Seismic tomography of Shatsky Rise by adaptive importance sampling

J. Korenaga[1] and W. W. Sager[2]

Received 22 February 2012; revised 1 June 2012; accepted 7 June 2012; published 9 August 2012.

[1] We present results from a wide-angle seismic refraction survey over Shatsky Rise, a large oceanic plateau in the northwestern Pacific. A new Monte Carlo sampling scheme is developed to explore comprehensively the model space of crustal velocity structure by joint refraction and reflection seismic tomography. The new scheme, which is founded on the notion of adaptive importance sampling, is made possible by the combination of several independent developments, including the introduction of effective model parameters, the implementation of automated model regularization, and a polynomial expansion of the probability density function. From 21 ocean-bottom instruments deployed along a 420-km-long refraction transect, we collect 16446 *Pg* and 4735 *PmP* travel times, which are inverted to construct a two-dimensional compressional velocity model across the major axis of Shatsky Rise. The convergence of Monte Carlo sampling is tested by running parallel sampling chains, and the effective dimensions of the model space are estimated to be $\sim$10 by principal component analysis. The thickness of the rise crust varies from $\sim$9 km to $\sim$30 km, and the average velocity of the lower crust ranges from $\sim$7.0 km s^{-1} to $\sim$7.2 km s^{-1}. One standard deviation uncertainty for whole-crustal thickness and average lower-crustal velocity is <1 km and $\sim$0.05 km s^{-1}, respectively, for most of the model domain. Shatsky Rise crustal structure exhibits a negative correlation between crustal thickness and velocity, which implies that a chemically anomalous mantle may have been responsible for the formation of the rise.

Citation: Korenaga, J., and W. W. Sager (2012), Seismic tomography of Shatsky Rise by adaptive importance sampling, *J. Geophys. Res.*, *117*, B08102, doi:10.1029/2012JB009248.

1. Introduction

[2] Oceanic plateaus are among the most prominent features of the Earth's ocean basins, yet their origins remain uncertain. These giant submarine mountains represent a large flux of magma from the mantle, thereby holding a key to understanding large-scale mantle dynamics in the past. The most popular explanation has been the plume head hypothesis [*Richards et al.*, 1989; *Campbell and Griffiths*, 1990], which calls for rapid plateau construction by high-degree partial melting associated with the upwelling of a large blob of hot mantle (a plume head) from a thermal boundary layer (usually considered as the core-mantle boundary region). The plume head hypothesis was developed in part to explain the ocean's largest plateaus, Kerguelen Plateau in the Indian Ocean and Ontong Java Plateau in the Pacific Ocean. Geochronologic data imply that both formed mainly during a brief interval of time within the Cretaceous [*Tarduno et al.*, 1991; *Mahoney et al.*, 1993]. Recent ocean drilling legs on these plateaus [*Coffin et al.*, 2000; *Mahoney et al.*, 2001], however, revealed complications that do not fit a simple plume head model; the bulk of Ontong Java Plateau, for example, apparently never reached sea level, contrary to the expectation from the buoyancy of a plume head [*Korenaga*, 2005]. Current research efforts include the modification of the original plume head model by incorporating additional complexities [e.g., *Farnetani and Samuel*, 2005; *Lin and van Keken*, 2006] as well as the development of alternative ideas [e.g., *Anderson*, 2000; *Korenaga*, 2005].

[3] Despite being extraordinary large features, the remote locations of oceanic plateaus means that they are generally poorly sampled both geophysically or geochemically, and the paucity of observational constraints is mainly responsible for the current rampancy of theoretical conjectures. Accurate information about crustal structure, for example, can help us to estimate the total volume of melt and its major-element composition, which in turn could distinguish between various scenarios of mantle melting [*White and McKenzie*, 1989; *Kelemen and Holbrook*, 1995; *Korenaga et al.*, 2002]. Unfortunately, such geophysical data are surprisingly rare for oceanic plateaus. Even Ontong Java Plateau, the one most frequently probed by seismic surveys, still suffers from

[1]Department of Geology and Geophysics, Yale University, New Haven, Connecticut, USA.
[2]Department of Oceanography, Texas A&M University, College Station, Texas, USA.

Corresponding author: J. Korenaga, Department of Geology and Geophysics, Yale University, PO Box, 208109, New Haven, CT 06520-8109, USA. (jun.korenaga@yale.edu)

©2012. American Geophysical Union. All Rights Reserved.
0148-0227/12/2012JB009248

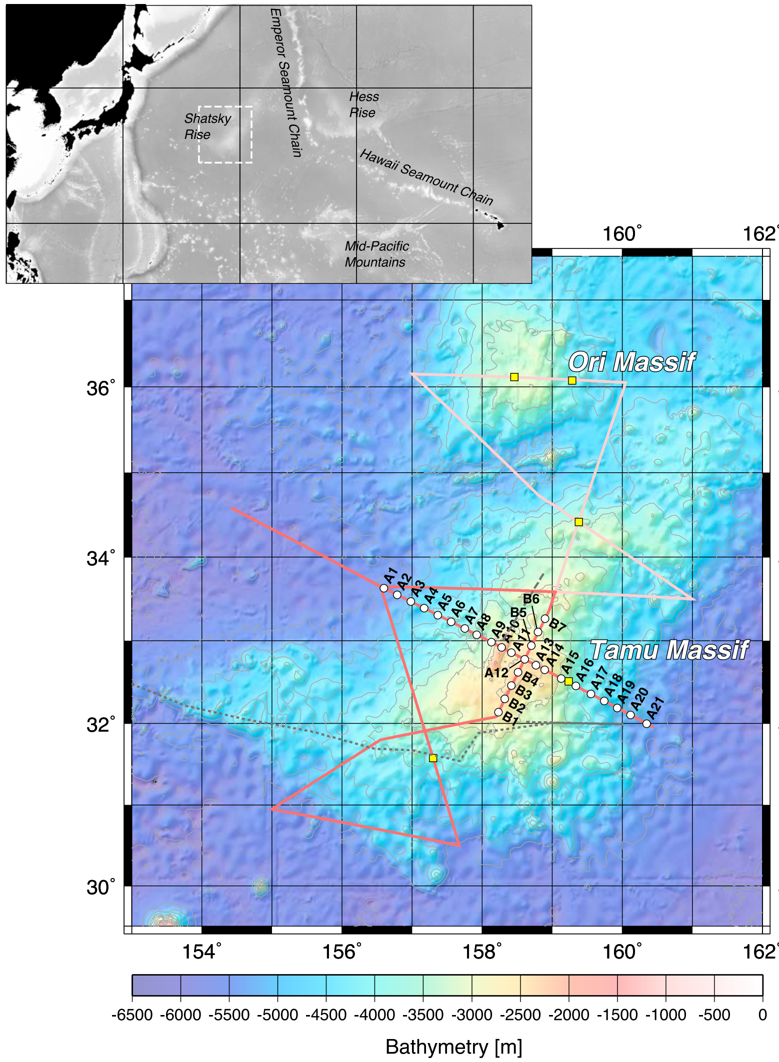

Figure 1. Configuration of the Shatsky Rise seismic survey, with MCS reflection lines (red), OBS locations (open circles), and IODP sites (boxes). Lines with OBSs were shot twice, with different shot distances: 50 m for MCS profiling and 162 m for refraction. Light red indicates additional reflection lines collected during the spring of 2012. Dashed lines denote seismic lines from previous studies: *Den et al.* [1969] (short dash) and *Gettrust et al.* [1980] (long dash). Bathymetric contours are drawn at 500-m interval. Inset shows the location of Shatsky Rise with respect to other major features in the northwestern Pacific, such as Hawaii-Emperor Seamount Chain, Hess Rise, and Mid-Pacific Mountains.

considerable ambiguity in its crustal structure [*Korenaga*, 2011]. Many seismic surveys on oceanic plateaus were done in the 1960s and 1970s [e.g., *Den et al.*, 1969; *Murauchi et al.*, 1973; *Furumoto et al.*, 1976; *Hussong et al.*, 1979; *Gettrust et al.*, 1980], and such vintage data do not reliably image deep crustal layers and the base of the crust (Moho). Modern seismic data do exist for some plateaus [e.g., *Operto and Charvis*, 1996; *Miura et al.*, 2004], but crustal structure models are often published without a thorough uncertainty analysis.

[4] In this paper, we present new deep-crustal seismic data collected over Shatsky Rise, which is another massive plateau in the Pacific (Figure 1), along with the development of a new tomographic inversion scheme. Among large oceanic plateaus, Shatsky Rise, the total area of which amounts to $\sim 4.8 \times 10^5$ km^2 (about the same size as California [*Sager et al.*, 1999]), is unique owing to a combination of factors that make it optimal for the study of oceanic plateau origin. First of all, it was formed at the right time, during the late Jurassic and Early Cretaceous when the magnetic field was reversing, so magnetic anomalies were recorded in ocean crust to show the locations of nearby, coeval spreading ridges [*Nakanishi et al.*, 1999]. This is how we know that Shatsky Rise formed at a ridge-ridge-ridge triple junction, and magnetic data allow us to reconstruct the tectonic history of the rise and adjacent ridges. In contrast, many large oceanic plateaus (e.g., Ontong Java Plateau, Kerguelen Plateau, Manihiki Plateau, and Hess Rise) formed during the Cretaceous Quiet Period when the magnetic field did not reverse, so their tectonic environment is hard to resolve. Second, the

spreading of the ridges adjacent to Shatsky Rise was relatively rapid, and the rise volcanism was spread laterally, rather than stacked vertically, so that the history of the rise is more easily interpreted. Third, sediments on Shatsky Rise flanks are generally thin [*Ewing et al.*, 1966; *Ludwig and Houtz*, 1979; *Sliter and Brown*, 1993], so the morphology of the rise, which can be measured by bathymetry, is a direct reflection of its structure and evolution. Finally, the rapidly spreading triple junction indicates the formation of Shatsky Rise on thin lithosphere, which should have had minimal influence on melt generation and migration. Indeed, the thin lithosphere of the on-ridge setting makes Shatsky Rise ideal for using geochemical and geophysical data to test whether the source mantle was thermally or chemically anomalous.

[5] This paper is organized as follows. After describing the acquisition of seismic data and their processing, we propose a new strategy to invert seismic travel time data to estimate crustal velocity structure. This strategy is built on joint refraction and reflection tomography developed by *Korenaga et al.* [2000], with a few important modifications such as the implementation of automatic smoothing, the introduction of a second model measure, and the use of adaptive importance sampling. The crustal velocity model of Shatsky Rise is then presented, together with the quantification of model accuracy, uniqueness, and resolution. We close with the implications of the derived crustal structure for the origin of Shatsky Rise.

2. Data Acquisition and Processing

[6] According to existing magnetic and bathymetry data, Shatsky Rise volcanism displays a progression in both age and volume along the trace of the triple junction, which was migrating northeastward relative to the Pacific plate during the rise formation [*Sager et al.*, 1988; *Nakanishi et al.*, 1989; *Sager et al.*, 1999; *Nakanishi et al.*, 1999]. This progression is reflected in three large volcanic constructs: Tamu Massif, Ori Massif, and Shirshov Massif (the first two shown in Figure 1). Among these three, Tamu Massif is the oldest and largest volcano with an estimated volume of 2.5×10^6 km^3 [*Sager et al.*, 1999], and this is the main target of our seismic survey, which was conducted during the summer of 2010 aboard R/V *Marcus G. Langseth.* This survey consists of $\sim$2000 km of multichannel seismic (MCS) reflection lines that run along and across the major axis of Shatsky Rise and two crossing perpendicular refraction transects over Tamu Massif (Figure 1).

[7] MCS profiling was conducted using a 36-element, 6600 cubic inch air-gun array fired every 50 m ($\sim$20 s with a cruising speed of $\sim$4.5 knots) towed at a water depth of 9 m and a 6.0-km hydrophone streamer with 468 channels. Shot gathers were recorded at a sampling interval of 2 ms. Streamer group spacing was 12.5 m, and with the 50-m shot spacing, the common midpoint fold is 58 with a spacing of 6.25 m. The details of MCS data processing and scientific interpretation will be published elsewhere. Two-way travel times to the top of igneous basement are converted to the thickness of a sedimentary layer assuming an interval velocity of 2 km s^{-1}, which is used when constructing initial models for crustal tomography. As already mentioned, the sedimentary layer is generally thin over Shatsky Rise; it is typically a few hundred meters, with a maximum of $\sim$1 km over the summit of the rise [*Sliter and Brown*, 1993; *Sager et al.*, 1999].

[8] The same air-gun array was used for two perpendicular refraction transects, but fired every 162 m to have randomized time interval of 70 ± 2 s and towed at a water depth of 12 m to enhance the low-frequency components of the source signal. Twenty eight ocean bottom seismometers (OBS) from Woods Hole Oceanographic Institution were deployed on the transects; 21 on a 420-km-long transect (Transect A) and 7 on a 162-km-long transect (Transect B). All instruments were recovered successfully, and they all returned good data. Among four components (vertical, two horizontal, and hydrophone), the hydrophone component turned out to be consistently of high quality (Figure 2), so our analysis is based on this component. The sampling rate was 200 Hz, and 5–20 Hz band-pass filtering and predictive deconvolution were applied to the data. Direct water arrivals, together with shot locations and multibeam bathymetry data, were used to relocate the instruments. Best-fit locations were found to be all within $\sim$300 m from the planned site locations. Offsets between the instruments and the shots were recalculated with the WGS-84 ellipsoid using the formula of *Vincenty* [1975].

[9] The travel times of the refraction (*Pg*) and reflection (*PmP*) phases were then picked manually, and half a period of the first cycle of an arrival was used when assigning a picking error. Source-to-receiver reciprocity was utilized to ascertain the internal consistency of phase identification among different instruments (Figure 2). Picking errors vary from 50 ms to 150 ms, depending on the clarity of arrivals. The mantle refraction phase (*Pn*) was also observed for some instruments on deep seafloor where crustal thickness is expected to be close to normal (Figure 2a), but because of their paucity we did not incorporate it in our subsequent analysis. Most instruments exhibit clear *Pg* and *PmP* arrivals, with the source-receiver distance often exceeding 200 km (Figures 2c and 2g). In total, 16446 *Pg* and 4735 *PmP* travel times were collected for Transect A, and 4926 *Pg* and 377 *PmP* travel times for Transect B (Figure 3).

3. Joint Refraction and Reflection Tomography

3.1. Some Long-Standing Issues and Overview of the New Strategy

[10] Two-dimensional (2-D) joint refraction and reflection tomography developed by *Korenaga et al.* [2000] has been applied to a range of active-source marine seismic data during the last decade, with the original inversion strategy and uncertainty analysis relatively unmodified [e.g., *Hosford et al.*, 2001; *Canales et al.*, 2003; *Sallares et al.*, 2005; *Hooft et al.*, 2006; *Holmes et al.*, 2008; *White and Smith*, 2009; *Contreras-Reyes et al.*, 2010; *Shulgin et al.*, 2011]. There are, however, at least a few unsatisfactory aspects in the original formulation. To explain the issues to be improved in a self-contained manner, we first provide a brief summary for the inversion strategy of *Korenaga et al.* [2000] in the following.

