Imaging of structure at and near the core-mantle boundary using a generalized radon transform: 2. Statistical inference of singularities

P. Ma,[1] P. Wang,[2] L. Tenorio,[3] M. V. de Hoop,[4] and R. D. van der Hilst[2]

[1]   We present the second part of our approach to high-resolution imaging of deep Earth's interfaces with large volumes of broadband, three-component seismograms. We focus on the lowermost mantle, also referred to as D″ region, but the methodology can be applied more generally. The first part describes the generalized radon transform (GRT) of broadband *ScS* data (comprising main arrival, precursors, and coda). The GRT produces "image gathers," which represent multiple images of medium contrasts at the same image point near the base of the mantle. With a method for statistical inference we use this redundancy (1) to enhance the GRT images through improved recovery of weak contrasts and through suppression of spurious oscillations in the GRT image gathers and (2) to provide uncertainty estimates that can be used to identify the robust features in the images. Using the image gathers from paper 1 as input, we use mixed effects statistical modeling to produce the best estimates of reflectivity along with their uncertainty. In this framework, random noise in the signal is separated into white and coherent components using the geometry of the (GRT) imaging operators and a generalized cross-validation method. With synthetic data we show that conventional GRT images deteriorate substantially, in some cases to the point at which weak reflectors can no longer be detected, due to effects of uneven sampling, wave phenomena that are not accounted for in the underlying single scattering approximation, or errors in the assumed background wave speed model. We demonstrate that even in these circumstances, statistical analysis can yield adequate estimates of the true model. GRT imaging produces robust images of the core-mantle boundary (CMB) beneath Central America and suggests the presence of several structures in the D″ region, in particular between 100 and 200 and between 270 and 320 km above the CMB proper. Most of these structures are significant at the 1σ (that is, 68%) level, but at 2σ (95%) confidence the images show, at various depths above the CMB, intermittent instead of laterally contiguous features.

1. Introduction

[2]   The remote sensing of deep mantle discontinuities, for instance the core-mantle boundary (CMB), is a challenge in part because the seismic waves used to probe it propagate (at least twice) through Earth's heterogeneous mantle before they are observed and in part because the diagnostic seismological signals are often too small for direct observation and (forward) waveform modeling. Since the pioneering work by *Lay and Helmberger* [1983], many seismologists have tried to image the so-called D″ region of enhanced heterogeneity in the bottom 300 km or so of the mantle (see *Garnero* [2000] and *Helmberger et al.* [2005] for extensive reviews). Recently, mineral physicists presented compelling evidence for a phase transition in the mantle silicate $(Mg, Fe)SiO_3$, from perovskite (pv) to the so-called postperovskite (ppv) phase, at pressures that roughly coincide with the changes in elastic parameters inferred from seismic imaging [*Murakami et al.*, 2004; *Oganov and Ono*, 2004; *Shim et al.*, 2004]. The presence of a phase transition is consistent with earlier seismological observa-

---

[1]Department of Statistics, University of Illinois, Champaign, Illinois, USA.
[2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[3]Center for Wave Phenomena, Colorado School of Mines, Golden, Colorado, USA.
[4]Center for Computational and Applied Mathematics, Purdue University, West Lafayette, Indiana, USA.

tions and geodynamical arguments [*Sidorin et al.*, 1999; *van der Hilst et al.*, 2006].

[3] These are exciting developments for studies of Earth's deep interior. However, much uncertainty remains. On the one hand, estimates from theoretical and experimental mineral physics of the pressure at which the transition occurs show a large uncertainty, and the temperature and composition dependencies are not yet precisely known [*Shim*, 2005]. Consequently, the depth at which the pv-ppv transition occurs has an uncertainty of several 100 km. On the other hand, the seismological detection and characterization of such subtle and remote changes in elasticity faces formidable observational and theoretical (and computational) challenges. Various types of noise and scatter from three-dimensional (3-D) heterogeneity can mask weak signals in seismic data. Along with the massive size of modern data sets, this poses severe limitations on forward (waveform) modeling. The increasing availability of large volumes of densely sampled broadband data has begun to allow application of subsurface imaging methods based on inverse scattering, which exploits more efficiently the rich information contained in seismic waveforms.

[4] To meet the challenge of imaging and characterizing structure at and near remote interfaces and boundary layers, we are developing techniques for the automated identification, extraction, and interpretation of structural signal pertinent to subtle medium contrasts. Our approach differs in several important ways from forward modeling: First, we exploit the redundancy in large data volumes. Second, we make only a few restrictive a priori assumptions about the structures of interest. Third, we use data from a wide range of incidence angles. A simple analysis of reflections at an interface that marks a wave speed increase shows that the reflection coefficient is very small for near vertical incidence and increases dramatically toward critical incidence. For this reason, waveform modeling is usually restricted to observations made near (and beyond) the critical incidence, that is, large epicentral distances. In addition to the obvious reduction in target regions that can be studied, and the need to deal with triplicated waveforms, one should also realize that the radial resolution to interface depth degrades markedly with decreasing vertical slowness. Indeed, the wide-angle reflections considered in forward modeling have, in general, rather poor sensitivity to the depth of the contrasts being studied. Narrow angle data provide better resolution of discontinuity depth, but because of their small amplitudes they are rarely used in forward modeling. A further difference is that we can estimate formal uncertainties on the estimates of interface properties.

[5] We combine concepts from inverse scattering and modern statistics into a two-step strategy. As the first step, *Wang et al.* [2006] (hereinafter referred to as paper 1) developed a generalized radon transform (GRT) of global seismic network data in heterogeneous, anisotropic elastic media to map tens of thousands of seismograms to a set of multiple images of the same target structure. These "common image point gathers" reveal multiscale variations in elastic properties. For a detailed discussion and a historical perspective we refer readers to paper 1.

[6] In the second step, which is the focus of this paper, we obtain estimates of variations in Earth's deep interior from the image gathe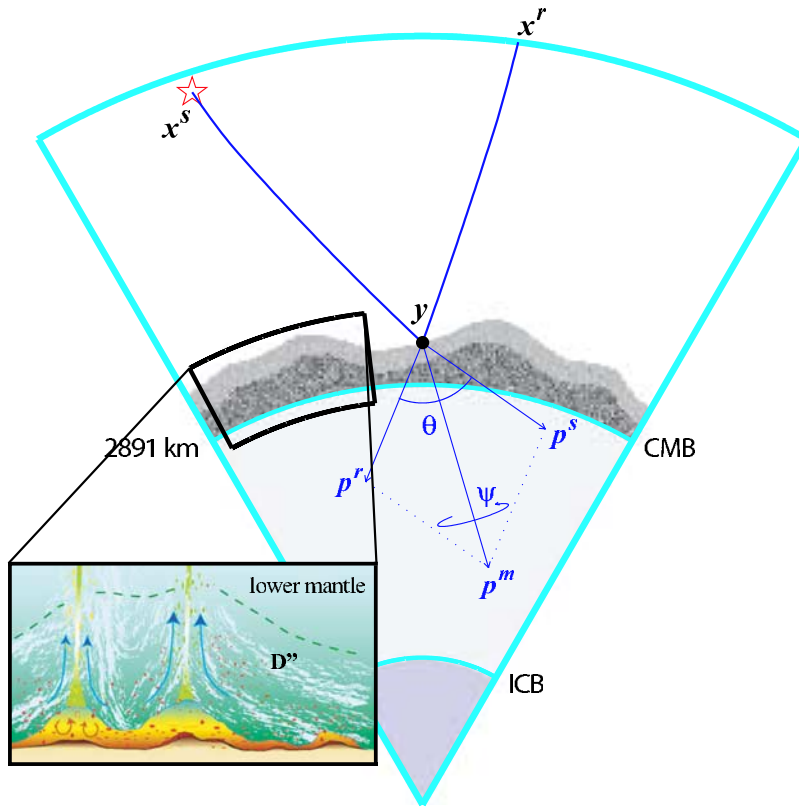rs using a statistical approach, in which the image gathers are modeled nonparametrically using mixed effects statistical models. In this framework, the random noise in the signal is allowed to have white and coherent components, and the latter are estimated from the data through prediction error minimization (also known as generalized cross validation). This methodology, a flexible type of Tikhonov regularization, can be used with different types of correlated noise and with the typically sparsely and unevenly sampled image gathers owing to the geographic distribution of sources and receivers.

[7] The mixed effects estimate of the reflectivity profile replaces the linear stack in the conventional GRT. This achieves three specific goals: first, it enhances signal-to-noise in the image gathers; second, it adapts to and mitigates effects of error in the background wave speed model, and third, it provides quantitative uncertainty estimates, which are more satisfactory than the ones from ubiquitous bootstrapping of slowness stacks. Of key importance is the "prestack" aspect of the analysis: the common image point gathers (and not the 2-D image profiles) are subjected to statistical analysis, and the optimal gathers are combined into the stack for that image point. This allows us to exploit the additional information contained in the dependence on scatter angle and to identify and remove poorly constrained gathers as well as artifacts due to, for instance, incorrect back ground velocities and presence of signal that cannot be modeled with single scattering.

[8] In section 2 we briefly summarize the (geometrical) aspects of the GRT that we need for the statistical analysis. In particular, we mention how three-component broadband global network data can be transformed to so-called common image point gathers. In section 3, we explain the concept of mixed effects statistical models and describe how we transform the image gathers to statistical estimates of discontinuities. Technical aspects of parameter estimation in mixed effects models are presented in Appendix A. In section 4 we test the effectiveness of the methodology on synthetic data against the presence of various types of noise, the imprint of source-receiver geometry, and imperfections of the mantle (wave speed) model. Finally, in section 5, we apply the method to the *ScS* wavefield (containing ∼80,000 broadband records) for the purpose of imaging the CMB and overlying D″ region beneath Central America. In addition to a laterally continuous image of the core-mantle boundary, the resulting 2-D image profile reveals intriguing structure, including multiple interfaces, in the few hundred kilometers above the CMB.

