

Validated 3D Velocity Models in Asia Using Joint Regional Body- and Surface-Wave Tomography

**Delaine T. Reiter
William L. Rodi**

**Weston Geophysical Corporation
181 Bedford Street, Suite 1
Lexington, MA 02420**

Final Report

17 February 2009

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.



**AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Rd
AIR FORCE MATERIEL COMMAND
HANSCOM AFB, MA 01731-3010**

NOTICES

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release and is available to the general public, including foreign nationals. Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>). All others should apply to the National Technical Information Service.

AFRL-RV-HA-TR-2009-1009 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

ROBERT J. RAISTRICK
Contract Manager

//Signature//

PAUL TRACY, Acting Chief
Battlespace Surveillance Innovation Center

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 17-02-2009		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 21-Jul-04 to 31-Dec-08	
4. TITLE AND SUBTITLE Validated 3D Velocity Models in Asia from Joint Regional Body- and Surface-Wave Tomography				5a. CONTRACT NUMBER FA8718-04-C-0027	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 1010	
6. AUTHOR(S) Delaine T. Reiter and William L. Rodi				5d. PROJECT NUMBER SM	
				5e. TASK NUMBER A1	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Weston Geophysical Corporation 181 Bedford St., Suite 101 Lexington, MA 02420				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory AFRL/RVBYE 29 Randolph Rd. Hanscom AFB, MA 01731				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RV-HA-TR-2009-1009	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT We describe a method to jointly invert body-wave travel times and surface wave dispersion data for a model of compressional and shear velocities in the crust and upper mantle. The nonlinearity in the problem is handled by iterating over linearized inversion steps, with the aid of finite-difference ray tracing techniques to perform the necessary forward modeling in the updated Earth model. At each step of the iteration, we perform linearized body-wave and surface-wave inversion as separate procedures. The body-wave data, comprising first-arrival P-wave times, are used to update the P velocity model. The surface-wave dispersion data are used to update the S-wave model in a two-step procedure, ignoring the small dependence of Rayleigh wave dispersion on P velocity. The separate inversions are coupled through two forms of prior information applied to their respective velocity models: constraints on the size and spatial smoothness of the velocity perturbations and bounds on the velocities themselves that are determined in part from bounds on Poisson's ratio. The new method has been applied on a 1 x 1 degree grid covering a large region of Asia, using first arriving P wave travel times from 0 to 18 degrees epicentral distance and fundamental mode Rayleigh group velocities at periods from 10 to 150 seconds. We validate the new model using standard techniques with a high quality set of travel-time observations from well located explosions and earthquakes.					
15. SUBJECT TERMS Regional event location, Joint P and S velocity inversion, Crust and upper mantle structure					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 55	19a. NAME OF RESPONSIBLE PERSON Robert Raistrick
a. REPORT UNC	b. ABSTRACT UNC	c. THIS PAGE UNC			19b. TELEPHONE NUMBER (include area code) 781-377-3726

Table of Contents

1. INTRODUCTION	1
2. MODELING AND INVERSION METHODS.....	3
2.1 Model Parameterization	3
2.2 Forward Modeling.....	3
2.3 Joint Inversion Approach	4
2.4 Factoring the Surface-Wave Inverse Problem	5
2.5 Linearization.....	6
2.6 Updating Procedure.....	7
2.7 Bayesian Inversion Method.....	8
2.8 Crustal Thickness Inversion.....	10
2.9 Factored vs. Unfactored Surface-Wave Inversion	10
2.10 Velocity Correlation.....	11
2.11 Parameter Bounds	11
3. APPLICATION TO DATA.....	14
3.1 Travel-Time Observations.....	15
3.2 Group-Velocity Dispersion Measurements.....	15
4. RESULTS OF JOINT INVERSION	19
4.1 Velocity Heterogeneities	19
4.2 Resolution Tests	22
4.3 Model Correlations.....	23
4.3 Data Variance Reduction	25
5. DISCUSSION AND CONCLUSIONS	26
5.1 Tests for Predictive Capability.....	26
5.2 Model Uncertainty.....	30
5.3 Summary	32
6. ACKNOWLEDGMENTS	33
7. REFERENCES	35
8. APPENDIX A: Factored Inversion.....	39
9. APPENDIX B: Conditional Inversion.....	41
10. APPENDIX C: Additional Inversion Results.....	43

1. INTRODUCTION

A variety of new three-dimensional models of the Earth's seismic velocity structure have been generated in the last decade, using more sophisticated numerical techniques and an ever increasing amount of high quality data. Some of these velocity models are designed to illuminate the dynamic structure of the Earth's internal mechanics, while others are primarily used for their predictive qualities in various seismic wavefield phenomena. For example, accurate travel-time predictions for regional seismic phases are essential for locating seismic events with the accuracy needed for nuclear monitoring decisions. The events of interest to monitoring seismologists are those at lower magnitudes, which will more than likely only be recorded on a sparse network of stations at regional distances less than approximately 20° . Under these conditions, systematic biases caused by inadequately modeled Earth structures cause errors in the estimation of travel times and amplitudes of seismic phases. More accurate and reliable estimates of these quantities (especially in aseismic regions) have great potential to improve nuclear monitoring efforts to detect, locate and discriminate regional events. Travel times calculated through a three-dimensional (3D) Earth model have the best chance of achieving acceptable prediction errors, if the model is constrained by sufficient data. Previous work has shown that incorporating predictions from 3D models significantly improves the location of regional events. For instance, Ritzwoller *et al.* (2003) demonstrated that a shear-wave model developed from broadband surface-wave measurements could be used to improve the predictions of Pn and P travel times, and subsequent regional event locations, using a thermoelastic conversion from shear to compressional velocities. Flanagan *et al.* (2007) also demonstrated regional location improvement using an *a priori* model of Western Eurasia and North Africa (WENA 1.0; Pasyanos *et al.*, 2004).

The difficulty of converting an S velocity model to a P velocity model, or vice versa, in order to get a complete seismic velocity model, has been a persistent problem that has hampered previous efforts to develop regional crust and upper mantle models. In some instances, empirical scaling relations (e.g. Poisson's ratio) are assumed for the crust and upper mantle, and the S model is then converted to a P model. However, in regions with sedimentary basins or anomalous mantle structures, the assumed relationship may be wrong, leading to errors in the velocity model derived from the conversion. In other cases, additional geophysical data sets, such as receiver functions or gravity, can be used to complement the results derived from surface waves (e.g. Ammon *et al.*, 1990; Julia *et al.*, 2000). A significant drawback for these efforts can be the discrepancy between the resolution or spatial coverage of the various data sets. At the global teleseismic scale, several researchers have utilized body-wave travel times and surface-wave dispersion measurements (among other data) jointly to produce P and S models of the Earth's mantle (e.g., Su and Dziewonski, 1997; Masters *et al.*, 2000; Antolik *et al.*, 2003). However, a joint inversion approach for detailed regional crust and upper-mantle P and S velocity models has not yet been attempted.

With this motivation, we have developed a self-consistent 3D P and S velocity model of the crust and upper mantle in a large region of central and southern Asia. The new model is the

result of a nonlinear, joint body-wave/surface-wave inversion method applied to Pn travel times and Rayleigh group-velocity measurements. The body- and surface-wave data used in the inversion were specifically chosen for their complementary sensitivity to the velocity structures at crust and upper mantle depths. Consistency between the P and S velocities is achieved by imposing bounds on Poisson's ratio and by invoking a regularization constraint that correlates variations in P and S velocity from an initial model. We report on the application of the new inversion technique to data observed in the broad region shown in Figure 1. This region, extending from 0-60° N and 30-120° E, covers some of the most tectonically complex areas on Earth. The inversion model extends from the surface to approximately 300 km beneath the surface.

To evaluate the usefulness of the new model for locating small regional events we have performed several validation exercises using a set of earthquakes and explosions with well-known epicentral locations. These tests include travel-time predictions and relocations of ground-truth events in the study area using both P and S regional phase arrivals. Our validation results indicate that in many cases our 3D model achieves excellent travel-time prediction and epicentral accuracy.

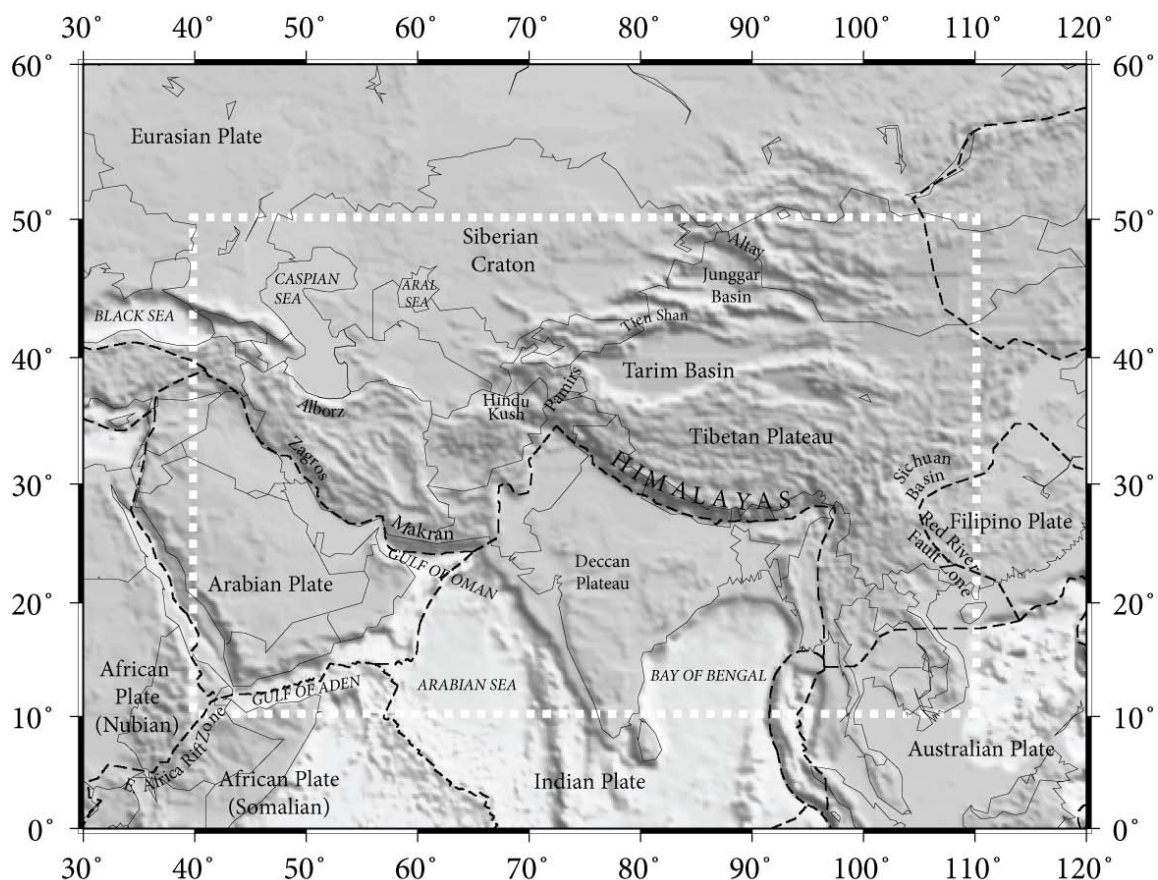


Figure 1: A topographic map of the study region, which encompasses most of central and southern Asia and portions of the Middle East. Dashed black lines indicate major plate boundaries, and labels denote some major tectonic features mentioned in the text. The thick white dashed line outlines the area covered by the new inversion model.

2. MODELING AND INVERSION METHODS

2.1 Model Parameterization

A complete 3D model for our purposes comprises five 3D parameter functions: α (P velocity), β (S velocity), ρ (density), Q_α and Q_β (quality factors for P and S velocity). We represent these as a latitude/longitude grid of vertical profiles, all sharing a common parameterization with respect to depth. Namely, the crust at each latitude/longitude point is divided into six, vertically homogeneous layers corresponding to those of the CRUST2.0 model (Laske *et al.*, 2001): water overlying two sediment layers (soft and hard sediments) overlying three igneous/metamorphic layers (upper, middle and lower crust). The thicknesses of the crustal layers are allowed to vary between profiles. The upper mantle is represented vertically by piecewise linear parameter functions sampled at 17 nodes distributed between the Moho discontinuity and a depth of 410 km. The parameters are discontinuous across the Moho.

2.2 Forward Modeling

The forward problem in our inversion methods requires prediction of first-arriving body-wave travel times and frequency-dependent surface-wave group delays in a three-dimensional (3D) Earth model. We calculate body-wave travel times with the finite-difference method of Podvin and Lecomte (1991) as implemented by Lomax *et al.* (2000). We extended the Podvin-Lecomte (P-L) software with algorithms for mapping Earth models in spherical coordinates to flat-Earth, Cartesian models, and for generating the partial derivatives (sensitivities) of the calculated travel times with respect to the P-wave velocity model parameters.

Our method for modeling group delays assumes that the *phase* delay of a Rayleigh wave at a fixed frequency ω is a line integral of a local phase velocity along a path connecting the source and receiver locations. Denoting the local phase-velocity function, or *map*, as $c(\theta, \phi; \omega)$, the propagation path as Γ_ω , and phase delay as $t^{ph}(\omega)$, we have

$$t^{ph}(\omega) = \int_{\Gamma_\omega} \frac{ds}{c(\theta, \phi; \omega)}, \quad (1)$$

where s is the distance along the path. This equation assumes there is no static phase attributable to source excitation or receiver effects. Previous workers in surface-wave tomography have generally taken Γ_ω to be the great-circle arc connecting the source and receiver (e.g. Woodhouse and Dziewonski, 1984). Bukchin *et al.* (2006) have proposed that a better approximation, that accounts for the focusing effects due to lateral variations in local phase velocity, is achieved with Γ_ω taken as the geometrical ray (Fermat path) through the phase-velocity map, although the authors stress that the effects of strong lateral variations can only be accurately modeled with a numerical wave-equation or mode-coupling approach. Our modeling technique adopts the Fermat approach, implemented with a 2D version of the Podvin-Lecomte finite-difference method to obtain Γ_ω .

The group delay between a source and receiver is defined as

$$t^{gr}(\omega) = \frac{d}{d\omega} (\omega t^{ph}(\omega)). \quad (2)$$

We can apply Equation (2) to equation (1) – ignoring the dependence of Γ_ω on ω since $t^{ph}(\omega)$ is stationary with respect to path perturbations---and obtain

$$t^{gr}(\omega) = \int_{\Gamma_\omega} \frac{ds}{u(\theta, \phi; \omega)}. \quad (3)$$

where $u(\theta, \phi; \omega)$ is the group-velocity map at frequency ω , given by

$$\frac{1}{u(\theta, \phi; \omega)} = \frac{d}{d\omega} \left(\frac{\omega}{c(\theta, \phi; \omega)} \right). \quad (4)$$

We obtain the phase-velocity and group-velocity maps for a 3D Earth model with point-wise dispersion calculations in 1D Earth models (Herrmann, 2002). That is, for a fixed geographic location (θ, ϕ) , $c(\theta, \phi; \omega)$ and $u(\theta, \phi; \omega)$ are calculated from the spherically symmetric Earth model whose depth-dependent velocities and density are given by the vertical profile at (θ, ϕ) : $\alpha(\theta, \phi, z)$, etc. Implicit in these calculations are an earth-flattening transformation that corrects for the Earth's sphericity and corrections for anelasticity using the relationships developed by Liu *et al.* (1976) and Yu and Mitchell (1979).

2.3 Joint Inversion Approach

Our inversion procedure iteratively updates an initial, or prior, model with the objective of fitting both P-wave travel-time and Rayleigh-wave group-delay observations, subject to smoothness and other constraints on the model parameters. The prior model is a composite 3D model consisting of CRUST2.0 for the crust and the 1-D AK135 reference model (Kennett *et al.*, 1995) for the mantle. The CRUST2.0 Pn and Sn velocities were ignored in favor of the AK135 velocities ($\alpha = 8.04$ km/s, $\beta = 4.48$ km/s at the top of the mantle). However, the CRUST2.0 variable Moho depth was retained and accommodated by vertical compression or extension of the AK135 mantle thickness to a depth of 210 km.

The free parameters in the inversion include the velocities, α and β , for three crustal layers (upper, middle, lower crust) and 16 of the 17 vertical nodes in the upper mantle; the parameters at 410 km are held fixed to AK135 values. Additionally, we vary the Moho depth, or crustal thickness, as an explicit unknown parameter function (of latitude and longitude). As in the construction of the prior model, changes to Moho depth are implemented as a proportional extension or compression of the crustal layers and upper mantle nodal depths. The inversion allows changes to density, ρ , by linking them directly to changes in α through a version of Birch's Law (Birch, 1961). We do not allow Q_α and Q_β to change from their initial values.

We can thus formulate our inversion method with four vectors of unknown model parameters. We let vectors **a** and **b** contain the layer or nodal values of the P-wave and S-wave velocity functions, $\alpha(\theta, \phi, z)$ and $\beta(\theta, \phi, z)$, respectively, and let **r** contain the corresponding values of the 3D density function, $\rho(\theta, \phi, z)$. The fourth vector, **h**, contains nodal values of the crustal thickness function, which we denote $H_{cr}(\theta, \phi)$. The data vectors involved in the problem are **t**, containing observed P-wave travel times for various source-receiver paths; and **g**, containing observed group delays for various paths and frequencies. We express the joint inverse problem in terms of these vectors as

$$\mathbf{g} = F_g(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}) + \mathbf{e}_g \quad (5)$$

$$\mathbf{t} = F_t(\mathbf{a}, \mathbf{h}) + \mathbf{e}_t, \quad (6)$$

where F_g and F_t are nonlinear forward modeling functions, and $\mathbf{e}_g$ and $\mathbf{e}_t$ are error vectors. While not shown, it is understood that F_g also depends on the fixed parameters Q_α and Q_β . Equations (5) and (6) are coupled through the dependence of $\mathbf{g}$ and $\mathbf{t}$ on common parameters, $\mathbf{a}$ and $\mathbf{h}$.

2.4 Factoring the Surface-Wave Inverse Problem

We decompose the surface-wave inverse problem in Equation (5) to mirror the two-stage procedure we use in forward modeling group delays, described in Section 1.2. Let the vectors $\mathbf{u}$ and $\mathbf{c}$, respectively, contain values of the local group-velocity and phase-velocity maps at the nodes of our geographic grid and at the observation frequencies. The first stage in calculating group delays for a model $(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h})$ ---dispersion calculations in 1D media---defines functions U and C , such that

$$\mathbf{c} = C(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}) \quad (7)$$

$$\mathbf{u} = U(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}). \quad (8)$$

The second stage – integrating group-velocity maps along Fermat paths found by 2D raytracing through phase-velocity maps – defines a function H_g , such that

$$F_g(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}) = H_g(\mathbf{u}, \mathbf{c}). \quad (9)$$

The dependence of H_g on $\mathbf{u}$ arises from the integrand of Equation (3), while the dependence on $\mathbf{c}$ is from the integration path. Equations (7) – (9) allow us to factor the nonlinear inverse problem of Equation (5) as

$$\mathbf{g} = H_g(\mathbf{u}, \mathbf{c}) + \mathbf{e}_g \quad (10)$$

$$\mathbf{c} = C(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}) + \mathbf{e}_c \quad (11)$$

$$\mathbf{u} = U(\mathbf{a}, \mathbf{b}, \mathbf{r}, \mathbf{h}) + \mathbf{e}_u, \quad (12)$$

where, strictly, $\mathbf{e}_u = \mathbf{e}_c = 0$.

While our notation does not show it, it is important to point out that, as a consequence of our forward modeling assumptions, each of Equations (10) – (12) actually embodies a set of decoupled problems. Equation (10) defines one inverse problem for each frequency ω since there is no cross-frequency dependence of group delays on group-velocity maps. Equations (11) and (12) decouple geographically since, by assumption, the local phase and group velocities at a latitude/longitude point (θ, ϕ) depend only on the 1D velocity and density profiles associated with that point.

2.5 Linearization

One step of our iterative inversion procedure solves linear approximations to the surface-wave and body-wave inverse problems with certain parameter dependences ignored in order to decouple the problems. Let $(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}_{ref})$ denote the parameters of a reference model. Starting with the body-wave problem, we linearize Equation (6) as

$$\mathbf{t} = F_t(\mathbf{a}_{ref}, \mathbf{h}_{ref}) + \mathbf{A}_t(\mathbf{a} - \mathbf{a}_{ref}) + \mathbf{e}_t, \quad (13)$$

where $\mathbf{A}_t$ is the Jacobian matrix (partial derivative matrix) of F_t with respect to $\mathbf{a}$, evaluated at the reference model. In this case, we ignore the dependence on crustal thickness $\mathbf{h}$, assumed to be fixed at its reference value.

