Methods for inversion of body-wave waveforms for localized three-dimensional seismic structure and an application to D" structure beneath Central America

Kenji Kawai, Kensuke Konishi, Robert J. Geller, Nobuaki Fuji

Geophysical Journal International

Geophys. J. Int. (2014) **197,** 495–524
Advance Access publication 2014 February 4

doi: 10.1093/gji/ggt520

Methods for inversion of body-wave waveforms for localized three-dimensional seismic structure and an application to D″ structure beneath Central America

Kenji Kawai,[1,2] Kensuke Konishi,[3] Robert J. Geller[4] and Nobuaki Fuji[5]

[1] *Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Ookayama* 2-12-1, *Meguro-ku, Tokyo,* 152-8551, *Japan.*
E-mail: kenji@geo.titech.ac.jp
[2] *Earth-Life Science Institute, Tokyo Institute of Technology, Ookayama* 2-12-1, *Meguro, Tokyo* 152-8551, *Japan*
[3] *School of Earth and Environmental Sciences, Seoul National University, Gwanak, Seoul* 151-742, *Republic of Korea*
[4] *Department of Earth and Planetary Science, Graduate School of Science, University of Tokyo, Hongo* 7-3-1, *Bunkyo, Tokyo,* 113-0033, *Japan*
[5] *Institut de Physique du Globe de Paris,* 1 *rue Jussieu F*-75328 *Paris Cedex* 05, *France*

Accepted 2013 December 23. Received 2013 December 19; in original form 2013 November 4

SUMMARY

We formulate the inverse problem of waveform inversion for localized 3-D seismic structure, computing partial derivatives of waveforms with respect to the elastic moduli at arbitrary points in space for anisotropic and anelastic media. In this study we minimize computational requirements by using the Born approximation with respect to a laterally homogeneous model, but this is not an inherent limitation of our approach. We solve the inverse problem using the conjugate gradient (CG) method, using Akaike's Information Criterion (AIC) to truncate the CG expansion. We apply our method to invert for 3-D shear wave structure in the lowermost mantle beneath Central America using a total of 2154 waveforms at periods from 12.5 to 200 s recorded at stations near the Pacific coast of North America for 29 deep and intermediate-depth events beneath South America. The resulting model shows lateral heterogeneity in the E–W direction, which may be associated with a subducted cold slab surrounded by hotter materials with slower velocities. Various tests show that our model is robust.

Key words: Mantle processes; Body waves; Seismic tomography.

1 INTRODUCTION

Forward modelling studies have made many important contributions to studies of the Earth's interior. However, as the volume of data increases, it will become increasingly difficult for analysts, however skilled, to determine which model provides the best overall fit to the data. Since many more seismic observatories such as arrays are now being deployed, it is desirable to be able to conduct objective and quantitative inversion of a large amount of seismic waveform data for detailed seismic structure in the Earth.

In order to conduct waveform inversion, computation of the partial derivatives is required. Tarantola (1984) conducted pioneering work to develop algorithms for efficient computation of steepest descent vectors for acoustic wavefields for application to waveform inversion. Geller & Hara (1993) extended these results to derive efficient algorithms for computation of partial derivatives in both matrix and wavefield formulations. While the former is suitable for inversion for global structure (e.g. Hara *et al.* 1993), the latter is appropriate for inversion for local structure (e.g. Kawai & Geller 2010a; Fuji *et al.* 2012). Many other studies have also presented theoretical formulations of waveform inversion and/or inversion results for synthetic or model data, including Pratt (1999), Sirgue & Pratt (2004) and Tromp *et al.* (2005).

Several groups have conducted waveform inversion of observed data, sometimes called full-waveform inversion. Chen *et al.* (2007) inverted phase-delay measurements of *P* and *S* body waves to determine the 3-D structure of the Los Angeles region, using the finite-difference method to compute synthetic seismograms. Tape *et al.* (2009, 2010) inverted for crustal structure beneath southern California using the spectral element method (SEM) to compute synthetic seismograms. They used waveform data to obtain traveltime or phase data for finite frequency kernels (Dahlen & Tromp 1998). Waveform inversion was conducted using a data set consisting primarily of surface waves by Fichtner *et al.* (2008, 2009, 2010), who used the SEM for computation of synthetics. They weighted phase data and amplitude data and created penalty functions to investigate not only laterally heterogeneous but also anisotropic structure in the upper mantle. Except for studies on shallow structure (such as the crust and upper mantle), it is currently difficult to conduct waveform inversion following the above-mentioned approaches, because SEM is computationally costly.