2. Common Image Point Gathers

[9] For a detailed discussion of the construction of the GRT we refer the readers to paper 1, but for completeness sake we mention here the aspects that are relevant for the development of the statistical methods. In essence, The GRT enables the automated transformation of a large number of seismic waveform data into a set of multiple images of the same structure in the vicinity of a target region. In order to do so, one has to account for variations in volumetric wave speed or, in general, density normalized stiffness (in case of an anisotropic model) and such geometrical aspects as the focal depth and radiation patterns of the earthquakes considered, the various move outs due to
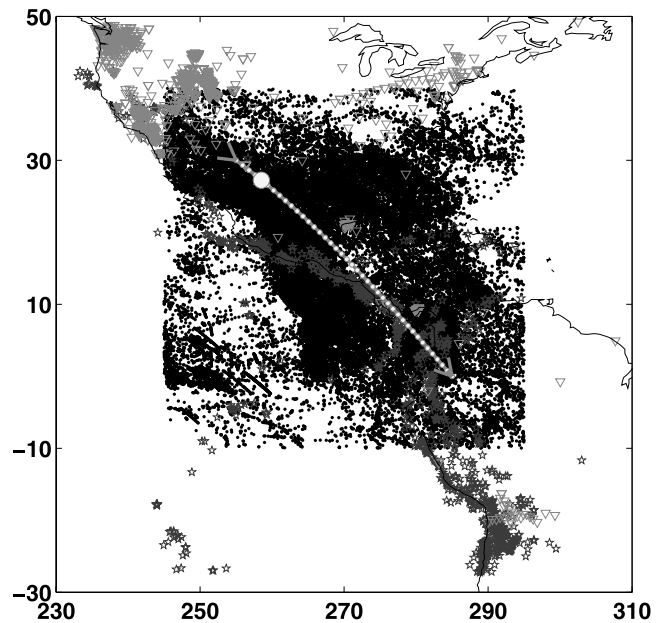
Figure 1a. Geometry of the GRT with ScS precursors and coda waves. For illustration purposes, the ray geometry and associated imaging parameters are shown for scattering at image point **y**. The objective of the work presented here, and by *Wang et al.* [2006], is the high-resolution imaging of the structures in the lowermost mantle, also referred to as D'' region, that may arise from boundary layer processes (e.g., flow), lateral variations in composition, and pressure-induced phase changes (as depicted in the inset [after *Garnero*, 2000]).

the large range of slownesses and epicentral distances of the data used, and the effects of uneven sampling. The geometry involved in the reconstruction is illustrated in Figure 1a. Figure 1b depicts the study region considered here as well as the distribution of *ScS* reflections at the CMB associated with the broadband wavefield used to construct the image profiles presented later in this paper.
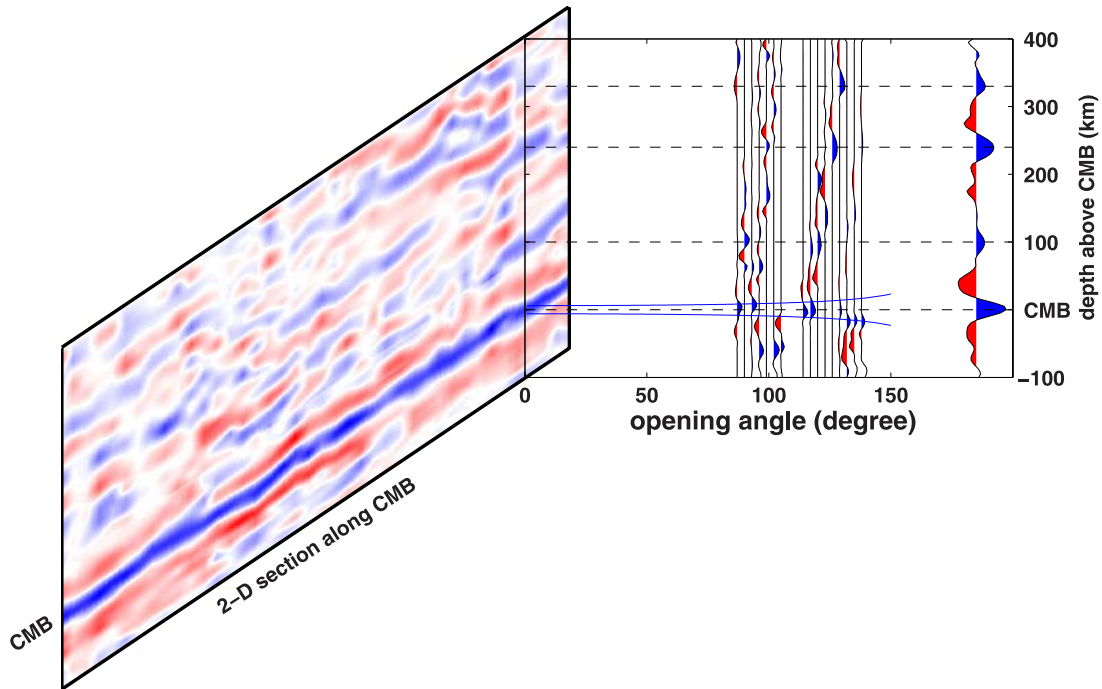
[10] With $\mathbf{y} = (y_1, y_2, y_3)$ the image point and superscripts s and r the association with a ray from a source and a receiver, respectively, the "two-way" traveltime for a particular diffraction branch associated with a raypath connecting $\mathbf{x}^r$ with $\mathbf{x}^s$ via $\mathbf{y}$ is denoted by $T = T(\mathbf{x}^s, \mathbf{x}^r, \mathbf{y})$. The slowness vector of the ray connecting a source point $\mathbf{x}^s$ with image point $\mathbf{y}$, evaluated at $\mathbf{y}$, is given by $\mathbf{p}^s(\mathbf{y})$, and $\mathbf{p}^s(\mathbf{x}^s)$ indicates the slowness along this ray evaluated at the source. Furthermore, we introduce the phase direction $\boldsymbol{\alpha}^s = \mathbf{p}^s/|\mathbf{p}^s|$ and the phase velocity V^s according to $|\mathbf{p}^s| = 1/V^s$. A similar notation is employed for the slowness vector related quantities along the ray connecting the receiver with the image point, namely $\mathbf{p}^r(\mathbf{y})$, $\mathbf{p}^r(\mathbf{x}^r)$, and $\boldsymbol{\alpha}^r$ and V^r.

[11] A key element in the GRT, the migration dip, $\boldsymbol{\nu}^m(\mathbf{y})$, is the direction $i\boldsymbol{\nu}^m(\mathbf{y}) = \mathbf{p}^m(\mathbf{y})/|\mathbf{p}^m(\mathbf{y})|$ of the migration slowness vector, $\mathbf{p}^m(\mathbf{y}) = \mathbf{p}^s(\mathbf{y}) + \mathbf{p}^r(\mathbf{y})$. Together, the migration dip and the phase directions of incoming and scattered rays define the scattering vector,

$$\psi = (\boldsymbol{\alpha}^s \times \boldsymbol{\alpha}^r) \times i\boldsymbol{\nu}^m \quad \text{at } \mathbf{y}. \qquad (1)$$

Figure 1b. Geographical map of study region. Small black dots depict ∼80,000 (specular) CMB reflection points associated with the broadband *ScS* wavefield used to construct the image profiles presented in this paper.

Figure 2. Image volume generated in paper 1 [*Wang et al.*, 2006]. (right) Common image point gathers in (scatter, or opening) angle for a selected geographic location at the CMB. For each scatter angle the best estimate of the (azimuth-integrated) reflectivity as a function of depth is shown. To the right of these gathers we show the GRT trace, which represents the optimal estimate of the radial contrasts in reflectivity at that particular CMB location. We note that the statistical inference described in this paper is used to produce the best possible estimate of such an GRT trace for a specific location; as such, it replaces direct (non) stacking linear stacking over azimuth and scatter angle. (left) A 2-D image profile that results from lateral juxtaposition of 40 of such GRT images. Peaks in contrast as revealed by the GRT produce blue "events" in the seismic section to the left. We note that except for interpolation between the GRT traces, no lateral smoothing or other image processing (or statistical inference) is used to produce such 2-D image profiles.

For a particular traveltime diffraction branch away from caustics at $\mathbf{x}^r$ or $\mathbf{x}^s$, the opening (or scattering) angle, θ, between incoming and scattered rays is related to the scattering vector according to

$$\sin\theta = |\boldsymbol{\psi}| \quad \text{at } \mathbf{y}; \quad \theta = \theta(\mathbf{x}^s, \mathbf{x}^r, \mathbf{y}). \tag{2}$$

The scattering azimuth, ψ, is the angular displacement of the scattering vector, normalized to one: $\boldsymbol{\psi}/|\boldsymbol{\psi}|$. An image gather at $\mathbf{y}$ can now be obtained by integrating, for each (θ, ψ), the preprocessed global network data over migration dip $\boldsymbol{\nu}^m$ (Figure 1a).

[12] With a generalized radon transform the data are transformed to an extended image volume, $\mathcal{I}(\mathbf{y}; \theta, \psi)$. Typically, one distinguishes the depth coordinate from the other coordinates representing the image point $\mathbf{y}$; an image gather is formed by plotting the image (or reflectivity) as a function of depth, which forms a radial reflectivity profile, against scattering angle and azimuth (θ, ψ). Integration over scattering azimuth, ψ, then yields azimuth-integrated (normalized) reflectivity profiles as a function only of scattering angle. In Figure 2 (right) we plot for each opening angle the best estimate of reflectivity as a function of depth (using statistical models as discussed in section 3) along with the best estimate of radial variations in reflectivity inferred from

the angle-dependent traces. Lateral juxtaposition of this estimate then produces 2-D image profiles.
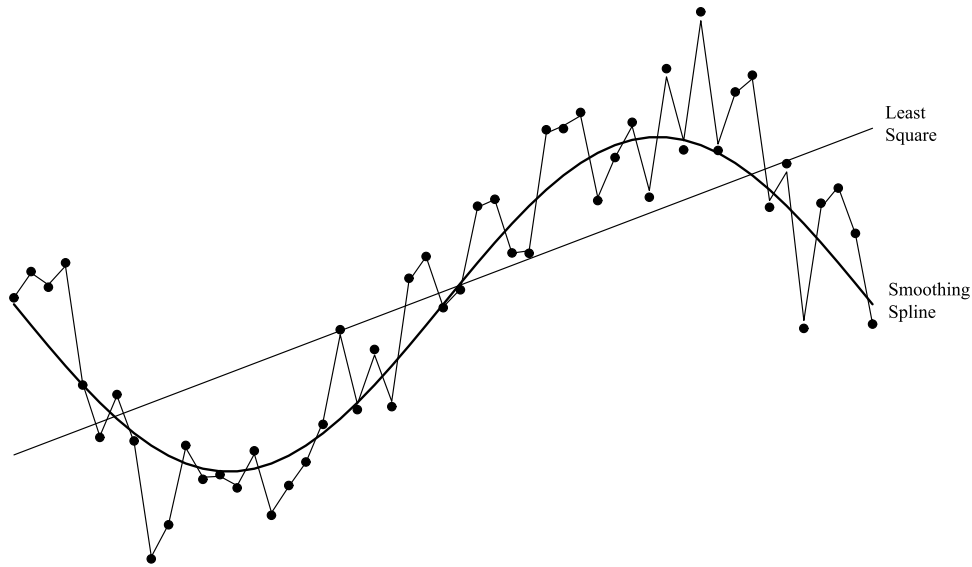
[13] The fact that image gathers form multiple images (namely, one for each pair of scatter angle and azimuth (θ, ψ)) of the same image point represents redundancy in the data. This is exploited in the statistical development of the GRT but should be accounted for with care. First, in paper 1 we noticed and analyzed the dilation with scattering angle θ as well as the possible presence of a polarity flip at a particular scattering angle. Second, artifacts (with residual moveout in (θ, ψ)) may appear due to the presence of caustics. (These can be removed by extending the GRT as in the work by *Stolk and De Hoop* [2002].) Third, imperfections in the wave speed model will lead to residual moveout with (θ, ψ). Fourth, the limited acquisition footprint leads to small shifts in (radial) depth in reflector images, different for each (θ, ψ). Fifth, scattered phases different from the phases scattered off the neighborhood of the CMB can lead to artifacts (locally) in the image gathers.