[11] For an initial velocity model $\mathbf{m_0}$, the following tomographic equation is constructed by ray-tracing through the given model:

$$\begin{bmatrix} \delta\mathbf{d} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{B} \end{bmatrix} \delta\mathbf{m}, \tag{1}$$

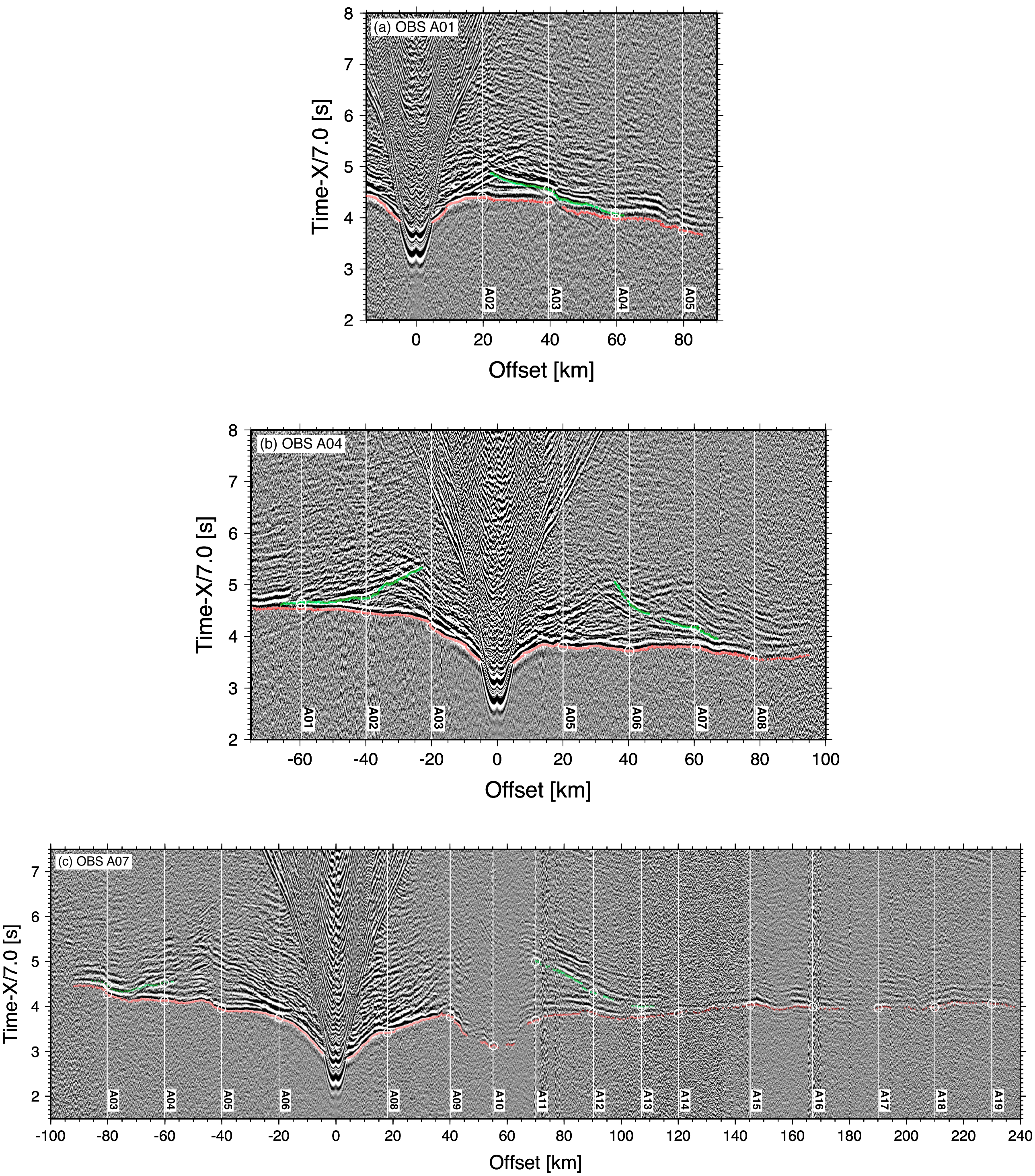

Figure 2. Processed seismogram for selected OBSs, plotted with a reduction velocity of 7.0 km s^{-1}. Semi-transparent markings denote the picked travel times of *Pg* (red) and *PmP* (green). White vertical lines denote the locations of other instruments, and circles correspond to their travel time picks at reciprocal relations (corrected for water-depth difference between instruments), demonstrating the consistency of phase identification among different instruments. (a) A1, (b) A4, (c) A7, (d) A10, (e) A13, (f) A16, (g) A19, (h) B1, (i) B4, and (j) B6.

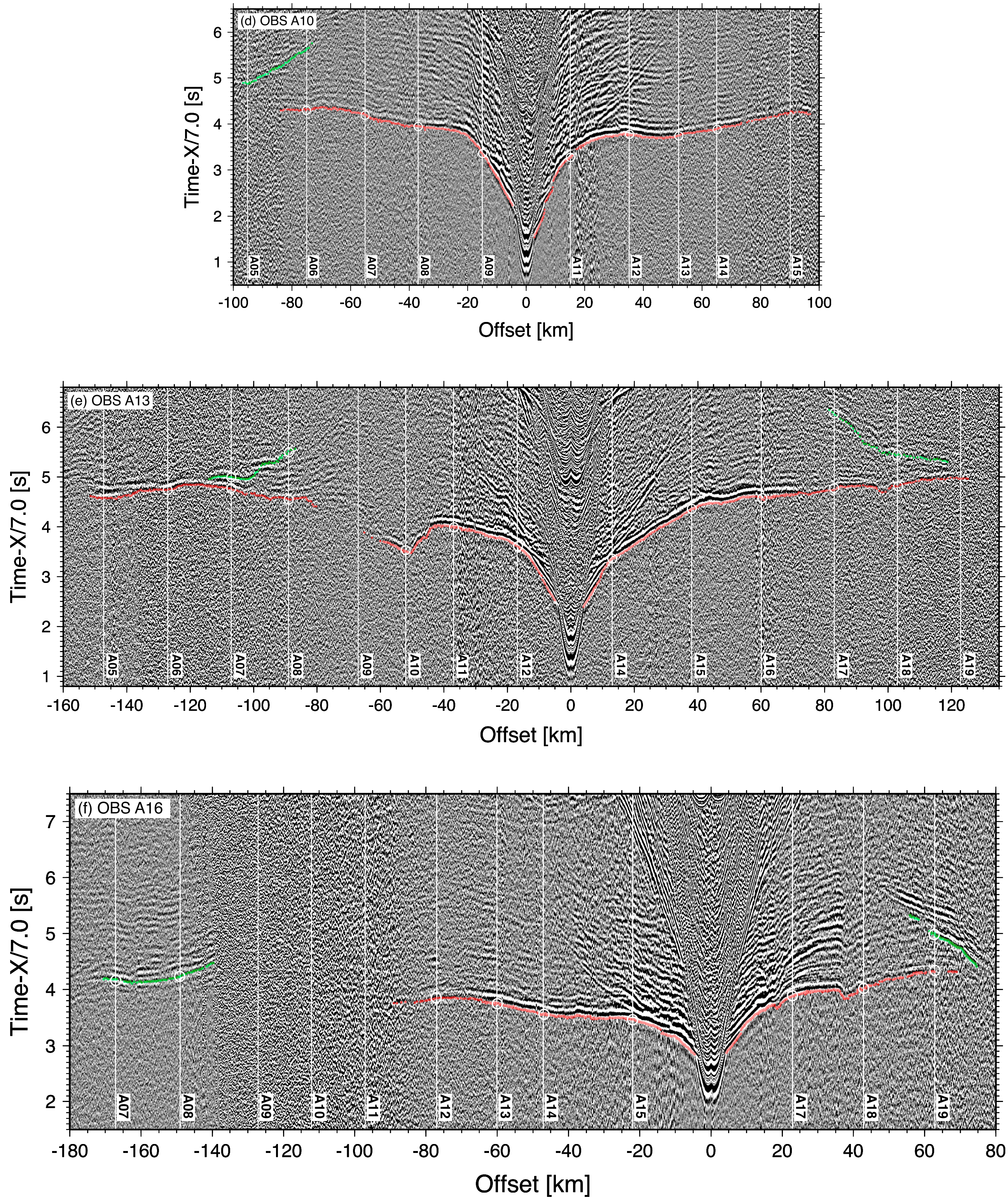

Figure 2. (continued)

where $\delta\mathbf{d}$ is a vector containing difference between observed travel times and theoretical travel times based on the assumed velocity model, $\delta\mathbf{m}$ is a vector dictating how to modify the assumed model to reduce the travel time misfit, $\mathbf{G}$ is a sensitivity kernel matrix that relates data misfit and model perturbation, and $\mathbf{B}$ is a regularization matrix that imposes smoothing and damping. Solving this equation and updating the initial model by $\mathbf{m_0} + \delta\mathbf{m}$ does not eliminate the travel time

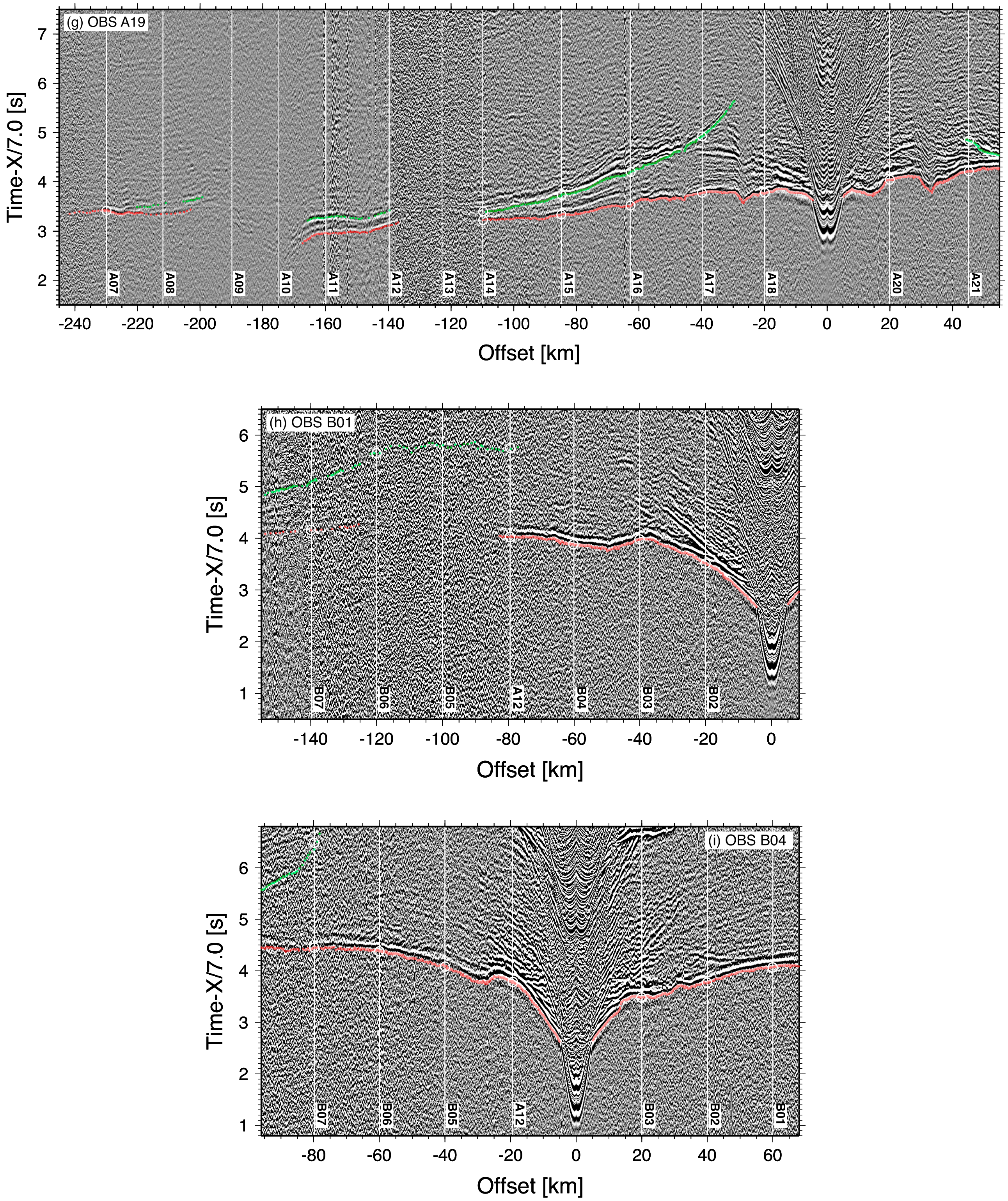

Figure 2. (continued)

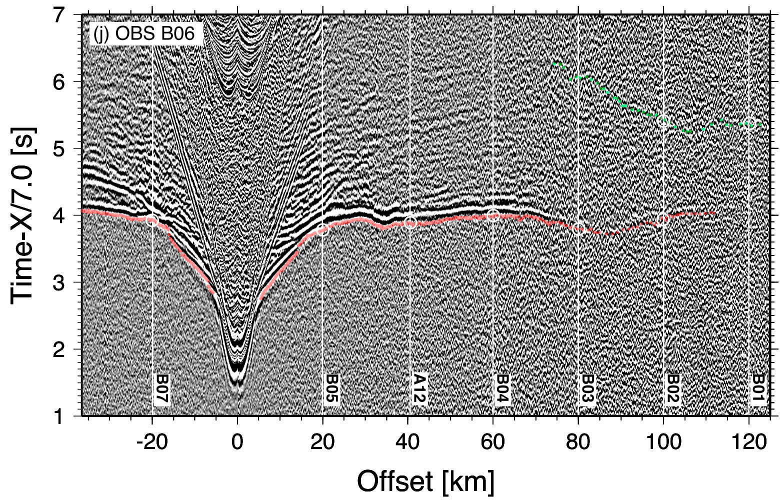

Figure 2. (continued)

misfit entirely unless the initial model happens to be sufficiently close to a true model, which is usually not the case, so we need to repeatedly solve the tomographic equation by updating a velocity model until the misfit becomes sufficiently small. Several important control parameters become apparent by writing out each component of equation (1) as

$$\delta\mathbf{d} = \begin{bmatrix} \delta\mathbf{d}^R \\ \delta\mathbf{d}^L \end{bmatrix}, \tag{2}$$

$$\delta\mathbf{m} = \begin{bmatrix} \delta\mathbf{m}_v \\ \frac{1}{w}\delta\mathbf{m}_d \end{bmatrix}, \tag{3}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_v^R & \mathbf{0} \\ \mathbf{G}_v^L & w\mathbf{G}_d^L \end{bmatrix}, \tag{4}$$

and

$$\mathbf{B} = \begin{bmatrix} \lambda_v \mathbf{L}_{Hv} & \mathbf{0} \\ \lambda_v \mathbf{L}_{Vv} & \mathbf{0} \\ \mathbf{0} & w\lambda_d \mathbf{L}_d \\ \alpha_v \mathbf{D}_v & \mathbf{0} \\ \mathbf{0} & w\alpha_d \mathbf{D}_d \end{bmatrix}, \tag{5}$$

where the superscripts R and L correspond to refraction and reflection phases, respectively, the subscripts v and d correspond to velocity nodes and reflector depth nodes, respectively, w is the depth-kernel scaling parameter, $\mathbf{L}_{Hv}$ and $\mathbf{L}_{Vv}$ are horizontal and vertical smoothing matrices for velocity nodes, respectively, $\mathbf{L}_d$ is a smoothing matrix for depth nodes, $\mathbf{D}_v$ and $\mathbf{D}_d$ are damping matrices for velocity and depth nodes, respectively, and the parameters λ_v, λ_d, α_v, and α_d control the relative weights of these smoothing and damping matrices with respect to the sensitivity kernel $\mathbf{G}$. The depth-kernel scaling parameter w is important to explore potential velocity-depth ambiguity inherent in using reflection data, smoothing is essential for tomography because of its generally underdetermined nature, and damping is necessary to stabilize iterative inversion. The weighting parameters for damping can be determined automatically by imposing the upper bound on model perturbation per each step of inversion [*Korenaga et al.*, 2000], but both the construction of the smoothing matrices, which depend on assumed correlation lengths, and the choice of their weighting parameters are left to be determined by trial and error. Because this iterative inversion scheme depends critically on an initial model, and because there can be a number of different models that explain observed travel times equally well, it is important to use a range of initial models and explore the model space. Model uncertainty can then be estimated by taking the statistics of such equally valid models.