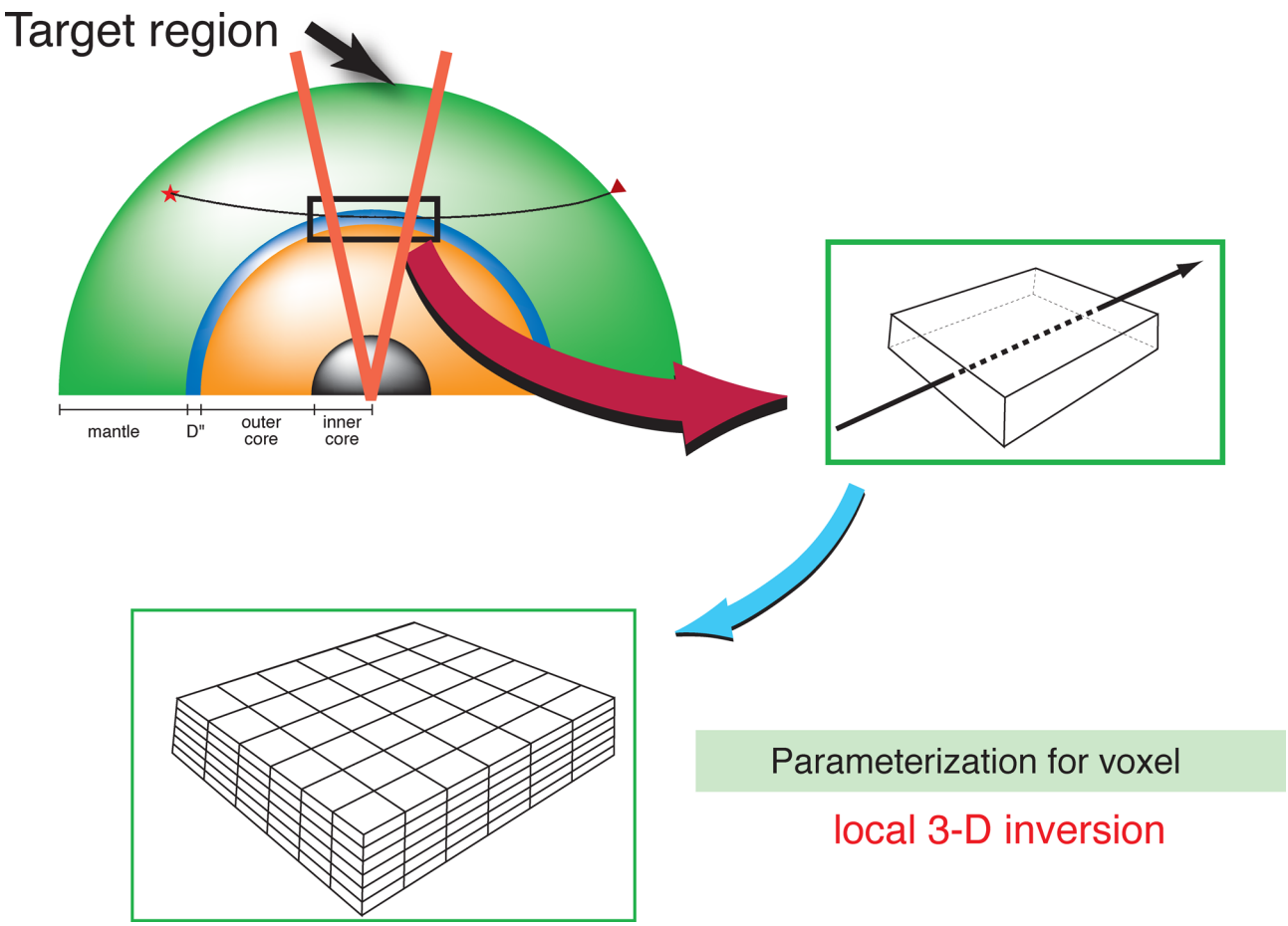

Figure 1. Schematic explanation of waveform inversion for localized structure in the Earth's deep interior. We select target regions which are well sampled by waveforms excited in regions where intermediate or deep focus earthquakes (to minimize complications in the waveforms due to surface reflections) occur frequently and are observed by dense receiver networks such as arrays; we then invert for 3-D local structure of the target region parametrized as voxel structure. In order to conduct waveform inversion for localized structure, it is necessary to be able to make static corrections stably and to efficiently compute partial derivatives for voxel perturbations.