Linearizing (10) at the reference model we get

$$\mathbf{g} = H_g(\mathbf{u}_{ref}, \mathbf{c}_{ref}) + \mathbf{B}_g(\mathbf{u} - \mathbf{u}_{ref}) + \mathbf{e}_g \quad (14)$$

where

$$\mathbf{u}_{ref} = U(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}_{ref}) \quad (15)$$

$$\mathbf{c}_{ref} = C(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}_{ref}). \quad (16)$$

The Jacobian of H_g with respect to $\mathbf{u}$, shown as $\mathbf{B}_g$, contains 2D raypath sensitivities, analogous to the 3D raypath sensitivities for body-wave travel times that comprise $\mathbf{A}_t$. The Jacobian of H_g with respect to $\mathbf{c}$ is zero since the Fermat raypaths are stationary with respect to perturbations in $\mathbf{c}$ from $\mathbf{c}_{ref}$. For this reason, we do not consider the linearization of Equation (11). Linearizing (12) yields

$$\mathbf{u} = U(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}) + \mathbf{B}_u(\mathbf{h})(\mathbf{b} - \mathbf{b}_{ref}) + \mathbf{e}_u. \quad (17)$$

Here we ignore the dependence of U on $\mathbf{a}$ and $\mathbf{r}$, but retain its nonlinear dependence on $\mathbf{h}$. The Jacobian matrix $\mathbf{B}_u(\mathbf{h})$ is derived from the group-velocity partial derivatives implied by surface-wave theory in 1D media (Harkrider and Anderson, 1966; Harkrider, 1968).

Under the same assumptions, we can also linearize Equation (5), the unfactored surface-wave inverse problem, to obtain

$$\mathbf{g} = F_g(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}) + \mathbf{A}_g(\mathbf{h})(\mathbf{b} - \mathbf{b}_{ref}) + \mathbf{e}_g. \quad (18)$$

Applying the chain rule to the composite function $F_g = H_g(U, C)$, we find that

$$\mathbf{A}_g(\mathbf{h}) = \mathbf{B}_g \mathbf{B}_u(\mathbf{h}). \quad (19)$$

We are quick to point out that ignoring certain parameter sensitivities will clearly lead to suboptimal solutions of the linearized inverse problems. What is not clear is whether, or how, doing so affects our final solution to the nonlinear joint problem obtained by the iterative updating scheme we use. We note, however, that the sensitivities we ignore are relatively small so that Equations (13) – (18) do retain most of the information about the unknown parameters: on α from the P-wave data and on β and H_{cr} from the Rayleigh-wave data. Moreover, all

dependencies of data on parameters are accounted for in the final solution since nonlinear forward modeling is performed after each round of reference model updating.

The three linearized inverse problems – Equations (13), (14) and (17) – can be written collectively as

$$\mathbf{d} = F^{lin}(\mathbf{m}) + \mathbf{e}, \quad (20)$$

where $\mathbf{d}$ is a data vector, $\mathbf{m}$ is a model vector, and $\mathbf{e}$ is an error vector. The function F^{lin} is the linearization of a nonlinear function F at a reference model, i.e.

$$F^{lin}(\mathbf{m}) = F(\mathbf{m}_{ref}) + \mathbf{A}(\mathbf{m} - \mathbf{m}_{ref}). \quad (21)$$

We allow Equation (20) to stand for the *decoupled* versions of Equations (14) and (17). Thus, there is one instance of (20) for each frequency in the surface-wave data set, as well as one for each geographic point in the group-velocity maps. Therefore, in all instances, the model vector $\mathbf{m}$ becomes a sampled version of a model function, $m(\mathbf{x})$, where $\mathbf{x}$ denotes spatial position and is variously 3D position (θ, ϕ, z) , 2D position (θ, ϕ) or 1D position z . Table 1 identifies the data vector and model function for each of the linearized inverse problems we solve. Note that the model function is taken to be slowness in the body-wave and surface-wave tomographic problems, while it is velocity in the dispersion inverse problem.

Table 1: Data and Model Associations for Linearized Inverse Problems

<i>Equation</i>	<i>Data Vector (d)</i>	<i>Model Function</i>
13	P-wave travel time vs. source-receiver path	$m(\theta, \phi, z) = \alpha^{-1}(\theta, \phi, z)$
14	Rayleigh-wave group delay vs. path (fixed ω)	$m(\theta, \phi, z) = u^{-1}(\theta, \phi; \omega)$
17	Group velocity vs. ω (fixed θ, ϕ)	$m(z) = \beta(\theta, \phi, z)$

2.6 Updating Procedure

Our iterative inversion procedure starts with the reference model ($\mathbf{a}_{ref}$, etc.) set to the initial model ($\mathbf{a}_{pri}$, etc.). Each step of the iteration solves the linearized inverse problems of Equations (14), (17) and (13) in sequence, using the solutions to update the reference model. Specifically, we

- i. Solve (14) to obtain the solution $\tilde{\mathbf{u}}$.
- ii. Solve (17), with $\mathbf{u} = \tilde{\mathbf{u}}$, to obtain $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{h}}$. Then set
$$\mathbf{h}_{ref} = \tilde{\mathbf{h}}$$

$$\mathbf{b}_{ref} = \tilde{\mathbf{b}}$$
- iii. Solve (13) to obtain $\tilde{\mathbf{a}}$. Then set
$$\mathbf{a}_{ref} = \tilde{\mathbf{a}}$$

$$\begin{aligned}
\mathbf{r}_{ref} &= \mathbf{r}_{pri} + 0.3 (\mathbf{a}_{ref} - \mathbf{a}_{pri}) \\
\mathbf{c}_{ref} &= \mathcal{C}(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}_{ref}) \\
\mathbf{u}_{ref} &= U(\mathbf{a}_{ref}, \mathbf{b}_{ref}, \mathbf{r}_{ref}, \mathbf{h}_{ref})
\end{aligned}$$

Step (i) is done frequency-by-frequency and Step (ii) is done point-wise geographically. After iterating this sequence several times, we take the latest reference model to be the solution of the nonlinear joint inverse problem.

Even though we do not solve the linearized inverse problems simultaneously, the regularization technique and constraints we apply to the solutions end up coupling the solutions for $\mathbf{a}$ and $\mathbf{b}$, as we explain later.

2.7 Bayesian Inversion Method

We adopt a Bayesian approach (e.g. Tarantola, 2005) to solving each instance of Equation (20). We describe the approach first in the absence of parameter bounds, which we discuss later.

We assume that $\mathbf{e}$ and $\mathbf{m}$ are each Gaussian random variables, uncorrelated with each other, with prior means zero and $\mathbf{m}_{pri}$, respectively, and with prior variance matrices $\mathbf{V}_e$ and $\mathbf{C}_m$, respectively. We take the solution of Equation (20) to be the maximum *a posteriori* (MAP) estimator of $\mathbf{m}$, which minimizes the objective function given by

$$\Psi(\mathbf{m}) = (\mathbf{d} - F^{lin}(\mathbf{m}))^T \mathbf{V}_e^{-1} (\mathbf{d} - F^{lin}(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_{pri})^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{pri}). \quad (22)$$

We base $\mathbf{C}_m$ on geostatistical concepts (e.g. Deutsch and Journel, 1998), parameterizing it in terms of the second-order spatial statistics of the Gaussian random function $m(\mathbf{x})$. We specify these statistics in terms of a variance function, $\sigma_m^2(\mathbf{x})$, and a correlation kernel, $C_0(\mathbf{x}, \mathbf{x}')$, such that, for any two points $\mathbf{x}$ and $\mathbf{x}'$,

$$\text{Cov}[m(\mathbf{x}), m(\mathbf{x}')] = \sigma_m(\mathbf{x}) \sigma_m(\mathbf{x}') C_0(\mathbf{x}, \mathbf{x}'). \quad (23)$$

We require $C_0(\mathbf{x}, \mathbf{x}) = 1$. Letting the matrix $\mathbf{C}_0$ be a discretized version of $C_0(\mathbf{x}, \mathbf{x}')$, and letting $\mathbf{\Sigma}_m$ be a diagonal matrix containing samples of $\sigma_m(\mathbf{x})$ on its diagonal, we can write a discrete version of Equation (23) as

$$\mathbb{E}[(\mathbf{m} - \mathbf{m}_{pri})(\mathbf{m} - \mathbf{m}_{pri})^T] \equiv \mathbf{C}_m = \mathbf{\Sigma}_m \mathbf{C}_0 \mathbf{\Sigma}_m \quad (24)$$

where $\mathbb{E}[\]$ denotes the mathematical expectation operator on random variables.

We choose the kernel $C_0(\mathbf{x}, \mathbf{x}')$ to be a decaying function of the spatial separation, $\mathbf{x} - \mathbf{x}'$, with the decay rate controlled by a specified *correlation length* in each direction. This is achieved by letting C_0 be the Green's function of a differential operator D_0 – i.e.

$$D_0 C_0(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \quad (25)$$

– with D_0 defined as

$$D_0 = \frac{1}{K_1 \det \Lambda(\mathbf{x})} \left[I - \frac{1}{K_2} \nabla \cdot \Lambda^2(\mathbf{x}) \cdot \nabla \right]^\ell. \quad (26)$$

Here, I denotes the identity operator; ∇ is the gradient operator on the appropriate spatial domain (depth, geographic position, or 3D position); and K_1 , K_2 and ℓ are constants which depend on the dimensionality of the spatial domain (see Table 2). The correlation tensor, Λ , is allowed to vary with position but we restrict its structure to be (in the case of three dimensions)

$$\Lambda(\mathbf{x}) = \lambda_z(z) \mathbf{n}_z \mathbf{n}_z + \lambda_h (\mathbf{n}_\theta \mathbf{n}_\theta + \mathbf{n}_\phi \mathbf{n}_\phi), \quad (27)$$

where $\mathbf{n}_z$, etc. are unit vectors in a spherical coordinate system, and where λ_z and λ_h , respectively, are correlation lengths in the vertical and horizontal directions.

Table 2: Constants for Regularization Operator in a Whole-Space

<i>Dimension of x</i>	<i>ℓ</i>	<i>K_1</i>	<i>K_2</i>
1	1	2	1
2	2	2π	2
3	2	8π	1

In the 3D inversion for α (Equation (13)) and in the 1D inversions for β (Equation (17)), we associate D_0 with Neumann boundary conditions (zero normal gradient) at the Earth's surface and from each side of the Moho discontinuity. The Moho condition leads to the de-correlation of $m(\mathbf{x})$ between the crust and upper mantle, i.e. $C_0(\mathbf{x}, \mathbf{x}') = 0$ when $\mathbf{x}$ and $\mathbf{x}'$ are on different sides of the Moho. Table 3 lists the values of the geostatistical parameters we used in each of the three linearized inversions.

Numerically, we implement the geostatistical approach by constructing a finite-difference matrix, $\mathbf{D}_0$, that approximates the differential operator D_0 . Given $\mathbf{C}_0 = \mathbf{D}_0^{-1}$ and Equation (24), the objective function in Equation (22) can be written explicitly in terms of $\mathbf{D}_0$ as

$$\Psi(\mathbf{m}) = (\mathbf{d} - F^{lin}(\mathbf{m}))^T \mathbf{V}_e^{-1} (\mathbf{d} - F^{lin}(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_{pri})^T \Sigma_m^{-1} \mathbf{D}_0 \Sigma_m^{-1} (\mathbf{m} - \mathbf{m}_{pri}). \quad (28)$$

The numerical algorithms we use to minimize Ψ will be mentioned later.

Table 3: Geostatistical Parameters Used in Inversions

Model Parameter	α^{-1} (slowness)	β (velocity)	u^{-1} (slowness)
Equation in Text	13	17	14
Std. Dev. (σ_m)	—	—	3%
Crust	3%	3%	—
Upper Mantle	2%	2%	—
Horiz. correlation length (λ_h)	350 km	—	350 km
Vertical correlation length (λ_z)			
Crust	$0.5 H_{cr}$	$0.5 H_{cr}$	—
Upper Mantle	70 km	70 km	—

2.8 Crustal Thickness Inversion

The linearized surface-wave inverse problem, Equation (17), differs from the other two problems in that it treats the crustal thickness vector, $\mathbf{h}$, as a free parameter. The objective function for this problem can be written as

$$\begin{aligned} \Psi(\mathbf{b}; \mathbf{h}) = & (\mathbf{u} - U^{lin}(\mathbf{b}; \mathbf{h}))^T \mathbf{V}_e^{-1} (\mathbf{u} - U^{lin}(\mathbf{b}; \mathbf{h})) \\ & + (\mathbf{b} - \mathbf{b}_{pri})^T \boldsymbol{\Sigma}_b^{-1} \mathbf{D}_0 \boldsymbol{\Sigma}_b^{-1} (\mathbf{b} - \mathbf{b}_{pri}). \end{aligned} \quad (29)$$

Letting the minimum of Ψ with respect to $\mathbf{b}$, with $\mathbf{h}$ fixed, be achieved at $\mathbf{b} = \tilde{\mathbf{b}}(\mathbf{h})$, we can write an objective function for $\mathbf{h}$ as

$$\tilde{\Psi}(\mathbf{h}) = \Psi(\tilde{\mathbf{b}}(\mathbf{h})). \quad (30)$$

We take the solution for $\mathbf{h}$ to be the value minimizing $\tilde{\Psi}$. Since the inverse problem decouples geographically, the minimization is performed separately for each (θ, ϕ) in our model grid, yielding $\tilde{H}_{cr}(\theta, \phi)$ for each point. Numerically, we find the solution with a grid search over crustal thickness values spanning a range of $\pm 2.5\%$ from the current reference value.

It is worth noting that each trial crustal thickness in a grid search for $H_{cr}(\theta, \phi)$ entails forward modeling calculations for group velocities at all the observed frequencies, including the computation of shear-wave sensitivities, as well as the inversion calculations required to obtain the solution shear-wave profile, $\tilde{\beta}(\theta, \phi, z)$. In other words, allowing crustal thickness to be an inversion parameter comes at a high computational price.

2.9 Factored vs. Unfactored Surface-Wave Inversion

Appendix A shows that appropriate choices of prior model means and prior error and model variances make the solution of a nonlinear inverse problem, and of a factored version of that problem, the same within a linear approximation. The equivalence is exact for linearized problems. The requirement for prior means is easily met; the prior mean of $\mathbf{u}$ in solving Equation (14) is

$$\mathbf{u}_{pri} = U(\mathbf{a}_{ref}, \mathbf{b}_{pri}, \mathbf{r}_{ref}, \mathbf{h}_{ref}). \quad (31)$$

However, the required choice of error and model variances in the factored problems is complicated and difficult to implement. In particular, the variance of the error vector $\mathbf{e}_u$ in Equation (17) would have to be set to the posterior variance of $\mathbf{u}$ in Equation (14), which is a non-diagonal matrix and, in any case, difficult to compute. Additionally, it is not practical to select prior model variances in the factor problems that equate to a desirable prior variance on $\mathbf{b}$ in the composite problem, e.g. one implied by a 3D geostatistical treatment of $\mathbf{b}$ analogous to that assumed for $\mathbf{a}$ in the body-wave inversion.

Nonetheless, the geostatistical assumptions we did use for $\mathbf{u}$ and $\mathbf{b}$ in the factor problems *attempted* to achieve parity with our 3D geostatistical treatment of $\mathbf{a}$ in the travel-time tomography problem. As discussed above and summarized in Table 3, we used a 2D correlation

kernel for $\mathbf{u}$ and a 1D correlation kernel for $\mathbf{b}$ that shared the variances and correlation lengths of the 3D correlation kernel assumed for $\mathbf{a}$.

2.10 Velocity Correlation

In a Bayesian inversion methodology it is straightforward to make the *a priori* assumption that the P and S model velocities are correlated. Considering both α and β as functions of 3D position, we can express this as

$$\text{Cov}[\alpha(\mathbf{x}), \alpha(\mathbf{x}')] = \sigma_\alpha(\mathbf{x}) \sigma_\alpha(\mathbf{x}') C_0(\mathbf{x}, \mathbf{x}') \quad (32)$$

$$\text{Cov}[\beta(\mathbf{x}), \beta(\mathbf{x}')] = \sigma_\beta(\mathbf{x}) \sigma_\beta(\mathbf{x}') C_0(\mathbf{x}, \mathbf{x}') \quad (33)$$

$$\text{Cov}[\alpha(\mathbf{x}), \beta(\mathbf{x}')] = r(\mathbf{x}) \sigma_\alpha(\mathbf{x}) \sigma_\beta(\mathbf{x}') C_0(\mathbf{x}, \mathbf{x}'), \quad (34)$$

where $C_0(\mathbf{x}, \mathbf{x}')$ specifies the spatial correlation structure of each velocity function, and $r(\mathbf{x})$ specifies the correlation coefficient between velocities at each spatial point. Separating the spatial correlation and velocity correlation in this way requires that

$$r(\mathbf{x}) C_0(\mathbf{x}, \mathbf{x}') = r(\mathbf{x}') C_0(\mathbf{x}, \mathbf{x}'), \quad (35)$$

which forces r to be piece-wise constant, differing between points $\mathbf{x}$ and $\mathbf{x}'$ only when $C_0(\mathbf{x}, \mathbf{x}') = 0$.

A simultaneous inversion for α and β can incorporate Equations (32) – (34) in an appropriate objective function of $\mathbf{a}$ and $\mathbf{b}$. Appendix B shows that the same assumptions can be used in each step of our sequential approach with an appropriate modification of the individual objective functions for each inversion. The modifications are particularly simple under the restriction that

$$\alpha_{\text{pri}}(\mathbf{x})/\sigma_\alpha(\mathbf{x}) = \beta_{\text{pri}}(\mathbf{x})/\sigma_\beta(\mathbf{x}), \quad (36)$$

i.e. the prior means and standard deviations of α and β are in the same ratio everywhere. Then, for example, in solving Equation (13), the objective function to be minimized would be Equation (28) with $\mathbf{m} = \mathbf{a}$ and $\mathbf{m}_{\text{pri}}$ and Σ_m set as discretized versions of

$$m_{\text{pri}}(\mathbf{x}) = [1 - r(\mathbf{x})] \alpha_{\text{pri}}(\mathbf{x}) + r(\mathbf{x}) \frac{\alpha_{\text{pri}}(\mathbf{x})}{\beta_{\text{pri}}(\mathbf{x})} \beta_{\text{ref}}(\mathbf{x}) \quad (37)$$

$$\sigma_m(\mathbf{x}) = [1 - r(\mathbf{x})^2]^{1/2} \sigma_\alpha(\mathbf{x}). \quad (38)$$

Similar expressions determine the mean and standard deviation of $\mathbf{b}$ used in solving Equation (17).

2.11 Parameter Bounds

We constrain the solutions of Equations (13) and (17) to be within prior upper and lower bounds on α , β and Poisson's ratio ν , requiring

$$\alpha_{\min}(\theta, \phi, z) \leq \alpha(\theta, \phi, z) \leq \alpha_{\max}(\theta, \phi, z) \quad (39)$$

$$\beta_{\min}(\theta, \phi, z) \leq \beta(\theta, \phi, z) \leq \beta_{\max}(\theta, \phi, z) \quad (40)$$

$$v_{min}(\theta, \phi, z) \leq v(\theta, \phi, z) \leq v_{max}(\theta, \phi, z). \quad (41)$$

In general, the constraints at each point (θ, ϕ, z) define a six-sided region of admissible points in α - β space, as illustrated in Figure 2. However, the bounds in the separate inversion for each velocity comprise a simple velocity interval, derived by using reference values of the other velocity to translate the bounds on Poisson's ratio. That is, we augment Equation (20), which represents each of the linearized inverse problems we solve, with the constraint

$$m_{min}(\mathbf{x}) \leq m(\mathbf{x}) \leq m_{max}(\mathbf{x}). \quad (42)$$

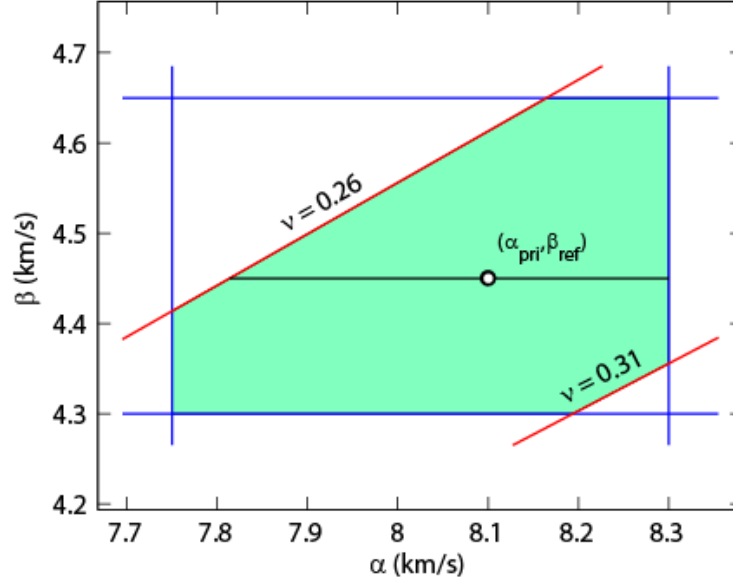


Figure 2: The default bounds for the upper mantle lid on P velocity (α), S velocity (β) and Poisson's ratio (v) are displayed in α - β space. The region shaded green contains velocity pairs satisfying the constraints. The horizontal black line displays the interval of admissible P velocities that results when the S velocity is held to a reference value indicated by the black circle.