[12] Though often overlooked, the most important but perhaps most difficult part of estimating a velocity model is quantifying its uncertainty. Finding just one model that can explain given travel times, without estimating its uncertainty, is not very useful. For example, whether the P-wave velocity of oceanic lower crust is 7.0 km s^{-1} or 7.3 km s^{-1} corresponds to very different scenarios of parental mantle dynamics [e.g., *White and McKenzie*, 1989], so we need to make sure that a velocity model to be interpreted has sufficiently small uncertainty (e.g., <0.1 km s^{-1}); otherwise it would be impossible to judge the reliability of interpretation. In contrast to global mantle tomography [e.g., *Romanowicz*, 2003], however, active-source crustal studies generally lack a good reference model because crustal structure can be regionally highly heterogeneous, and raypaths through an initial model can be drastically different from those through a final model. Owing to the lack of a reasonably accurate reference model, crustal tomography is a strongly nonlinear inverse problem, and with typical data acquisition geometry, it can suffer from a large degree of nonuniqueness. Our task is therefore to explore the model space extensively and delineate an entire subspace that corresponds to given observational constraints. The strength of a tomographic approach, as opposed to a more traditional forward or semi-forward approach using coarse model parameterization [e.g., *Zelt and Smith*, 1992], is that we can explore the model space more freely and find alternative velocity models more easily, thereby being able to estimate model uncertainty with greater confidence.

[13] Exploring the model space in the formulation of *Korenaga et al.* [2000], however, is still bounded by smoothing

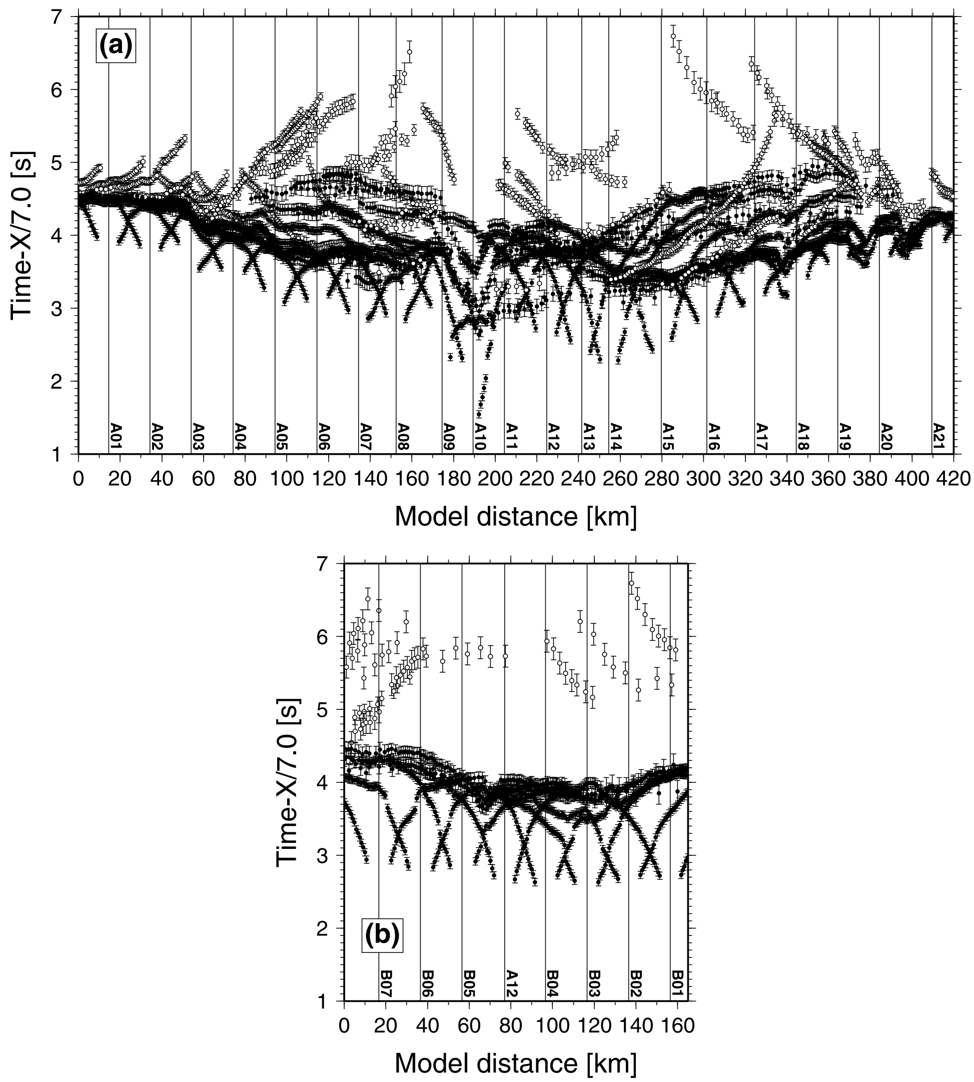

Figure 3. Picked travel times from all instruments are shown with their uncertainty as a function of model distance. Vertical lines denote OBS locations. Solid and open circles are for *Pg* and *PmP*, respectively, and data are shown at every 5 points for clarify. (a) Transect A. (b) Transect B.

constraints, which are imposed rather arbitrarily. When constructing a velocity model for the southeast Greenland margin, for example, they defined the correlation length function as follows: the horizontal correlation length for velocity nodes ranges from 10 km at seafloor, linearly increasing to 20 km at the model bottom, and the vertical correlation length also varies linearly from 50 m at seafloor to 4 km at the bottom. Correlation lengths for depth nodes were sampled from a 2-D function for horizontal correlation length. These correlation lengths and corresponding smoothing weights were determined by trial and error; too large lengths or weights led to too smooth models with poor data fits, whereas too small values made velocity models exhibit small-scale features that were unlikely to be required by data. Finding an acceptable combination of smoothing parameters that leads to 'reasonable-looking' models is not difficult for an experienced user of the tomographic code of *Korenaga et al.* [2000] (called TOMO2D), but it has some room for subjective judgments, which are better to be avoided as much as possible. We may also wonder how the estimate of model uncertainty is affected by a particular choice of smoothing parameters. In this regard, it is important to test a range of correlation lengths, but doing so would be time-consuming if we have to find corresponding smoothing weights by trial and error.

[14] Another nagging issue in the uncertainty analysis is the number of models to be collected. We randomly generate initial models to collect a number of alternative models, but when can we stop this Monte Carlo sampling and declare the convergence of model statistics? Owing to the limitation of available computational resources then, *Korenaga et al.* [2000] were able to collect only 100 models, but it is now easy to collect >1,000 models in a reasonable time frame (e.g., within a few days) using a commodity cluster [*Korenaga*, 2011]. A related question is how many truly different models exist. Even though we can generate 1,000 models easily today, it does not necessarily mean that all of these 1,000 models look very different to each other. In other words, the effective dimension of the model subspace compatible to observational constraints can be much smaller than the

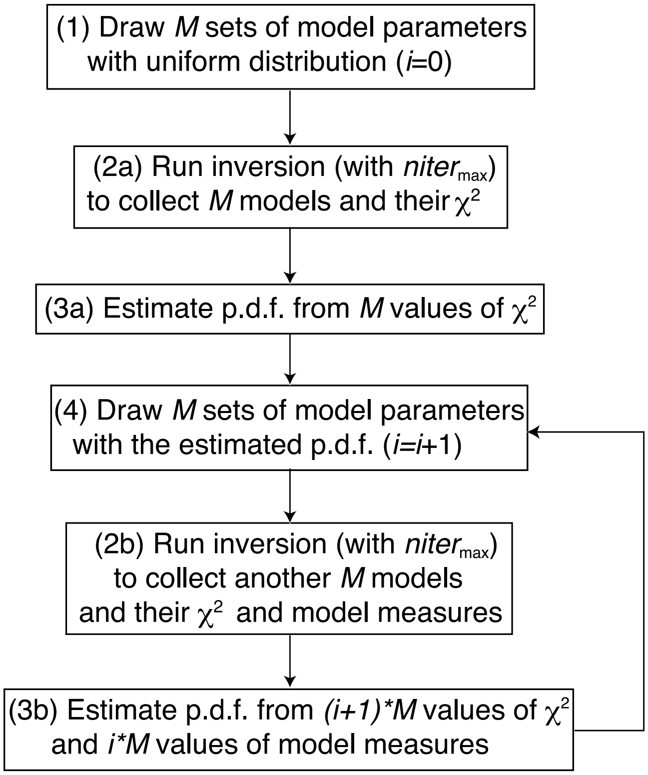

Figure 4. Flowchart for the new Monte Carlo sampling scheme. The entire procedure is one long sampling 'chain', which is composed of several 'steps'(shown as boxes). Each inversion step (2a or 2b) contains M inversion 'runs', and each inversion run conducts $niter_{\max}$ iterations of tomographic inversion. For a model measure, the diagnostic R is used in this study. The original Monte Carlo scheme of *Korenaga et al.* [2000] consists of steps 1 and 2a only.

number of model samples. It is also noted that not every initial model leads to a model with good data fit. Iterative inversion could be trapped in one of local minima with large data misfits. In the uncertainty analysis of *Korenaga et al.* [2000], only good models were considered, so failed inversion runs were simply discarded. Also, when finding a proper combination of smoothing parameters by trial and error, a number of inversion runs ended up with geologically unrealistic models, which were again simply discarded. It would be desirable if we could exploit these failed attempts and make the whole inversion process more effective.

[15] These issues are motivations for us to suggest a new inversion strategy in this study. The fundamental components of tomographic inversion remain the same as in *Korenaga et al.* [2000]; a crustal model is parameterized with a sheared velocity grid and a floating reflector, forward ray-tracing is done by the graph method with bending correction, and the tomographic equations (1)–(5) are still used. What is new is an overall framework encapsulating these components (Figure 4), which is based on the notion of adaptive importance sampling [e.g., *Liu*, 2001]. In the new strategy, correlation lengths used for smoothing are part of 'effective' model parameters, which are randomly sampled from their a priori ranges. Thus, we no longer need to determine them by trial and error. For this approach to work properly, however, smoothing weights have to be determined in an automated manner, so the TOMO2D code has been modified to determine both damping and smoothing weights by iteration. Purely random sampling of model parameters is not very efficient, because some combinations of parameters may lead to unacceptable local minima with too large data misfit, so we estimate the probability density function of the model space based on all inversion runs for more efficient random sampling. We periodically update the probability density function with more inversion runs, so the efficiency of sampling is expected to improve progressively. All of the inversion runs, including failed ones, are thus utilized in this framework. Furthermore, we introduce a second model measure, in addition to data misfit, to avoid sampling geologically unrealistic models. The convergence of Monte Carlo sampling can be judged by repeating the entire procedure with different random seeds, and we also suggest how to estimate the effective dimensions of the model space using the principal component analysis. In the following sections, the key elements of the new inversion strategy are described in detail.

3.2. Effective Model Parameters

[16] The number of model parameters in 2-D crustal tomography is on the order of $10^4 - 10^5$. Randomly sampling a model space with such vast dimensions is impractical unless we can ray-trace a given model (for all source-receiver pairs, which is typically on the order of $10^3 - 10^4$) in a fraction of a second. Previous studies, however, indicate that we can search the model space more efficiently with randomized 1-D initial models, for which we need only several parameters [e.g., *Korenaga et al.*, 2000; *Sallares et al.*, 2003; *Parkin and White*, 2008]. This is because inversion brings any given model to one of nearby local minima, so in effect, even one initial 1-D model can represent a variety of models in the vicinity of the corresponding local minimum. In this study, a 1-D velocity profile is constructed from upper-crustal velocity V_{UC}^0, mid-crustal velocity V_{MC}^0, lower-crustal velocity V_{LC}^0, upper-crustal thickness H_{UC}^0, and lower-crustal thickness H_{LC}^0:

$$V(z) = \begin{cases} V_{UC}^0 + \left(V_{MC}^0 - V_{UC}^0\right)\dfrac{z}{H_{UC}^0}, & z \le H_{UC}^0 \\ V_{MC}^0 + \left(V_{LC}^0 - V_{MC}^0\right)\dfrac{z - H_{UC}^0}{H_{LC}^0}, & z > H_{UC}^0 \end{cases} \quad (6)$$

where z is the depth measured from the base of the sedimentary layer, V_{UC}^0 is randomly chosen from a range between 3 km s^{-1} and 5 km s^{-1}, V_{MC}^0 between 5 km s^{-1} and 7 km s^{-1}, V_{LC}^0 between 7 km s^{-1} and 8 km s^{-1}, H_{UC}^0 between 3 km and 9 km, and H_{LC}^0 between 15 km and 25 km. The reflector nodes are initially set at the same depth, Z_M^0 (i.e. a horizontal reflector), which is randomly chosen from a range between 20 km and 30 km. The number of parameters for initial models is thus six.

[17] Correlation lengths are set similarly. Horizontal correlation length, for example, assumes the following 1-D function:

$$L_h(z) = \begin{cases} L_{h,UC} + \left(L_{h,MC} - L_{h,UC}\right)\dfrac{z}{H_{UC}^L}, & z \le H_{UC}^L \\ L_{h,MC} + \left(L_{h,LC} - L_{h,MC}\right)\dfrac{z - H_{UC}^L}{H_{LC}^L}, & z > H_{UC}^L \end{cases} \quad (7)$$

where $L_{h,UC}$, $L_{h,MC}$, and $L_{h,MC}$ are horizontal correlation lengths at $z = 0$, H_{UC}^L, and $H_{UC}^L + H_{LC}^L$, respectively. Correlation lengths are expected to become progressively greater at greater depths, so instead of randomly choosing the above three lengths, we first sample $L_{h,UC}$ from a range between 2 km and 20 km, and then set $L_{h,MC} = r_{Lh}L_{h,UC}$ and $L_{h,LC} = r_{Lh}L_{h,MC}$, where r_{Lh} is another random variable sampled from a range between 1 and 3. We also sample randomly H_{UC}^L between 3 km and 9 km and H_{LC}^L between 15 km and 25 km (note that these thicknesses are different from H_{UC}^0 and H_{UC}^0), so even with a single scaling constant r_{Lh}, we can generate a variety of depth-dependencies. Vertical correlation length is set similarly, with $L_{v,UC}$ and r_{Lv}, which are sampled between 1 km and 3 km and between 1 and 3, respectively. Correlation lengths for the sedimentary layer are set as $L_h = 10$ km and $L_v = 50$ m. Unlike a velocity model, correlation lengths do not change during an inversion run.

[18] It is also important to test a range of the depth-kernel scaling parameter w to fully explore the extent of velocity-depth ambiguity [*Korenaga*, 2011], so we randomly sample it from a range between 10^{-2} and 10^2. The total number of effective model parameters is therefore 13 in this study. The definition of effective model parameters is expected to vary for different tectonic settings. Our parameterization of an initial velocity model is, for example, suitable for oceanic plateaus, but more complicated ones would be appropriate for active margins. What is important is to devise a reasonably small number of effective model parameters that can still allow us to explore the model space extensively.

3.3. Automated Regularization

[19] There are four weighting parameters in the regularization matrix **B** (equation (5)): two for smoothing (λ_v and λ_d) and two for damping (α_v and α_d). The strategy of *Korenaga et al.* [2000] is to first determine the smoothing weights by systematically testing velocity and depth smoothing with a preliminary single-step inversion and then to fix these weights throughout all subsequent iterations. When model perturbations are found to be too large, damping is turned on, and optional sweeps on the damping weights are done by the secant and bisection method to satisfy user-defined upper limits on average velocity and depth perturbations.