Takeuchi (2012) conducted full-waveform inversion for the shear velocity structure of the whole mantle expanding the laterally heterogeneous structure in spherical harmonics up to angular order 18, using the direct solution method (DSM; Geller & Ohminato 1994; Kawai *et al.* 2006) to compute synthetics. As for exploration seismology, it has long been suggested that waveform inversion will eventually become one of the most important tools. Some basic algorithms have already been developed (e.g. Tarantola 1984, 1986; Pratt 1999), and techniques for waveform inversion are being widely discussed in exploration seismology (see Virieux & Operto 2009, for a review). However, published applications to real data (e.g. Sirgue *et al.* 2010; Prieux *et al.* 2011) are still relatively rare and are generally limited to frequency domain applications for acoustic-wave approximations.

Our group has conducted waveform inversion for seismic structure in the deep Earth's interior using the DSM to compute synthetics and their partials. Kawai *et al.* (2007) determined the fine structure of depth dependence of shear wave velocity structure within D″, whereas traveltime tomography can resolve only average structure over a depth range of three to four hundred kilometres. Fuji *et al.* (2010) directly inverted for 1-D localized anelastic structure in the mantle transition zone, whereas previous studies investigated anelastic structure using intermediate values such as t^*, an integrated value which is frequency-dependent. Kawai & Geller (2010b) showed that resolution of the velocity of *SV* shear waves very close to the CMB is inherently limited due to the boundary condition of zero tangential traction at the CMB and that no method, including shear wave splitting studies, can resolve anisotropy immediately above the CMB. They also showed that the elastic constants in D″, including anisotropy, can be determined straightforwardly by waveform inversion for depths greater than one or two wavelengths above the CMB. In this study we extend our methods to invert for 3-D shear velocity structure in D″ beneath Central America.

2 METHODS FOR WAVEFORM INVERSION FOR 3-D LOCALIZED STRUCTURE

In this section we summarize our methods (Fig. 1) for inversion of waveforms for localized 3-D structure. We discuss the applicability, robustness and resolution of these methods in Section 3.

2.1 Synthetics

In this study we use the DSM to compute full-wave synthetic seismograms for a 1-D transversely isotropic model. We then compute the partials for 3-D structure using the Born approximation. The

DSM obtains the solution of the weak form of the equation of motion by directly solving the Galerkin weak form of the equation of motion:

$$(\omega^2\mathbf{T} - \mathbf{H} + \omega\mathbf{R})\boldsymbol{c} = -\boldsymbol{g}, \tag{1}$$

where $\boldsymbol{c}$ is the vector of coefficients of the trial functions, $\mathbf{T}$ is the mass matrix, $\mathbf{H}$ is the stiffness matrix, $\boldsymbol{g}$ is the force vector and $\mathbf{R}$ enforces continuity conditions at fluid–solid boundaries, which exist only for the spheroidal case. We do not use geometrical optics or Earth-flattening approximations in computing the synthetics and their partial derivatives. The publicly available DSM software efficiently computes highly accurate synthetic seismograms in a spherically symmetric transversely isotropic earth model (Kawai *et al.* 2006). Although the publicly released software is only for 1-D media at present, this approach can also be applied to 3-D problems (e.g. Cummins *et al.* 1997). In this study, although only toroidal–toroidal coupling is taken into account, we compute synthetics for both the toroidal and spheroidal wavefield to remove the effects of non-physical waves (see Kawai *et al.* 2006, for details).

2.2 Partials

We use the methods of Geller & Hara (1993), which are an extension of the methods of Tarantola (1984), to compute the partial derivatives. Geller & Hara (1993) give both matrix and wavefield formulations of the partials. While the former are more suitable for investigation of global scale 3-D structure, the latter are more suitable for inversion for localized 3-D structure. The latter approach is therefore adopted in this study.

We express the i-component of the displacement for the k-th earthquake at the point $\boldsymbol{r}$ as $u_i^{(k)}(\boldsymbol{r})$, and the i-component of the backpropagated displacement at $\boldsymbol{r}$ excited by a point force in the j-direction at the p-th station as $\eta_{ij}^{(p)}(\boldsymbol{r})$. Note that $\boldsymbol{r}$ will in general be a point in the Earth's interior (the 'target' of the inversion).