3. Statistical Inference of Singularities

3.1. From Image Gathers to Medium Contrasts

[14] How can we best extract (robust) information about contrasts in elastic (medium) parameters, that is, radial reflec-

Figure 3. Curve fitting with an arbitrary function showing a simple linear interpolation (over fitting), a least squares fit, and a smoothing spline estimate.

tivity profiles $\mathcal{I}(\mathbf{y})$, from the above mentioned image gathers (IGs)? Traditional methods involve stacking over (θ, ψ),

$$\mathcal{I}(\mathbf{y}) = \int\int \mathcal{I}(\mathbf{y}; \theta, \psi) d\theta d\psi. \qquad (3)$$

Indeed, a structural image could be obtained by performing a GRT, followed by an integration over scattering angle and azimuth. With array observations, the signal-to-noise ratio could be improved, for instance, by means of phase-weighted stacking, which is nonlinear in the data [*Schimmel and Paulssen*, 1997]. However, this approach does not fully benefit from structural information that might be contained in the "noise." Moreover, it would not be clear how to assess the uncertainty of the final estimate.

[15] Singularities in one-dimensional signals could be detected by means of wavelets, but the resolution and uneven coverage of the (multidimensional) image gathers make an approach based on the wavelet transform unfeasible.

[16] The analysis that we develop here can be viewed as a focusing procedure, in which the geometry and statistics of noise in the data is used to enhance the scanning for singularities or discontinuities through common image point gathers. In the data we distinguish and model separately the systematic (nonrandom) and random components, hence the name "mixed effects" statistical modeling. (We note that this approach is similar to Tikhonov regularization with random effects.)

[17] Assuming that we have IGs in the angle domain (that is, $\mathcal{I}(\mathbf{y}; \theta, \psi)$), we estimate an optimal reflectivity profile $\mathcal{I}(\mathbf{y})$ through mixed effects statistical inference instead of (linear or nonlinear) stacking, as in (3). This allows us to mitigate more effectively the systematic errors due to uncertainty in background velocity, source location and origin time, and for imaging artifacts, artifacts owing to phase misinterpretation and, e.g., near source scattering, and errors introduced during preprocessing (e.g., filtering).

3.2. Mixed Effects Models

3.2.1. Motivation and Strategy

[18] To motivate the methodology, we recall how a line is fitted to observed data using the linear model $Y_i = \alpha + \beta x_i + \varepsilon_i$ $i = 1, \ldots, n$, where x_i are fixed design points and ε_i are independently and identically distributed (usually Gaussian) with mean zero and variance σ^2. Typically we estimate the slope β and intercept α using a least squares approach based on minimizing the residual sum of squares $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \alpha - \beta x_i)^2$.

[19] Suppose that instead of fitting a straight line we wish to fit a curve to the data; thus we write our model as $Y_i = \eta(x_i) + \varepsilon_i$. The residual sum of squares that is to be minimized can be written as

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \eta(x_i))^2. \qquad (4)$$

By fitting a curve we mean that not only do we want to estimate the value of $\eta(x)$ at $x = x_i$ but, in fact, at any x in the domain. Clearly, this problem is ill-posed as there are many functions that pass through all the observed data points with a zero residual sum squares. We need to impose some restrictions on $\eta(x)$ in order to regularize the problem and to obtain a reasonable function estimate (see Figure 3).

[20] One approach is to assume that the function $\eta(x)$ is a linear combination of polynomials. However, polynomials over large intervals often display undesirable oscillations and other artifacts, especially when such polynomials are of order greater than three. Another approach is to find a balance between the residual sum of squares and a measure of the smoothness of the unknown function, measured, for example, by a functional $J(\eta)$. To obtain such a functional, we use the norm of the first derivative: $J(\eta) = \int \eta'(x)^2 dx$. In our case, smoothness is justified by the smoothing effect of a convolution of the reflector with a deterministic resolution filter, which reflects the acquisition imprint.

[21] To fit the data with smooth function, instead of (4), we minimize

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \eta(x_i))^2 + \lambda J(\eta), \qquad (5)$$

where λ is a smoothing parameter that controls the trade-off between the goodness of fit and smoothness of η. This method is commonly called penalized least squares or Tikhonov regularization. The minimization of (5) is performed in a space of functions where the evaluation of a function at a point is a continuous linear functional (i.e., a reproducing kernel Hilbert space). Remarkably, the function $\eta(x)$ that minimizes (5), η_λ, turns out to be a finite linear combination of particular basis functions. The minimization problem is thus reduced to solving a linear system for the coefficients of this linear combination. The parameter λ is estimated through a refined leave-one-out cross validation. Penalized least squares has been studied extensively in the literature; see, e.g., *Wahba* [1990] and *Gu* [2002] for comprehensive treatments of the subject.

[22] A disadvantage of the described method is that it does not perform well with correlated noise, which limits its applicability in a variety of settings. To overcome this difficulty, we use mixed effects statistical models, which explicitly distinguish systematic (nonrandom) from random components. They provide a unified framework for modeling a variety of correlated data [*Vonesh and Chinchilli*, 1997; *Wang*, 1998; *Pinheiro and Bates*, 2000]. For our purposes, an important characteristic of these models is that it can flexibly accommodate angular dependence and account for coherent noise and artifacts.

[23] Azimuth-integrated angle gathers are functions of depth and angle that can be modeled as

$$G_{ij} = \int \mathcal{I}(y_j; \theta_i, \psi) d\psi = g(\theta_i, \ y_j) + \varepsilon_{ij},$$

where $g(\theta_i, y_j)$ is the ideal noiseless angle gather at angle θ_i and (radial) depth y_j, and ε_{ij} are random noise terms that are usually coherent. Our goal is to estimate $g(\theta_i, y_j)$, which should be a singular function of depth (smoothed by the convolution with some pulse) that varies slowly with angle (including a dilation). To achieve this goal, we make additional assumptions on g and the noise. For g, we use a flexible functional representation in terms of some basis functions φ_ℓ in depth. The noise is modeled as a sum of a discrete harmonic process and white noise. The angle gather is thus modeled as

$$G_{ij} = g(\theta_i, y_j) + h(\theta_i, y_j) + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \ \ j = 1, \ldots, b. \qquad (6)$$

Here, $h(\theta_i, y_j) = \sum_{k=1}^{m} h_k(\theta_i) \ \phi_k(y_j)$ represent a harmonic process that models coherent noise through sinusoid functions $\phi_k(y)$ in the depth coordinate. For a fixed angle θ_i, the coefficients $h_k(\theta_i)$ are assumed to be uncorrelated, but $h_k(\theta_i)$ and $h_k(\theta_j)$ may be correlated for $\theta_i \neq \theta_j$. The errors ε_{ij} are independent Gaussian random variables with mean zero and variance σ^2. A refined model for the noiseless IG that include angle dilation is given by $g(\theta_i, \alpha_i y_j)$, which replaces $g(\theta_i, y_j)$.

[24] To estimate g, we need estimates of the covariance parameters of the random effects (i.e., h_k), including sinusoid frequencies determining the ϕ_k. We will consider three particular cases of (6) that can be analyzed efficiently. We use the same notation for random errors, ε_{ij}, for the different models below.

3.2.2. Random Intercept Model

[25] For weak angle dependence we can approximate $g(\theta_i, y_j)$ in (6) by

$$g(\theta_i, \alpha_i y_j) \approx b_i + g(\alpha_i y_j). \qquad (7)$$

We start with the simple model in (7) that uses a single parameter to aggregate higher-order approximation terms and which also ignores the angle dilation effect:

$$G_{ij} = b_i + g(y_j) + \varepsilon_{ij}, \quad i = 1, \ldots, a, \ \ j = 1, \ldots, b, \qquad (8)$$

where the b_i are Gaussian $N(0, \ \sigma_s^2)$. Clearly, (8) is a particular case of (6) obtained by setting $g(\alpha_i y_j) = g(y_j)$ and $h(\theta_i, y_j) = b_i$.
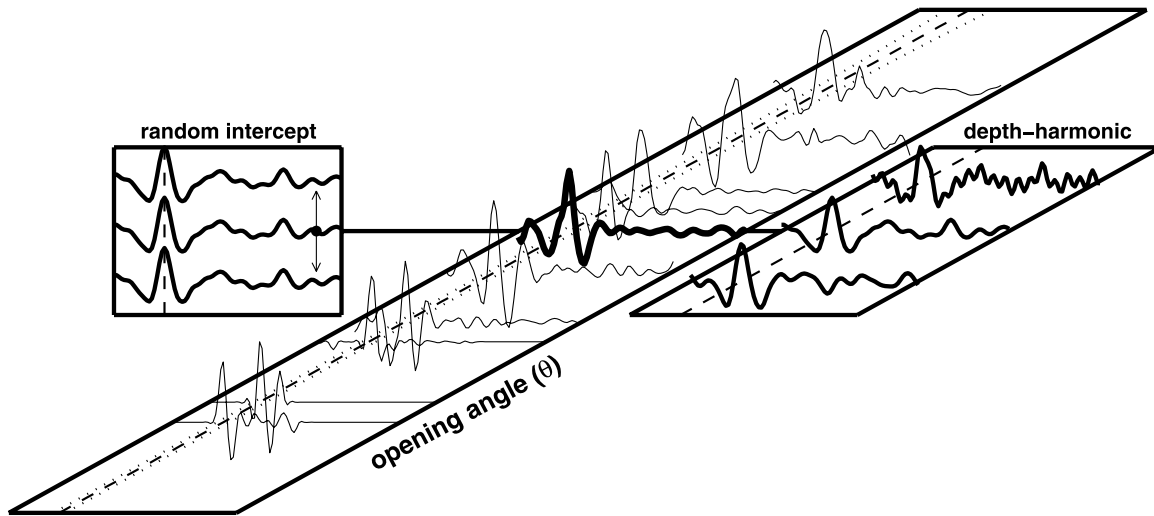
[26] We estimate $g(y)$ by minimizing a penalized sum of squares similar to (5) [*Robinson*, 1991]:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (G_{ij} - g(y_j) - b_i)^2 + \frac{\sigma^2}{\sigma_s^2} \sum_{i=1}^{a} b_i^2 + n \lambda J(g), \qquad (9)$$

where $J(g) = \int g''(y)^2 dy$ quantifies the smoothness of g and the smoothing parameter λ controls the trade-off between the goodness of fit and smoothness of g. Since the parameters λ and σ^2/σ_s^2 provide a large family of possible estimates, one has to choose values that lead to good estimates as measured by the goodness of fit and smoothness of the estimate. To choose the parameters, we use a method known as generalized cross validation (GCV), a method that is widely used in Tikhonov regularization [*Wahba*, 1990]. The basic idea is as follows. We start by fixing the angle (optimal parameters are chosen for each fixed angle, i.e., each fixed i). For each j and choice of parameters we obtain estimates $\widehat{b}_{i,-j}$ and $\widehat{g}_{-j}(y_j)$ of b_i and $g(y_j)$, respectively, using all the data but G_{ij}. That is, we predict the value of G_{ij} using the data $G_{i\ell}$ for $\ell \neq j$ (the notation $-j$ means that the jth observation has been deleted from the data set). These estimates in turn provide the prediction $\widehat{G}_{ij,-j} = \widehat{b}_{i,-j} + \widehat{g}_{-j}(y_j)$ of G_{ij}. It has been shown [see *Wahba*, 1990] that by using the deleted estimates one obtains better estimates of the prediction error which is what we want to minimize. Ordinary cross validation chooses the parameters that lead to the smallest value of $\sum_j (G_{ij} - \widehat{G}_{ij,-j})^2$. In Appendix A this is discussed in more detail.