[20] In our new approach, the smoothing weights are also determined by the secant and bisection method. We suggest to judge the success of intended smoothing by calculating the following model deviation after each inversion step:

$$\Delta m_i = \frac{\sum_{j \neq i} \beta_{ij} \delta m_j}{\sum_{j \neq i} \beta_{ij}} - \delta m_i, \tag{8}$$

where β_{ij} are averaging weights (used in the smoothing matrices) applied to nodes surrounding the i-th node. The first term in the right-hand side denotes an expectation for the i-th node by the weighted average of its surrounding nodes. As small Δm_i indicates that imposed smoothing constraints are successfully reflected in model perturbations, so during each iteration, the secant and bisection method is used to determine proper smoothing weights so that the following constraint is satisfied:

$$\max(|\Delta m_i / m_i|) < C, \tag{9}$$

where C is a user-defined upper bound on relative model deviation. As in *Korenaga et al.* [2000], optional sweeps on the damping weights are conducted only when model perturbations are found to be too large. That is, we first determine λ_v and λ_d with $\alpha_v = \alpha_d = 0$, and if model perturbations are within their limits, damping will not be applied. In this study, we use $C = 0.01$ for both velocity and depth nodes, and the upper limits on the average velocity and depth perturbations per iteration are set to 5% and 10%, respectively.

3.4. Model Evaluation

[21] The smoothing constraints as formulated in equation (5) are applied only for model perturbations, so a final model could still exhibit non-smooth features by accumulating small perturbations through iterations. Figure 5 shows a few sample inversion results for Transect A obtained after 20 iterations, starting with randomly generated effective model parameters. All of these models are compatible with the travel time data; their normalized χ^2 (i.e., χ^2/N where N is the number of data) is less than unity where

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{t_i - t_i^p}{\sigma_i} \right)^2. \tag{10}$$

Here t_i and σ_i are observed travel time and its uncertainty, respectively, and t_i^p is predicted travel time. These examples suggest that we need more than data misfit to quantify what "acceptable" models are. Small-scale features as seen in some models (Figures 5b, 5c, 5d, and 5f), which appear to concentrate on raypaths and do not seem to be geologically realistic, usually result from too small correlation lengths. Also, if the depth-kernel scaling parameter is too small, inversion tries to fit data by modifying velocity nodes as much as possible while keeping reflector depth nodes relatively unchanged, and this may sometimes lead to unrealistic velocity structure.

[22] As one possible measure of overall model roughness, we propose the following diagnostic:

$$R \equiv \left(\int_\Omega \left(\frac{V(x,z) - V_{\mathrm{ref}}(z/h(x))}{V_{\mathrm{ref}}(z/h(x))} \right)^2 dxdz \Big/ \int_\Omega dxdz \right)^{1/2} \times 100, \tag{11}$$

where Ω denotes the entire crustal domain, $h(x)$ is crustal thickness, and $V_{\mathrm{ref}}(\cdot)$ is a horizontally averaged velocity profile calculated as

$$V_{\mathrm{ref}}(z') = \int V(x, h(x)z')dx \Big/ \int dx. \tag{12}$$

Here the normalized coordinate z' ranges from 0 to 1. The quantity R measures an average relative deviation from the self-similar 1-D reference model that is calculated from a given model itself. In this study, we found that models with $R > 5$ are those which are typically judged as geologically unrealistic (e.g., Figure 5).

[23] This diagnostic is obviously just one possibility, and it may not be appropriate for different tectonic environments. The choice of this particular diagnostic may be deemed subjective, but it is an easily reproducible measure that tries to quantify our visual impressions. This reproducibility is

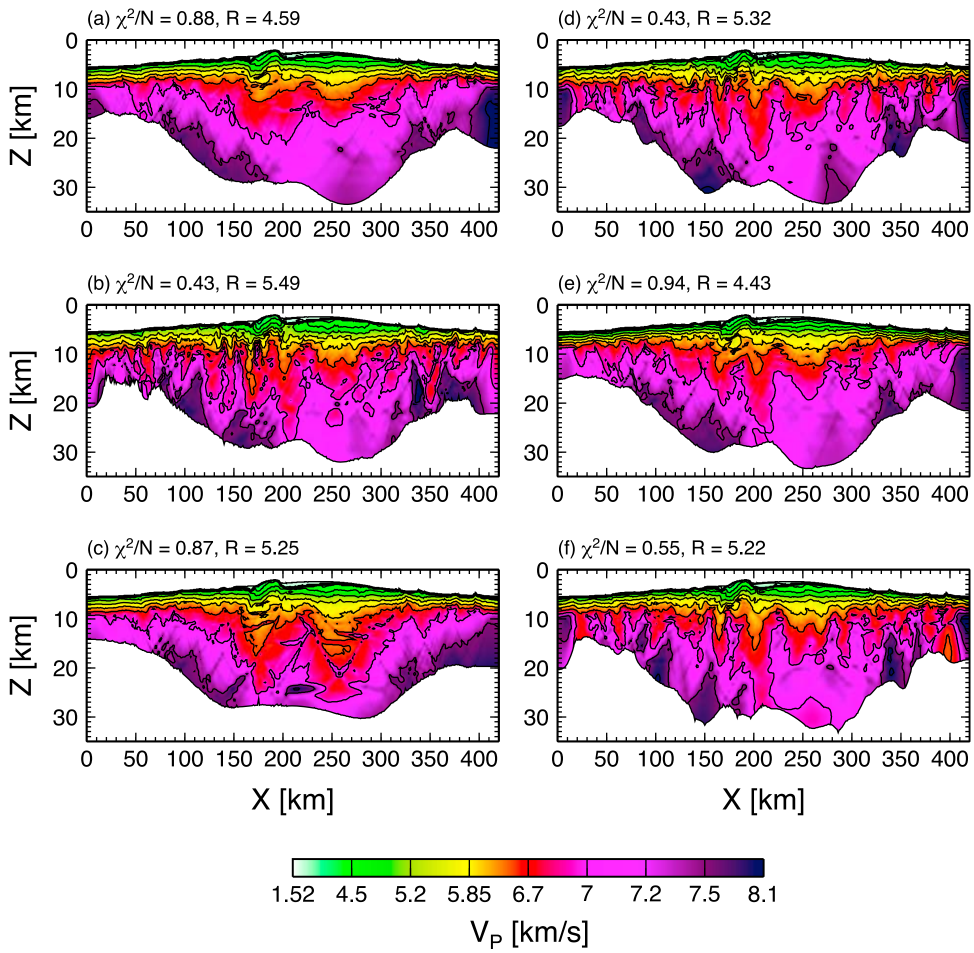

Figure 5. Examples of *P*-wave velocity models for Transect A, obtained after 20 iterations, starting with different combinations of randomly chosen effective model parameters. Normalized χ^2 and diagnostic R are also shown for each model. They are all successful in terms of data fit, but some of them (those with $R > 5$) indicate insufficient model regularization.

what frees us from the 'art' of finding a proper set of correlation lengths by trial and error.

3.5. Adaptive Importance Sampling

[24] At the very first step of the new inversion strategy, we prepare M sets of effective model parameters by randomly sampling from their a priori ranges (Figure 4, step 1). For each set, we conduct iterative inversion by a fixed number of iterations, $niter_{\max}$, and measure the χ^2 of a final model (step 2a). We then try to quantify the relation between the effective model parameters and their consequence (step 3a) by modeling the normalized χ^2 as

$$\log(\chi^2/N) = f(\{m_i^{\text{eff}}\}) = a_0 + \sum_i a_i m_i^{\text{eff}} + \sum_{i,j} b_{i,j} m_i^{\text{eff}} m_j^{\text{eff}} + .., \quad (13)$$

where a polynomial expansion is employed. This is followed by preparing another M sets of effective model parameters (step 4), but this time a set of parameters with its predicted χ^2/N exceeding unity is discarded (so we need to sample more than M times to prepare M sets of parameters). For each of these new parameter sets, we repeat iterative inversion and collect its χ^2 and R values (step 2b). Similarly to χ^2, the diagnostic R is modeled as a function of effective model parameters (step 3b):

$$R = p_0 + \sum_i p_i m_i^{\text{eff}} + \sum_{i,j} q_{i,j} m_i^{\text{eff}} m_j^{\text{eff}} + . \quad (14)$$

The reason for not estimating this functionality right after the first inversion step (i.e., step 2a) is that final models from the first set of runs have a wide range of χ^2 because the effective model parameters are sampled purely randomly; many models can be rejected solely on the basis of data fit, and their R values do not carry much significance. It is important to wait after the second inversion step so that we can collect R values from a large number of models with good data fits. In this step 3b, we also update our functional estimate of equation (13) incorporating additional M values of χ^2.

[25] We then return to the guided sampling step (step 4) to prepare another M sets of effective model parameters and repeat steps 2b and 3b (Figure 4). In addition to χ^2, we start using predicted R values for screening, i.e., we reject a set of parameters if its predicted R exceeds a given threshold. This

sequence of steps 4, 2b, and 3b is repeated several times, and the functional approximations for $\log(\chi^2/N)$ and R are updated in step 3b by using all inversion results up to the point (excluding those from the step 2a for R, as already explained). By periodically updating the functional approximations of equations (13) and (14), the sampling of the model space becomes more and more efficient, yielding progressively a larger number of successful models.

3.6. Convergence and Effective Model Dimensions

[26] The chain of Monte Carlo sampling outlined in Figure 4 is based on a pseudorandom number generator, which provides a series of random numbers from a given seed. The convergence of such Monte Carlo sampling is typically judged by running multiple sampling chains with different starting seeds and comparing the statistics of these parallel chains [e.g., *Robert and Cassela*, 2004]. That is, we need to repeat the entire sampling strategy of Figure 4 by at least a few times, to see whether the statistics of the first attempt is stable or not.

[27] The results of crustal tomography are usually summarized by the mean and variance of final velocity models obtained from Monte Carlo sampling. The variance is, however, just the diagonal component of the covariance matrix, and the rest of the covariance matrix is usually not exploited. One way to make use of the covariance matrix, C_{ij}, is to calculate corresponding correlation coefficients, ρ_{ij}, as [e.g., *Tarantola*, 1987]:

$$\rho_{ij} = \frac{C_{ij}}{C_{ii}^{1/2} C_{jj}^{1/2}}. \tag{15}$$

Correlation coefficients vary from -1 to 1 and quantify how strongly the uncertainty of a certain parameter is correlated with that of the other; for example, $\rho_{ij} \sim 1$ means that if the i-th parameter actually happens to be higher than its mean value, the j-th parameter is very likely to be higher than its mean value as well, i.e., their uncertainties are positively correlated. Examples in crustal tomography can be seen in the work of *Zhang and Toksöz* [1998] and *Zhang et al.* [1998].

[28] Showing correlation coefficients for just one parameter (one velocity or depth node in our case) takes the same space as that for one velocity model, so it is quite cumbersome to visualize correlation coefficients for more than a few parameters; the vast majority of the covariance matrix is thus still hard to appreciate. Also, even though it has been suggested that correlation coefficients may be used to diagnose the spatial resolution of a velocity model [e.g., *Tarantola*, 1987], such interpretation of correlation may not be valid in case of highly nonlinear inversion, as discussed later in section 4.2. Instead of directly showing correlation coefficients, therefore, we suggest that it is more informative to present the results of the principal component analysis [e.g., *Gershenfeld*, 1998], which is based on the eigenvalue decomposition of the covariance matrix. Because the covariance matrix is semipositive definite, all of its eigenvalues are nonnegative real numbers, and we can always decompose it as

$$C_{ij} = A \cdot \mathrm{diag}\{\lambda_1 \cdots \lambda_m\} \cdot A^T, \tag{16}$$

where eigenvalues λ_i are ordered such that

$$\lambda_1 \geq \lambda_2 \geq . \geq \lambda_m \geq 0, \tag{17}$$

and corresponding eigenvectors $\mathbf{a_i}$ are collectively denoted by

$$A = [\mathbf{a}_1\, \mathbf{a}_2 \cdots \mathbf{a_m}]. \tag{18}$$

Eigenvectors associated with large eigenvalues represent the dominant part of the covariance matrix, and the rest of eigenvectors with $\lambda_i/\lambda_1 \ll 1$ may be safely disregarded. We suggest to use the number of large enough eigenvalues as a proxy for the effective dimension of the model space compatible with observational constraints.

4. Results

[29] Travel time data from both transects have been analyzed by the new inversion strategy. Results from Transect A will be described in detail in this section; a brief summary for those from Transect B will be given in section 5.

[30] The model domain for Transect A is 420 km wide and 40 km deep from seafloor, with a horizontal grid space of 1 km and a vertical grid spacing gradually increasing from 50 m at the seafloor to 1 km at the model bottom, amounting to over 33,000 velocity nodes. The number of reflector nodes is 421 with a uniform horizontal grid spacing of 1 km. As described in section 3, correlation lengths and the depth-kernel scaling parameter are sampled randomly from their a priori ranges, and initial models are constructed by combining randomly generated 1-D velocity profiles with randomly placed flat Moho discontinuities. As mentioned in section 2, a priori information on the sedimentary layer based on MCS data is incorporated in those initial models. For each inversion run, travel time data are also randomized as in *Korenaga et al.* [2000], with random common receiver errors with a maximum amplitude of 50 ms and random travel time errors with the same maximum amplitude.

[31] For the strategy illustrated in Figure 4, we used $M = 2^{10}$ and $niter_{\max} = 20$ and repeated the sampling phase up to $i = 4$. We found that $\log(\chi^2/N)$ and R can be approximated reasonably well by 2nd-order and 3rd-order polynomials, respectively (Figure 6); for the 13 effective model parameters, the number of unknown coefficients to be determined by linear regression is 105 for 2nd-order polynomials and 560 for 3rd-order. A threshold on the diagnostic R (in step 4 for $i > 1$) is set to 5. The efficiency of adaptive importance sampling using these function approximations may be seen by how the distributions for χ^2 and R change after estimating or updating their functionality (Figures 7a and 7b). For the initial purely random sampling from the given a priori ranges (step 1), the χ^2/N values, for final models obtained after 20 iterations, are broadly distributed from $\sim\exp(-1)$ to $\sim\exp(4)$, with most models having $\chi^2/N > 1$. Soon after the first estimate of the χ^2 functionality, however, the distribution of χ^2 becomes tightly clustered in the range of $\chi^2/N < 1$. Similar progressive improvement may be observed for the diagnostic R. Note that the number of iterations in each run is always equal to $niter_{\max} = 20$, i.e., we do not terminate a run even when χ^2/N becomes lower than unity during iteration. If we

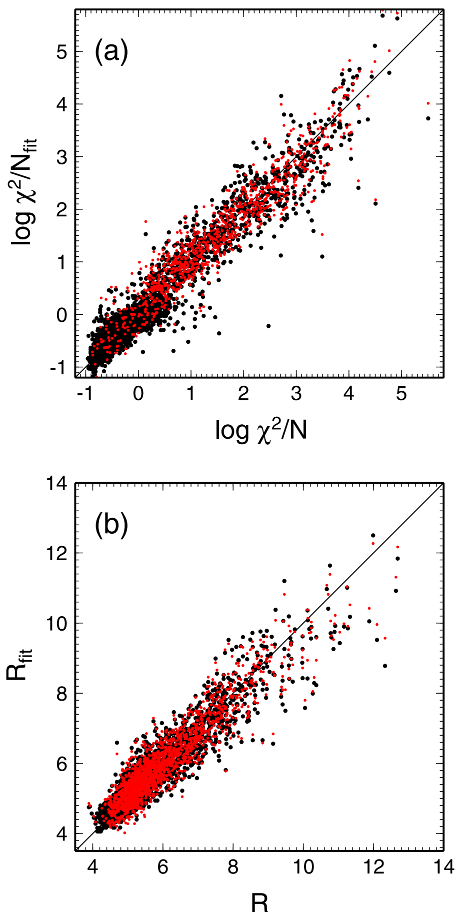

Figure 6. (a) Comparison of actual log χ^2/N values and modeled values based on a polynomial expansion with effective model parameters. Red dots for the first 1,000 runs (at $i = 0$), and black dots for all 5,000 runs (at $i = 4$). (b) Same as Figure 6a but for the diagnostic R.

do, the range of collected χ^2/N values would be limited by a lower bound of ~1, but a broader range is necessary to better estimate the behavior of the χ^2 function around $\chi^2/N \sim 1$. A model with $\chi^2/N < 1$ may be regarded as over-fit, but from a run with such over-fit final model, we can always extract an intermediate model with $\chi^2/N \sim 1$, which we may call an optimal model.