The partial derivatives in the wavefield formulation are expressed as follows:

$$\left\{\frac{\partial u_i^{(k)}\left[\boldsymbol{r}^{(p)}\right]}{\partial m_l}\right\}^* = \int_V (\omega^*)^2 \left[u_j^{(k)}\right]^* \left[\rho^{(l)}\right]^* \eta_{ji}^{(p)} \mathrm{d}V - \int_V \left[u_{j,q}^{(k)}\right]^* \left[C_{jqrs}^{(l)}\right]^* \eta_{ri,s}^{(p)} \mathrm{d}V\,, \tag{2}$$

where $\rho^{(l)} = \delta\rho\delta(\boldsymbol{r} - \boldsymbol{r}_0)$ or $C_{ijpq}^{(l)} = \delta C_{ijpq}\delta(\boldsymbol{r} - \boldsymbol{r}_0)$ and $u_{i,j} = \frac{\partial u_i}{\partial x_j}$ is a locally Cartesian derivative (see Geller & Hara 1993, for details). Since the transverse component of waveforms is used in this study, we take into account toroidal–toroidal coupling and neglect other coupling.

2.2.1 3-D synthetics based on the Born approximation

We compute 3-D synthetics for voxel perturbations including anisotropy and anelasticity using the first-order Born approximation as follows:

$$\boldsymbol{u}_{3\mathrm{D}}(\boldsymbol{r}) = \boldsymbol{u}_{1\mathrm{D}}(\boldsymbol{r}) + \sum_l \frac{\partial \boldsymbol{u}(\boldsymbol{r})}{\partial m_l}\delta m_l, \tag{3}$$

where $\boldsymbol{u}_{3\mathrm{D}}$ is a synthetic seismogram for the 3-D earth model and $\boldsymbol{u}_{1\mathrm{D}}$ is a synthetic for a 1-D earth model such as PREM (Dziewonski & Anderson 1981). We have developed efficient software for computation of synthetics for a 3-D earth model using the above formulation.

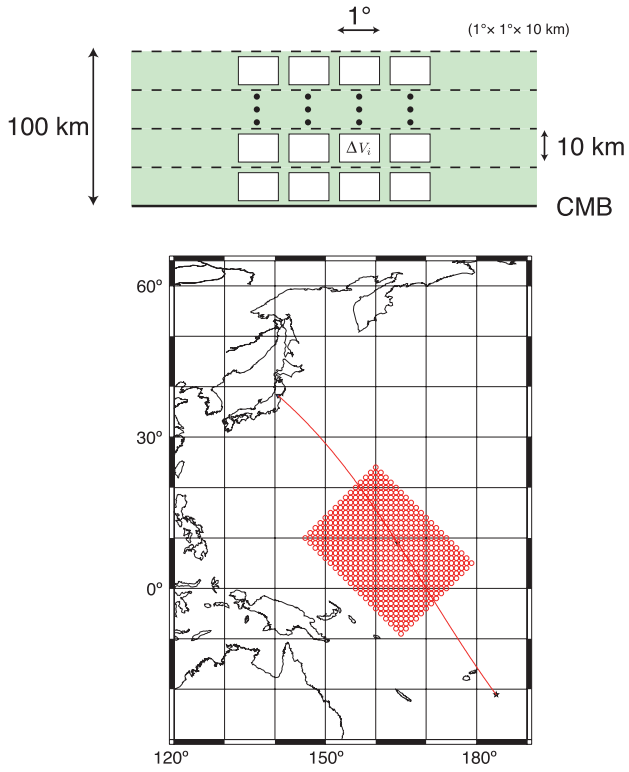

Figure 2. We compare synthetics computed for a finite perturbation to those computed using the partial derivatives to confirm the absence of bugs. The green layer, which has a thickness of 100 km, is perturbed finitely and spherically symmetrically. Partial derivatives for each white block $1^\circ \times 1^\circ \times 10$ km are computed. The sum of the voxel perturbations is compared to the spherically symmetric perturbation to confirm the reasonableness of the voxel perturbations.

In order to validate these methods, we compare finite perturbation waveforms to the partial derivatives. We simulate a locally laterally homogeneous perturbation to a 1-D model by uniformly perturbing a large number of adjacent voxels, as shown in Fig. 2. We compare the difference of the synthetics for the initial and final 1-D models to the perturbations computed by summing the Born perturbations for all of the voxels for several different finite perturbations. Since the 1- and 3-D models are not directly comparable this is not a rigorous test of the accuracy of the partials, but rather just a check to confirm the absence of a serious bug. In all cases the results were reasonable.