For the P-wave travel-time problem, for example, we have (for each $\mathbf{x}$)

$$m_{min} = \max \left\{ \alpha_{min}, \beta_{ref} \sqrt{\frac{1-v_{min}}{1/2-v_{min}}} \right\} \quad (43)$$

$$m_{max} = \min \left\{ \alpha_{max}, \beta_{ref} \sqrt{\frac{1-v_{max}}{1/2-v_{max}}} \right\}. \quad (44)$$

This conditional constraint is illustrated in Figure 2 by the horizontal black line spanning the green region of admissible velocities.

We developed spatially variable velocity and Poisson's ratio bounds in a two-step analysis. The first step established default bounds, independent of latitude and longitude, for each crustal layer and as a function of depth in the upper mantle. These values, listed in Table 4, reflect a modest variation from the AK135 model (in the upper mantle) and from typical values from continental regions in CRUST2.0 (in the crust). The second step found the largest and smallest

velocities in CRUST2.0 as a function of tectonic regime. Worldwide, there are 26 such regimes (e.g. “Platform”, “Continental Shelf”). The velocity bounds at each point in the model were expanded to accommodate these extrema, for the appropriate regime, with an additional 0.05 km/s buffer.

Table 4: Default Parameter Bounds

<i>Depth</i>	<i>α (km/s)</i>		<i>β (km/s)</i>		<i>v</i>	
	<i>min</i>	<i>max</i>	<i>min</i>	<i>max</i>	<i>min</i>	<i>max</i>
Upper Crust	5.50	6.20	3.00	3.45	0.23	0.34
Middle Crust	6.10	6.80	3.35	3.80	0.24	0.33
Lower Crust	6.70	7.40	3.70	4.10	0.25	0.32
z_{moho}	7.75	8.30	4.30	4.65	0.26	0.31
$z \approx 120$	7.80	8.35	4.35	4.70	0.26	0.31
$z = 210$	8.05	8.55	4.40	4.70	0.26	0.31
$z = 410$	8.75	9.25	4.65	5.00	0.26	0.31

To minimize $\Psi(\mathbf{m})$ in Equation (28) subject to the constraint (42), we generate $m(\mathbf{x})$ from an unconstrained function, $q(\mathbf{x})$, with the following transformation:

$$m(\mathbf{x}) = m_{pri}(\mathbf{x}) + \Delta m(\mathbf{x}) \operatorname{erf} \left\{ \frac{\sqrt{\pi}}{2} \frac{\sigma_m(\mathbf{x})}{\Delta m(\mathbf{x})} q(\mathbf{x}) \right\} \equiv M(q(\mathbf{x})), \quad (45)$$

where

$$\Delta m(\mathbf{x}) = \begin{cases} m_{\max}(\mathbf{x}) - m_{pri}(\mathbf{x}), & \text{if } q(\mathbf{x}) \geq 0; \\ m_{\min}(\mathbf{x}) - m_{pri}(\mathbf{x}), & \text{if } q(\mathbf{x}) < 0. \end{cases} \quad (46)$$

It is easily shown that as the bounds become wide compared to $\sigma_m(\mathbf{x})$ ($|\Delta m(\mathbf{x})| \gg \sigma(\mathbf{x})$), then

$$m(\mathbf{x}) \rightarrow m_{pri}(\mathbf{x}) + \sigma_m(\mathbf{x}) q(\mathbf{x}). \quad (47)$$

The linear inverse problem of Equation (20), in terms of $q(\mathbf{x})$, becomes the nonlinear problem

$$\mathbf{d} = F^{lin}(M(\mathbf{q})) + \mathbf{e}, \quad (48)$$

where the vector $\mathbf{q}$ is a sampled version of the function $q(\mathbf{x})$. However, we do not simply substitute Equation (45) into (28) to obtain an objective function for $\mathbf{q}$. Instead, we assume $\mathbf{q}$ to be a zero-mean random variable with unit variance and the same spatial correlation kernel as $\mathbf{m}$:

$$\operatorname{Cov}[q(\mathbf{x}), q(\mathbf{x}')] = C_0(\mathbf{x}, \mathbf{x}') \quad (49)$$

or

$$\mathbf{C}_q = \mathbf{C}_0. \quad (50)$$

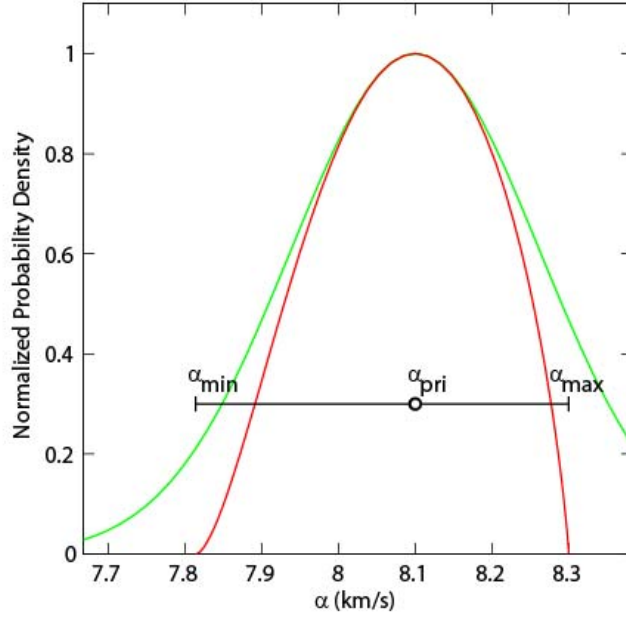


Figure 3: Velocity bounds have the effect of modifying a Gaussian prior probability distribution on P velocity (green line) to a non-Gaussian distribution (red line).

Assuming $q(\mathbf{x})$ to be Gaussian has the effect of making $m(\mathbf{x})$ non-Gaussian, with a probability density function that peaks at $m_{pri}(\mathbf{x})$ and is coerced to zero at the bounds, illustrated in Figure 3. However, in the limit of wide bounds, Equation (47), $\mathbf{m}$ approaches Gaussian with mean $\mathbf{m}_{pri}$ and the variance matrix $\mathbf{C}_m$ given in (24).

The objective function for $\mathbf{q}$ becomes

$$\Psi_q(\mathbf{q}) = (\mathbf{d} - F^{lin}(M(\mathbf{q})))^T \mathbf{V}_e^{-1} (\mathbf{d} - F^{lin}(M(\mathbf{q}))) + \mathbf{q}^T \mathbf{D}_0 \mathbf{q}. \quad (51)$$

The numerical algorithm we used for minimizing Ψ_q is different for the three linearized inverse problems we solve. In the group-velocity and travel-time tomography problems (Equations (13) and (14)), we perform the minimization of Ψ_q with the nonlinear conjugate gradients algorithm described by Rodi and Mackie (2001). In the surface-wave dispersion inverse problem (Equation (17)) we use a Gauss-Newton technique with iterative reweighting.

3. APPLICATION TO DATA

In our study we made use of observed travel times from regional P waves and frequency-dependent Rayleigh-wave group delays at short to intermediate distances to determine the crust and upper-mantle seismic structure at depth ranges between the surface and approximately 300 km depth. In the following sections we provide details about the data sets we utilized and the parameterization of the inversion model.

3.1 Travel-Time Observations

Our primary source of P-wave travel times was a subset of the International Seismic Centre (ISC) catalog, processed according to the methodology developed by Engdahl, van der Hilst and Buland (EHB; Engdahl *et al.*, 1998). We extracted arrivals from the years 1982 – 2004 having event and station locations within a latitude and longitude box defined as 0-60° N and 30-120° E and event depths between 0 and 200 km. We included only first-arriving phases denoted Pg, Pb or Pn and which were defining phases for the EHB locations. We did not relocate the events as part of this study, and therefore filtered the events carefully to ensure small epicentral mislocations. In particular, we required the secondary azimuth gap, defined as the largest azimuthal gap closed by a single station, for a given event to be less than or equal to 130° (Bondár *et al.*, 2004b) and the number of teleseismic arrivals to be at least 15. In addition to this set of data we added EHB bulletin picks over the Tibetan Plateau and southwestern China using data from temporary arrays (Li *et al.*, 2006). The data set satisfying our criteria comprised 167,384 arrivals from 7,681 events and 944 stations.

We compressed this data set by forming summary events on a regular grid having 0.5-degree spacing in latitude and longitude and containing 13 nodes in depth between 0 and 200 km, with the depth spacing per node increasing from 5 to 20 km. For each summary-event node and each station/phase type, a summary travel-time residual (relative to the AK135 Earth model) was formed by averaging the individual residuals for the events near that node. Following this compression, stations containing fewer than 25 arrivals were dropped from the data set. The use of summary events acknowledges the redundant sensitivity of individual data to the Earth model (which is on a 1-degree grid) and, combined with the station-dropping rule, reduces the ray-tracing requirements for the inversion substantially.

The final database used in the body-wave tomography contained 104,065 arrivals for 3,689 summary events and 603 stations. The data spanned epicentral distances to 18.7 degrees, and the travel-time residuals (relative to AK135) ranged from -9.8 to 9.9 s with a root-mean-square (RMS) error of 2.6 s.

3.2 Group-Velocity Dispersion Measurements

The surface-wave dispersion database primarily consists of measurements donated by Ritzwoller and Levshin (1998) and the Lawrence Livermore National Laboratory. Different portions of the database have been used in the derivation of recent models (Ritzwoller *et al.*, 2002; Shapiro and Ritzwoller, 2002; Pasyanos, 2005). Some measurements in the database were also made by the Weston Geophysical research staff using the Ritzwoller and Levshin (1998) procedures for measuring group velocities. To eliminate potential outliers in the combined data set, we performed a period-by-period grooming of the data, in which we retained a group velocity measurement if it was within two standard deviations of the mean group velocity for that period. This exercise resulted in a database of 98,889 fundamental-mode Rayleigh group velocity picks from 3,993 events, 172 stations and 23 periods: $T = 10\text{-}16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 125$ and 150 seconds. The surface-wave database is nonuniformly distributed across the various periods and in paths lengths. The distribution of the

group velocity data is shown in Figure 4 as a function of period and path length. The data are dominated by path lengths between 5 and 35 degrees periods below 100 seconds. Using the accepted relation that surface waves can resolve velocity changes down to approximately 1/3 of their wavelength, the shear-wave portion of our model should be well constrained to over 200 km.

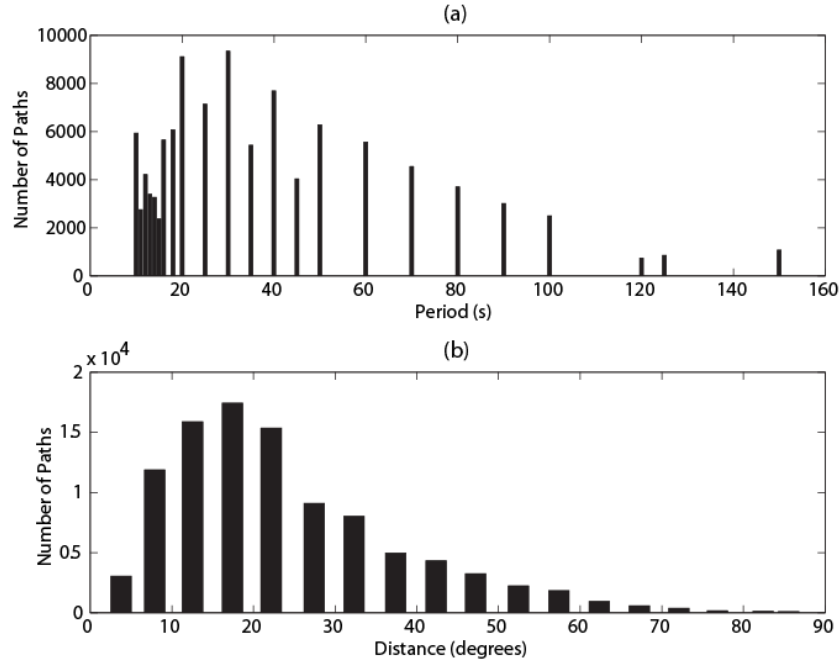


Figure 4: The distribution of the group velocity database as a function of period and station-to-event epicentral distance. (a) The number of group velocity measurements (paths) at the discrete periods in the database; (b) the number of paths at distance ranges between 0 - 90°, in 5° bins.

In Figure 5 we show examples of the spatial coverage of our data sets. To generate these maps we used our extended version of the P-L algorithms to calculate sensitivities of the body-wave travel times (right column) and frequency-dependent surface-wave delay times (left column) in our data set, calculated with our final inversion model. We then summed up a measure of 'hits' in the model with respect in specific depth intervals or at specific periods at each geographic point in a 0.5 x 0.5 degree grid in latitude and longitude. Figure 5 demonstrates that the data sets we used to constrain the inversion model provide excellent coverage of our study region.

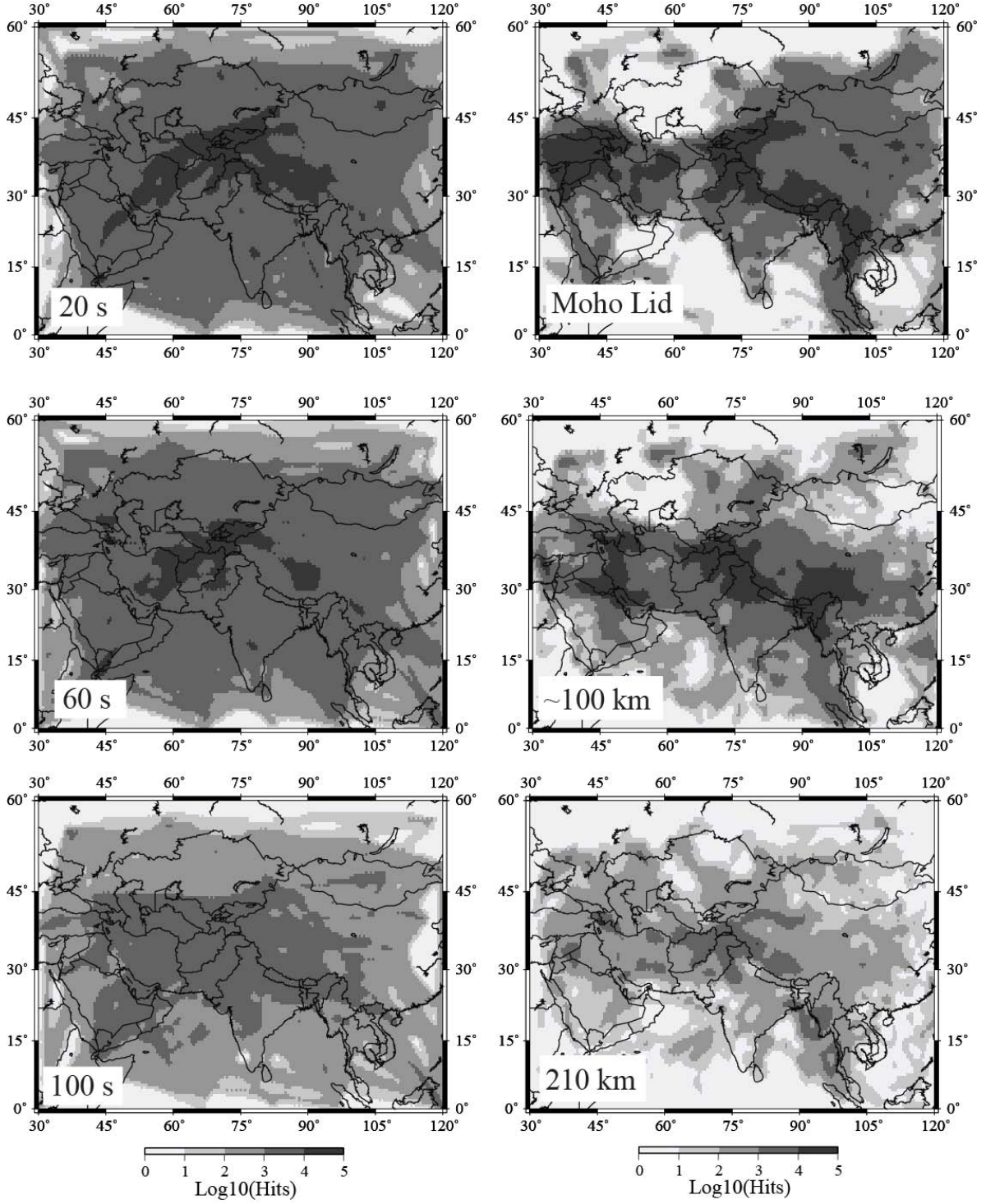


Figure 5: Examples of data sensitivity distribution, expressed as $\log_{10}(\text{hits})$ for the surface- and body-wave data sets used in the joint inversion. Left column: Maps of the Rayleigh-wave group-delay sensitivity hits at 20s, 60s, and 100s; Right Column: Maps of the summary-event P-wave sensitivity coverage at depths just beneath the Moho Lid, at approximately 100 km, and 210 km.

To quantify the importance of applying a 2-D Fermat approach to the forward modeling of the group delays in our surface-wave data set, we calculated a measure of 'path wander' for a given source-receiver pair as the maximum deviation from the curved ray path to the great-circle path (Ritzwoller *et al.*, 2002). This quantity is somewhat tricky to define in a finite-difference ray tracing approach since a raypath is only implicitly defined by a sensitivity distribution that has a finite width determined by such factors as the source-receiver distance and the degree of heterogeneity in the phase-velocity map for a given frequency. To prevent overestimating the maximum deflection along a given path, we calculated the path wander from sensitivities above a threshold chosen to isolate the equivalent geometrical ray between an event and station. Using this formulation we found the maximum lateral deviation from a great-circle path (a straight line between the station and the event in a Cartesian grid) for all the arrivals in our group velocity database through the predicted phase-velocity maps calculated with the final inversion model.

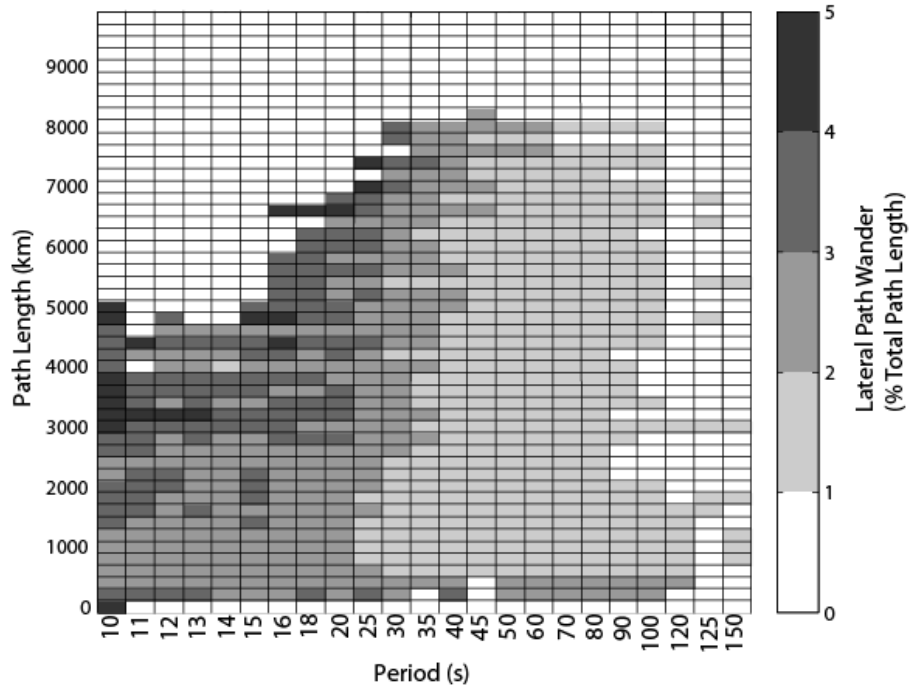


Figure 6: Depiction of mean lateral path wander as a function of path-length range (vertical axis) and period (horizontal axis). For each group delay in the database, the maximum lateral deflection of the Fermat path from the great-circle path, normalized as a percentage of the source-receiver distances, was computed and assigned to a 200-km path-length bin. The mean (normalized) deviation for each bin, as a function of period, is displayed in gray scale. Darker blocks indicate significant path wander, which occurs primarily at shorter periods and longer path lengths. White blocks indicate that a particular bin either had limited measurements or low lateral deviation.