[32] The a posteriori distributions of effective model parameters (Figures 7c–7o) show what parameters need to be carefully chosen for the success of inversion. Regarding an initial 1-D velocity profile, for example, inversion runs would likely be successful if the top velocity V^0_{UC} and the mid-crust velocity V^0_{MC} are in the ranges of 4–4.5 km s^{-1} (Figure 7c) and 6–6.7 km s^{-1} (Figure 7d), respectively, and if the thickness of the upper crust H^0_{UC} is thinner than ~5 km (Figure 7f), whereas other parameters related to the lower crust exhibit more uniform distributions (Figures 7e, 7g, and 7h). This indicates the importance of having the velocity gradient of the upper crust in a proper range. Inversion would be more successful when the horizontal correlation length at the top of crust $L_{h,uc}$ is smaller than ~7 km (Figure 7i) with $r_{Lh} < \sim 2$ (Figure 7k), but the vertical correlation function does not seem to be very selective (Figures 7j and 7l). Last, small depth-kernel weighting parameters (<~0.1) are found to lead to unacceptable models (Figure 7o), probably because they suppress the perturbation of depth nodes more than necessary during iterative inversion.

[33] To test the convergence of the adaptive importance sampling, we repeated the entire sampling procedure by nine more times, using different seeds for a pseudorandom number generator, for which we used Numerical Recipes' ran2 function [*Press et al.*, 1992]. Each sampling chain resulted in >2,500 acceptable runs with final models having $\chi^2/N < 1$ and $R < 5$. From each of these runs, we extracted an optimal model with its χ^2/N closest to unity, and the collection of such models is used for subsequent statistical analyses. Figure 8 compares the cumulative moving average of the lower-crustal velocity at the model offset km 270 among the ten parallel sampling chains. It is apparent that an estimate for the lower-crustal velocity at this location converges to the value of 7.12 ± 0.04 km s^{-1} after the number of models exceeds ~1000. Similar convergence is observed for other parts of the model. The difference in the mean value among parallel chains is much smaller than the estimate of standard deviation, indicating that the model space has already been explored comprehensively by the first sampling chain. The following sections are thus based on ~2,600 optimal models collected during the first chain.

4.1. Mean and Variance

[34] The mean of initial crustal models for the successful inversion runs is shown in Figure 9a, and its standard deviation in Figure 9b. The mean and standard deviation of corresponding optimal models are shown in Figures 9c and 9d, respectively. Parts of the model within ~30–40 km from the edges suffer from large uncertainties, as expected from the geometry of data acquisition, so they are excluded from model interpretation. Standard deviation for velocity nodes in the optimal models is generally less than 0.1–0.2 km s^{-1}, except for the mid-crustal region just below the rise summit (around km 180), and that for depth nodes is less than ~1 km. The Transect A crustal model exhibits a slight asymmetry with respect to the rise axis, which is located at around km 220; the northwestern part (with smaller model offsets) is generally thinner, with the crustal thickness varying from ~9 km to ~25 km, whereas the southeastern part (with model offsets more than 220 km) has crustal thickness ranging from ~13 km to ~30 km. Throughout the model domain, vertical velocity gradients show a marked change at a velocity of ~6.5 km s^{-1}, and we divide the crust into the upper- and lower-crustal sections by the velocity contour of 6.5 km s^{-1}. The northwestern part of the transect appears to have slightly higher lower-crustal velocity, with the lowermost part exceeding 7.5 km s^{-1} more frequently than at the southeastern part.

[35] Also shown in 9e is the derivative weighted sum (DWS) calculated by ray-tracing through the average model of Figure 9c, which may be regarded as a proxy for the number of rays passing through each velocity node [*Toomey and Foulger*, 1989]. The DWS is a qualitative measure of linear sensitivity, and its comparison with the model uncertainty quantified by our Monte Carlo sampling (Figure 9d) points to the importance of nonlinear sensitivity achieved through iterative inversion [e.g., *Zhang and Toksöz*, 1998; *Korenaga*, 2011]. For example, the lowermost crust for km 240–300 is only lightly sampled by reflection rays, but its

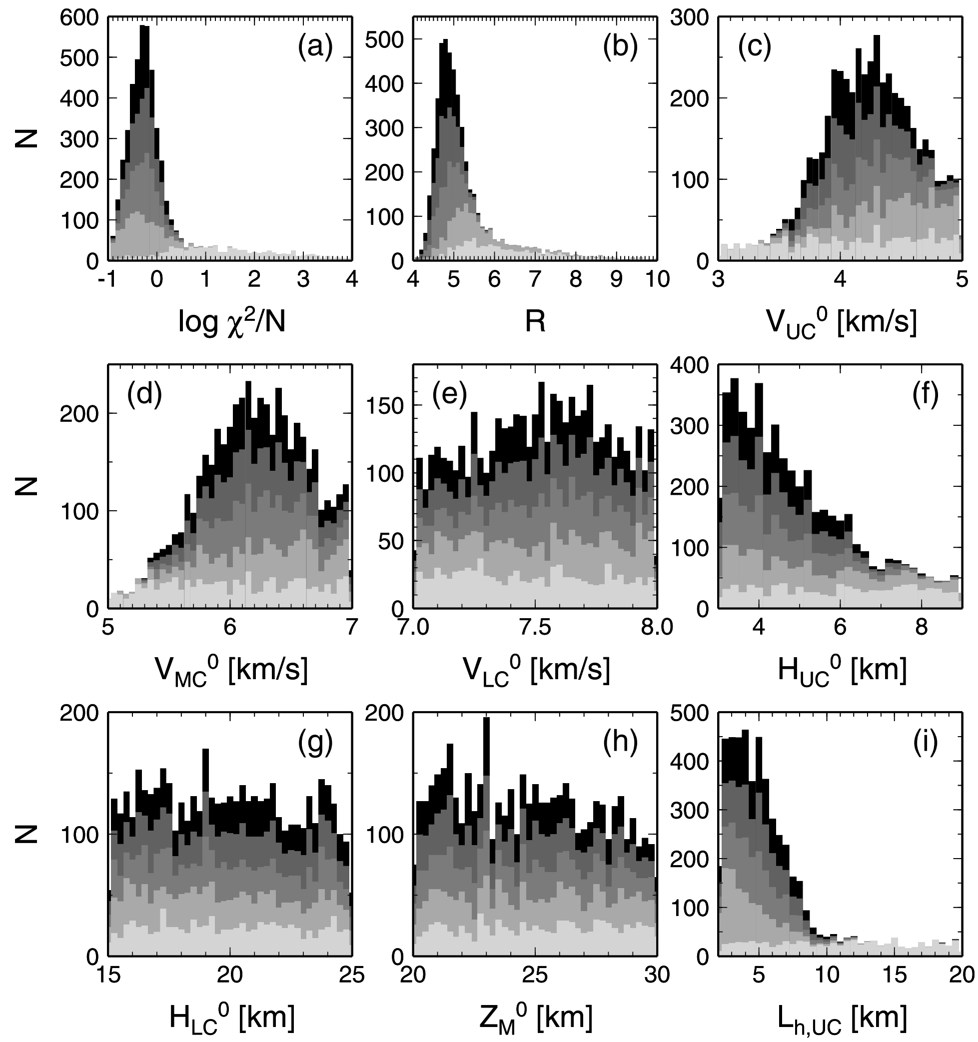

Figure 7. A posteriori distributions of model diagnostics and effective model parameters. The shade of each histogram changes gradually from the lightest ($i = 0$) to darkest ($i = 4$) to clarify progression through a sampling chain. (a) Normalized χ^2, (b) diagnostic R, (c) initial upper-crustal velocity, (d) initial mid-crustal velocity, (e) initial lower-crustal velocity, (f) initial upper-crustal thickness, (g) initial lower-crustal thickness, (h) initial Moho depth, (i) horizontal correlation length for upper crust, (j) vertical correlation length for upper crust, (k) scaling constant for horizontal correlation length, (l) scaling constant for vertical correlation length, (m) upper-crustal thickness for correlation-length function, (n) lower-crustal thickness for correlation-length function, and (o) depth-kernel scaling parameter.

standard deviation is lower than other parts of the lower crust with much higher ray density. If one perturbs the lower-crustal velocity for km 240–300 too much, the region would then be sampled by more rays, leading to greater data misfit, and subsequent inversion would reduce such velocity perturbations. Calculating DWS is easy, but showing it without estimating the standard deviation may thus give a false impression on how a model is constrained by data. In addition to this nonlinear sensitivity, it is also worth pointing out that the model uncertainty shown in Figure 9d reflects smoothing constraints as well as the a priori range of effective model parameters. The edges of the velocity model, for example, have high but still finite standard deviations even though these parts of the model are never sampled by seismic rays. The reason for finite uncertainty is simply that these edges are still under smoothing constraints. Also, the very shallow part of the model is sampled only sparsely by nearly vertical rays, but this region is nevertheless characterized by relatively small uncertainties because initial velocity models are bounded by geologically reasonable velocity values.

4.2. Covariance, Correlation, and Eigenmodes

[36] In addition to the standard deviation (which is the square root of variance), we also calculate the full covariance matrix using the ensemble of optimal models from the first sampling chain. Using the original model parameters (33,259 velocity nodes and 421 depth nodes), however, results in a dense matrix with dimensions of 33680×33680. Even by exploiting its symmetry and storing it to single-precision floating point, the matrix requires more than 2 GB of memory, which we regard too excessive. Given the smooth nature of tomographic models, we do not expect

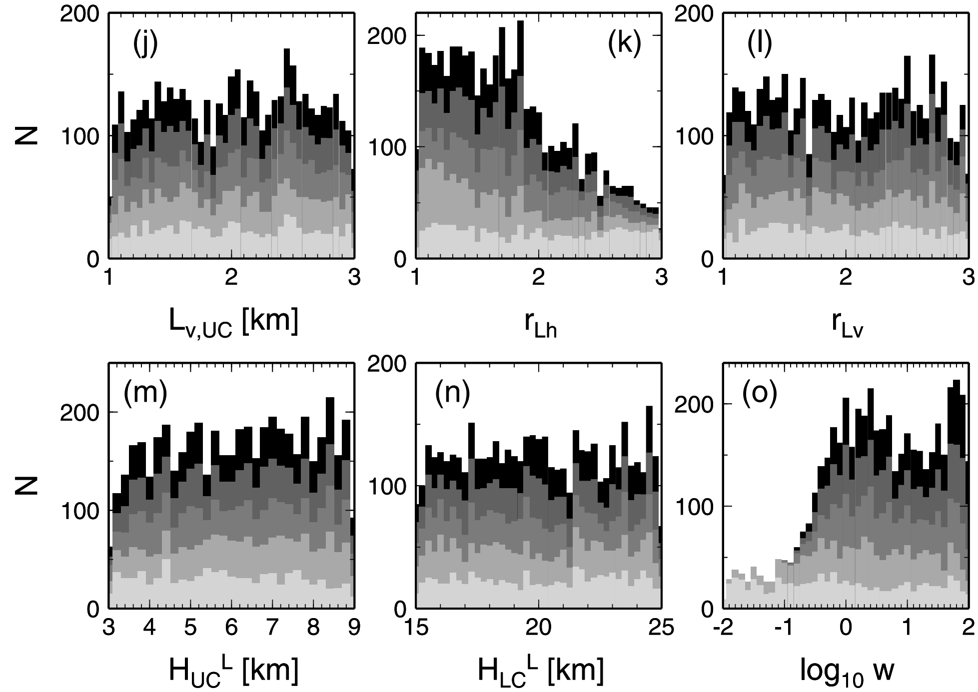

Figure 7. (continued)

that coarse-graining the model mesh would result in a considerable loss of information contents, so we decimated the velocity mesh down to 106×20 (the total number of velocity nodes is now only 2,120). Prior to this decimation, velocity nodes below the Moho reflector were replaced uniformly with a mantle velocity of 8.2 km s^{-1}, to assimilate the information of depth nodes into the velocity mesh.

[37] Correlation coefficients calculated from the covariance matrix (equation (15)) are shown in Figure 10 for a few selected velocity nodes. Nodes in the upper crust exhibit strongly positive correlation with other nodes at similar depths throughout the model extent and negative correlation with other parts of the upper crust (Figures 10a and 10b). This reflects the uncertainty of a velocity gradient in the upper crust. Even though the upper-crustal velocity nodes are characterized with relatively small errors (Figure 9d), errors are still non-zero. In the model space, therefore, when the shallower part of the upper crust deviates slightly from its mean value, the deeper part of the upper crust has to deviate to an opposite direction to conserve travel times through the upper crust, which are constrained mostly by the refraction phase *Pg*. Deep-crustal nodes tend to have correlation more in the vertical direction than in the horizontal direction (Figures 10c and 10d), reflecting trade-offs associated with the reflection phase *PmP*.

[38] Displaying correlation coefficients is thus somewhat informative, but what is shown in Figure 10 represents just four columns out of 2,120, i.e., $\sim$0.2% of the matrix. The vast majority of the covariance matrix is thus not presented. We tried to select representative nodes, but it is always possible for us to discount some nodes that may be interesting to others. Also, because correlation coefficients are normalized to vary from -1 to 1, the significance of correlation, or whether it is associated with large model uncertainty or not, cannot be read from them. In this regard, the principal component analysis outlined in section 3.6 offers an attractive alternative. The eigenvalues of the covariance matrix are shown in Figure 11a, and we define the 'scaled eigenmode' as $\lambda_i^{1/2}\mathbf{a_i}$ to illustrate the most important kinds of parameter correlation and trade-off (Figures 11b–11i). These eigenvalues represent variance in the transformed coordinates spanned by eigenvectors, and it can be seen that $\lambda_i/\lambda_1 \ll 1$ for $i > \sim10$, indicating that the effective dimension of the model space, compatible with observed travel times and with our smoothness constraints, is no more than 10. In other words, a large number of models collected through Monte Carlo sampling can be reproduced reasonably well by a linear combination of just $\sim$10 eigenvectors. The scaled eigenmodes are similar to correlation coefficients, but they have the dimension of velocity, and they can be listed in order of their significance. The first scaled eigenmode represents the uncertainty in the velocity gradient through the upper crust (Figure 11b), the second eigenmode corresponds to the uncertainty in the lower-crustal velocity beneath the rise summit (Figure 11c), and the third eigenmode reveals trade-off between the upper-crustal velocity and the mid-crustal velocity (Figure 11d). Higher-order eigenmodes are characterized by progressively smaller amplitudes, so the parameter correlations represented by them are less significant. The information contents of the covariance matrix can therefore be succinctly visualized by the first few scaled eigenmodes.