[27] We use Bayesian confidence intervals as a benchmark for assessing the uncertainties in the estimate of $g(y)$. We calculate posterior means and variances of $g(y)$ under the Bayes model and use them to construct 95% and 99% confidence interval estimates. An important feature of these intervals is that they approximately have the correct across-the-function coverage probability. That is,

$$\frac{1}{n} \sum_{i=1}^{n} P[g(y_i) \in C_\beta(y_i)] \approx \beta,$$

Figure 4. An illustration of how the noise adds to an image trace in the (left) random intercept, (right) depth harmonic, and angle dilation depth harmonic (center, indicated by dotted dilation curves) models. (middle) Actual image gathers for different opening (or scatter) angles θ. The random intercept model estimates the best "common component" in theses traces (Figure 4, left, surface, middle trace), which is allowed to move up-and-down in order to find the best fit to the gathers. The depth harmonic models detect and correct for spurious oscillations in radial direction: at the right, the middle trace depicts the best image estimate, and the adjacent traces show this trace with a (randomly chosen) harmonic either added to or subtracted from it. The third model, the angle dilation depth harmonic model, is similar to the depth harmonic model proper, but it also accounts for the angle dilation (depicted with dotted lines in Figure 4, middle) that results from the geometry of the GRT image problem at hand (see paper 1).

where $C_\beta(y_i)$ is the Bayesian confidence interval and β is the coverage (we use $\beta = 0.95$ and 0.99). Although these intervals were originally derived under the independence assumption, there are straightforward extensions that can be used with correlated data. See *Wahba* [1983], *Nychka* [1988], and *Ma* [2003] for details.

3.2.3. Depth Harmonic Model

[28] Now we take a different approach to model the noise in the data. Instead of combining all the effects in a simple random effect, we model the harmonic process explicitly. We extend the random intercept model (8) to

$$G_{ij} = \sum_k a_{ik} \cos\left(\omega_{ik} y_j + \beta_{ik}\right) + b_i + g\left(y_j\right) + \varepsilon_{ij}, \qquad (10)$$

where the coefficients a_{ik} and phases β_{ik} are random and independent and the frequencies ω_{ik} are fixed but unknown.

[29] Since the harmonic component is fixed for each angle, we may think of it as a sinusoid signal contaminating $g(y)$. Thus, although it could be included as a random effect, it is easier to fit and remove it from each trace. This fitting is done using ARMA time series techniques described in Appendix A. The corrected G_{ij} is then analyzed using the simpler random intercept model.

3.2.4. Angle Dilation Depth Harmonic Model

[30] To account for the angle dilation (see paper 1) as well as a harmonic process of coherent noise, we use the following modification of (10) to fit the data:

$$G_{ij} = \sum_k a_{ik} \cos\left(\omega_{ik} y_j + \beta_{ik}\right) + b_i + g\left(\alpha_i y_j\right) + \varepsilon_{ij},$$

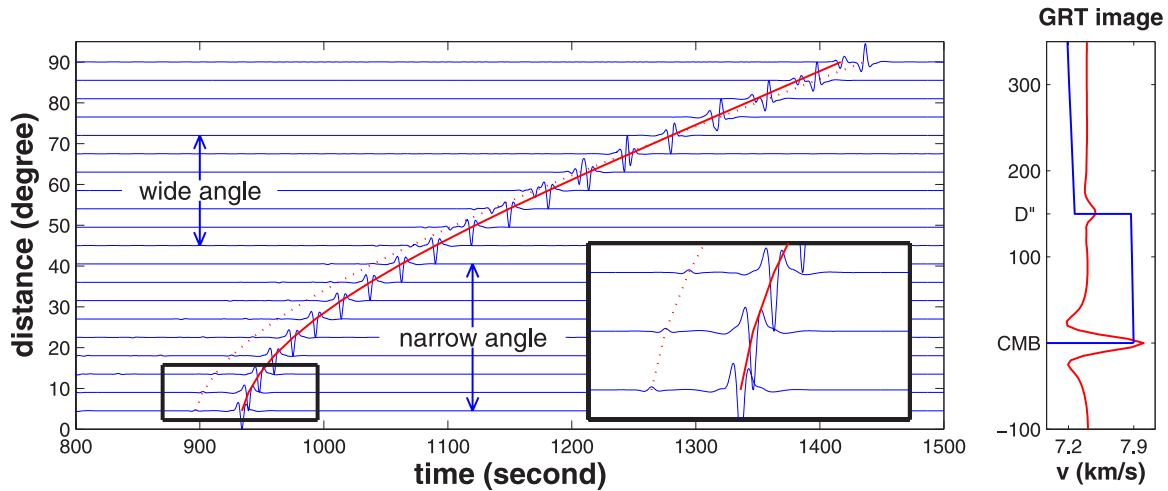$$i = 1, \ldots, a, \ \ j = 1, \ldots, b, \qquad (11)$$

where α_i is the angle dilation effect. As the remaining noise ε_{ij} is assumed white, the term b_i has to account for higher-order angle dilation effects not modeled by $g(\alpha_i y_j)$. The lack of structure in the box plots of the residuals shows that this assumption is reasonable.

[31] The parameters are estimated in two stages: The first stage is the same harmonic correction as before. In the second stage the parameters α_i and b_i are simultaneously estimated with other parameters through penalized least squares (see Appendix A for more details).

[32] In Figure 4, we illustrate the properties of the different mixed effects models. In Figure 4 (middle), we plot an actual image gather. On Figure 4 (left) we illustrate the random intercept effects: the best image estimate (middle curve) is randomly shifted up or down. On Figure 4 (right) we illustrate the depth harmonic effects: A randomly chosen harmonic is added (back) or subtracted (front) to the best image estimate (middle curve). The angle dilation of the third model is indicated by two parting (dotted) curves in the image gather in Figure 4 (middle).

3.3. Model Validation

[33] The adequacy of a statistical model is checked by comparing, at each scattering angle, the data G_{ij} to the estimates $\widehat{G}_{ij}$ under the corresponding model. We study plots of the residuals defined as $e_{ij} = G_{ij} - \widehat{G}_{ij}$ normalized by the model estimate $\widehat{\sigma}$ of the noise standard deviation σ (i.e., $e^*_{ij} = e_{ij}/\widehat{\sigma}$). box plots of the angle-dependent residuals provide information about the shape of the residual distribution (that is, its median, interquartile range and presence of outliers). For example, inadequacies in the background elastic properties would lead to biased residuals not centered

Figure 5. Synthetic data and illustration of the robustness of the GRT in the presence of random noise in the data. (left) Synthetic (WKBJ) records of *ScS* and *SdS*, that is the signal from a weak reflector at 150 km above the CMB. The inset shows the weak precursory energy for narrow angle reflections. At wide angles this reflection becomes stronger and crosses over with, and eventually arrives in the coda of, direct *ScS*. Red lines are theoretical traveltime curves for *ScS* (solid) and *SdS* (dashed). (right) GRT image trace (solid red line) constructed from the synthetic data shown on the left and the wave speed profile used to generate the synthetic waveforms (blue curve).
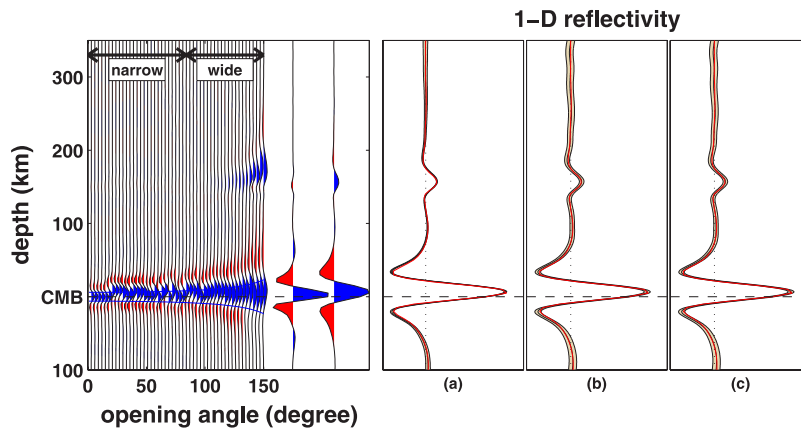
around zero, with a magnitude that depends on how the structure is sampled (i.e., the scatter angle). Residuals centered at zero without discernible patterns indicate that the estimate model is reasonable. Examples of box plot analysis are given below.

4. Study of Synthetic *ScS* Data

[34] In paper 1 we used WKBJ synthetics to explore certain aspects of image recovery with GRT, including its performance on random, additive noise in the data. Here we use the same synthetic data (Figure 5, left) to test the performance of our statistical inference method on nonran-
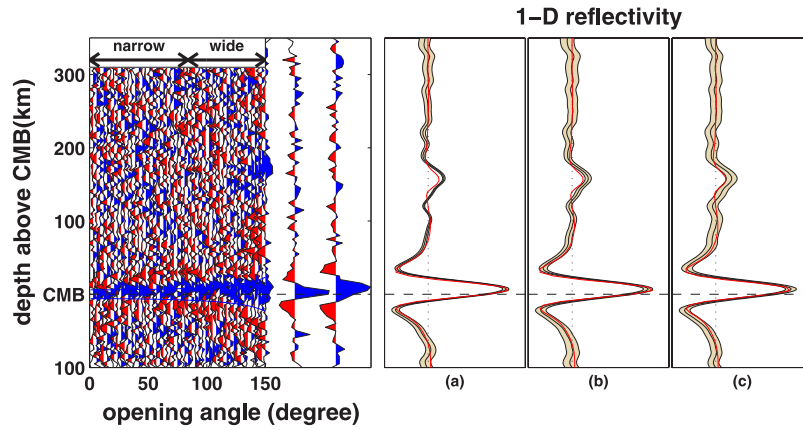
dom noise in the image gathers. The data are generated from a 1-D, spherically symmetric background model with a wave speed increase at 150 km above the CMB (Figure 5, right). Figure 5 also illustrates the distance ranges associated with "narrow" and "wide" scattering angle data. At narrow angles, the signal from the top of D″ is a precursor to *ScS*, see inset in Figure 5 (left), but because of a small reflection coefficient it has a small amplitude and will be difficult to detect in raw data. At larger scatter angles the reflection becomes stronger, and it arrives closer to and, eventually, after the direct *ScS* arrival.