In Figure 6 we show the summary results of these calculations, which are plotted in matrix form, after they have been grouped as a function of period (horizontal axis) and 200-km-wide path-length bins (vertical axis). The results are presented as the mean absolute deviation, normalized to be a percentage of the total path length, among all the paths in a given distance/period bin. A white colored block indicates either no measurements existed or there was

insignificant lateral deviation. The figure demonstrates that there are patches of significant path wander, mainly for measurements made at shorter periods (up to perhaps 40 seconds) and longer path lengths. At longer periods the predicted phase-velocity maps do not exhibit large lateral variation, which is the primary factor that deviates rays from a great-circle path. At shorter path lengths, the positions of the event and station are the most important factors, since if they sample regions with intervening strong lateral variability the ray path will deviate significantly. These regions of strong lateral variability exist in our model region, but their effect is somewhat masked in Figure 6 when we combine the results from many paths into a single mean value at a given period. Overall, the results indicate that only certain regions of our model (e.g. the Himalayan front) require the Fermat approach, but we use it everywhere, since there is a small computational price exacted by using the same algorithm for both body- and surface-wave data.

4. RESULTS OF JOINT INVERSION

We applied the nonlinear joint inversion method presented earlier in Section 2 to the data set described in Section 3 to construct a new compressional and shear velocity model, which we have dubbed the JWM model (for **J**oint **W**eston/ **M**IT). The new model comprises a set of P and S velocity profiles on a $1^\circ \times 1^\circ$ grid that is well resolved within a latitude and longitude box of (10° - 50° N, 40° - 110° E). In the following sections we highlight some of the features in the new model, particularly in the upper mantle where the sensitivity of the data sets is the highest.

4.1 Velocity Heterogeneities

In Figure 7 we show the P and S velocity variations with respect to the AK135 model for the new JWM model at three depths in the upper mantle: 90, 150 and 210 km. Some features of note in the new model include:

- Crustal thickening beneath the major orogenic zones in the region, including the Hindu Kush, Tibetan Plateau and Altai.
- The presence of a high-velocity lithospheric root to the south of the Himalayan front.
- Stronger lows in the crust and uppermost mantle shear velocities under the Tibetan Plateau than in the compressional velocities.
- Signatures of the sedimentary basin structures across the region (e.g. Tarim, Ordos) persisting into the uppermost mantle.

We show additional slices of the JWM model at 30 km intervals starting at 30 km depth in Appendix C.

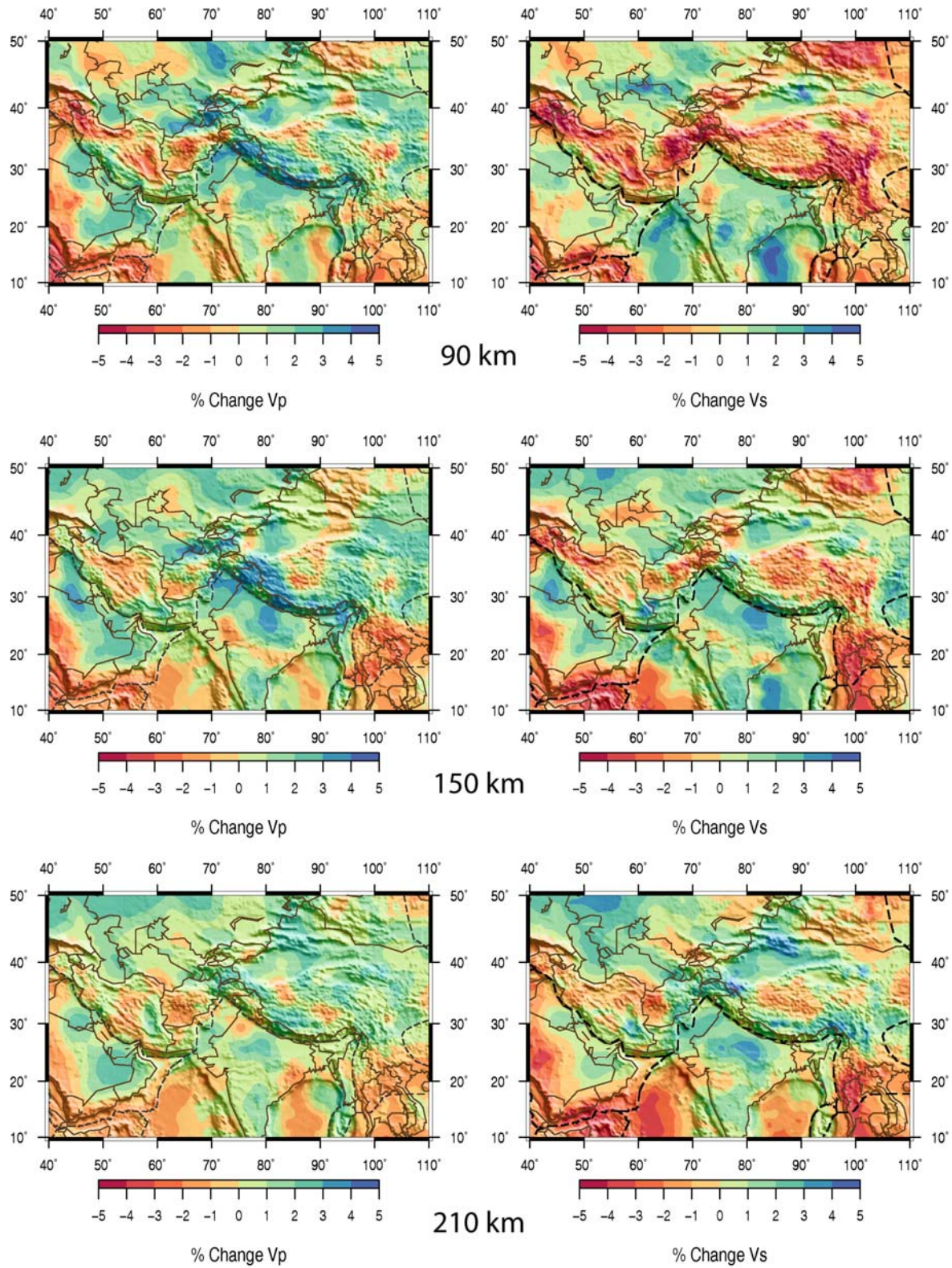


Figure 7: Compressional (left) and shear (right) velocity structure in the upper mantle of the JWM inversion model (horizontal slices at 90 km, 150 km and 210 km depths). The velocities are displayed in percent deviation from the AK135 model.

In Figure 8 we show two vertical, or depth, slices through the JWM model along great-circle paths. On the left is a longitudinal slice (at 55°E) across the Saudi Arabian Peninsula to the north across southern Iran and into southern Turkmenistan. One notable feature in this slice is the low velocity area with respect to the background model beneath central Iran, which may have implications for the active subduction processes occurring beneath the Eurasian continental collision zone. The slice on the right at 85°E cuts across the Himalayan Front, from northeastern India into central China. Here the P and S velocity models both show the Himalayan front and the thick lithospheric root beneath the Tibetan Plateau. We find that the Poisson's ratio in the lower crust of the Plateau is strongly elevated, which agrees with some previous studies in the region (e.g. Owens and Zandt, 1997). However, in contrast with some travel-time based studies (e.g. Chang *et al.*, 2008), we find velocity highs in the uppermost mantle across larger portions of the Tibetan Plateau.

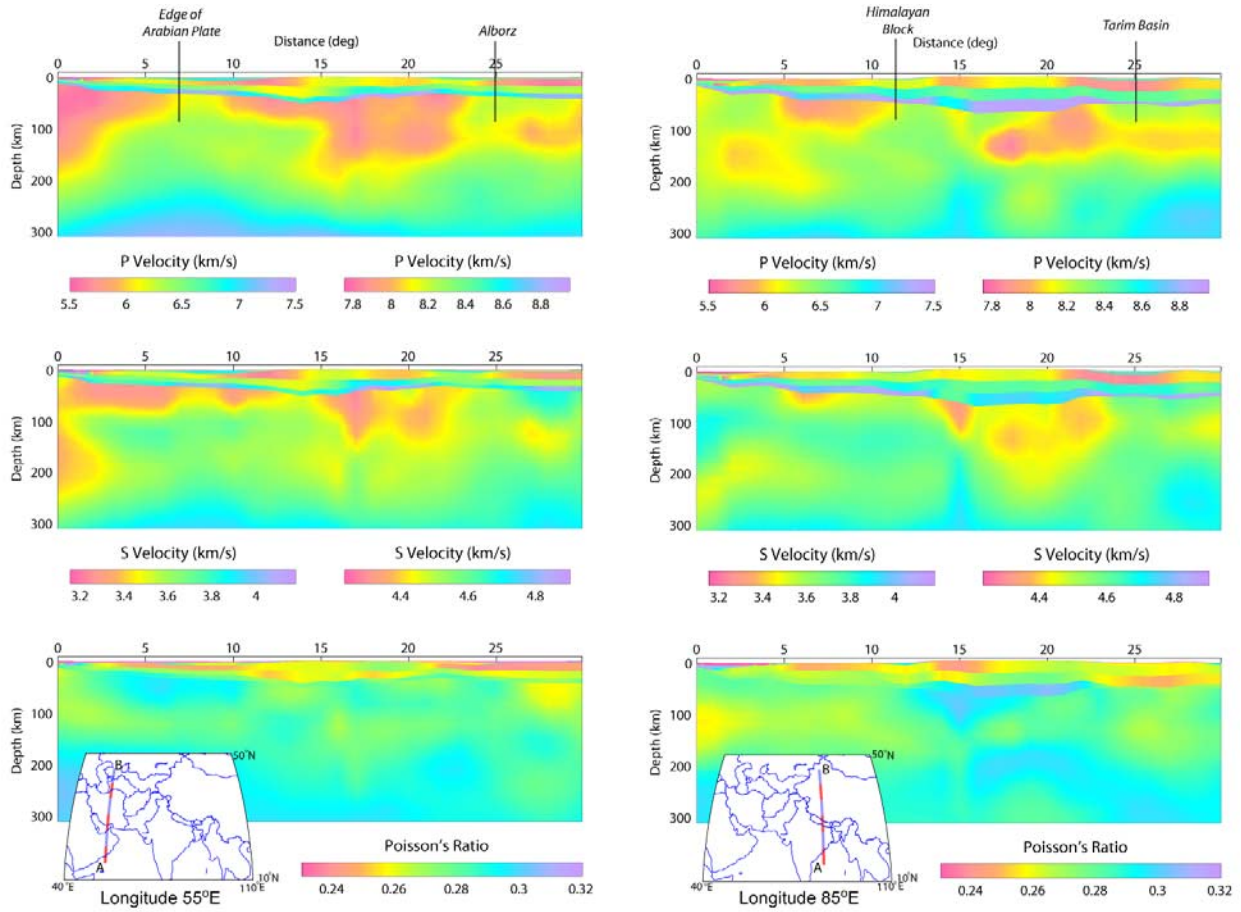


Figure 8: Depth slices along two great-circle paths (see inset maps for path AB) in the JWM inversion model, including P velocity (top), S velocity (middle) and Poisson's ratio (bottom). Note that a different color scale is used for crust (left scale) and upper mantle (right scale) velocities. Some prominent tectonic features are marked by vertical black lines on the top of each cross-section.

4.2 Resolution Tests

In this section we show the results of performing some tests using a checkerboard pattern to assess which features of the new model are real and which might be potential artifacts. Checkerboard tests are well known to have limitations (Nolet *et al.*, 1999; Leveque *et al.*, 1993), and in our case they are only diagnostic of how the coupled inversion behaves in the linearized sense. The input model for our checkerboard experiments consisted of a lateral pattern of alternating positive and negative velocity perturbations centered on $5^\circ \times 5^\circ$ squares. We perturbed the crustal layers by $\pm 3\%$ and the upper mantle layers by $\pm 2\%$ in both P and S velocity. The checkerboard pattern also varied with depth, varying in the same direction across the three crustal layers, and then alternating at a five-node interval (approximately every 100 km) in the upper mantle down to 410 km. We defined a checkerboard model by overlaying this pattern onto our model determined after the fifth iteration of the nonlinear inversion. First-order data perturbations induced by the checkerboard pattern were added to the observed data. We then performed five additional linearized inversion steps with the perturbed data, holding the raypath sensitivities to the values calculated with the fifth-iteration model. The resolution test was completed by comparing the resulting inversion model to that obtained with the unperturbed data. The top plots in Figure 9 show the recovered P and S checkerboard pattern as a depth slice at 90 km. The bottom of this figure shows the profiles of the fifth-iteration, perturbed checkerboard and recovered pattern velocities at the geographic point (30° N, 60° E). Both inversions, with and without the checkerboard-induced data perturbations, used the same regularization scheme, velocity bounds, and variance parameters. Additional results for the checkerboard resolution test are provided as depth-horizon, or map-view, slices in Appendix C.

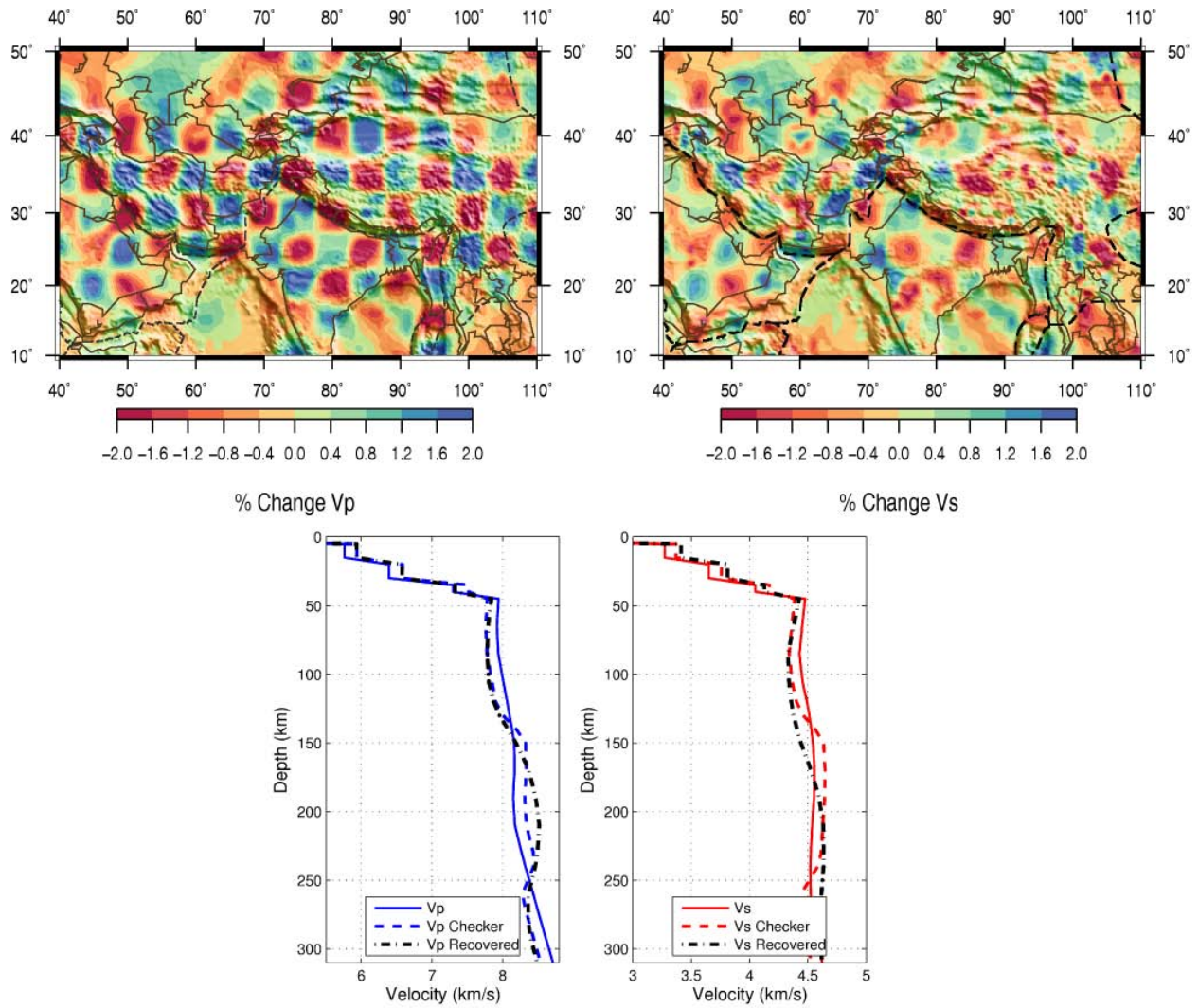


Figure 9: Selected results from a checkerboard resolution test. Top: slices of the recovered checkerboard pattern at 90 km. Bottom: the P and S velocity profiles at the geographic point (30° N, 60° E) for the fifth iteration, target checkerboard and recovered checkerboard models.

4.3 Model Correlations

In Figure 10 we examine features of the JWM model across the study region. Figure 10a shows the RMS amplitude change in the compressional and shear velocities with respect to the initial model as a function of depth, averaged over the regional box defined by (10 – 50°N, 40 – 110° E). The crust between the surface and approximately 50 km exhibits the most change from the initial model (CRUST2.0 over an AK135 upper mantle). This is primarily due to significantly higher P velocities in the lower crust across the Tibetan Plateau and much lower P and S velocities in the entire crust along the orogenic belts crossing Iran and Pakistan, near the northern edge of the Caspian Sea, and along the eastern edge of the Indian plate. In some areas the changes can range as high as $\pm 12\%$ for P and $\pm 14\%$ for S from the values in the CRUST2.0 model, even in places where the crustal thickness did not change significantly over the course of

the inversion. Peak variations from the initial model in the upper mantle reach $\pm 3.6\%$ for P and $\pm 4.8\%$ for S velocities.

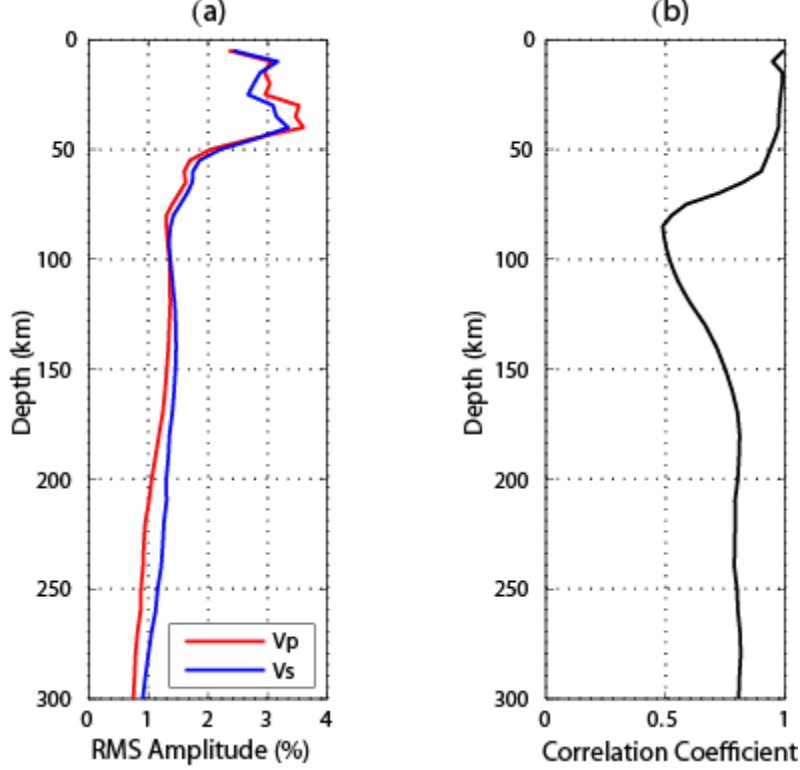


Figure 10. Features of JWM across the study region. (a) RMS amplitude (in percent) of the changes to the initial model for the compressional and shear wave velocities, averaged over the study region; and (b) the correlation coefficient between compressional and shear velocity as a function of depth within the regional limits of the new JWM model.