4.3. Spatial Resolution

[39] The spatial resolution of a model has sometimes been discussed by calculating correlation coefficients [e.g.,

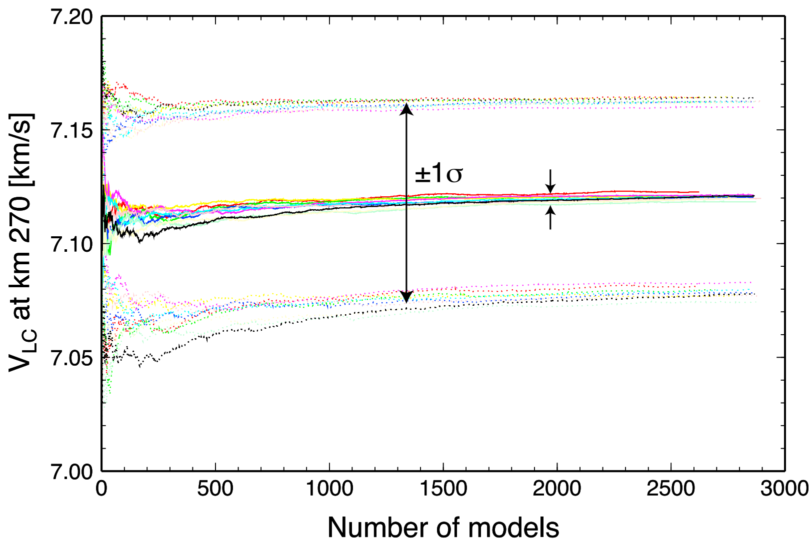

Figure 8. Example of convergence test with ten parallel sampling chains. Cumulative moving average for lower-crustal velocity at a model offset of 270 km is shown as a function of the number of optimal models used for averaging. Different colors indicate different chains. Dotted curves denote the range of one standard deviation.

Tarantola, 1987; *Zhang and Toksöz*, 1998]. Horizontally extending correlations observed for upper-crustal nodes (e.g., Figure 10a) might then indicate that the upper crust is characterized by extremely poor spatial resolution. This sort of interpretation conflicts with our expectation that the upper crustal structure would be resolved reasonably well by the dense *Pg* coverage from a number of OBSs, and indeed this interpretation is not correct. Parameter correlation, delineated by correlation coefficients or scaled eigenmodes, is simply the correlation between parameter uncertainties, and spatial resolution is a different issue. In case of Figure 10a, for example, a correct interpretation is that if a true velocity of this upper-crustal node happens to be higher than its mean, all other nodes in the shallow upper crust would also deviate to higher velocities and deeper nodes in the upper crust would deviate in the opposite direction, and there is nothing more than this. If such deviations are smaller than the amplitude of some local feature, this small-scale feature would still be considered to be well resolved. Parameter correlation and spatial resolution are thus not so simply related.

[40] A popular exercise regarding spatial resolution is to conduct checkerboard tests; one creates synthetic data for a velocity model with a checkerboard-like pattern of perturbations, using the same sources and receivers as in the actual observation, adds random noise to the data, and invert them with an initial model free of perturbations to see how well given perturbations are recovered. This is a good test of linear sensitivity, so it is expected to work well when an initial model is close to the final model. However, in case of crustal tomography, where a reliable reference model is often unavailable, initial models are usually far from a true crustal structure, and nonlinear sensitivity becomes important (section 4.1). We need something more robust than conventional checkerboard tests to discuss the spatial resolution of highly nonlinear tomography.

[41] To appreciate why linear inverse theory does not work well, consider calculating a resolution matrix, which is another popular exercise in seismic tomography. For any linear system, we may relate a true model **m** and observation **d** as

$$\mathbf{Gm} = \mathbf{d}, \tag{19}$$

where **G** is a sensitivity kernel. Using singular value decomposition, the kernel can be decomposed as

$$\mathbf{G} = \mathbf{U}\Sigma\mathbf{V}^*, \tag{20}$$

where * denotes complex conjugate, and the generalized inverse of **G** is defined as

$$\mathbf{G}_g^{-1} = \mathbf{V}\Sigma^{-1}\mathbf{U}^*. \tag{21}$$

A generalized inverse solution is then expressed as

$$\mathbf{m}_g = \mathbf{G}_g^{-1}\mathbf{d}, \tag{22}$$

and by combining with equation (19), we can see that

$$\mathbf{m}_g = \mathbf{G}_g^{-1}\mathbf{Gm} = \mathbf{VV}^*\mathbf{m}. \tag{23}$$

The matrix product **VV*** is what is called a resolution matrix **R**, which quantifies how closely the generalized inverse solution approximates the true model [e.g., *Aki and Richards*, 1980; *Aster et al.*, 2005]. The resolution matrix is thus built entirely from the sensitivity kernel **G**. In the iterative linearized inversion described in section 3.1, the sensitivity kernel is modified at each iteration as a model improves by inversion, and a resolution matrix corresponding to a final model would be built from a sensitivity kernel that is constructed by raytracing through a model at one iteration earlier. This model at one iteration earlier is, however, already very close to the final model and is usually very different from an initial model. Such resolution matrix thus represents only a very small fraction of the total inversion

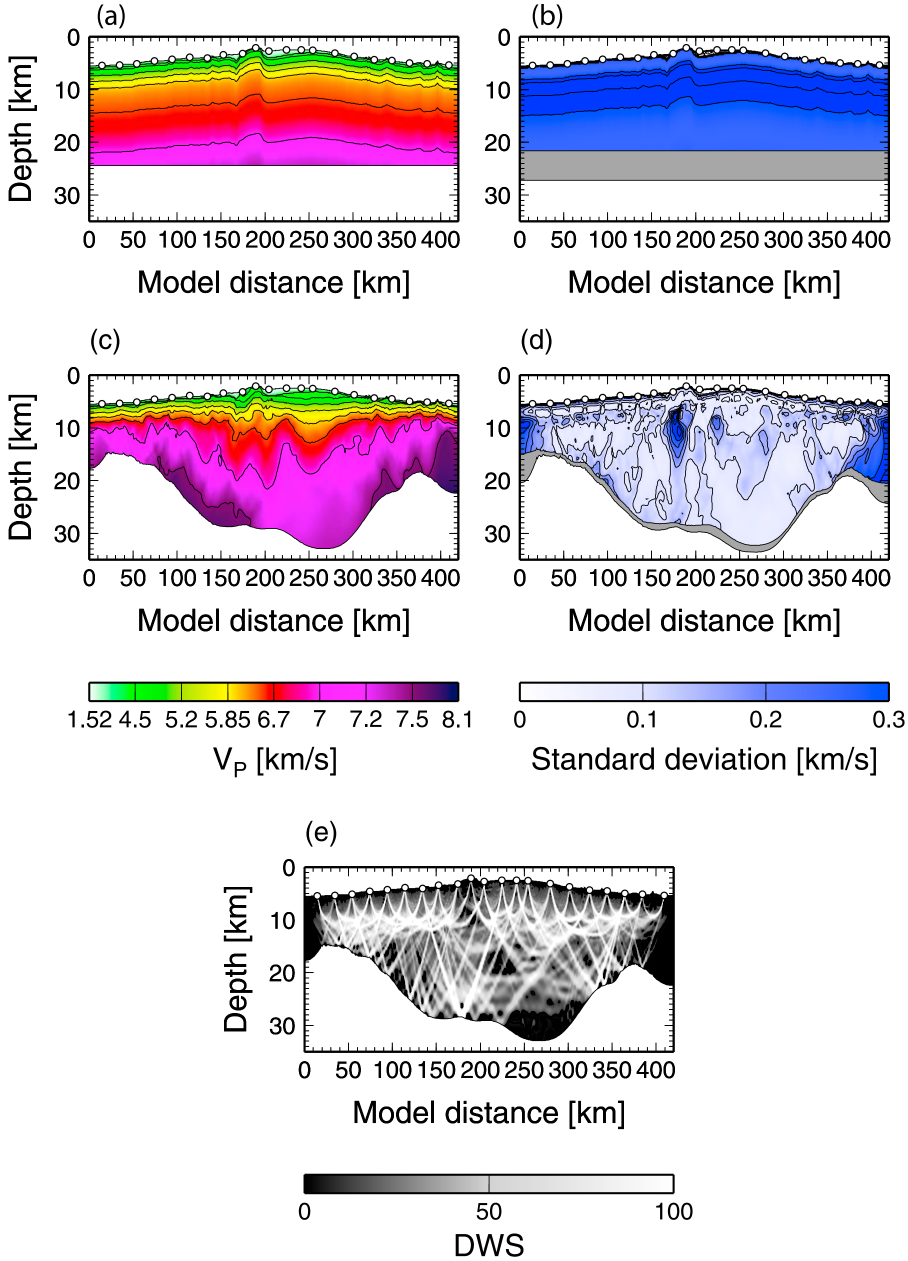

Figure 9. Summary of inversion results for Transect A. (a) The average of initial models corresponding to ~2,600 successful runs (i.e., with the final $\chi^2/N < 1$ and $R < 5$). (b) The standard deviation of those initial models. Gray region denotes the range of one standard deviation for initial reflector depths. (c) The average of optimal models ($\chi^2/N \sim 1$) chosen from the successful runs. (d) The standard deviation of those models. (e) Derivative weight sum, which may be regarded as a proxy for ray density, for the average model shown in Figure 9c. Open circles along seafloor denote the location of OBSs.

process and neglects the bulk of nonlinear relation between model and data, acquired by extensively sampling the model space starting from a number of initial models. It is important to remember that concepts in linear inverse theory are useful only when an initial model is close enough to a true model, but such a fortunate situation is rare in crustal tomography. In other words, doing a checkerboard test or calculating a resolution matrix is meaningful when we already know a crustal structure reasonably well (within the uncertainty of a few percent) from previous studies.

[42] As an alternative approach, we suggest that information on spatial resolution may be extracted directly from the ensemble of optimal models, by taking the statistics of the relative model deviation defined as

$$\delta V^*(x,z) = \frac{V(x,z)}{V_{\mathrm{ref}}(z/h(x))} - 1. \quad (24)$$

The mean of the relative deviation is shown in Figure 12a. Regions with positive or negative deviations occur where

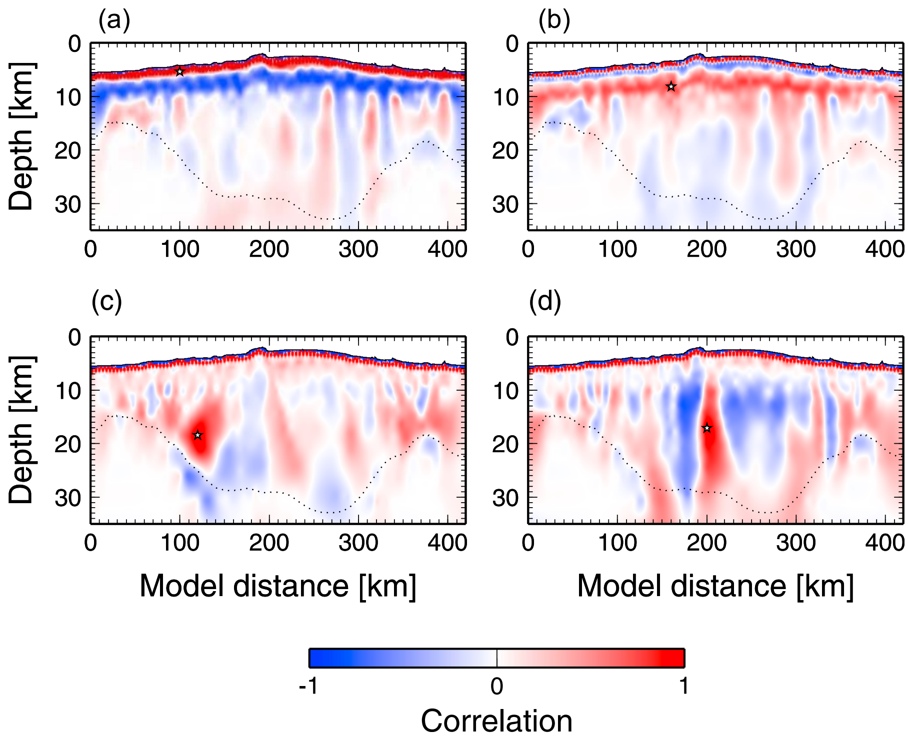

Figure 10. Correlation coefficients for selected velocity nodes (denoted by stars) from (a) upper crust, (b) mid-crust, and (c, d) lower crust. Dotted curve denotes the average Moho as shown in Figure 9c.

velocity values consistently deviate from a self-similar reference model, either positively or negatively. The upper crust beneath the rise summit (around km 190) is, for example, characterized by highly localized negative deviation (region A in Figure 12a). The robustness of such local features can be judged by the standard deviation of the relative deviation (Figures 12b and 12c). A pair of positive and negative patches in the upper crust around km 330 (region C), for example, can be persistently seen in these figures, so this small-scale feature may be deemed well resolved. In contrast, a weak positive deviation observed in the mid-crust beneath the rise axis (region E) is not robust as its positiveness diminishes in Figure 12c. These mid-crustal anomalies (regions E and F) are also seen as velocity contour undulations in Figure 9c, and their robustness can be understood from the statistics of relative deviation. Unlike checkerboard tests, we do not have to conduct another set of inversion in this approach; we have already conducted a fairly large number of highly nonlinear inversion runs to explore the model space extensively, so in principle, all of what we need can be obtained by exploiting the model ensemble in hand. The relative deviation defined above is also useful when discussing regional variations in detail.

5. Discussion

5.1. Nature of Source Mantle

[43] Most of the crustal velocity model for Transect A, excluding its edges, has low enough standard deviations for petrological interpretation to be meaningful (Figure 9d). As a preliminary attempt, we calculate average lower-crustal velocity and whole-crustal thickness from km 50 to km 350 at 20 km intervals with a 20 km wide averaging window, using all of ∼2,600 optimal models from the first sampling chain (Figure 13a). Taking a spatial average usually results in the cancellation of parameter trade-offs, so the uncertainty of average lower-crustal velocity tends to be somewhat smaller than that of individual velocity nodes. The reason for focusing on lower-crustal velocity is that the upper-crustal velocity is largely controlled by porosity and does not carry useful compositional information, and given the cumulative nature of the oceanic lower crust, we can use the lower-crustal velocity as the upper bound on the whole-crustal velocity [*Korenaga et al.*, 2002]. As mentioned in section 4.1, the lower-crustal section is the region below the velocity contour of 6.5 km s^{-1}, as velocity gradients change shapely at this velocity. Each pair of lower-crustal velocity and crustal thickness may be compared with theoretical predictions based on single-stage mantle melting. In general, the melting of a hotter mantle results in a thicker crust with higher crustal velocity, and if the mantle rises faster than surface divergence (i.e., active upwelling), it results in even thicker crust with little change in crustal velocity. The thinner part of the transect (with crustal thickness less than ∼17 km) exhibits a positive correlation between crustal thickness and velocity, as expected from a thermal origin for thick crust, but the thicker part, which occupies the bulk of the Shatsky Rise crust, exhibits a peculiar negative correlation: Crustal velocity is lower for thicker crust. At face value, this observation implies that the thicker part of the rise crust was formed by a more vigorous upwelling of *colder* mantle, which seems dynamically unlikely.