[35] We explore how each of the statistical models described in section 3 handles the following constituent

Figure 6. Effect of the acquisition imprint, that is, source-receiver distribution.(left) Image gathers for a range of opening (scatter angles) and, right next to them, the result of the GRT restricted to narrow (left) and wide (right, with arrow) scattering angles. The narrow and wide-angle data are illustrated in Figure 5. (right) Estimates of the reflectivity profiles using (a) the random intercept, (b) depth harmonic, and (c) angle dilation depth harmonic statistical models. The bands around the mean estimate (black line) correspond to 95% Bayesian confidence intervals. For reference, the red line depicts the true signal.
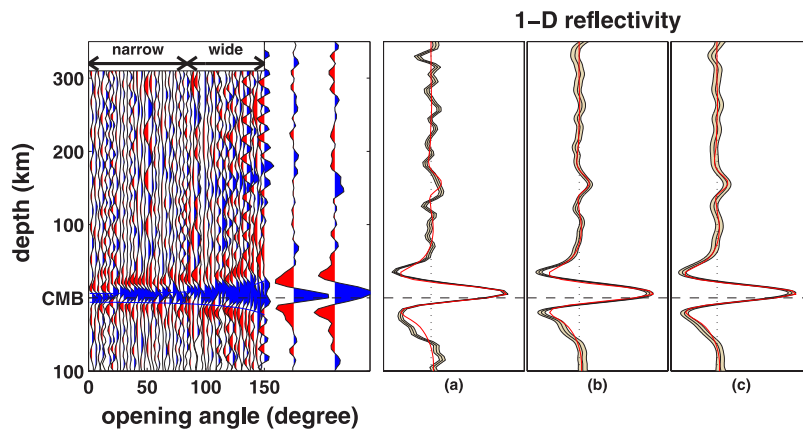
Figure 7. Estimation in the presence of random noise in the image gathers. For this purpose, we added random noise to the image gathers shown in Figure 6, that is, for a realistic acquisition imprint. Both the narrow and wide-angle GRT stacks reveal significant jitter and neither suggests the presence of a reflector at 150 km above CMB. In contrast, the top reflector is detected in the statistical estimates, even though the images of it are slightly distorted compared to the true model (red line). In the absence of a harmonic noise component, all three models detect the contrast at the CMB.

effects in the estimation: highly irregular sampling (due to the actual, uneven distribution of stations and events), random additive noise in the image gathers (for example, due to scattering not explained by the single scattering approximation), and random harmonic noise in the image gathers (for example, due to isolated spectral components generated by the imaging). For each test, we show the GRT image (that is, the substack over the image gathers for different scatter angle) both for the narrow and the wide-angle data. However, the statistical results discussed and shown here are based on data associated with the large scattering angles (in Figures 6–9 the GRT image restricted to large scattering angles is indicated by a vertical arrow). In Figures 6–9, the black curves represent the statistical estimate of $g(y)$ and the light green bands depict the 95%
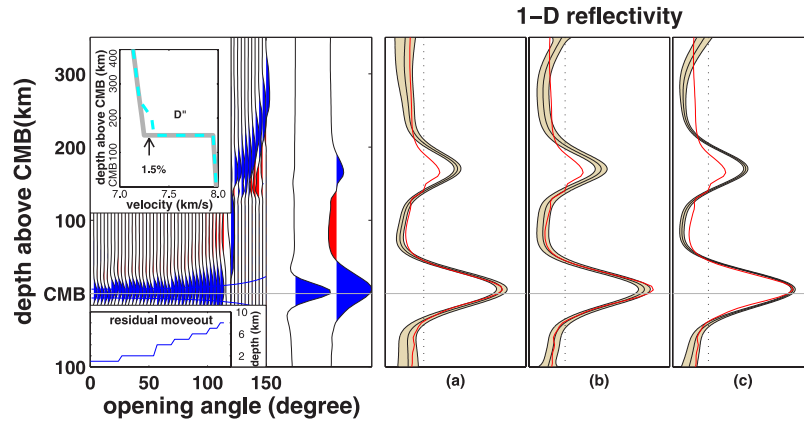
Bayesian confidence intervals. For comparison, we also plot the true model (red line).

4.1. Effects of "Station-Event" Sampling

[36] First, we analyze the effect of acquisition imprint. Acquisition is here viewed as the spatial distribution of stations and events. If the stations and events were regularly spaced, aliasing would be an immediate concern. This is not quite the situation, but the effects due to the actual acquisition, locally, are related. In Figure 6 we illustrate how our approach treats these effects, using a realistic acquisition geometry. The image gathers on the left reveal significant scatter associated with nonrandom sampling. The two traces directly to the right of the image gathers illustrate the action of the GRT restricted to small scattering angles (left trace) and large scattering angles (right trace, with arrow). The

Figure 8. Same as Figure 7 but for harmonic instead of random noise. While performing better than the GRT without statistical estimation, the random intercept model begins to break down in the presence of a strong harmonic component in the noise structure; indeed, trace a reveals much spurious structure and the estimates of the contrasts at CMB and 150 km above it deviate significantly from the actual model (red line). Both harmonic models, traces b and c, retrieve the model well (that is, within 2σ they are the same as the true model), but the angle dilation depth harmonic model c performs slightly better than the standard depth harmonic b.

Figure 9. Same as Figure 6 but for image estimation with an inaccurate wave speed model. As before, we use synthetic data generated from a model with a simple stepwise increase in wave speed at 150 km above the CMB (solid gray line in inset, top left), but as background model for the GRT we assumed a model with a slightly different D″ structure (blue dashed line in inset, top left).

image is distorted in that the phase of the two events associated with the reflectors has changed. Figure 6 (right) shows the mixed effects model estimates for random intercept (Figure 6a), depth harmonic (Figure 6b), and angle dilation depth harmonic (Figure 6c). The black line in the middle of the 95% confidence band shows the mean estimate, and the red line shows the signal to be recovered. We see that the three different methods are all capable of undoing the effects of irregular event-receiver sampling, and restore the phase.

4.2. Effects of Noise in the Image Gathers

[37] In Figure 7 we have added random noise (energy not explained by the single scattering approximation) to the image gather obtained from the noise-free data in Figure 5 (left), subjected to a realistic acquisition geometry. The GRT image traces (small and large scattering angles) directly to the right of the image gather are strongly affected by the noise; in fact, the top reflector is no longer visible, whereas multiple weak, false reflectors appear. The random intercept model is capable of providing a clean estimate of the image (Figure 7a), including the top reflector. The estimate is consistent with the true model (red line). Because the added noise did not have a harmonic component, there is no perceptible difference in performance between the random intercept model and either of the harmonic models (Figures 7b and 7c).

[38] In Figure 8 we have added random noise with a harmonic component to the image gather obtained from the noise-free data in Figure 5 (left) again subjected to a realistic acquisition geometry. The GRT image traces (small and large scattering angles) directly to the right of the image gather are both affected by the noise; the top reflector has disappeared, and more spurious signals are visible. If untreated, the latter can produce spurious events in the image profiles. The random intercept model estimate (Figure 8a) is beginning to have problems recovering the top reflector as seen by the true signal (red line) being mostly outside the error band. We also see remnant oscillations of the harmonic components that the random intercept model could not remove. The depth harmonic models (Figures 8b–8c) clearly perform better; many of the harmonic oscillations

have been removed and the estimate (black line), with the error bars, is consistent with the true signal.

4.3. Effects of an Inaccurate Wave Speed Model

[39] An important assumption in the application of the GRT as developed in paper 1 is that we have a reasonable estimate of the elastic properties, say the wave speed, of the background. Incorrect properties of the back ground model would produce (scatter) angle-dependent artifacts in the image gathers. We recall that the model validation operates on the prestack image gathers, which helps us recognize and correct such artifacts. In order to demonstrate this premise, we again form GRT images using the noise free, WKBJ modeled data (Figure 5, left), but now we assume background properties that are different than those used to produce the synthetic data. The perturbed wave speed model assumed in the GRT and the one used to produce the data are shown in the inset to Figure 9.
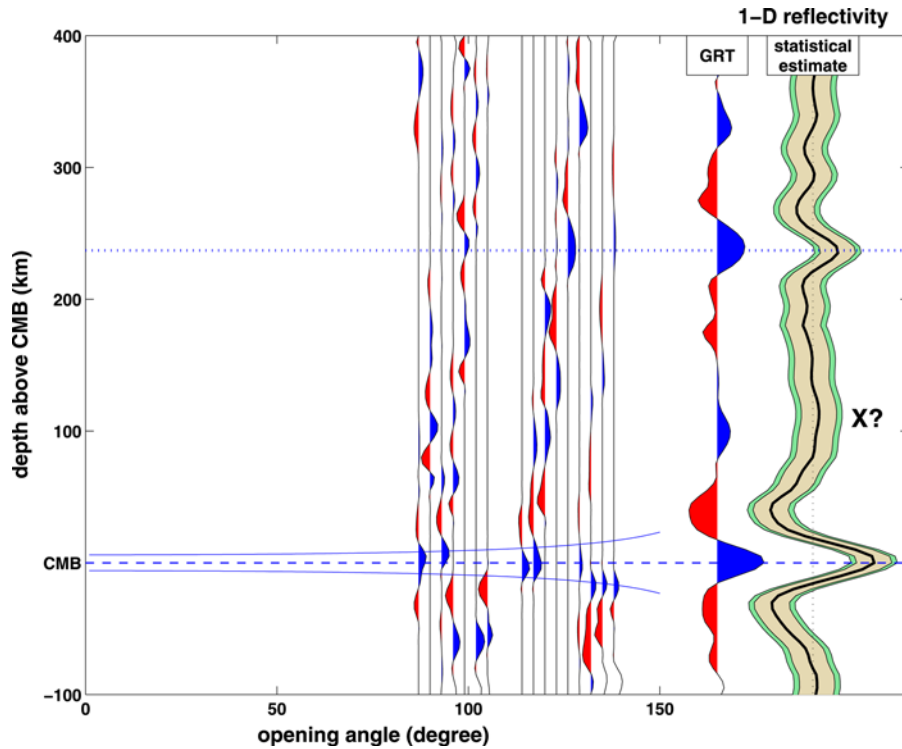
[40] The GRT image traces are affected by the use of the incorrect model. In fact, for small scattering angles, the top reflector has become almost invisible. Figures 9a–9c demonstrate that the different mixed effects models are all capable of recovering the image of both reflectors. However, we observe a clear deterioration in spatial resolution, and the depth estimate of the reflectors have decreased. Tests like these show that not knowing the background wave speeds well does not prevent us from detecting interfaces, but it may produce artificial boundary topography.

5. Imaging the Lowermost Mantle Beneath Central America

[41] We illustrate the proposed methodology with a study of inference of singularities over a $50° \times 50°$ patch of the core-mantle boundary (CMB) beneath Central America, using the broadband data from paper 1. The data selection and preprocessing is explained in section 4.1 of paper 1.