The compressional and shear velocities are strongly correlated in the crust and upper mantle, which is an expected consequence of the constraints that we implemented in the joint inversion. Figure 10c shows the mean correlation coefficient as a function of depth for the JWM model. We calculate the correlation coefficient across the defined model region using the traditional linear correlation formula

$$r_k = \frac{n \sum \alpha_k \beta_k - \sum \alpha_k \sum \beta_k}{\sqrt{n \sum \alpha_k^2 - (\sum \alpha_k)^2} \sqrt{n \sum \beta_k^2 - (\sum \beta_k)^2}}, \quad (51)$$

where k is a specific depth and the sum is over the set of latitudes and longitudes in the model region. The results in Figure 10c indicate a strong positive correlation between the P and S velocities at most depths. The weakest correlation occurs in the uppermost mantle, where the correlation between P and S drops close to 0.5. The depths between approximately 70 and 120 km are generally where both body and surface wave data sets have high sensitivity. To take advantage of this, our inversion constraints in the upper mantle regions allowed changes in the P and S velocities that were less well correlated with one another.

4.3 Data Variance Reduction

In Figure 11 and Table 5 we demonstrate the improved fit to the data sets achieved by the new JWM inversion model. Figure 11 shows histograms of the residuals for the travel-time and group-delay data used in the inversion. In both cases the joint inversion removed the mean in the data and lowered the variance around the mean. The improved data fit produced by the inversion is more dramatic for the travel-time data, and illustrates the inherent noisiness of the group-velocity data.

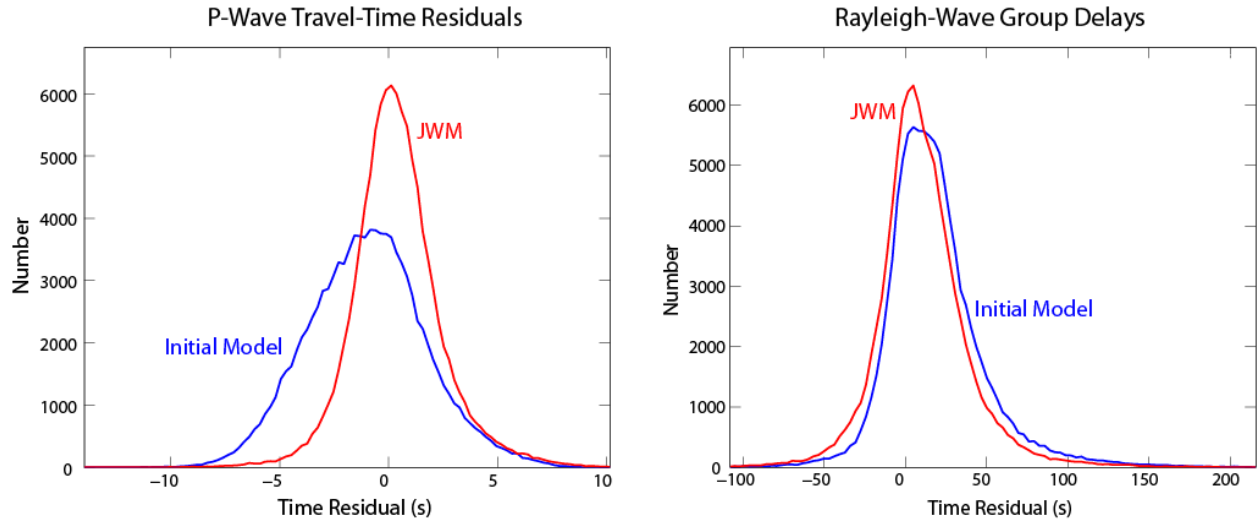


Figure 11. Distribution of the time residuals for the P-wave travel-time data (left) and Rayleigh-wave group delays (right) for the initial (blue lines) and final JWM (red lines) inversion models.

The data fit results are presented in another fashion in Table 5, which lists the values of the root-mean-square (RMS) error measured for the P-wave travel times and Rayleigh-wave group delays with respect to the new JWM model. For comparison purposes we also include the fits to the initial and AK135 models. The variance reduction from these models is a measure of the success of our inversion method; in addition, the variance reduction compared to AK135 reveals how well we constructed our prior model. Table 5 illustrates that the JWM inversion model significantly lowers the data variance for the travel-time and group velocity data sets with respect to the other models. For example, the percent variance reduction with respect to the initial model is 51% for the P-wave travel times and 26% for the Rayleigh-wave group delays.

When we utilize subsets of the data in the calculation of the RMS error values, we reveal some interesting features in the fits of the various models. The variance reduction for the far-regional travel times is significantly higher (57%) than for the near-regional (41%), which reflects the more significant changes to the upper mantle velocities that are due to the increased sensitivity to the velocity structure at those depths in our data.

The group delays show considerable variance reduction with respect to both the initial and AK135 models, especially at shorter periods ($T \leq 80$ s). The AK135 model has an especially poor

fit to the group-delay data at periods between 25 – 45 seconds, which reflects the lack of a reasonable crustal structure in our study region and the likely presence of noise in the group-velocity data at those periods. Our prior model corrects this bias even before we perform the inversion; however, the JWM model produces even better fits to the group velocity data compared to the initial model across the entire band of periods, especially at the short periods in our data set ($T \leq 20$ s).

Table 5: RMS error of the data sets with respect to the AK135, initial and final inversion models; various subsets of the data are also shown.

<i>Data Set</i>		<i>RMS Error AK135 (s)</i>	<i>RMS Error Starting Model (s)</i>	<i>RMS Error JWM (s)</i>
P Travel Times	all distances	2.50	2.87	2.02
	$\Delta = 0 - 12^\circ$	2.27	2.43	1.86
	$\Delta = 12 - \sim 18.6^\circ$	2.81	3.42	2.24
Group Delays	all periods (T)	60.0	35.2	30.2
	T = 10 - 20 s	46.9	37.5	28.2
	T = 25 - 45 s	82.2	37.6	36.2
	T = 50 - 80 s	45.7	28.8	24.1
	T = 90 - 150 s	28.3	27.8	24.7

5. DISCUSSION AND CONCLUSIONS

5.1 Tests for Predictive Capability

As we noted in the introduction, a primary objective of the study was to improve the quality of event locations using data from the region covered by our 3D model. We have employed two techniques to assess the improvement in location performance of the JWM crust/upper mantle velocity model. First, we have assessed the ability of JWM to predict regional travel-time observations for a set of ground-truth (GT) events. Our GT database of explosions and shallow earthquakes was derived from several sources, including the EHB bulletin, the Group2 Reference Event List (REL; Bondár *et al.*, 2004a) and a small supplemental list developed for an IASPEI location workshop (Engdahl, 2006). In Figure 12 we show the events in our GT database that are within the resolved boundaries of the JWM model (10-50 N, 40-110 E). The GT database contains high-quality epicenters, but the Group2 events do not include a large number of regional P and S observations. Therefore, we cross-referenced the Group2 REL events to the EHB bulletin (Engdahl *et al.*, 1998) to retrieve a larger set of regional observations. This filtering exercise produced a validation dataset of 183 explosions and 225 earthquakes, with 7,795 Pg, Pb, or Pn first arrivals and 2,252 Sg, Sb, or Sn arrivals observed at stations in the box (0-60 N, 30-120 E). The great-circle paths of the P and S GT observations are shown on the right side of

Figure 12. The region of the inversion model that is sampled by the GT data raypaths is limited, which illustrates the difficulty of validating a model in this manner.

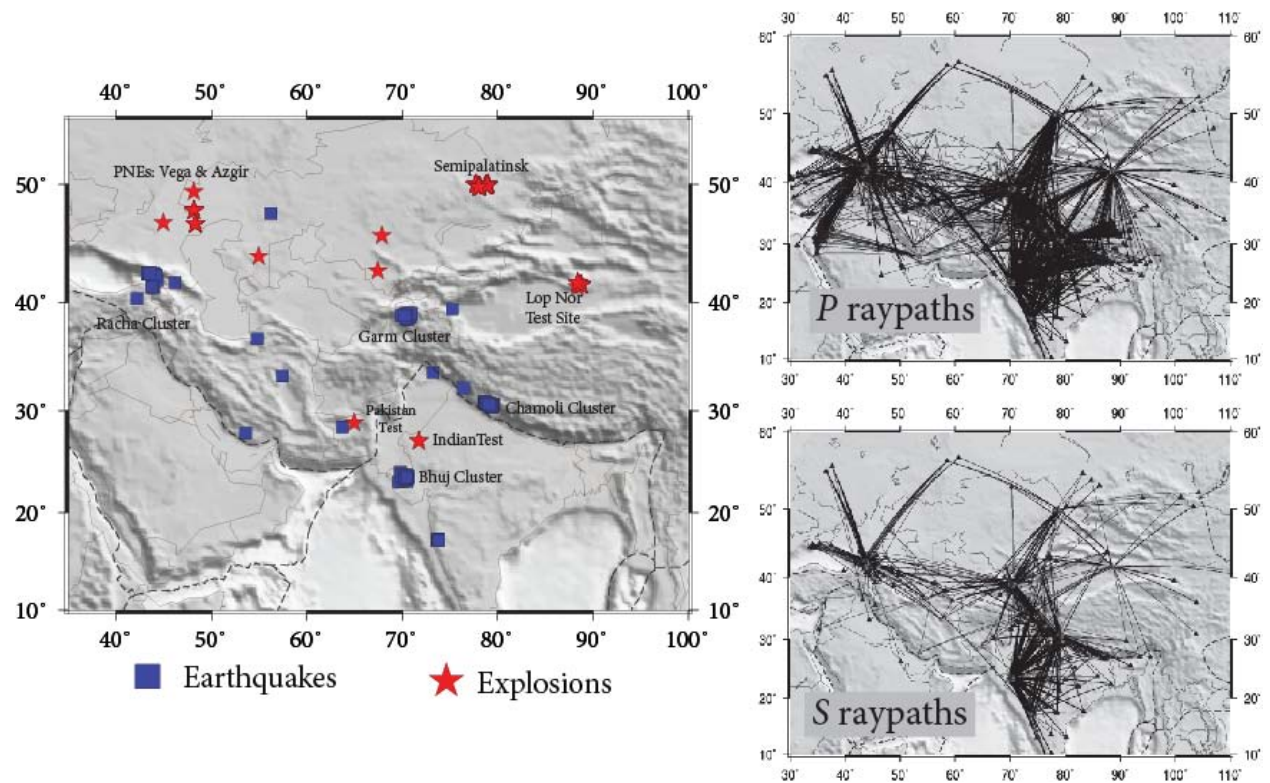


Figure 12: Ground-truth epicenters (GT0-GT7) in the resolved regions of the JWM model. Left: explosions (red stars) and earthquakes (blue squares) within our region; top right: great-circle raypaths for the EHB P observations associated with events in the GT dataset; and bottom right: great-circle S raypaths.

We then predict the travel times of our regional GT observations through both the AK135 and JWM models using our 3D finite-difference technique (Podvin and Lecomte, 1991) and determine residuals with respect to the GT observed data. We note that the travel-time residuals for the AK135 model were computed using the same finite-difference approach as for the JWM model, to minimize numerical effects that could influence the comparisons. In Figure 13 we show the results of this exercise, using residual density plots in which we binned and averaged the residuals as a function of station-to-event epicentral distance. The residuals were binned at 0.25° in epicentral distance and 0.5 seconds in time. The top two subplots in Figure 13 show the binned residuals relative to the AK135 model for the GT P (left column) and S (right column) travel-time data, while the bottom subplots show the results of the same calculation for the JWM model. We also overlaid the estimated residual median and spread at 1° increments in distance (blue dots and vertical lines, respectively). The specific bin counts are shown in the color scale at the bottom of each column; these were chosen to help emphasize the tails of the residual distributions.

The results for the P data clearly indicate the improved fit produced by the JWM model across all epicentral distances. The AK135 model is in general too fast at the shorter distances and too slow at the longer distances. The spread values in the residual plots also indicate that there are likely some arrivals in the groomed GT database that have been incorrectly included, especially at crossover distances between Pg and Pn, or between Pn and P.

The GT S data appear to suffer more significantly from the presence of bad picks in the arrival bulletin, especially at distances greater than approximately 8° . At the shorter distances it is clear, however, that the JWM model produces a better fit to the GT data. The scatter of the S GT data is not unexpected since regional S arrivals are often difficult to identify and pick, being often downweighted in location procedures as a result. We did not perform a rigorous outlier analysis of the GT S data, mainly because there is not enough of them in our ground-truth database. We relied instead on a simple hard cutoff of ± 10 seconds as a criterion for rejecting outlier residuals. This issue needs to be addressed in future efforts, since the GT S data are a valuable model validation resource.

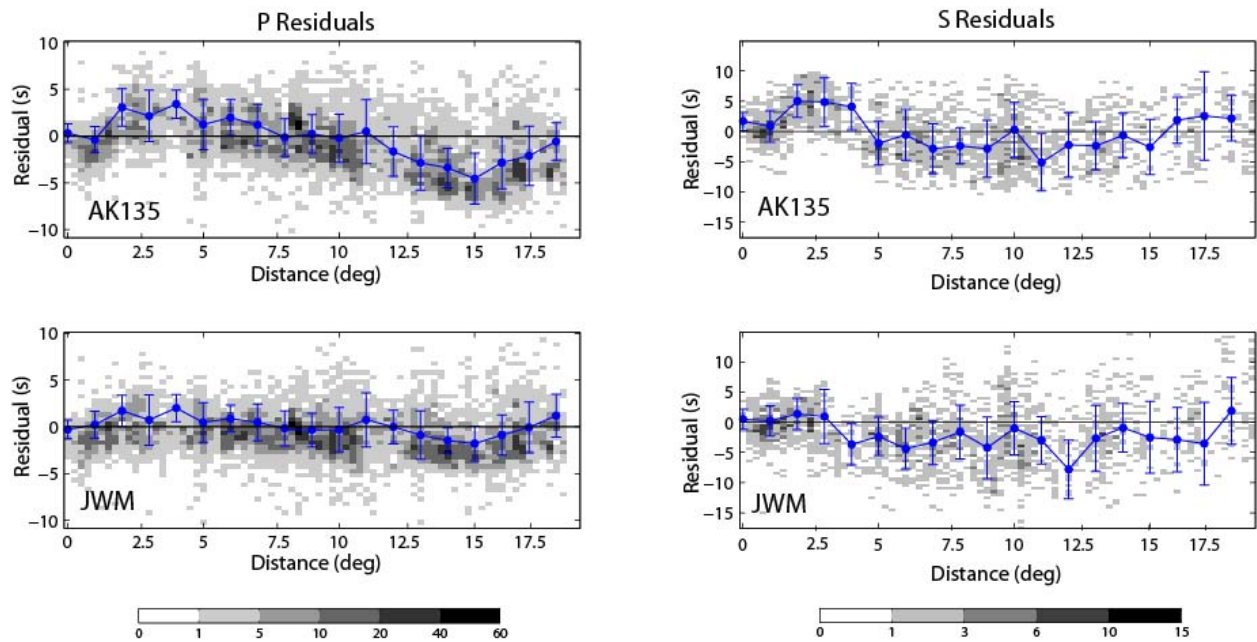


Figure 13: GT travel-time residuals binned as a function of epicentral distance for the AK135 (top) and JWM (bottom) models. Residual densities are binned at 0.25° intervals in distance and 0.5s in time. The bin hit counts are shown in the color scale at the bottom of each column, and the median and corresponding spread at 1° intervals are shown as blue circles and vertical lines. Subplots on the left show the P residuals, and subplots on the right show the S residuals.

In a second validation test we relocated a subset of events in our GT database to test the epicentral location accuracy of the JWM model. To eliminate some of the effects that network geometry can have on the solutions, we filtered the GT data to include only those events whose regional station distribution within our model has a secondary azimuthal gap less than 180° . Secondary azimuth gap (the largest azimuth gap when a single station of the network is

removed) is a good proxy for the quality of the network coverage (Bondár *et al.*, 2004b). We also removed stations that were within 2.5° of the event. This filtering reduced the testing data set to a list of 18 explosions and 112 earthquakes within our region, with 7,111 regional P-wave arrivals and 1,935 regional S-wave arrivals. Figure 14 shows the locations of our relocation events, color-coded according to the secondary azimuth gap of the regional arrivals.

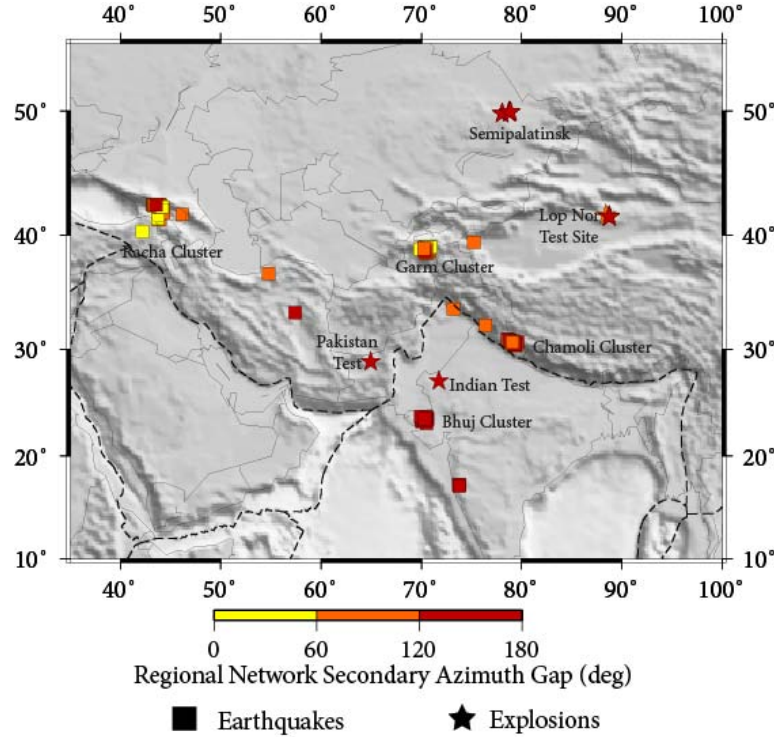


Figure 14: Map of events (112 earthquakes, 18 explosions) from our GT database that meet regional network criteria for the epicenter relocation exercise. The events are color-coded by their regional secondary azimuth gap. Several clusters of events are labeled by location.

We used the Grid-search Multiple-Event Location (GMEL) algorithm (Rodi, 2006) to relocate the events shown in Figure 14 with subsets of the regional arrival bulletin. The GT relocation experiments were done with the event depths fixed to their reported bulletin values and origin times allowed to vary. The arrival-time errors were set to 1.0 s for P and 2.0 s for S observations. Figure 15 directly compares the epicenter mislocations achieved using AK135 versus JWM travel-time predictions. Events that fall above the black unity line indicate that the JWM model relocates an event closer to the published GT value. In Figure 15a the results show that, for event locations based on P arrivals alone, JWM performs better than AK135 for all of the explosions and most of the earthquakes. We also performed relocations including both P and S arrivals, and the results in Figure 15b show that JWM still outperforms the AK135 model, although not as decisively as when only P arrivals are used. It is clear that adding the S arrivals has a slightly more harmful effect on the locations for JWM than for AK135. However, we note that JWM produces smaller mislocations overall.

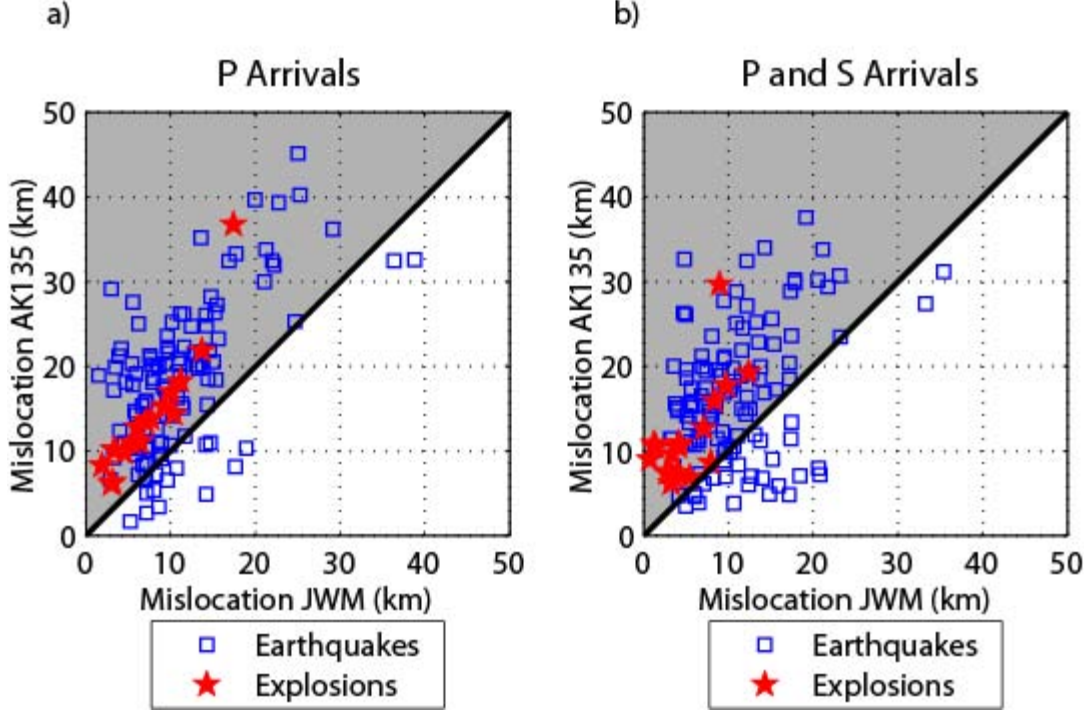


Figure 15: Epicenter mislocations resulting with the AK135 versus JWM models for the set of events (red stars: explosions; blue squares: earthquakes) shown in Figure 14. Events that fall in the gray shaded region above the black unity line in each plot indicate 'wins' for the JWM model. (a) Relocation results using regional P arrivals; (b) Relocation results using regional P and S arrivals.