[44] There are a few complications to be considered, however, before discussing the implications of this negative correlation. First of all, the theoretical prediction for crustal

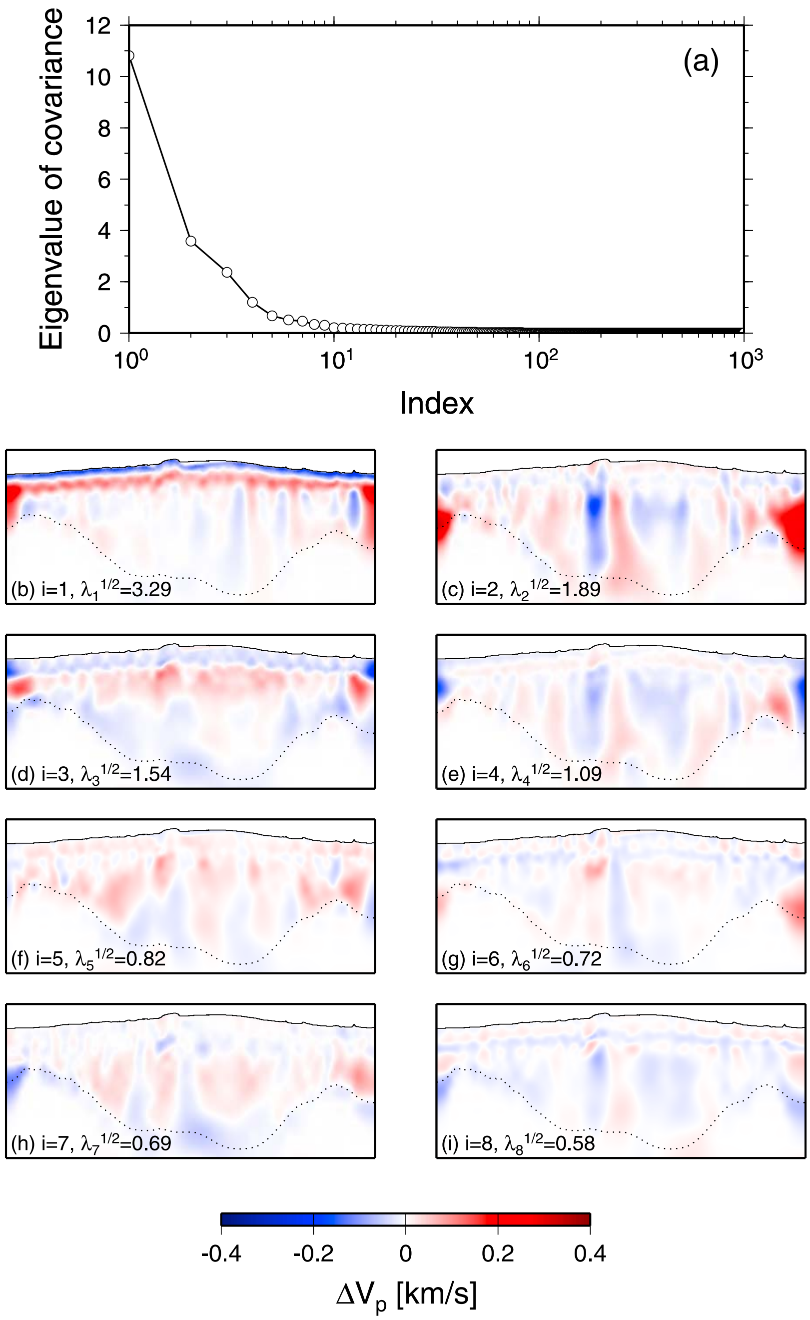

Figure 11. Results of principal component analysis of the ensemble of optimal models. (a) Eigenvalues of the covariance matrix shown in order of decreasing magnitude, and (b–i) scaled eigenmodes corresponding to the first to eighth eigenvalues. Dotted curve denotes the average Moho as shown in Figure 9c.

velocity and thickness as shown in Figure 13 assumes a temperature of 400°C and a pressure of 600 MPa [*Korenaga et al.*, 2002] when calculating the velocity of crustal rocks, so the velocity of our crustal model has to be corrected for this reference state. We thus applied a pressure correction of 0.2×10^{-3} km s^{-1} MPa^{-1} and a temperature correction of -0.4×10^{-3} km s^{-1} °C^{-1} [*Korenaga et al.*, 2002], using a linear conductive geotherm with a thermal gradient of 11°C km^{-1} and a surface temperature of 0°C, which is appropriate for ~140 Ma old oceanic lithosphere. As seen in Figure 13b, this correction results in an overall reduction in crustal velocity, but the sense of correlation is not affected.

[45] Another complication is the possibility of lower-crustal velocity not serving as an upper bound on whole-crustal velocity. As seen in Figure 13, passive upwelling of normal mantle (i.e., with a potential temperature of 1350°C) beneath mid-ocean ridges is predicted to create crust with the thickness of ~7 km and velocity of ~7.15 km s^{-1}, but the lower-crustal velocity of normal oceanic crust is typically around 6.9–7.0 km s^{-1} [*White et al.*, 1992]. *Korenaga et al.*

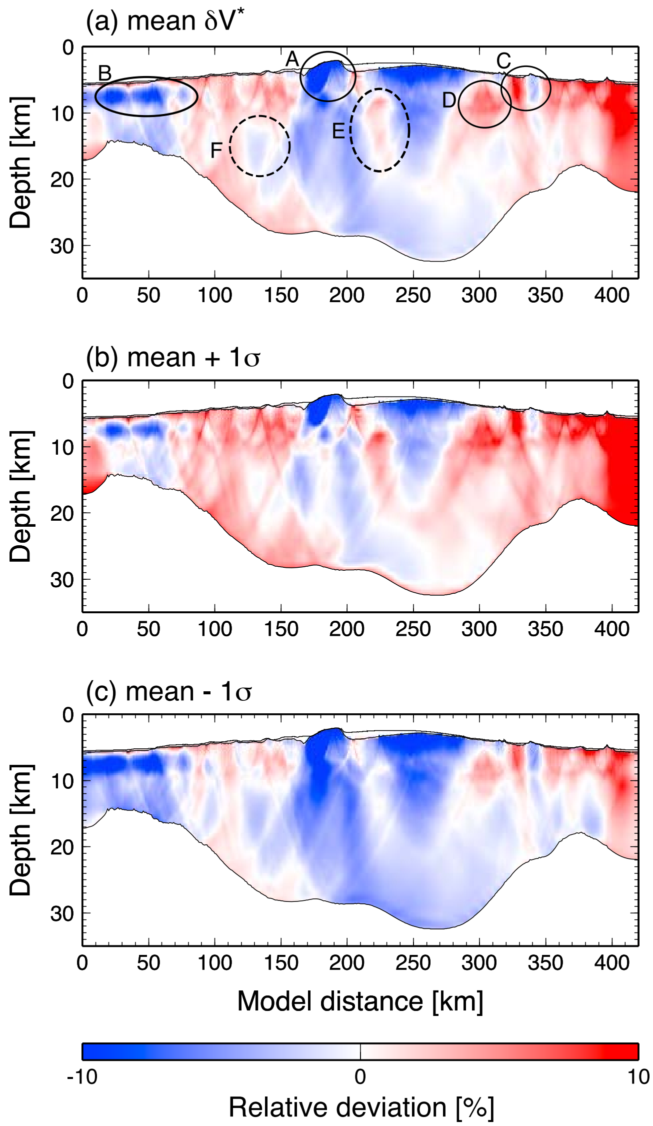

Figure 12. Statistics of relative model deviation δV^*: (a) mean, (b) mean plus one standard deviation, and (c) mean minus one standard deviation.

[2002] argued that the effect of seawater alteration was too small to explain this discrepancy and that the only viable mechanism was velocity reduction owing to crack-like residual porosity, which could be formed by thermal cracking [*Korenaga*, 2007]. The residual porosity effect is, however, expected to become less important as crustal thickness (thus pressure) increases, so it fails to explain the negative correlation between thickness and velocity.

[46] Finally, we have so far implicitly assumed that any vertical column through the Shatsky Rise crust was formed simultaneously so that individual pairs of crustal velocity and thickness correspond to different parts of the source mantle. This assumption may not be valid if the emplacement of igneous materials is laterally extensive, or if Shatsky Rise was formed on a preexisting oceanic crust. At least for the southeastern section of Transect A, however, the latter possibility appears to be unlikely because this part of the transect is nearly parallel to a paleo-ridge axis suggested by magnetic lineations [*Nakanishi et al.*, 1999], so upwelling mantle melt was probably emplaced directly at the spreading center. In any case, it will be important to take into account the details of tectonic reconstruction and consider a range of possible emplacement scenarios for more thorough petrological interpretation.

[47] A fundamental assumption behind the theoretical relation between crustal structure and mantle melting is the pyrolitic composition of a source mantle. Though a pyrolitic mantle is appropriate for the average mantle composition [e.g., *McDonough and Sun*, 1995; *Lyubetskaya and Korenaga*, 2007], the composition of the convecting mantle can be perturbed locally or regionally by plate-tectonic processes such as the subduction of oceanic crust [e.g., *Takahashi et al.*, 1998; *Korenaga and Kelemen*, 2000; *Korenaga*, 2004; *Sobolev et al.*, 2007], and indeed, the negative correlation between crustal thickness and velocity could be explained by a non-pyrolitic source mantle [*Korenaga et al.*, 2002]. Fortunately, Shatsky Rise has been investigated by a recent drilling expedition [*Sager et al.*, 2010, 2011], so a future synthesis with petrological and geochemical data obtained from drilling is expected to offer more insight into the origin and formation mechanism of Shatsky Rise.

5.2. Transect B Results

[48] Travel time data from Transect B have been analyzed with the identical inversion procedure applied for Transect A, and results are shown in Figure 14. Because of considerably fewer *PmP* travel times from the shorter transect (Figure 3), the lower-crustal structure suffers from much greater uncertainty (Figure 14d). Its petrological interpretation is too uncertain to be useful (Figure 14e), though it does verify that the two transects are consistent at their crossing point within uncertainty.

[49] Shooting for the seismic refraction survey was conducted after all of OBSs were deployed on both transects, so OBSs on Transect B recorded shooting over Transect A, and vice varsa. There are also additional shooting lines around the crossing point (not shown in Figure 1) to increase the three-dimensionality of the shooting geometry. The true value of Transect B will thus become clear when investigating the central volcanic system of Shatsky Rise by 3-D seismic tomography.

6. Summary and Outlook

[50] We have devised and implemented a new sampling strategy for joint refraction and reflection tomography. This was motivated by a long-standing desire to overcome the various shortcomings of the Monte Carlo uncertainty analysis originally formulated by *Korenaga et al.* [2000], all of which are addressed in this study. Model regularization has become more objective by randomly sampling correlation lengths from their broadly defined a priori ranges and by modifying the TOMO2D code to determine both smoothing and damping weights in an automated manner. The model space is sampled efficiently as well as extensively by defining the compact set of effective model parameters, which includes initial velocity models, correlation lengths, and the depth-kernel scaling parameter. In order to quantify

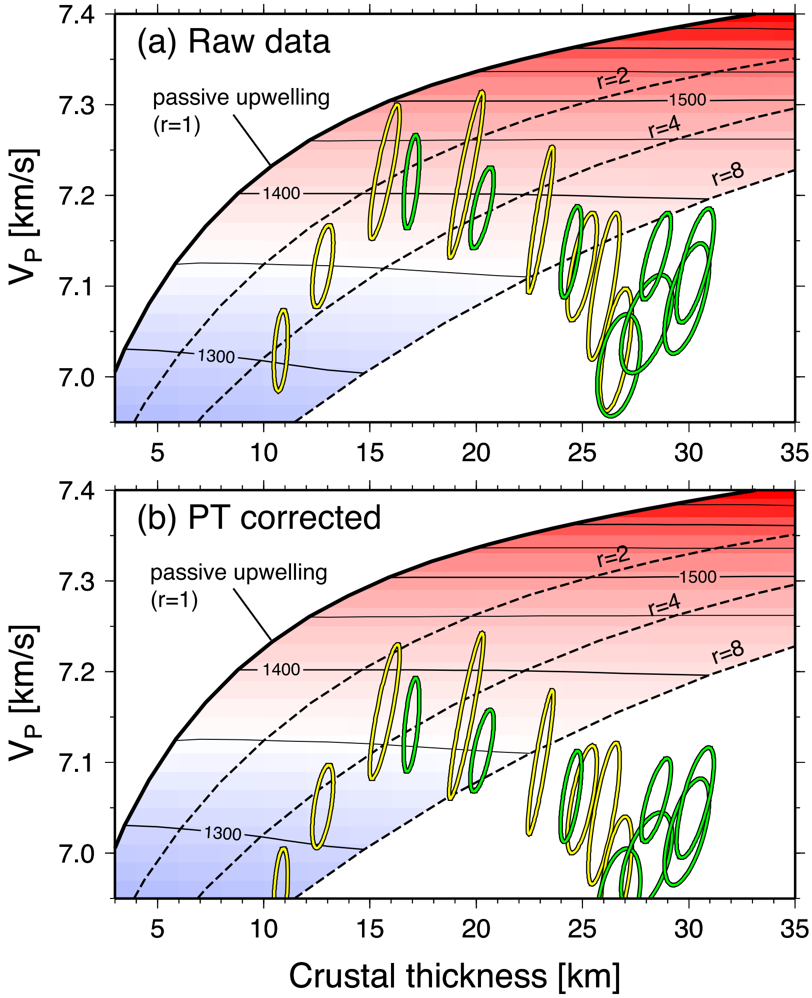

Figure 13. Covariation of crustal thickness and *P*-wave velocity at Shatsky Rise and its petrological interpretation based on the method of *Korenaga et al.* [2002]. Nearly horizontal contours are for mantle potential temperature in °C, which is also shown in color shading (white corresponding to the present-day ambient mantle temperature, 1350°C [*Herzberg et al.*, 2007]). Other more diagonal contours correspond to different degrees of active mantle upwelling (r), and thick curve represents the standard case of passive upwelling beneath a mid-ocean ridge ($r = 1$). Theoretical crustal velocities are values expected at a pressure of 600 MPa and temperature of 400°C. Shown over these predictions are Shatsky Rise data from the optimal model ensemble for Transect A. Ellipses denote the 68% confidence region of whole crustal thickness and lower-crustal velocity. Yellow and green ellipses represent the northwestern and southeastern sections of the transect. Velocity data are taken from the ensemble of (a) original optimal models and (b) optimal models corrected for pressure and temperature effects (see text for details).

the geological fitness of a velocity model, we introduced a new diagnostic based on the deviation from a self-similar reference, which can be computed in a bootstrap manner. Random sampling becomes progressively more efficient by periodically updating the estimate of the probability density function, and in this adaptive importance sampling, even failed inversion runs are utilized to constrain the shape of the density function. The convergence of Monte Carlo sampling can be assessed by running parallel sampling chains and comparing them, and the effective dimensions of the model space can be estimated by principal component analysis. As correlation coefficients or checkerboard tests are not suitable for spatial resolution tests of highly nonlinear tomography, we suggested the statistics of relative model deviation as a possible alternative.

