hydrous remnant of sub-continental lithosphere, *Earth Planet. Sci. Lett.*, *268*, 450–462, doi:10.1016/j.epsl.2008.01.038.
- Tseng, T.-L., W.-P. Chen, and R. L. Nowack (2009), Northward thinning of Tibetan crust revealed by virtual seismic profiles, *Geophys. Res. Lett.*, *36*, L24304, doi:10.1029/2009GL040457.
- Unsworth, M. J., et al. (2005), Crustal rheology of the Himalaya and Southern Tibet inferred from magnetotelluric data, *Nature*, *438*, 78–81, doi:10.1038/nature04154.
- VanDecar, J. C., and R. S. Crosson (1990), Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seismol. Soc. Am.*, *80*, 150–169.
- Wagner, L. S., M. L. Anderson, J. M. Jackson, S. L. Beck, and G. Zandt (2008), Seismic evidence for orthopyroxene enrichment in the continental lithosphere, *Geology*, *36*, 935–938, doi:10.1130/G25108A.1.
- Wiens, D. A., and S. Stein (1983), Age dependence of oceanic intraplate seismicity and implications for lithospheric evolution, *J. Geophys. Res.*, *88*, 6455–6468, doi:10.1029/JB088iB08p06455.
- Wolfe, C. J., S. C. Solomon, G. Laske, J. A. Collins, R. S. Detrick, J. A. Orcutt, D. Bercovici, and E. H. Hauri (2009), Mantle shear-wave velocity structure beneath the Hawaiian hot spot, *Science*, *326*, 1388–1390, doi:10.1126/science.1180165.
- Yang, H.-Y., and S.-H. Hung (2005), Validation of ray and wave theoretical travel times in heterogeneous random media, *Geophys. Res. Lett.*, *32*, L20302, doi:10.1029/2005GL023501.
- Yang, T., Y. Shen, S. van der Lee, S. C. Solomon, and S.-H. Hung (2006), Upper mantle structure beneath the Azores hotspot from finite-frequency seismic tomography, *Earth Planet. Sci. Lett.*, *250*, 11–26, doi:10.1016/j.epsl.2006.07.031.
- Yin, A., and T. M. Harrison (2000), Geologic evolution of the Himalayan-Tibetan orogen, *Annu. Rev. Earth Planet. Sci.*, *28*, 211–280, doi:10.1146/annurev.earth.28.1.211.
- Zhang, H., and C. H. Thurber (2003), Double-difference tomography: The method and its application to the Hayward fault, California, *Bull. Seismol. Soc. Am.*, *93*, 1875–1889, doi:10.1785/0120020190.
- Zhang, P., et al. (2004), Continuous deformation of the Tibetan plateau from global positioning system data, *Geology*, *32*, 809–812, doi:10.1130/G20554.1.
- Zhao, D., and H. Kanamori (1993), The 1992 Landers earthquake sequence: Earthquake occurrence and structural heterogeneities, *Geophys. Res. Lett.*, *20*, 1083–1086, doi:10.1029/93GL01239.
- Zhao, L., T. H. Jordan, and C. H. Chapman (2000), Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, *141*, 558–576, doi:10.1046/j.1365-246x.2000.00085.x.
- Zhou, Y., F. A. Dahlen, G. Nolet, and G. Laske (2005), Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.*, *163*, 1087–1111, doi:10.1111/j.1365-246X.2005.02780.x.
- Zhu, L., and D. V. Helmberger (1998), Moho offset across the northern margin of the Tibetan plateau, *Science*, *281*, 1170–1172, doi:10.1126/science.281.5380.1170.
- Zienkiewicz, O. C., and R. L. Taylor (1989), *The Finite Element Method: Basic Formulation and Linear Problems*, vol. 1, 4th ed., McGraw-Hill, New York.
- W.-P. Chen, Department of Geology, University of Illinois-Urbana, Urbana, IL 61801, USA.
L.-Y. Chiao, Institute of Oceanography, National Taiwan University, Taipei, 10617, Taiwan.
S.-H. Hung, Department of Geosciences, National Taiwan University, Taipei, 10617, Taiwan. (shung@ntu.edu.tw)