5.2 Model Uncertainty

A final task in our project was to find techniques that can be used to calculate the uncertainty of the new JWM model. The Bayesian framework on which our joint inversion method is based provides a formal framework for model uncertainty, whereby model uncertainty is characterized by the posterior probability distribution on model parameters inferred from the prior distributions on the parameters and on data errors. A number of factors make it difficult to implement this framework in a complete way. First, for large, nonlinear inverse problems in general, it is only practical to calculate posterior parameter distributions under the approximation of linearization. In addition, our particular choice of algorithms presents further roadblocks to a complete calculation, even for the linearized problem. In particular, in solving the body-wave and surface-wave linearized problems sequentially, rather than simultaneously, we preclude a full Bayesian uncertainty analysis for the joint body-wave/surface-wave inverse problem. Moreover, our factoring of the surface-wave problem stymies a full Bayesian uncertainty analysis of that problem, even considered alone. For these reasons, our uncertainty analysis as of this writing is restricted to the linearized body-wave travel-time tomography problem, without consideration of the information the surface-wave data may provide about P velocity or of the potential trade-offs among P velocity, S velocity and crustal thickness.

As described in Section 2.7, our application of the Bayesian framework employs a geostatistical parameterization of the prior model variance, involving a correlation kernel

$C_0(\mathbf{x}, \mathbf{x}')$ to specify the correlation coefficient between the model function – in this case P velocity – at any two spatial points $\mathbf{x}$ and $\mathbf{x}'$. The correlation kernel is parameterized in terms of correlation lengths in the horizontal and vertical directions. In addition, a function $\sigma_m(\mathbf{x})$ specifies the prior standard deviation (root variance) of the model function at each position. The Bayesian methodology for linear inverse problems straightforwardly defines posterior versions of these same quantities: $C_0^{pos}(\mathbf{x}, \mathbf{x}')$ and $\sigma_m^{pos}(\mathbf{x})$. As part of our linear inversion algorithms, we developed numerical techniques for computing the prior and posterior variance of the model function and a “slice” of the prior and posterior correlation kernels for a particular fiducial points $\mathbf{x}_0$, e.g. $C_0(\mathbf{x}, \mathbf{x}_0)$ as a function of $\mathbf{x}$.

Figures 16 and 17 show examples of model correlation slices for various fiducial points, computed for the JWM P-wave velocity inversion model. In Figure 16 we show prior (top) and posterior (middle) correlation slices for two fiducial points near the center of our study region (30° N, 75° E) at depths 60 km and 120 km. Note the de-correlation between the crust and upper mantle velocity in the prior slice, which is an intentional feature discussed in Section 2.7.

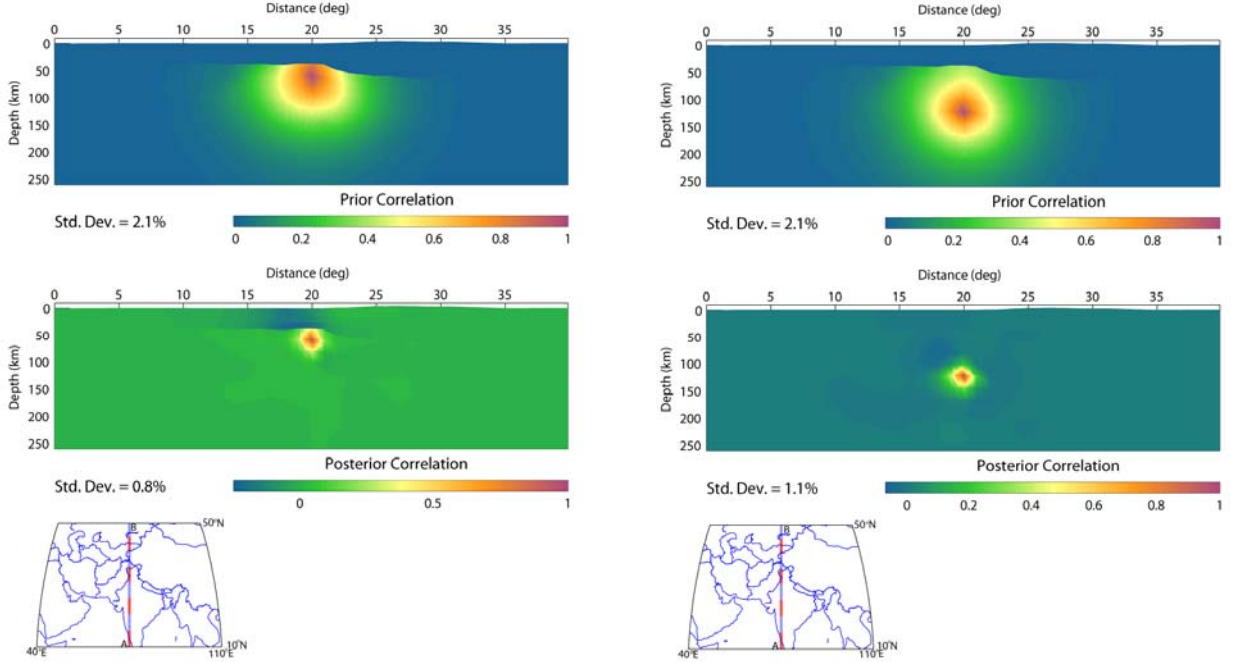


Figure 16: Vertical cross-sections through prior (top) and posterior (middle) correlation slices for the JWM P-wave model. Slices are shown for two fiducial points in the middle of the study region (30° N, 75° E) at depths of 60 km (left) and 120 km (right). Maps at the bottom of the subplots show the selected cross-sections as the AB lines. The number label “Std. Dev.” is the model standard deviation at the fiducial point.

The posterior correlation slice includes the information from travel-time data, leading to a reduction in model variance and in the correlation lengths implied by the correlation decay from the fiducial point. Strikingly, some posterior correlations are negative, especially between crust and mantle velocities (e.g. the left/middle plot).

In contrast to Figure 16, the correlation slices in Figure 17 correspond to fiducial points near the easternmost side of our study region (15° N, 105° E) and (45° N, 105° E) at 120 km depth. In these cases the prior and posterior correlations and standard deviations are more similar to each other because there is a dearth of travel-time data to constrain the model at these points.

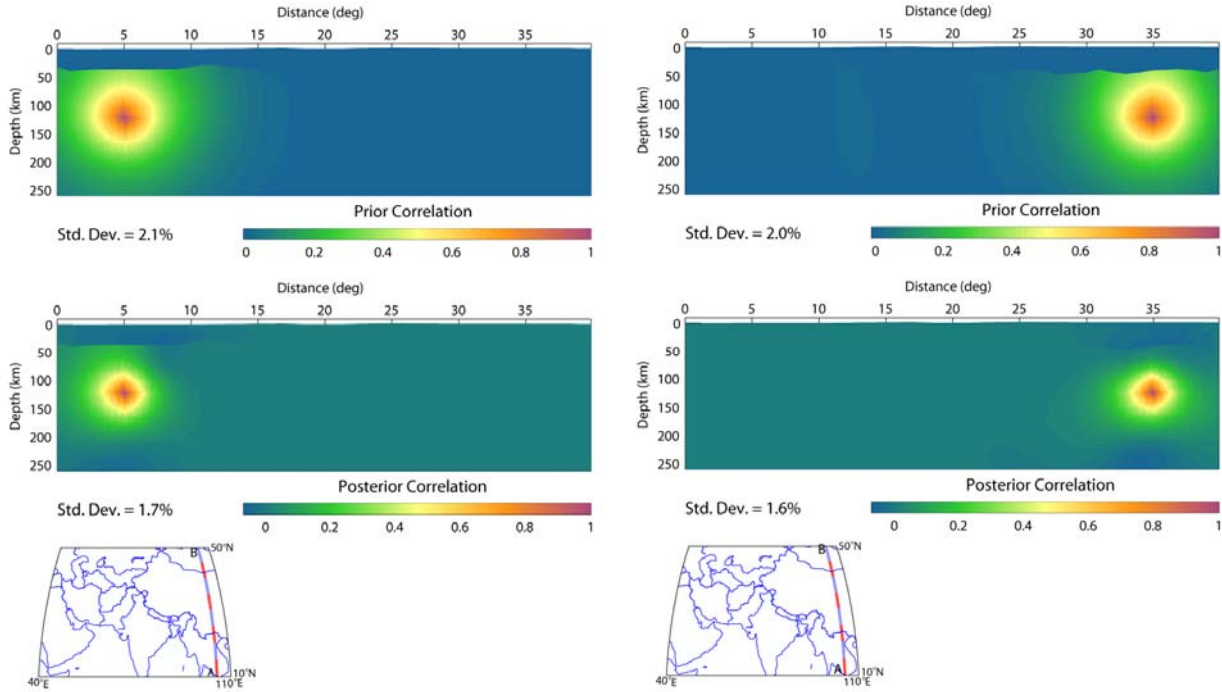


Figure 17: More examples of vertical cross-sections through prior (top) and posterior (middle) correlation slices for the JWM P-wave model. Slices are shown for two fiducial points (15° N, 105° E) and (45° N, 105° E) on the eastern edge of the study region at 120 km depth. Plotting conventions are the same as in Figure 16.

5.3 Summary

In this study we have demonstrated the application of a nonlinear joint inversion of body-wave travel times and surface-wave group velocities to data from a broad region in south-central Asia. The forward modeling incorporated in our inversion utilizes fully 3D ray tracing for the body-wave travel times, and a two-step procedure for the surface-wave dispersion data that includes 1-D dispersion modeling at a geographic point followed by 2-D ray tracing in the phase velocity maps. We numerically solved the inverse problem using a set of iterated inversion steps. Consistency between the P and S velocities was achieved by imposing bounds on Poisson's ratio and by invoking a regularization constraint that correlates variations in P and S velocity from an initial model. The resulting inversion model shows good agreement with the persistent features seen in previous seismic velocity models produced from separate body- or surface-wave data sets, as well as some intriguing differences between the compressional and shear wave structure.

Our primary goal in developing the new inversion approach and the JWM model was to improve regional seismic event location capability in a strongly heterogeneous crust and upper mantle. We have tested the JWM model for its predictive capabilities using data from a large database of ground-truth events, which were held out from the joint inversion. The tests include the relocation of the ground-truth events, using data sets of Pn-only and Pn/Sn arrivals, and the direct comparison of predicted Pn and Sn travel times to the ground-truth observations. Both types of tests indicate that our 3D inversion model has much better predictive capability than either a 1-D global model or our initial model, which comprised a 3D crustal structure overlying a 1-D mantle. Our S velocity model performed reasonably well in predicting Sn times, even though such data were not included in our joint inversion.

6. ACKNOWLEDGMENTS

There are many researchers to thank for their generous contribution of data to the project. These include Anatoli Levshin and Michael Pasyanos for group velocity measurements over our region of interest, Bob Engdahl for an updated version of his the EHB bulletin, Chang Li for the Tibetan experiment data, and Istvan Bondár for the ground-truth data used in the validation of the inversion model. All maps were generated using the Generic Mapping Tools (GMT) data processing and display package (Wessel and Smith, 1995). This work was supported and performed under Contract No. FA8718-04-C-0027 from the Air Force Research Laboratory.

7. REFERENCES

- Ammon, C. J., G. E. Randall, and G. Zandt (1990). On the nonuniqueness of receiver function inversions, *J. Geophys. Res.*, 95, 15,303 – 15,318.
- Antolik, M., Y. J. Gu, G. Ekström, and A. M. Dziewonski (2003). J362D28: a new joint model of compressional and shear velocity in the Earth's mantle, *Geophys. J. Int.*, 153, 443-466.
- Baig, A. M., and F. A. Dahlen (2004). Traveltime biases in random media and the S-wave discrepancy, *Geophys. J. Int.*, 158, 922 – 938.
- Birch, F. (1961). The velocity of compressional waves in rocks to 10 kilo bars, Part 2, *J. Geophys. Res.*, 66, 2,199 – 2,224.
- Bondár, I., E. R. Engdahl, X. Yang, H. A. A. Ghalib, A. Hofstetter, V. Kirichenko, R. Wagner, I. Gupta, G. Ekström, E. Bergman, H. Israelsson and K. McLaughlin (2004a). Collection of a reference event set for regional and teleseismic location calibration, *Bull. Seis. Soc. Amer.*, **94**, 1528 – 1545.
- Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004b). Epicentre accuracy based on seismic network criteria, *Bull. Seis. Soc. Amer.*, **156**, 483 – 496.
- Bukchin, B.G., T.B. Yanovskaya, J.-P. Montagner, A.Z. Mostinskiy and E. Beucler (2006). Surface wave focusing effects: Numerical modeling and statistical observations, *Phys. Earth. Planet. Int.* 155, 191 – 200.
- Deutsch, C.V. and A.G. Journel (1998). *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford University Press, Inc., New York, 369 pp.
- Engdahl, E.R. (2006), Application of an improved algorithm to high precision relocation of ISC test events, *Phys. Earth. Planet. Int.*, 158, 14 – 18.
- Engdahl E. R., R. van der Hilst, and R. Buland (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Geophys. J. Int.*, 88, 722 – 743.
- Flanagan, M. P., S. C. Myers, and K. D. Koper (2007). Regional travel-time uncertainty and seismic location improvement using a three-dimensional *a priori* velocity model, *Bull. Seis. Soc. Amer.*, 97, 804 – 825.
- Harkrider, D. G. (1968). The perturbation of Love wave spectra, *Bull. Seis. Soc. Am.*, 58, 861--880.
- Harkrider, D. G. and D. L. Anderson (1966). Surface wave energy from point sources in plane layered earth models, *J. Geophys. Res.*, 71, 2967--2980.
- Herrmann, R. B. (2002). *Computer Programs in Seismology*, Version 3.20, <http://www.eas.slu.edu/People/RBHerrmann/CPS32.html> (last accessed June 2003)
- Julia, J., C. J. Ammon, R. B. Herrmann, and A. M. Correig (2000). Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.*, 143, 99-112.
- Kennett, B. L.N. Engdahl, E. R., and Buland R. (1995). Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.*, 122, 108 – 124.

- Laske, G., Masters, G. and Reif, C. (2001). CRUST2.0 – a new global crustal model at 2×2 degrees, <http://mafi.ucsd.edu/Gabi/crust2.html> (last accessed 02/2009).
- Leveque, J. J., L. Rivera, and G. Wittlinger (1993). On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, *115*, 313 – 318.
- Li, C., R. D. van der Hilst, A. S. Meltzer, and E. R. Engdahl (2008). Subduction of the Indian lithosphere beneath the Tibetan Plateau and Burma, *Earth and Planet. Sci. Ltrs.*, *274*, 157 – 168.
- Li, C., R. D. van der Hilst, and M. N. Toksöz (2006). Constraining P wave velocity variations in the upper mantle beneath Southeast Asia, *Phys. Earth Planet. Int.*, *154*, 180 – 195.
- Liu, H.-P., D. L. Anderson, and H. Kanimori (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition, *Geophys. J. R. Astr. Soc.*, *47*, 41 – 58.
- Lomax, A., J. Virieux, Volant P., and Berge, C. (2000). Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location*, Thurber, C.H., and N. Rabinowitz (eds.), Kluwer, Amsterdam, 101–134.
- Masters, G., G. Laske, H. Bolton and A. Dziewonski (2000). The relative behavior of shear velocity, bulk sound speed and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale*, *Geophys. Monograph Ser.*, *117*, Karato et al. (eds.), AGU, Washington, DC, 63 – 87.
- Nolet, G., R. Montelli, and J. Virieux (1999). Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, *138*, 36 – 44.
- Mitra, S., K Priestley, A. J. K. Bhattacharyya, and V. K. Gaur (2005). Crustal structure and earthquake focal depths beneath northeastern India and southern Tibet, *Geophys. J. Int.*, *160*:11, 227 – 248.
- Pasyanos, M. E. (2005). A variable-resolution surface wave dispersion study of Eurasia, North Africa, and surrounding regions, *J. Geophys. Res.*, *110*: B12301, doi:10.1029/2005JB003749.
- Pasyanos, M. E., W. R. Walter, M. P. Flanagan, P. Goldstein and J. Bhattacharyya (2004a). Building and testing an a prior geophysical model for Western Eurasia and North Africa., *Pure Appl. Geophys.*, *161*, 235 – 281.
- Podvin, P., and I. Lecomte (1991). Finite difference computation of travel times in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.*, *105*, 271 – 284.
- Reiter, D., W. Rodi, and M. Johnson (2005). Development of a tomographic upper-mantle velocity model beneath Pakistan and northern India for improved regional seismic event location,, *95*, 926 – 940.
- Ritzwoller, M. H., and A. L. Levshin (1998). Eurasian surface wave tomography: Group velocities, *J. Geophys Res.*, *103*, B3, 4839 – 4878.

- Ritzwoller, M. H., N. M. Shapiro, A. L. Levshin, E. A. Bergman, and E. R. Engdahl (2003). Ability of a global three-dimensional model to locate regional events, *J. Geophys. Res.*, 108, doi:10.1029/2002JB002157.
- Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin and A. L. Levshin (2002). Global surface wave diffraction tomography, *J. Geophys. Res.*, 107, doi: 10.1029/2002JB001777.
- Rodi, W. (2006). Grid-search event location with non-Gaussian error models, *Phys. Earth Planet. Int.*, 158, 55 – 66.
- Rodi, W., and R. L. Mackie (2001). Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, 66, 174 – 187.
- Shapiro, N. M., and M. H. Ritzwoller (2002). Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.*, 151, 88 – 105.
- Su, W. and A. M. Dziewonski (1997). Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth Planet. Int.*, 100, 135-156.
- Tarantola, A. (2005). *Inverse Problem Theory*, Soc. Ind. Appl. Math., 342 pp.
- Tikhonov, A. N., and V. Y. Arsenin (1977). *Solutions of Ill-Posed Problems*. V.H. Winston and Sons (distributed by Wiley, New York), Washington, DC.
- Wessel, P., and W. H. F. Smith (1995). New Version of the Generic Mapping Tools Released, *EOS Trans. AGU*, 76.
- Woodhouse, J., and A. Dziewonski (1984). Mapping the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, 89(B7), 5953-5986.
- Yu, G.-Y., and B. J. Mitchell (1979). Regionalized shear velocity models of the Pacific upper mantle from observed Love and Rayleigh wave dispersion, *Geophys. J. R. Astr. Soc.*, 57, 311-341.

8. APPENDIX A: Factored Inversion

We consider an inverse problem for $\mathbf{m}$ of the form

$$\mathbf{d} = F(G(\mathbf{m})) + \mathbf{e}, \quad (\text{A1})$$

where $\mathbf{d}$ is a data vector, $\mathbf{m}$ is a model vector, $\mathbf{e}$ is an error vector, and F and G are nonlinear transformations whose composition is the forward model for the problem. We take the solution of (A1) to be the model minimizing the objective function

$$\Psi(\mathbf{m}) = (\mathbf{d} - F(G(\mathbf{m})))^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m}))) + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{D}(\mathbf{m} - \mathbf{m}_0). \quad (\text{A2})$$

In the context of stochastic inversion, $\mathbf{m}_0$ is the prior mean of $\mathbf{m}$, $\mathbf{D}$ is the inverse of the prior covariance operator of $\mathbf{m}$, and $\mathbf{W}$ is the inverse of the prior covariance operator of $\mathbf{e}$.