5.1. Statistical Analysis of an *ScS* Common Image Point Gather

[42] We select a particular location and image gather, and apply the analysis of section 4. In Figure 10 we show the

Figure 10. A typical image gather associated with large scattering angles and a GRT stack versus an image estimation based on the depth harmonic model. The bands of two colors in the depth harmonic estimate correspond to 95% and 99% confidence error bars. The GRT stack suggests a scatterer ~100 km above the CMB, but the statistical estimate only shows a weak, broad structure (marked by X), which may suggest that it is not (statistically) significant.

gather for large scattering angles (on the left), the associated GRT stack (first trace to the right), and an image estimate with the depth harmonic model (second trace to the right). The blue lines indicate the dilation derived from the GRT (see paper 1). The (linear) GRT stack contains multiple reflectors, but the (nonlinear) image estimate suggests that not all of them are real. Indeed, we note the general difference in appearance of the statistical estimate compared to the GRT stack. The estimate shows the CMB, symmetric, zero phase, unlike in the GRT stack, as well as a clear indication of a reflector about 240 km above it, within 95% confidence level. This suggests that the location and width of the two main peaks are consistent with the model of the lowermost mantle and the data resolution bounds used. The image estimate of the CMB appears on the coarsest resolution viewed with respect to the expected dilation in the image gather (blue lines). One may argue for the presence of a weak reflector about 115 km above the CMB, marked as X in Figure 10, but it remains within the error bars associated with 95% confidence.

5.2. The 2-D Image Profile

[43] We form an 2-D image profile (which in the context of this paper represents a synthesis of the mean statistical estimates) by repeating the analysis in section 5.1 for a large number of CMB points along a 2500 km long great circle path. We also address the validity of the different mixed effects models; for this purpose, we have selected six locations in the profile, indicated by image profiles a–f in Figure 11.

[44] In Figure 11 we present the results collectively. Figure 11 (top) is an image profile obtained with the random intercept model, whereas in Figure 11 (bottom), two image profiles are obtained with the angle dilation depth harmonic model. The Figure 11 (bottom) profile uses a dual color scale to display structure above the CMB (color) at an amplification by a factor of 5 compared to the structure around the CMB proper (grey tone). Superimposed on Figure 11 (bottom) are the statistical estimates at the mentioned six locations. Below these profiles, for these locations, we plot the image estimates accompanied with the 95% Bayesian confidence intervals, for the random intercept model (Figure 11, bottom left, top), the depth harmonic model (Figure 11, bottom left, middle), and the angle dilation depth harmonic model (Figure 11, bottom left, bottom).

[45] For location a we carry out a detailed validation of the entire image gather. The corresponding box plots are shown on the lower left. Each mixed effects model accounts for coherent noise in the traces in its own way. For the three mixed effects models introduced in this paper, we provide illustrations of the within-angle residuals e_{ij} as defined in section 3.3. Each box plot shows a box bounded by the 25% and 75% percentiles of the residual distribution (for e_{ij} with j fixed). This interval provides information about the spread of the distribution. The line within the box indicates the median, and the whiskers provide information about the symmetry and tails of the distribution. For all three models we observe that the residuals are close to centered at zero, which is an indication that the mean estimates seem
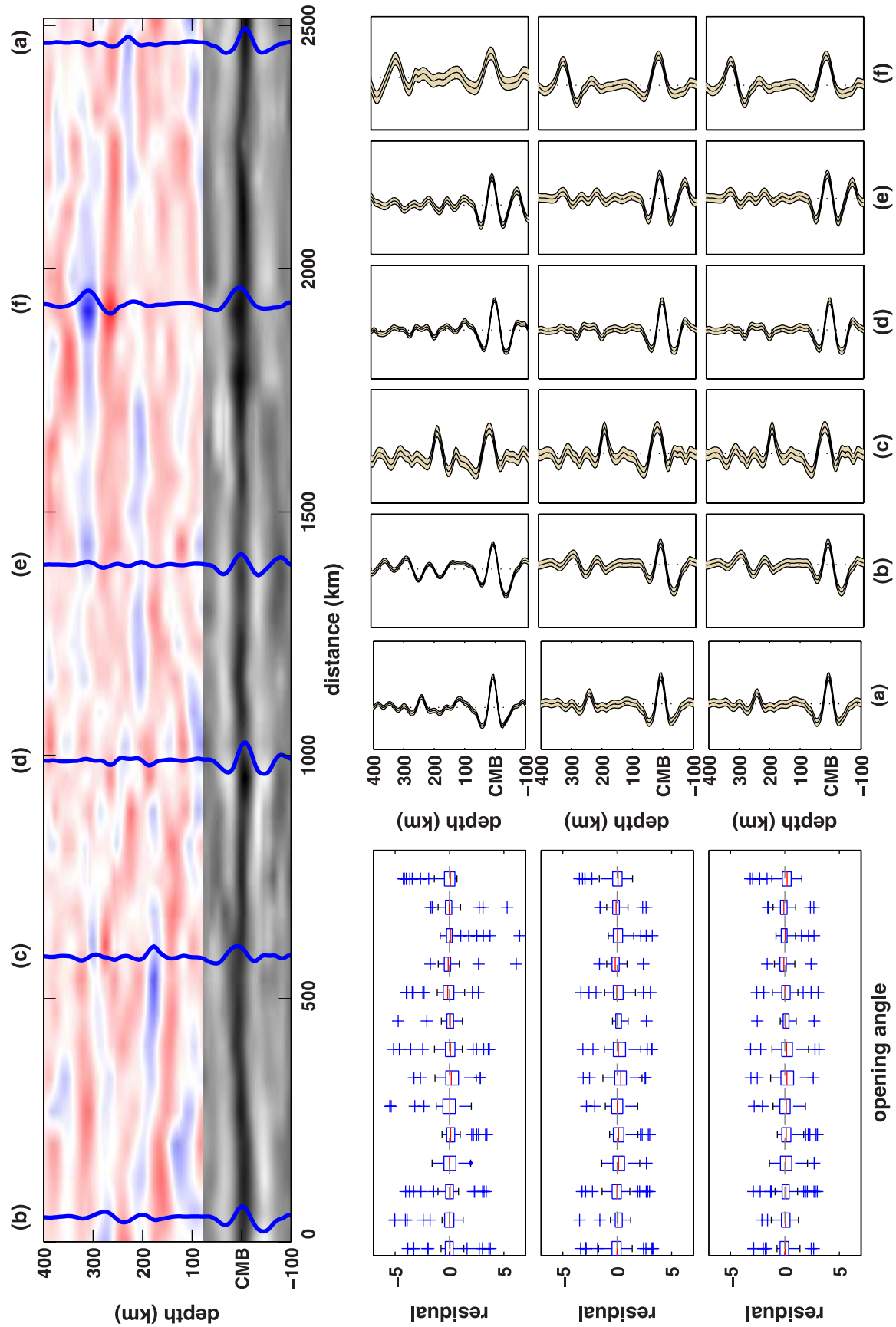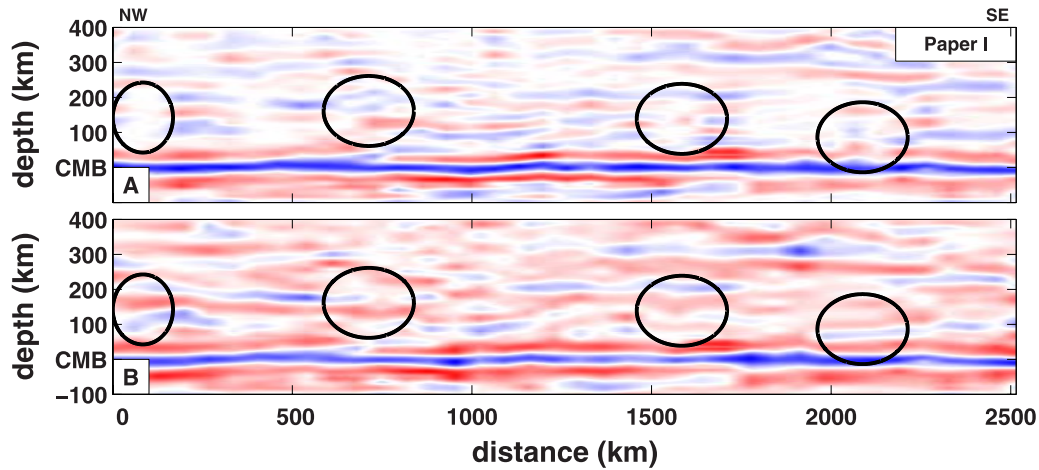
Figure 11

Figure 12.   (a) Comparison of original 2-D image profile (presented in paper 1), through (b) application of the (angle dilation) depth harmonic model. There is no exaggeration in the vertical scale.

reasonable. Their variability changes with angle, however, which is a measure of the degree of adequateness of the models. Such unmodeled variability may have an effect in uncertainty estimates.

[46]   As was the case in the examples presented in Figures 6–9, at first glance the three statistical models may seem to perform similarly. Upon closer inspection, however, we can see differences. We compare the models by viewing how the box plots of residuals are scattered around zero. The depth harmonic models show less structure and fewer outliers than the random intercept model, which may indicate that the they provide a better fit for our data. Furthermore, the spatial (depth) resolution has improved slightly (the peaks are sharper) by including angle dilation. With the depth harmonic models, however, we still see different dispersion across angles that may affect uncertainties. This may be caused by angle dependence in the depth of the singularities, an indication of errors in velocity model, perhaps associated with anisotropy in the lowermost mantle.

[47]   In Figure 12 we compare the best image estimate with the profile of paper 1 (also shown in Figure 2). Visually, the differences between results of GRT imaging with or without statistical analysis are fairly small. This is reassuring because it demonstrates that the structures are constrained by the data and not introduced by the statistical analysis. In detail, however, the effects of the statistical inference are visible (as marked). Statistical inference and validation lead to the suppression of rapid oscillations that are not required by the data, resulting in a smoother image. Furthermore, most of the seismic events in the image reveal more lateral continuity than in the original result presented in paper 1. As important as the visual effects, however, the statistical analysis presented here provides a means for estimating uncertainty, which will be key for subsequent interpretation of the structures that are visible in the image profiles. As an example, in Figure 13 we show the struc-

tures in Figure 12b at various levels of probability by muting structure that does not exceed the width of a particular confidence interval at that location. Figure 12b appears to be significant at 68% (1σ) confidence level, but as expected, only a few structures appear significant at 95% confidence. This example demonstrates how our analysis can be used not only to detect structure but also to identify and isolate the most robust features.
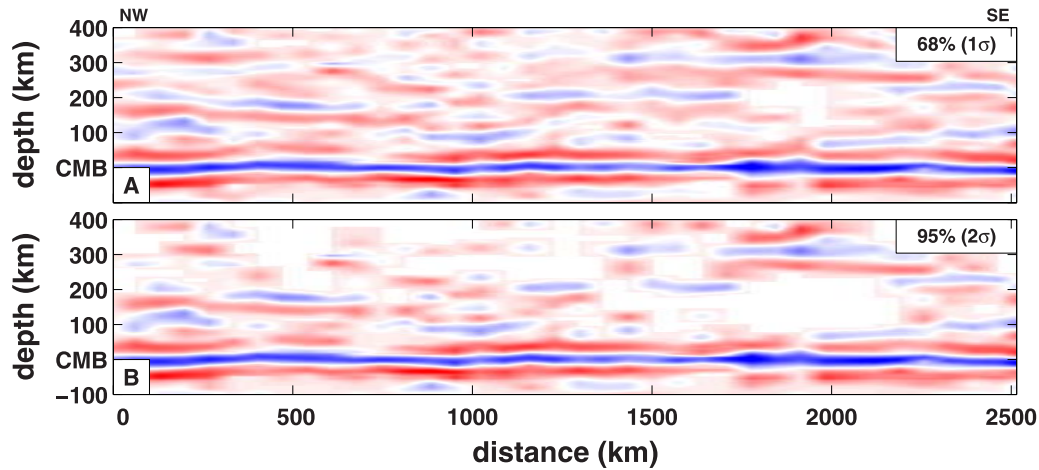
[48]   It should be noted that in the application introduced here mixed effects modeling is only used to estimate an optimal radial (reflectivity) profile at a particular image point at the CMB. In a similar vein, the lateral coherence in Figures 12 and 13 can be enhanced by applying the statistical models to the horizontal (space) distance parameter, but such image processing is not done here.