[51] This new inversion scheme was applied to wide-angle seismic data collected over Shatsky Rise, resolving a massive plateau crust with a maximum thickness of ~30 km. The crustal seismic structure reveals an intriguing negative correlation between crustal thickness and velocity, which is hard to explain by a standard mantle melting model. This finding suggests that further investigation is required by taking into account tectonic reconstruction as well as recent drilling data. Shatsky Rise formed on a rapidly spreading ridge-ridge-ridge triple junction, so in contrast to other large igneous provinces formed on or in the vicinity of continents,

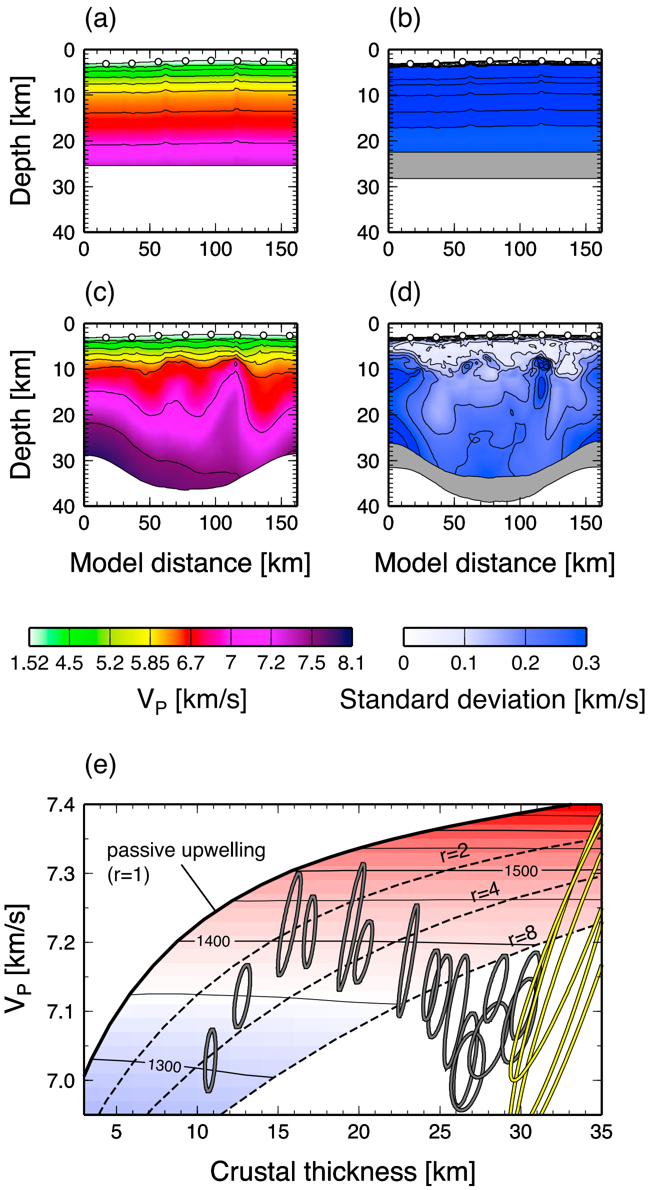

Figure 14. Inversion results for Transect B. As in Figure 9. (a) The average of initial models corresponding to ~3,200 successful runs (i.e., with the final $\chi^2/N < 1$ and $R < 5$). (b) The standard deviation of those initial models. (c) The average of optimal models ($\chi^2/N \sim 1$) chosen from the successful runs. (d) The standard deviation of those models. (e) Comparison of velocity-thickness measurements from Transect B (yellow ellipses) with those from Transect A (gray ellipses, which are taken from Figure 13a).

it offers an optimal opportunity to relate a crustal structure with its parental mantle dynamics. A future synthesis of our tomography results with other geological, geophysical, and geochemical constraints is thus expected to yield an incisive view on the origin of this gigantic oceanic plateau.

[52] **Acknowledgments.** We gratefully acknowledge the captain and crew of the R/V *Marcus G. Langseth* for their assistance in collecting seismic data on MGL1004. We thank Robert Steinhaus and his team for MCS data acquisition, Peter Lemmond, David Dubois, Timothy Kane, and James Elsenbeck for WHOI OBS data acquisition, and Jackie Floyd for onboard MCS data processing. This work was sponsored by the U.S. National Science Foundation under grants OCE-0926611 and OCE-0926945. This work was also supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. We appreciate constructive comments and suggestions from the Associate Editor, Jean-Xavier Dessa, and an anonymous reviewer.

References

Aki, K., and P. G. Richards (1980), *Quantitative Seismology*, W. H. Freeman, New York.

Anderson, D. L. (2000), The thermal state of the upper mantle; no role for mantle plumes, *Geophys. Res. Lett.*, *27*, 3623–3626.

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, Academic Press, Burlington, Mass.

Campbell, I. H., and R. W. Griffiths (1990), Implications of mantle plume structure for the evolution of flood basalts, *Earth Planet. Sci. Lett.*, *99*, 79–93.

Canales, J. P., R. S. Detrick, D. R. Toomey, and W. S. D. Wilcock (2003), Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, 8°15′N–10°5′N) from wide-angle seismic refraction profiles, *Geophys. J. Int.*, *152*, 766–794.

Coffin, M. F., F. A. Frey, P. J. Wallace, and the Shipboard Scientific Party (2000), *Proceedings of the Ocean Drilling Program Initial Report*, vol. 193, Ocean Drill., Program, College Station, Tex.

Contreras-Reyes, E., I. Grevemeyer, A. B. Watts, L. Planert, E. R. Flueh, and C. Peirce (2010), Crustal intrusion beneath the Louisville hotspot track, *Earth Planet. Sci. Lett.*, *289*, 323–333.

Den, N., et al. (1969), Seismic-refraction measurements in the northwest Pacific Basin, *J. Geophys. Res.*, *74*, 1421–1434.

Ewing, M., T. Saito, J. I. Ewing, and L. H. Burckle (1966), Lower Cretaceous sediments from the northwest Pacific, *Science*, *152*, 751–755.

Farnetani, C. G., and H. Samuel (2005), Beyond the thermal plume paradigm, *Geophys. Res. Lett.*, *32*, L07311, doi:10.1029/2005GL022360.

Furumoto, A. S., J. P. Webb, M. E. Odegard, and D. M. Hussong (1976), Seismic studies on the Ontong Java Plateau, *Tectonophysics*, *34*, 71–90.

Gershenfeld, N. (1998), *The Nature of Mathematical Modeling*, Cambridge, New York.

Gettrust, J. F., K. Furukawa, and L. W. Kroenke (1980), Crustal structure of the Shatsky Rise from seismic refraction measurements, *J. Geophys. Res.*, *85*, 5411–5415.

Herzberg, C., P. D. Asimow, N. Arndt, Y. Niu, C. M. Lesher, J. G. Fitton, M. J. Cheadle, and A. D. Saunders (2007), Temperatures in ambient mantle and plumes: Constraints from basalts, picrites, and komatiites, *Geochem. Geophys. Geosyst.*, *8*, Q02006, doi:10.1029/2006GC001390.

Holmes, R. C., M. Tolstoy, J. R. Cochran, and J. S. Floyd (2008), Crustal thickness variations along the Southeast Indian Ridge (100°–116°E) from 2-D body wave tomography, *Geochem. Geophys. Geosyst.*, *9*, Q12020, doi:10.1029/2008GC002152.

Hooft, E. E. E., B. Brandsdottir, R. Mjelde, H. Shimamura, and Y. Murai (2006), Asymmetric plume-ridge interaction around Iceland: The Kolbeinsey Ridge Iceland Seismic Experiment, *Geochem. Geophys. Geosyst.*, *7*, Q05015, doi:10.1029/2005GC001123.

Hosford, A., J. Lin, and R. S. Detrick (2001), Crustal evolution over the last 2 m.y. at the Mid-Atlantic Ridge OH-1 segment, 35°N, *J. Geophys. Res.*, *106*, 13,269–13,285.

Hussong, D. M., L. K. Wipperman, and L. W. Kroenke (1979), The crustal structure of the Ontong Java and Manihiki oceanic plateaus, *J. Geophys. Res.*, *84*, 6003–6010.

Kelemen, P. B., and W. S. Holbrook (1995), Origin of thick, high-velocity igneous crust along the U.S. East Coast Margin, *J. Geophys. Res.*, *100*, 10,077–10,094.

Korenaga, J. (2004), Mantle mixing and continental breakup magmatism, *Earth Planet. Sci. Lett.*, *218*, 463–473.

Korenaga, J. (2005), Why did not the Ontong Java Plateau form subaerially?, *Earth Planet. Sci. Lett.*, *234*, 385–399.

Korenaga, J. (2007), Thermal cracking and the deep hydration of oceanic lithosphere: A key to the generation of plate tectonics?, *J. Geophys. Res.*, *112*, B05408, doi:10.1029/2006JB004502.

Korenaga, J. (2011), Velocity-depth ambiguity and the seismic structure of large igneous provinces: A case study from the Ontong Java Plateau, *Geophys. J. Int.*, *185*, 1022–1036.

Korenaga, J., and P. B. Kelemen (2000), Major element heterogeneity of the mantle source in the North Atlantic igneous province, *Earth Planet. Sci. Lett.*, *184*, 251–268.

Korenaga, J., W. S. Holbrook, G. M. Kent, P. B. Kelemen, R. S. Detrick, H. C. Larsen, J. R. Hopper, and T. Dahl-Jensen (2000), Crustal structure of the southeast Greenland margin from joint refraction and reflection seismic tomography, *J. Geophys. Res.*, *105*, 21,591–21,614.

Korenaga, J., P. B. Kelemen, and W. S. Holbrook (2002), Methods for resolving the origin of large igneous provinces from crustal seismology, *J. Geophys. Res.*, *107*(B9), 2178, doi:10.1029/2001JB001030.
Lin, S.-C., and P. E. van Keken (2006), Dynamics of thermochemical plumes, 2, complexity of plume structures and its implications for mapping mantle plumes, *Geochem. Geophys. Geosyst.*, *7*, Q03003, doi:10.1029/2005GC001072.
Liu, J. S. (2001), *Monte Carlo Strategies in Scientific Computing*, Springer, New York.
Ludwig, W. J., and R. E. Houtz (1979), Isopach map of sediments in the Pacific Ocean Basin and marginal sea basins, *AAPG Map Ser.*, *647*.
Lyubetskaya, T., and J. Korenaga (2007), Chemical composition of Earth's primitive mantle and its variance, 1, methods and results, *J. Geophys. Res.*, *112*, B03211, doi:10.1029/2005JB004223.
Mahoney, J. J., M. Storey, R. A. Duncan, K. J. Spencer, and M. Pringle (1993), Geochemistry and age of the Ontong Java Plateau, in *The Mesozoic Pacific: Geology, Tectonics, and Volcanism*, edited by M. P. Pringle et al., *Geophys. Monogr. Ser.*, vol. 77, pp. 233–261, AGU, Washington, D. C.
Mahoney, J., J. Fitton, P. Wallace, and the Shipboard Scientific Party (2001), *Proceedings of the Ocean Drilling Program Initial Report*, vol. 192, Ocean Drill. Program, College Station, Tex.
McDonough, W. F., and S.-S. Sun (1995), The composition of the Earth, *Chem. Geol.*, *120*, 223–253.
Miura, S., K. Suyehiro, M. Shinohara, N. Takahashi, E. Araki, and A. Taira (2004), Seismological structure and implications of collision between the Ontong Java Plateau and Solomon Island Arc from ocean bottom seismometer-airgun data, *Tectonophysics*, *389*, 191–220.
Murauchi, S., W. J. Ludwig, N. Den, H. Hotta, T. Asanuma, T. Yoshii, A. Kubotera, and K. Hagiwara (1973), Seismic refraction measurements on the Ontong Java Plateau northwest of New Ireland, *J. Geophys. Res.*, *78*, 8653–8663.
Nakanishi, M., K. Tamaki, and K. Kobayashi (1989), Mesozoic magnetic anomaly lineations and seafloor spreading history of the northwestern Pacific, *J. Geophys. Res.*, *94*, 15,437–15,462.
Nakanishi, M., W. W. Sager, and A. Klaus (1999), Magnetic lineations within Shatsky Rise, northwest Pacific Ocean: Implications for hot spot-triple junction interaction and oceanic plateau formation, *J. Geophys. Res.*, *104*, 7539–7556.
Operto, S., and P. Charvis (1996), Deep structure of the southern Kerguelen Plateau (southern Indian Ocean) from ocean bottom seismometer wide-angle seismic data, *J. Geophys. Res.*, *101*, 25,077–25,103.
Parkin, C. J., and R. S. White (2008), Influence of the Iceland mantle plume on oceanic crust generation in the North Atlantic, *Geophys. J. Int.*, *173*, 168–188.
Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in C*, 2nd ed., Cambridge Univ. Press, New York.
Richards, M. A., R. A. Duncan, and V. E. Courtillot (1989), Flood basalts and hot-spot tracks: Plume heads and tails, *Science*, *246*, 103–107.
Robert, C. P., and G. Cassela (2004), *Monte Carlo Statistical Methods*, Springer, New York.
Romanowicz, B. (2003), Global mantle tomography: Progress status in the past 10 years, *Annu. Rev. Earth Planet. Sci.*, *31*, 303–328.
Sager, W. W., D. W. Handschumacher, T. W. C. Hilde, and D. R. Bracey (1988), Tectonic evolution of the northern Pacific plate and Pacific-Farallon-Izanagi triple junction in the Late Jurassic and Early Cretaceous (M21-M10), *Tectonophysics*, *155*, 345–364.
Sager, W. W., K. J. A. Klaus, M. Nakanishi, and L. M. Khakishieva (1999), Bathymetry of Shatsky Rise, northwest Pacific Ocean: Implications for ocean plateau development at a triple junction, *J. Geophys. Res.*, *104*, 7557–7576.
Sager, W. W., T. Sano, J. Geldmacher, and the Expedition 324 Scientists (2010), *Expedition 324: Shatsky Rise Formation*, *Proc. Integr. Ocean Drill. Program*, *324*.
Sager, W. W., T. Sano, and J. Geldmacher (2011), How do oceanic plateaus form? Clues from drilling at Shatsky Rise, *Eos Trans. AGU*, *92*, 37–38, doi:10.1029/2011EO050001.
Sallares, V., P. Charvis, E. R. Flueh, and J. Bialas (2003), Seismic structure of Cocos and Malpelo Volcanic Ridges and implications for hot spot-ridge interaction, *J. Geophys. Res.*, *108*(B12), 2564, doi:10.1029/2003JB002431.
Sallares, V., P. Charvis, E. R. Flueh, J. Bialas, and T. S. S. Party (2005), Seismic structure of the Carnegie ridge and the nature of the Galapagos hotspot, *Geophys. J. Int.*, *161*, 763–788.
Shulgin, A., H. Kopp, C. Mueller, L. Planert, E. Lueschen, E. R. Flueh, and Y. Djajadihardja (2011), Structural architecture of oceanic plateau subduction offshore Eastern Java and the potential implications for geohazards, *Geophys. J. Int.*, *184*, 12–28.
Sliter, W. V., and G. R. Brown (1993), Shatsky rise: Seismic stratigraphy and sedimentary record of Pacific paleoceanography since Early Cretaceous, *Proc. Ocean Drill. Program Sci. Results*, *132*, 3–13.
Sobolev, A. V., et al. (2007), The amount of recycled crust in sources of mantle-derived melts, *Science*, *316*, 412–417.
Takahashi, E., K. Nakajima, and T. L. Wright (1998), Origin of the Columbia River basalts: Melting model of a heterogeneous plume head, *Earth Planet. Sci. Lett.*, *162*, 63–80.
Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Sci., New York.
Tarduno, J. A., W. V. Sliter, L. Kroenke, M. Leckie, H. Mayer, J. J. Mahoney, R. Musgrave, M. Storey, and E. L. Winterer (1991), Rapid formation of Ontong Java Plateau by Aptian mantle plume volcanism, *Science*, *254*, 399–403.
Toomey, D. R., and G. R. Foulger (1989), Tomographic inversion of local earthquake data from the Hengill-Grensdalur central volcano complex, Iceland, *J. Geophys. Res.*, *94*, 17,497–17,510.
Vincenty, T. (1975), Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations, *Surv. Rev.*, *13*, 88–93.
White, R., and D. McKenzie (1989), Magmatism at rift zones: The generation of volcanic continental margins and flood basalts, *J. Geophys. Res.*, *94*, 7685–7729.
White, R. S., and L. K. Smith (2009), Crustal structure of the Hatton and the conjugate east Greenland rifted volcanic continental margins, NE Atlantic, *J. Geophys. Res.*, *114*, B02305, doi:10.1029/2008JB005856.
White, R. S., D. McKenzie, and R. K. O'Nions (1992), Oceanic crustal thickness from seismic measurements and rare earth element inversions, *J. Geophys. Res.*, *97*, 19,683–19,715.
Zelt, C. A., and R. B. Smith (1992), Seismic traveltime inversion for 2-D crustal velocity structure, *Geophys. J. Int.*, *108*, 16–34.
Zhang, J., and M. N. Toksöz (1998), Nonlinear refraction traveltime tomography, *Geophysics*, *63*, 1726–1737.
Zhang, J., U. S. ten Brink, and M. N. Toksöz (1998), Nonlinear refraction and reflection travel time tomography, *J. Geophys. Res.*, *103*, 29,743–29,757.