Define the objective functions

$$\Psi_1(\mathbf{u}) = (\mathbf{d} - F(\mathbf{u}))^T \mathbf{W}(\mathbf{d} - F(\mathbf{u})) + (\mathbf{u} - G(\mathbf{m}_0))^T \mathbf{D}_1(\mathbf{u} - G(\mathbf{m}_0)) \quad (\text{A3})$$

$$\Psi_2(\mathbf{m}) = (\mathbf{u} - G(\mathbf{m}))^T \mathbf{W}_2(\mathbf{u} - G(\mathbf{m})) + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{D}_2(\mathbf{m} - \mathbf{m}_0). \quad (\text{A4})$$

We will show that, under linear approximations to F and G and for suitable choices of the operators $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{W}_2$, the vector $\mathbf{m}$ minimizing Ψ_2 , where $\mathbf{u}$ minimizes Ψ_1 , is the same vector that minimizes Ψ in (A2). This in effect factors the inverse problem of Equation (A1) into two inverse problems having forward transformations F and G , respectively:

$$\mathbf{d} = F(\mathbf{u}) + \mathbf{e}_1 \quad (\text{A5})$$

$$\mathbf{u} = G(\mathbf{m}) + \mathbf{e}_2. \quad (\text{A6})$$

We see that $\mathbf{u}$ acts as the model vector in the first problem and the data vector in the second.

We start by expressing the composite inverse problem as the stationarity condition obtained by equating the gradient of Ψ to zero:

$$\mathbf{D}(\mathbf{m} - \mathbf{m}_0) = \mathbf{B}^T \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m}))), \quad (\text{A7})$$

where the linear transformation $\mathbf{A}$ is the Jacobian of F evaluated at $G(\mathbf{m})$, and $\mathbf{B}$ is the Jacobian of G evaluated at $\mathbf{m}$. The stationarity conditions for the factor problems are similarly given by

$$\mathbf{D}_1(\mathbf{u} - G(\mathbf{m}_0)) = \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(\mathbf{u})) \quad (\text{A8})$$

$$\mathbf{D}_2(\mathbf{m} - \mathbf{m}_0) = \mathbf{B}^T \mathbf{W}_2(\mathbf{u} - G(\mathbf{m})). \quad (\text{A9})$$

Presuming that the same value of $\mathbf{m}$ satisfies Equations (A9) and (A7), as it is our objective to prove, the Jacobian of G in each equation is the same, allowing us to denote both as $\mathbf{B}$. However, the value of $\mathbf{u}$ satisfying (A8) is not necessarily $G(\mathbf{m})$, in which case the Jacobian of F , shown as $\mathbf{A}$ in both equations, may differ. Therefore, we will replace F henceforth with its linear approximation, such that

$$F(\mathbf{u}) \approx F(G(\mathbf{m})) + \mathbf{A}(\mathbf{u} - F(G(\mathbf{m}))), \quad (\text{A10})$$

where, as in Equation (A7), $\mathbf{A}$ is the Jacobian of F evaluated at $G(\mathbf{m})$. Using (A10), the stationarity condition for the first factor inverse problem becomes

$$\mathbf{D}_1(\mathbf{u} - G(\mathbf{m}_0)) = \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m})) - \mathbf{A}\mathbf{u} + \mathbf{A}G(\mathbf{m})). \quad (\text{A11})$$

Equation (A11) is solved by

$$\begin{aligned} \mathbf{u} &= G(\mathbf{m}_0) + (\mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{D}_1)^{-1} \\ &\quad \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m})) + \mathbf{A}G(\mathbf{m}) - \mathbf{A}G(\mathbf{m}_0)) \end{aligned} \quad (\text{A12})$$

which implies

$$\begin{aligned} \mathbf{u} - G(\mathbf{m}) &= (\mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{D}_1)^{-1} \\ &\quad [\mathbf{D}_1(G(\mathbf{m}_0) - G(\mathbf{m})) + \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m})))]. \end{aligned} \quad (\text{A13})$$

Substituting this into (A9) yields

$$\begin{aligned} \mathbf{D}_2(\mathbf{m} - \mathbf{m}_0) &= \mathbf{B}^T \mathbf{W}_2 (\mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{D}_1)^{-1} \\ &\quad [\mathbf{D}_1(G(\mathbf{m}_0) - G(\mathbf{m})) + \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m})))]. \end{aligned} \quad (\text{A14})$$

Now let

$$\mathbf{W}_2 = \mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{D}_1. \quad (\text{A15})$$

Equation (A14) becomes

$$\mathbf{D}_2(\mathbf{m} - \mathbf{m}_0) = \mathbf{B}^T \mathbf{D}_1(G(\mathbf{m}_0) - G(\mathbf{m})) + \mathbf{B}^T \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m}))). \quad (\text{A16})$$

Up to this point we have not needed to approximate the function G , but now we do. Linearizing about $\mathbf{m}$, let

$$G(\mathbf{m}_0) \approx G(\mathbf{m}) + \mathbf{B}(\mathbf{m}_0 - \mathbf{m}). \quad (\text{A17})$$

Applying this in the first term of Equation (A16), we get

$$\mathbf{D}_2(\mathbf{m} - \mathbf{m}_0) = \mathbf{B}^T \mathbf{D}_1 \mathbf{B}(\mathbf{m}_0 - \mathbf{m}) + \mathbf{B}^T \mathbf{A}^T \mathbf{W}(\mathbf{d} - F(G(\mathbf{m}))). \quad (\text{A18})$$

This becomes Equation (A7), the stationarity condition for the composite inverse problem, by setting

$$\mathbf{D} = \mathbf{B}^T \mathbf{D}_1 \mathbf{B} + \mathbf{D}_2. \quad (\text{A19})$$

The equivalence of the composite inverse problem and factor inverse problems occurs when $\mathbf{W}_2$, $\mathbf{D}_1$ and $\mathbf{D}_2$ are chosen in accordance with (A15) and (A19). Equation (A15) sets the prior covariance operator of $\mathbf{e}_2$ in the second factor problem to the posterior covariance of $\mathbf{u}$ resulting from the first inverse problem. Equation (A19) relates the prior model covariances of the composite and factor problems.

9. APPENDIX B: Conditional Inversion

We consider an objective function for a nonlinear inverse problem for parameter vectors $\mathbf{a}$ and $\mathbf{b}$, given by

$$\begin{aligned}\Psi(\mathbf{a}, \mathbf{b}) &= (\mathbf{t} - F_t(\mathbf{a}))^T \mathbf{W}_t (\mathbf{t} - F_t(\mathbf{a})) + (\mathbf{g} - F_g(\mathbf{b}))^T \mathbf{W}_g (\mathbf{g} - F_g(\mathbf{b})) \\ &\quad + (\mathbf{a} - \mathbf{a}_0)^T \mathbf{D}_{aa} (\mathbf{a} - \mathbf{a}_0) + (\mathbf{a} - \mathbf{a}_0)^T \mathbf{D}_{ab} (\mathbf{b} - \mathbf{b}_0) \\ &\quad + (\mathbf{b} - \mathbf{b}_0)^T \mathbf{D}_{ba} (\mathbf{a} - \mathbf{a}_0) + (\mathbf{b} - \mathbf{b}_0)^T \mathbf{D}_{bb} (\mathbf{b} - \mathbf{b}_0).\end{aligned}\quad (\text{B1})$$

The vectors $\mathbf{a}_0$ and $\mathbf{b}_0$ play the role of prior means of $\mathbf{a}$ and $\mathbf{b}$, respectively. Denoting

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{aa} & \mathbf{D}_{ab} \\ \mathbf{D}_{ba} & \mathbf{D}_{bb} \end{pmatrix}, \quad (\text{B2})$$

$\mathbf{D}^{-1}$ plays the role of a variance/covariance matrix on $\mathbf{a}$ and $\mathbf{b}$. We will show that the conditional minimum of Ψ with respect to $\mathbf{a}$, holding $\mathbf{b}$ fixed, also minimizes the objective function

$$\Psi^*(\mathbf{a}) = (\mathbf{t} - F_t(\mathbf{a}))^T \mathbf{W}_t (\mathbf{t} - F_t(\mathbf{a})) + (\mathbf{a} - \mathbf{a}_0^*)^T \mathbf{D}_{aa}^* (\mathbf{a} - \mathbf{a}_0^*) \quad (\text{B3})$$

for appropriate choices of $\mathbf{a}_0^*$ and $\mathbf{D}_{aa}^*$.

While this can be shown in general, we will restrict attention to a parameter variance/covariance matrix of the form

$$\mathbf{D}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_b \end{pmatrix} \begin{pmatrix} \mathbf{C}_0 & \mathbf{R}\mathbf{C}_0 \\ \mathbf{R}\mathbf{C}_0 & \mathbf{C}_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_b \end{pmatrix}, \quad (\text{B4})$$

where $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ are diagonal matrices containing the standard deviations of the components of $\mathbf{a}$ and $\mathbf{b}$, respectively; $\mathbf{C}_0$ is a correlation matrix for both $\mathbf{a}$ and $\mathbf{b}$; and $\mathbf{R}$ is a diagonal matrix of correlation coefficients between $\mathbf{a}$ and $\mathbf{b}$. We require that $\mathbf{R}$ and $\mathbf{C}_0$ commute:

$$\mathbf{R}\mathbf{C}_0 = \mathbf{C}_0\mathbf{R}. \quad (\text{B5})$$

Further, we restrict $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ as indicated in Equation (B9) below. Given (B4), we have

$$\begin{pmatrix} \mathbf{D}_{aa} & \mathbf{D}_{ab} \\ \mathbf{D}_{ba} & \mathbf{D}_{bb} \end{pmatrix} = \mathbf{S}^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_a^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_b^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{D}_0 & -\mathbf{R}\mathbf{D}_0 \\ -\mathbf{R}\mathbf{D}_0 & \mathbf{D}_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_a^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_b^{-1} \end{pmatrix} \quad (\text{B6})$$

where $\mathbf{D}_0 = \mathbf{C}_0^{-1}$ and $\mathbf{S} = \mathbf{I} - \mathbf{R}^2$. Note that Equation (B5) ensures that $\mathbf{R}$ and $\mathbf{S}$ each commute with $\mathbf{D}_0$ (as they do with $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$).

To find $\mathbf{a}_0^*$ and $\mathbf{D}_{aa}^*$, we start by writing the stationarity condition implied by setting the gradient of Ψ with respect to $\mathbf{a}$ to zero:

$$\mathbf{D}_{aa}(\mathbf{a} - \mathbf{a}_0) + \mathbf{D}_{ab}(\mathbf{b} - \mathbf{b}_0) = \mathbf{A}_t^T \mathbf{W}_t (\mathbf{t} - F_t(\mathbf{a})). \quad (\text{B7})$$

The conditional minimum of Ψ , for any fixed $\mathbf{b}$, is achieved by solving (B7) for $\mathbf{a}$. Equation (B6) allows us to write

$$\mathbf{D}_{aa}(\mathbf{a} - \mathbf{a}_0) + \mathbf{D}_{ab}(\mathbf{b} - \mathbf{b}_0) = \mathbf{S}^{-1} \boldsymbol{\Sigma}_a^{-1} \mathbf{D}_0 [\boldsymbol{\Sigma}_a^{-1}(\mathbf{a} - \mathbf{a}_0) - \mathbf{R}\boldsymbol{\Sigma}_b^{-1}(\mathbf{b} - \mathbf{b}_0)]. \quad (\text{B8})$$

Assume the parameter standard deviations and prior means are in a common ratio:

$$\boldsymbol{\Sigma}_a^{-1}\mathbf{a}_0 = \boldsymbol{\Sigma}_b^{-1}\mathbf{b}_0. \quad (\text{B9})$$

Then

$$\mathbf{D}_{aa}(\mathbf{a} - \mathbf{a}_0) + \mathbf{D}_{ab}(\mathbf{b} - \mathbf{b}_0) = \mathbf{D}_{aa}^*(\mathbf{a} - \mathbf{a}_0^*) \quad (\text{B10})$$

where

$$\mathbf{a}_0^* = (\mathbf{I} - \mathbf{R})\mathbf{a}_0 + \mathbf{R}\boldsymbol{\Sigma}_a\boldsymbol{\Sigma}_b^{-1}\mathbf{b} \quad (\text{B11})$$

$$\mathbf{D}_{aa}^* = \mathbf{S}^{-1}\boldsymbol{\Sigma}_a^{-1}\mathbf{D}_0\boldsymbol{\Sigma}_a^{-1} = \mathbf{S}^{-1}\mathbf{D}_{aa}. \quad (\text{B12})$$

Plugging (B10) into the stationarity condition, (B7), gives

$$\mathbf{D}_{aa}^*(\mathbf{a} - \mathbf{a}_0^*) = \mathbf{A}_t^T \mathbf{W}_t(\mathbf{t} - F_t(\mathbf{a})), \quad (\text{B13})$$

which is the stationarity condition associated with the objective function Ψ^* in Equation (B3).

Equation (B11) states that $\mathbf{a}_0^*$ is a weighted average of $\mathbf{a}_0$ and the scaled version of $\mathbf{b}$ given by $\boldsymbol{\Sigma}_a\boldsymbol{\Sigma}_b^{-1}\mathbf{b}$. The relative weighting of the two vectors is determined by the correlation-coefficient matrix $\mathbf{R}$, with $\mathbf{R} = 0$ (no correlation between $\mathbf{a}$ and $\mathbf{b}$) resulting in $\mathbf{a}_0^* = \mathbf{a}_0$, as one should expect. Equation (B12) states that $\mathbf{D}_{aa}^*$ is $\mathbf{D}_{aa}$ inflated by $\mathbf{S}^{-1}$, where $\mathbf{S} = \mathbf{I}$ when $\mathbf{R} = 0$.

10. APPENDIX C: Additional Inversion Results

In the next four figures we show the map view slices of the JWM model at 30-km intervals, starting at 30 km below the surface. The results are provided in percent deviation from the AK135 model.

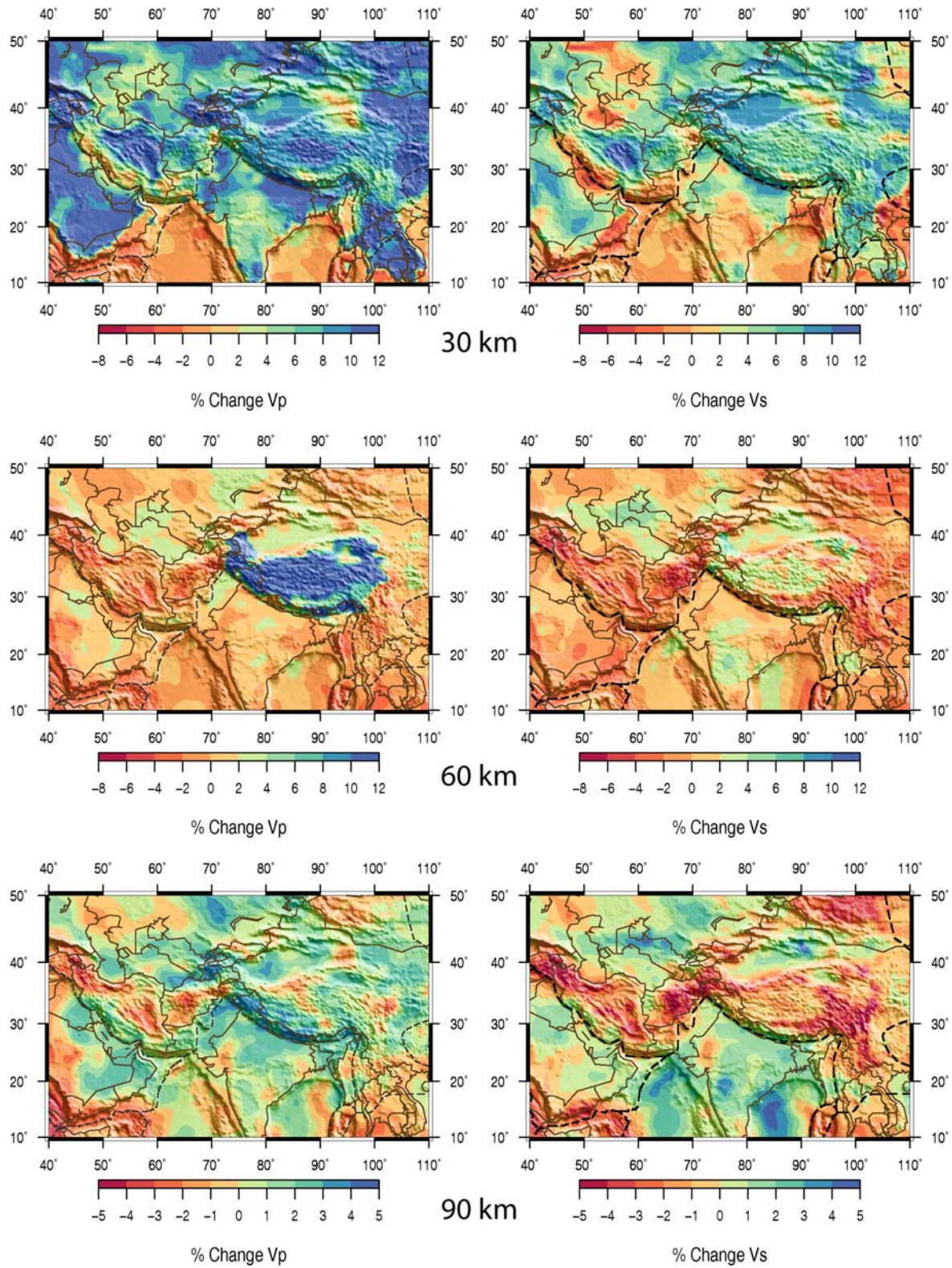


Figure C1. Map view slices of the JWM model in percent deviation with respect to AK135 at 30 km, 60 km, and 90 km depth. P velocities are shown on the left and S velocities on the right.

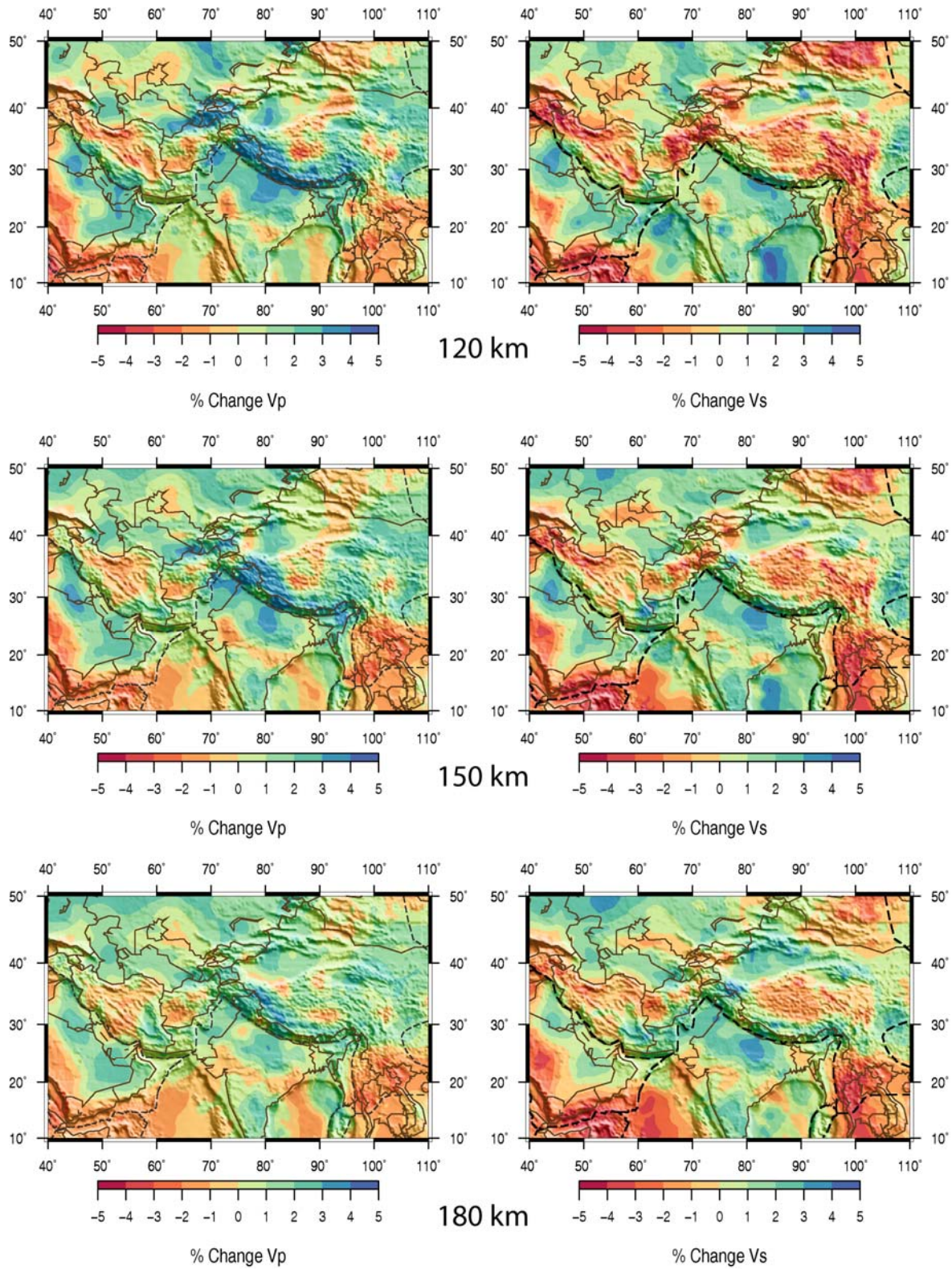


Figure C2. Map view slices of the JWM model in percent deviation with respect to AK135 at 120 km, 150 km, and 180 km depth. P velocities are shown on the left and S velocities on the right.