6.   Discussion and Concluding Remarks

[49]   The aim of the research described here and in paper 1 is to develop a novel approach to 3-D seismic imaging of the lowermost mantle using concepts from inverse scattering and modern statistics. In particular, we aim to exploit both the wide-angle reflections, which are used in most modeling studies because the large reflection coefficient produces arrivals that can be recognized in raw data, but which produce very poor depth resolution, as well as the narrow angle reflections, which are associated with weak reflections but which produce superior depth resolution. In fact, the recognition, and use, of this angle dependence of radial resolution, which is reflected in the angle dilation discussed in paper 1, allows high-resolution imaging and multiscale analysis of weak interfaces in Earth's deep interior.

[50]   The combined use of the generalized radon transform (GRT) and the mixed effect statistical inference presented here exploits the redundancy in the broadband data and allows the transformation of large volumes of global net-

Figure 11.   (top) A 2-D image profile resulting from processing with angle dilation depth harmonic model. A dual color scale is used to display structure above the CMB (color) at an amplification by a factor of 5 compared to the structure around the CMB proper (grey tone). Superimposed are the statistical estimates at six arbitrary positions. (bottom) Box plots used in the validation (see section 3.3) and the statistical estimates for six locations along the 2-D image profiles for (top) random intercept model; (middle) depth harmonic model, and (bottom) angle dilation depth harmonic model.

Figure 13. Structure in Figure 12b that is significant at the (a) 68% and (b) 95% confidence levels. This display is obtained by keeping only the features at each depth that are significantly different from zero at the chosen confidence level and muting the rest.

work data to statistical estimates and quantitative analysis of elastic singularities (such as discontinuities). Through the use of mixed effects models we can distinguish between and deal with true random noise in the data, random noise in the GRT images due to scatter that is not considered in the GRT theory, and artifacts between the assumed and real wave speeds of the medium. The examples discussed in section 4, and illustrated in Figures 6–9, demonstrate that this does, indeed, enhance the ability to extract weak signal from noisy data. In some cases the changes may seem subtle, but given the overall difficulty of, and interest in, imaging D″ structure, even small improvements are very welcome. Moreover, our statistical approach enables us to estimate uncertainties, in a Bayesian context, so that we can know with some confidence whether imaged structures are real or (statistically) insignificant. Also, in the future, it will allow a more rigorous analysis of the regularity (including such scaling parameters as roughness, sharpness, type of onset, and scale-dependent impedance contrast) of the transitions in elastic parameters detected here. This information will be gleaned from the wave number dependence of image gathers with scattering angle.

[51] In contrast to methods based on forward modeling, our method imposes few a priori assumptions about the geometry and nature of the structures that we attempt to image. Indeed, the structures revealed in the images are entirely controlled by the data themselves. The GRT only assumes that at a predefined image point the interfaces are contiguous, but in practice this is not a serious restriction. Furthermore, we assume to have a reasonable estimate of, or reference for, the smooth elastic properties of the medium. The latter assumption is not taken lightly; indeed, our statistical analysis quantifies the extent to which it is satisfied and, moreover, enables us to model and adjust for this type of noise in the image gathers. This is possible because of the prestack nature of our analysis: the statistical inference operates on image gathers at a large range of scatter angle, and not on the stack, or GRT image, itself.

[52] In the general context of mixed effect models, we have considered three particular cases: (1) a random intercept model, which assumes that the noise in the data, or in

the GRT images, is white, (2) a depth harmonic model, which identifies and removes spurious oscillations in radial direction, and (3) an angle dilation depth harmonic, which considers the dilation, and concomitant reduction in radial resolution, with increasing scatter angle that is inherent in the GRT imaging under consideration. With synthetic data we demonstrate that without a rigorous statistical approach, subtle wave speed contrast may be overlooked, or spurious ones introduced, with a traditional GRT.

[53] We found that the performance of the random intercept model is in many cases similar to that of the depth harmonic and angle dilation depth harmonic models. The good performance of this simple model is largely due to the data driven nature of the model fitting strategy. In particular, the use of a modified cross-validation procedure to select the tuning parameters of the model (see Appendix A for details) leads to robust optimal smoothing of the data that deviates significantly from the simple structure of the noise. On the other hand, the estimation procedure of the depth harmonic models (see Appendix A for details) could further reduce the noise level through integration of cross validation with estimation procedures similar to those studied by *Guo* [2002]; this is a topic of current research (P. Ma and W. Zhong, manuscript in preparation, 2007).

[54] We demonstrated the feasibility of the method with an application to the *ScS* wavefield that reflects off the CMB beneath Central America. The data are described in paper 1. Visual inspection suggests that the application of the mixed effects models leads to an image with better lateral definition of interfaces (in addition to information as to the statistical significance of each scatterer) but that the different mixed effect models used yield rather similar results. This may indicate that the level of harmonic noise in the real data is low or that the coherent noise is not truly harmonic. However, quantitative model validation suggests that the (angle dilation) depth harmonic models, which suppress spurious oscillations in the 2-D image profiles, produce slightly better data fits and uncertainty estimates.

[55] The image produced by ∼80,000 broadband *ScS* data (e.g., Figure 11) reveals strong contrasts in elastic parameters at about 0, that is, at the depth of the CMB and, locally,

near 150 and between 270 and 320 km above it. It is tempting to interpret the latter as the (fragments of the) "top" of the so-called D″ layer. Changes in elastic parameters near this depth have been the subject of many seismological studies [e.g., *Lay and Helmberger*, 1983; *Tromp and Dziewonski*, 1998; *Sidorin et al.*, 1999] (see *Garnero* [2000] for a comprehensive review), but there is as yet no consensus about this transition and its radial and lateral extent. Our results suggest that the structure may not be (laterally) continuous. The images also reveal significant scatter in between the presumed top of the D″ layer and the CMB proper. Collectively, our observations suggest that the D″ region is more complex than expected from models based on simple perovskite to postperovskite transitions. This implication will be explored elsewhere [e.g., *van der Hilst et al.*, 2006].

Appendix A: Estimating the Model Parameters

A1. Random Intercept Model

[56] In the standard formulation of penalized least squares regression, the minimization of (9) is performed in a reproducing kernel Hilbert space $\mathcal{H} \subseteq \{g : J(g) < \infty\}$ in which $J(g)$ is a square seminorm, and the solution resides in the space $\mathcal{N}_J \oplus \text{span} \{R_J(y_i, \cdot) : i = 1, \ldots, d\}$, where $\mathcal{N}_J = \{g : J(g) = 0\}$ is the null space of $J(g)$, and $R_J(\cdot, \cdot)$ is the so-called reproducing kernel in $\mathcal{H} \ominus \mathcal{N}_J$. The solution has an expression

$$g(y) = \sum_{\nu=1}^{k} \beta_\nu \phi_\nu(y) + \sum_{j=1}^{d} c_j R_J(y_j, y), \qquad (A1)$$

where $\{\phi_\nu\}_{\nu=1}^k$ is a basis of $\mathcal{N}_J$. For example, take a function g defined on [0, 1] with $J(g) = \int (g'')^2 dy$ and $\mathcal{N}_J = \{g : g(y) = \beta_1 + \beta_2 y\}$. In this case we get the popular cubic splines and the reproducing kernel is $R_J(x, y) = k_2(x)k_2(y) - k_4(x - y)$, where $k_\nu = B_\nu/\nu!$ are scaled Bernoulli polynomials. See *Wahba* [1990] and *Gu* [2002] for comprehensive treatments of the subject.

[57] Substituting (A1) into (9), one minimizes

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(G_{ij} - \phi(y_j)^t \beta - \varphi(z_j)^t \mathbf{c} - b_i \right)^2 + \frac{\sigma^2}{\sigma_s^2} \sum_{i=1}^{a} b_i^2 + n\lambda \mathbf{c}^t \mathbf{Q} \mathbf{c} \qquad (A2)$$

with respect to $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)^t$, $\mathbf{c} = (c_1, \ldots, c_d)^t$ and $\mathbf{b} = (b_1, \ldots, b_a)^t$, where $\varphi(y) = (R_J(y_1, y), \ldots, R_J(y_d, y))$, and $\mathbf{Q}$ is $d \times d$ with the (j, k)th entry $R_J(y_j, y_k)$. Estimates of $\boldsymbol{\beta}$, $\mathbf{c}$ and $\mathbf{b}$ are obtained by setting to zero the derivatives of (A5) with respect to $\mathbf{c}$ and $\mathbf{b}$. The minimizers of (A5) are solutions to the normal equation,

$$\begin{pmatrix} \mathbf{S}^t\mathbf{S} & \mathbf{S}^t\mathbf{R} & \mathbf{S}^t\mathbf{M} \\ \mathbf{R}^t\mathbf{R} & \mathbf{R}^t\mathbf{R} + (n\lambda)\mathbf{Q} & \mathbf{R}^t\mathbf{M} \\ \mathbf{M}^t\mathbf{S} & \mathbf{M}^t\mathbf{R} & \mathbf{M}^t\mathbf{M} + \tau\mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{c} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{S}^t\mathbf{G} \\ \mathbf{R}^t\mathbf{G} \\ \mathbf{M}^t\mathbf{G} \end{pmatrix}, \qquad (A3)$$

where $\mathbf{G} = (G_{11}, \ldots, G_{ab})^t$, $\mathbf{S}$ is $n \times k$ with the (i, ν)th entry $\phi_\nu(y_i)$, $\mathbf{R}$ is $n \times n$ with the (i, j)th entry $\varphi_j(y_i)$, $\mathbf{M}$ is $n \times a$

block diagonal identity matrix, $\mathbf{Q}$ is $n \times n$ with the (j, k)th entry $J(\varphi_j, \varphi_k) = R_J(v_j, v_k)$, and $\tau = \sigma^2/\sigma_s^2$ and $\mathbf{I}$ is identity matrix. The normal equation of (A3) can be solved by a Cholesky decomposition followed by backward and forward substitutions. Possible singularity of the matrix can be properly handled through pivoting in Cholesky decomposition; see, e.g., *Golub and Van Loan* [1989] and *Kim and Gu* [2004] for details.