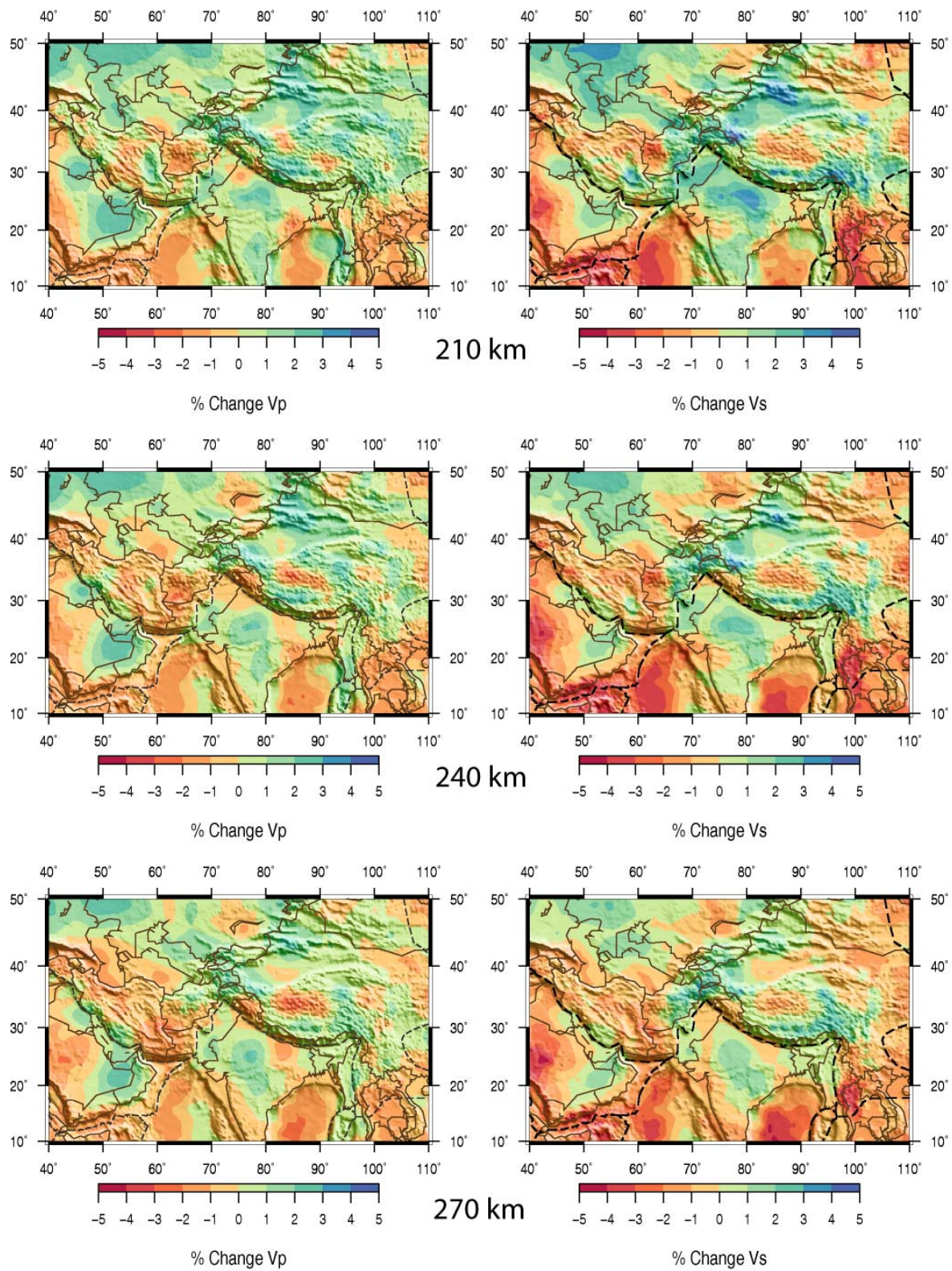


Figure C3. Map view slices of the JWM model in percent deviation with respect to AK135 at 210 km, 240 km, and 270 km depth. P velocities are shown on the left and S velocities on the right.

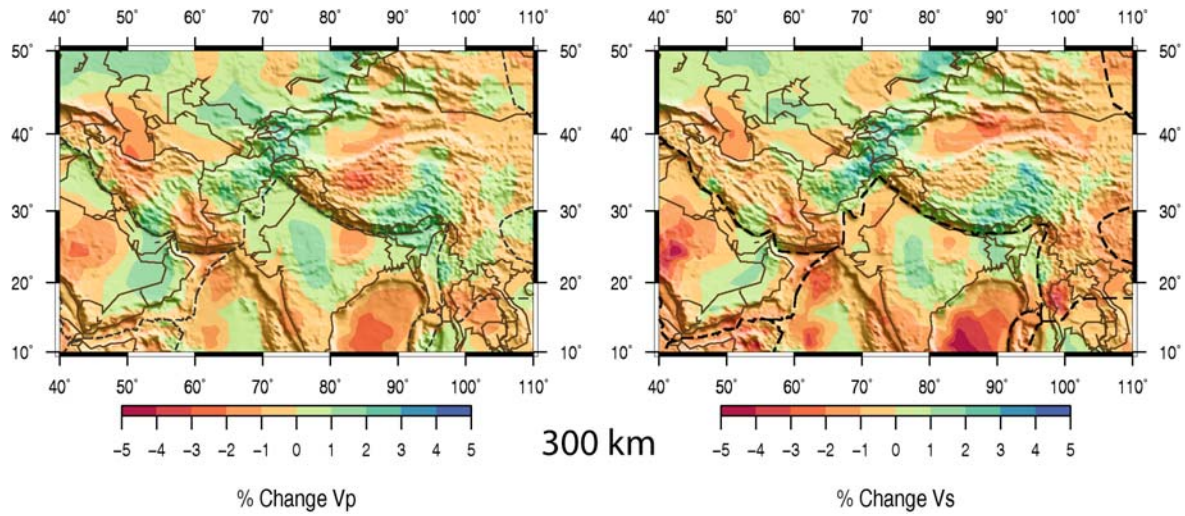


Figure C4. Map view slices of the JWM model in percent deviation with respect to AK135 at 300 km depth. P velocities are shown on the left and S velocities on the right.

Next we show the results of the checkerboard test, again by plotting map view slices of the retrieved checkerboard model at 30-km intervals, starting at 30 km below the surface. The results are provided in percent deviation from the final JWM model velocities.

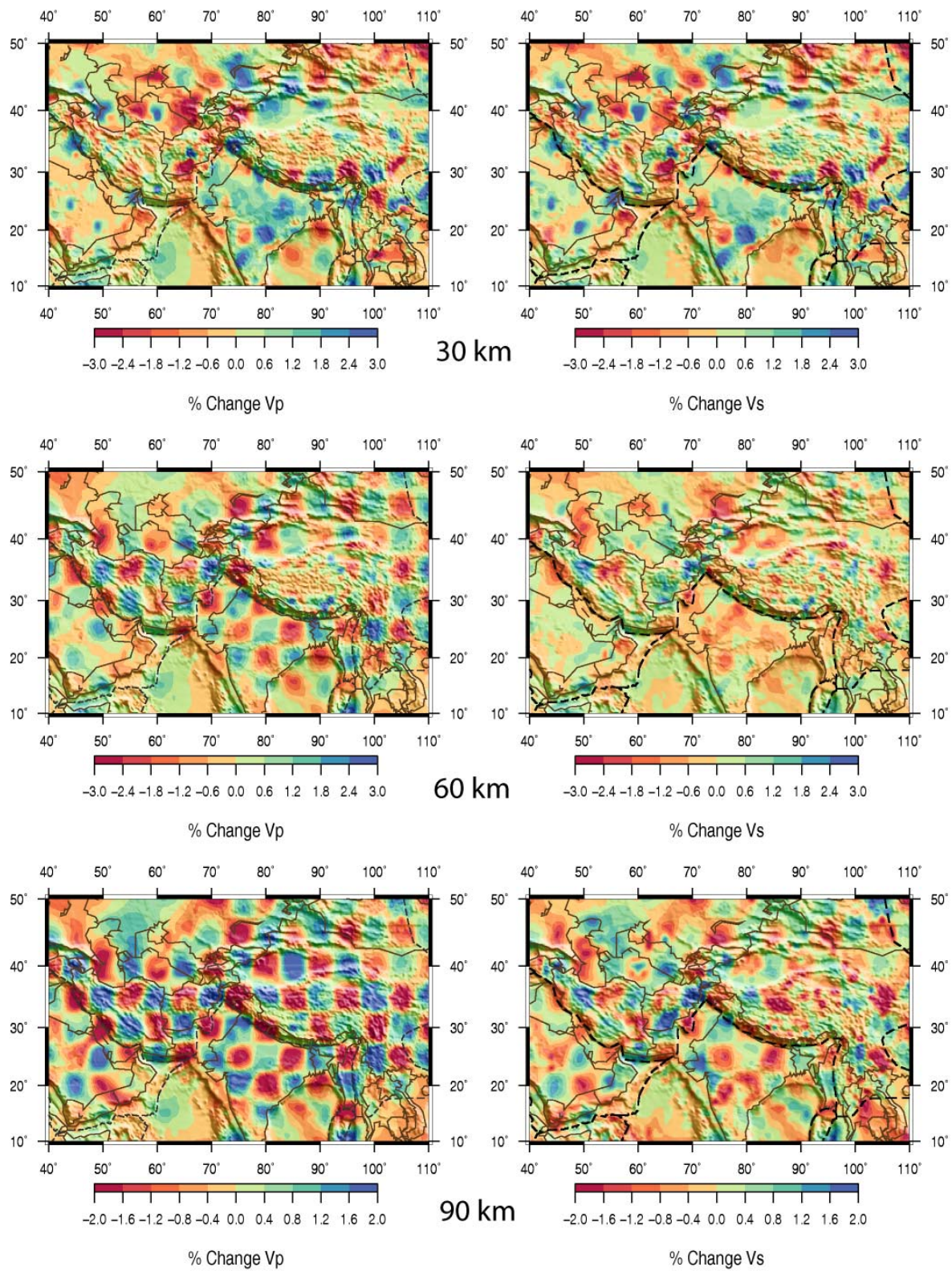


Figure C5. Map view slices of the recovered P (left) and S (right) checkerboard models at depths of 30 km, 60 km, and 90 km.

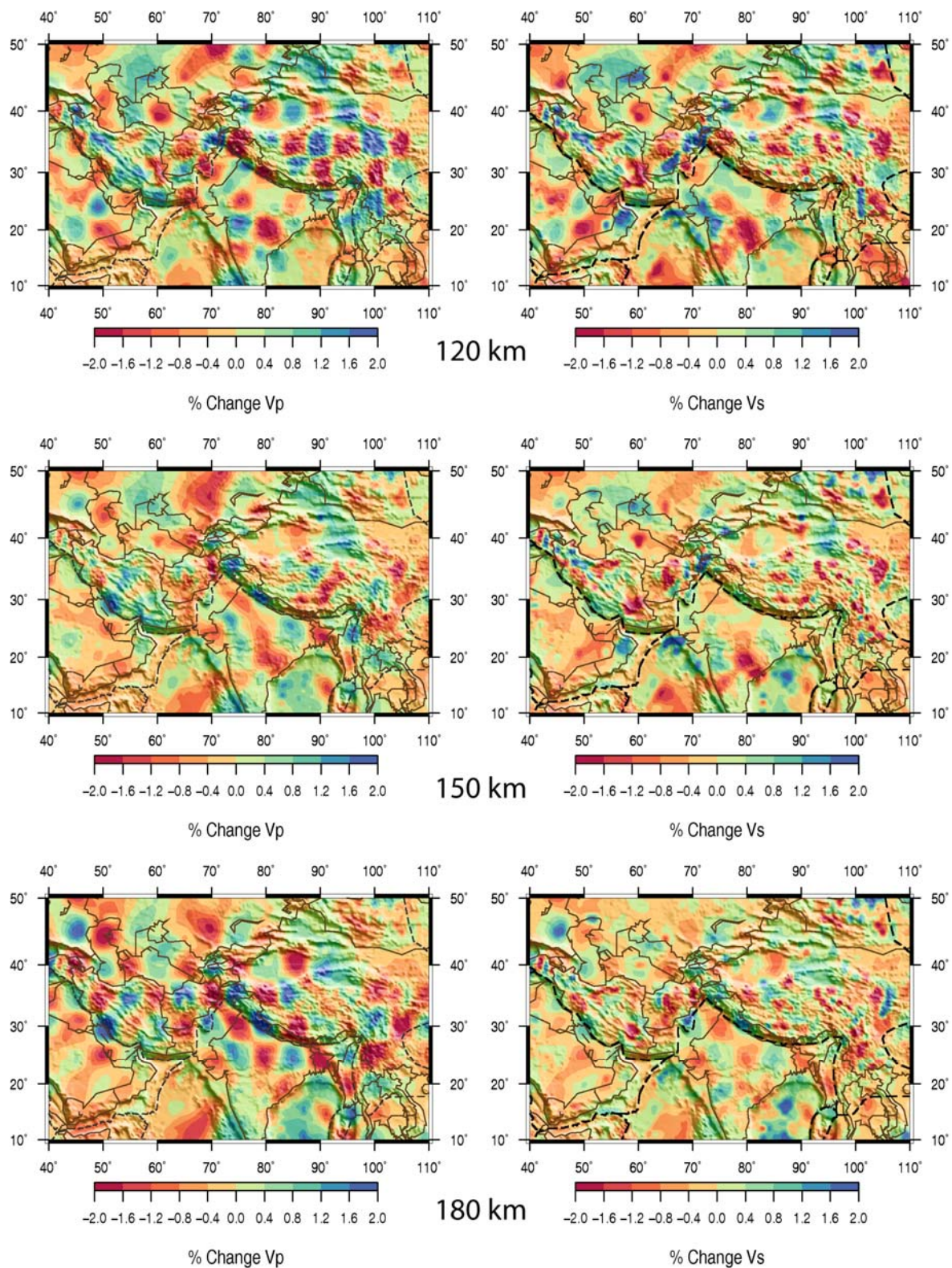


Figure C6. Map view slices of the recovered P (left) and S (right) checkerboard models at depths of 120 km, 150 km and 180 km.

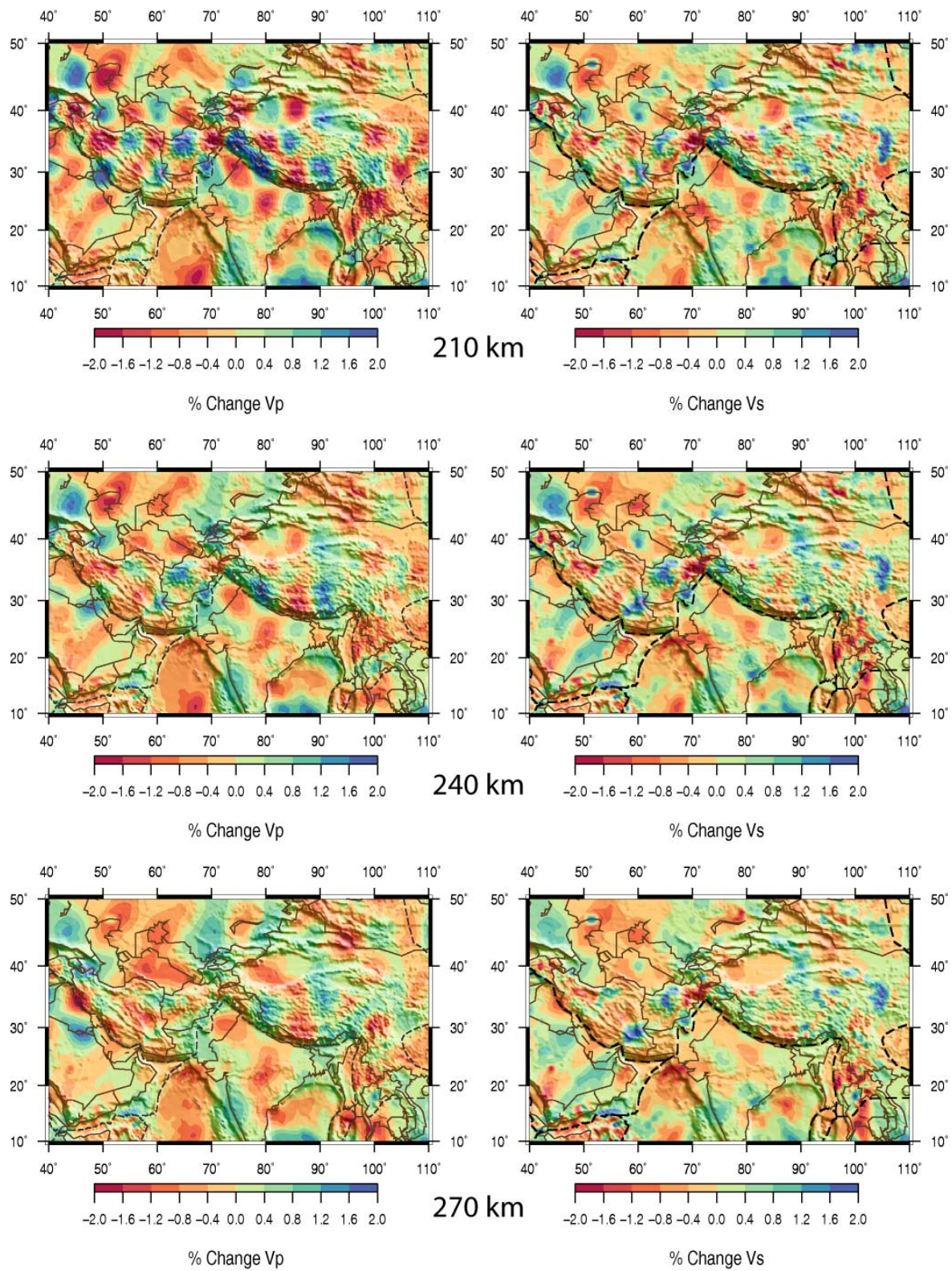


Figure C7. Map view slices of the recovered P (left) and S (right) checkerboard models at a depth of 210 km, 240 km and 270 km.

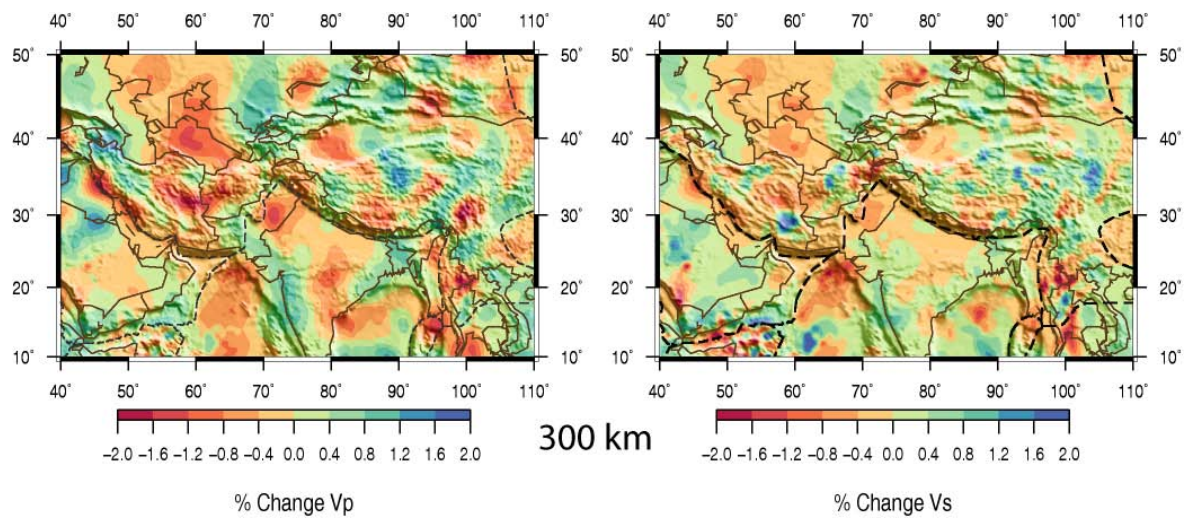


Figure C8. Map view slices of the recovered P (left) and S (right) checkerboard models at depths of 300 km.