[58] The fitted values $\widehat{\mathbf{G}} = \mathbf{S}\,\widehat{\boldsymbol{\beta}} + \mathbf{R}\,\widehat{\mathbf{c}} + \mathbf{M}\,\widehat{\mathbf{b}}$ can be written as $\widehat{\mathbf{G}} = \mathbf{U}(\lambda, \tau)\mathbf{G}$, where

$$\mathbf{U}(\lambda, \tau) = (\mathbf{S}, \mathbf{R}, \mathbf{M})$$
$$\cdot \begin{pmatrix} \mathbf{S}^t\mathbf{S} & \mathbf{S}^t\mathbf{R} & \mathbf{S}^t\mathbf{M} \\ \mathbf{R}^t\mathbf{R} & \mathbf{R}^t\mathbf{R} + (n\lambda)\mathbf{Q} & \mathbf{R}^t\mathbf{M} \\ \mathbf{M}^t\mathbf{S} & \mathbf{M}^t\mathbf{R} & \mathbf{M}^t\mathbf{M} + \tau\mathbf{I} \end{pmatrix}^{+} \begin{pmatrix} \mathbf{S}^t \\ \mathbf{R}^t \\ \mathbf{M}^t \end{pmatrix},$$

and $\mathbf{C}^+$ denotes a generalized inverse of $\mathbf{C}$ satisfying $\mathbf{CC}^+\mathbf{C} = \mathbf{C}$, $\mathbf{C}^+\mathbf{CC}^+ = \mathbf{C}^+$, $(\mathbf{CC}^+)^t = \mathbf{CC}^+$, and $(\mathbf{C}^+\mathbf{C})^t = \mathbf{C}^+\mathbf{C}$. $\mathbf{C}^+\mathbf{C}$. This inverse is also known as Moore-Penrose inverse.

[59] For different values of λ and τ, (A3) defines a family of possible solutions. Optimal values of these parameters are obtained by minimizing the generalized cross-validation score

$$V(\lambda, \tau) = \frac{n^{-1}\mathbf{G}^t(\mathbf{I} - \mathbf{U}(\lambda, \tau))^2\mathbf{G}}{[n^{-1}\,\text{tr}(\mathbf{I} - \mathbf{U}(\lambda, \tau))]^2}. \qquad (A4)$$

Gu and Ma [2005] showed that under very general conditions the minimizers of $V(\lambda, \tau)$ yield an optimal smoothing asymptotically.

A2. Depth Harmonic Model

[60] To correct for harmonic components of the noise, we proceed as follows. For a fixed angle, any profile $g(\theta, y)$ is a function of depth y that we will just denote by $g(y)$. These functions may be contaminated by coherent noise caused my unmodeled oscillations in the subsurface. One possible model for this oscillations is a harmonic process $\sum_i a_i \cos(\omega_i y + \phi)$, where a_i and ϕ_i are random amplitudes and phases, respectively, and the frequencies ω_i are fixed but unknown. Each realization of the process is just a sum of sinusoids and each profile is contaminated by a different harmonic process. Hence removal of a harmonic process from a profile is equivalent to correcting for sinusoid signals.

[61] To find sinusoids hidden in the data, we use a method developed by *Quinn and Fernandes* [1991]. To explain the main idea, we assume a single sinusoid. A profile is modeled as $\tilde{g}(y) = a \cos(\omega y + \phi) + g(y)$, where $g(y)$ is the clean signal we want to recover. A sinusoid can be annihilated with a second-order filter. Indeed, we note that $\tilde{g}(y)$ satisfies the difference equation

$$\tilde{g}(y_j) - 2\cos(\omega)\tilde{g}(y_{j-1}) + \tilde{g}(y_{j-2})$$
$$= g(y_j) - 2\cos(\omega)g(y_{j-1}) + g(y_{j-2}).$$

ARMA fitting techniques can then be used to estimate the frequency ω and then obtain amplitude and phase through least squares. The frequency estimate can be interpreted as a local maximizer of a smoothed periodogram. The case of

more than one frequency is treated in a similar way; there are difference operators that annihilate all the sinusoids. For more details on this methodology, see *Quinn and Fernandes* [1991].

[62] Once the harmonic components are estimated from data, we can obtain the harmonic component extracted profile: $\tilde{G}_{ij} = G_{ij} - \sum_k \hat{a}_{ik} \cos(\hat{\omega}_{ik} y_j + \hat{\phi}_{ik})$. Then the model can be fitted as before; that is, one minimizes

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(\tilde{G}_{ij} - \phi(y_j)^t \boldsymbol{\beta} - \boldsymbol{\varphi}(z_j)^t \mathbf{c} - b_i \right)^2 + \frac{\sigma^2}{\sigma_s^2} \sum_{i=1}^{a} b_i^2 + n\lambda \mathbf{c}^t \mathbf{Q} \mathbf{c}. \tag{A5}$$

A3. Angle Dilation Depth Harmonic Model

[63] As in the depth harmonic model, the harmonic components are estimated from the data and then subtracted from to obtain the corrected profile $\tilde{G}_{ij}$. We estimate (10) using penalized least squares

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(\tilde{G}_{ij} - g(\alpha_i y_j) - b_i \right)^2 + \frac{\sigma^2}{\sigma_s^2} \sum_{i=1}^{a} b_i^2 + n\lambda J(g). \tag{A6}$$

[64] The minimization is carried out iteratively as follows:

[65] In step 0, as the initial estimate we use the maximum likelihood estimates of the linear mixed effects model $\tilde{G}_{ij} = \alpha_i y_i + b_i + \varepsilon_{ij}$.

[66] In step 1, for estimated $\hat{\alpha}_i$, we minimize the following functional to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{c}}$:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(\tilde{G}_{ij} - \phi(\hat{\alpha}_i y_j)^t \boldsymbol{\beta} - \boldsymbol{\varphi}(\hat{\alpha}_i z_j)^t \mathbf{c} \right)^2 + \frac{\sigma^2}{\sigma_s^2} \sum_{i=1}^{a} b_i^2 + n\lambda \mathbf{c}^t \mathbf{Q} \mathbf{c}. \tag{A7}$$

[67] In step 2, for estimated $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{c}}$, we estimate α_i by minimizing

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \left(\tilde{G}_{ij} - \phi(\alpha_i y_j)^t \hat{\boldsymbol{\beta}} - \boldsymbol{\varphi}(\alpha_i z_j)^t \hat{\mathbf{c}} \right)^2. \tag{A8}$$

Steps 1 and 2 are iterated until convergence.

References

Garnero, E. J. (2000), Heterogeneity of the lowermost mantle, *Annu. Rev. Earth Planet. Sci.*, *28*, 509–537.

Golub, G., and C. Van Loan (1989), *Matrix Computations*, 2nd ed., Johns Hopkins Univ. Press, Baltimore, Md.

Gu, C. (2002), *Smoothing Spline ANOVA Models*, Springer, New York.

Gu, C., and P. Ma (2005), Optimal smoothing in nonparametric mixed-effect models, *Ann. Stat.*, *33*, 1357–1379.

Guo, W. (2002), Functional mixed effects models, *Biometrics*, *58*(1), 121–128.

Helmberger, D., T. Lay, S. Ni, and M. Gurnis (2005), Deep mantle structure and the postperovskite phase transition, *Proc. Natl. Acad. Sci. U.S.A.*, *102*, 17,257–17,263.

Kim, Y.-J., and C. Gu (2004), Smoothing spline Gaussian regression: More scalable computation via efficient approximation, *J. R. Stat. Soc., Ser. B*, *66*, 337–356.

Lay, T., and D. V. Helmberger (1983), A shear velocity discontinuity in the lower mantle, *Geophys. Res. Lett.*, *10*, 63–66.

Ma, P. (2003), Nonparametric mixed-effects models, Ph.D. thesis, Purdue Univ., West Lafayette, Indiana.

Murakami, M., K. Hirose, K. Kawamura, N. Sata, and Y. Ohishi (2004), Post-perovskite phase transition in MgSiO₃, *Science*, *304*, 855–858, doi:10.1126/science.1095932.

Nychka, D. (1988), Bayesian confidence intervals for smoothing splines, *J. Am. Stat. Assoc.*, *83*, 1134–1143.

Oganov, A., and S. Ono (2004), Theoretical and experimental evidence for a post-perovskite phase of MgSiO₃ in Earth's D″ layer, *Nature*, *430*, 445–448.

Pinheiro, J. C., and D. M. Bates (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York.

Quinn, B., and J. Fernandes (1991), A fast efficient technique for the estimation of frequency, *Biometrika*, *78*, 489–497.

Robinson, G. K. (1991), That BLUP is a good thing: The estimation of the random effects (with discussions), *Stat. Sci.*, *6*, 15–51.

Schimmel, M., and H. Paulssen (1997), Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.*, *130*, 497–505.

Shim, S. (2005), Stability of MgSiO₃ perovskite in the lower mantle, in *Earth's Deep Mantle: Structure, Composition, and Evolution*, Geophys. Monogr. Ser., vol. 160, edited by R. D. van der Hilst et al., pp. 261–282, AGU, Washington, D. C.

Shim, S.-H., T. S. Duffy, R. Jeanloz, and G. Shen (2004), Stability and crystal structure of MgSiO₃ perovskite to the core-mantle boundary, *Geophys. Res. Lett.*, *31*, L10603, doi:10.1029/2004GL019639.

Sidorin, I., M. Gurnis, and D. V. Helmberger (1999), Evidence for a ubiquitous seismic discontinuity at the base of the mantle, *Science*, *286*, 1326–1331.

Stolk, C. C., and M. V. De Hoop (2002), Microlocal analysis of seismic inverse scattering in anisotropic, elastic media, *Commun. Pure Appl. Math.*, *55*, 261–301.

Tromp, J., and A. M. Dziewonski (1998), Geoscience—Two views of the deep mantle, *Science*, *281*, 655–656.

van der Hilst, R. D., M. De Hoop, P. Wang, S.-H. Shim, P. Ma, and L. Tenorio (2006), Seismo-stratigraphy and thermal structure of Earth's core-mantle boundary region, *Science*, *315*, 1813–1817, doi:10.1126/science.1137867.

Vonesh, E. F., and V. M. Chinchilli (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.

Wahba, G. (1983), Bayesian "confidence intervals" for the cross-validated smoothing spline, *J. R. Stat. Soc., Ser. B*, *45*, 133–150.

Wahba, G. (1990), *Spline Models for Observational Data*, CBMS-NSF Reg. Conf. Ser. Appl. Math., vol. 59, SIAM, Philadelphia, Pa.

Wang, P., M. V. de Hoop, R. D. van der Hilst, P. Ma, and L. Tenorio (2006), Imaging of structure at and near the core mantle boundary using a generalized radon transform: 1. Construction of image gathers, *J. Geophys. Res.*, *111*, B12304, doi:10.1029/2005JB004241.

Wang, Y. (1998), Mixed-effects smoothing spline ANOVA, *J. R. Stat. Soc., Ser. B*, *60*, 159–174.

————————————

M. V. de Hoop, Center for Computational and Applied Mathematics, Purdue University, 150 N University Street, West Lafayette, IN 47907-2067, USA. (mdehoop@purdue.edu)

P. Ma, Department of Statistics, University of Illinois, 725 South Wright Street, Champaign, IL 61820, USA. (pingma@uiuc.edu)

L. Tenorio, Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA. (ltenorio@mines.edu)

R. D. van der Hilst and P. Wang, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 54-522, 77 Massachusetts Ave., Cambridge, MA 02139-4307, USA. (wangp@quake.mit.edu; hilst@mit.edu)