2.3 Static corrections

The observed waveforms sample both structure inside and outside the target region. The effects of the latter must be accounted for. We handle this by making static corrections (time-shifts) to the observed waveforms. We have developed three methods for static corrections: the direct inversion method (Kawai & Geller 2010a), the autopick method (time-shift before inversion) (Fuji *et al.* 2010) and the hybrid method (Konishi *et al.* 2012). While the second is suitable in cases where the reference phase and the phase of primary interest sample almost the same path except for the target region, the first is suitable in cases where a reference phase cannot be used because both phases sample the target regions. The third is suitable for both cases. Konishi *et al.* (2012) discuss the relation of these three methods. This study adopts the autopick method, since the

direct *S* phase can be used as a reference phase for the epicentral distances used in this study when the target region is D″.

2.4 Data processing

The synthetics and partial derivatives must both be processed in the same way as the observed data. We design a recursive digital Butterworth bandpass filter (e.g. Shanks 1967), following Saito (1978). As this paper is written in Japanese, we summarize this method in Appendix B.

2.5 Inverse problem formulation

In our various studies of 1-D structure, we have solved the waveform inversion inverse problem using DLS (damped least squares), SVD (singular value decomposition; Lanczos 1961; Wiggins 1972; Lawson & Hanson 1974; Menke 1984) and CG (conjugate gradient; Beckman 1960; Reeves 1964). Inversion using SVD and CG was discussed by Fuji *et al.* (2010) and Fuji (2010), respectively; we briefly summarize the main results in Appendix A.

2.6 Variance

When we use a relatively small amount of waveform data, we can directly compute $\mathbf{A}\delta\boldsymbol{m}$. However, note that we can easily compute the total variance $|\delta\boldsymbol{d} - \mathbf{A}\delta\boldsymbol{m}|^2$, as follows:

$$\begin{aligned}|\delta\boldsymbol{d} - \mathbf{A}\delta\boldsymbol{m}|^2 &= (\delta\boldsymbol{d} - \mathbf{A}\delta\boldsymbol{m})^T(\delta\boldsymbol{d} - \mathbf{A}\delta\boldsymbol{m})\\ &= (\delta\boldsymbol{d}^T - (\mathbf{A}\delta\boldsymbol{m})^T)(\delta\boldsymbol{d} - \mathbf{A}\delta\boldsymbol{m})\\ &= \delta\boldsymbol{d}^T\delta\boldsymbol{d} - \delta\boldsymbol{d}^T\mathbf{A}\delta\boldsymbol{m} - (\mathbf{A}\delta\boldsymbol{m})^T\delta\boldsymbol{d} + (\mathbf{A}\delta\boldsymbol{m})^T\mathbf{A}\delta\boldsymbol{m}\\ &= \delta\boldsymbol{d}^T\delta\boldsymbol{d} - (\mathbf{A}^T\delta\boldsymbol{d})^T\delta\boldsymbol{m} - \delta\boldsymbol{m}^T(\mathbf{A}^T\delta\boldsymbol{d})\\ &\quad + \delta\boldsymbol{m}^T(\mathbf{A}^T\mathbf{A})\delta\boldsymbol{m}. \qquad (4)\end{aligned}$$

Since $\mathbf{A}^T\delta\boldsymbol{d}$ and $\mathbf{A}^T\mathbf{A}$ are $M \times 1$ and $M \times M$ matrices, respectively, the above computation will always be feasible.

2.7 Model resolution matrix

The model resolution matrix, $\mathbf{R}$, is defined as follows by Menke (1984, p. 64):

$$\boldsymbol{m}^{\text{est}} = \mathbf{R}\boldsymbol{m}^{\text{true}}, \qquad (5)$$

where $\boldsymbol{m}^{\text{true}}$ and $\boldsymbol{m}^{\text{est}}$ are the solution of the inverse problem using a complete basis and a particular estimate of the model parameters obtained by an inversion using a truncated basis, respectively. The solution to the inverse problem

$$\mathbf{G}\boldsymbol{m}^{\text{true}} = \boldsymbol{d}^{\text{obs}} \qquad (6)$$

is

$$\boldsymbol{m}^{\text{est}} = \mathbf{G}^{-g}\boldsymbol{d}^{\text{obs}}, \qquad (7)$$

where $\mathbf{G}^{-g}$ is the generalized inverse matrix. The model resolution matrix is therefore written as:

$$\mathbf{R} = \mathbf{G}^{-g}\mathbf{G}. \qquad (8)$$

2.7.1 Model resolution matrix for SVD

In our linearized inversions $\mathbf{G} = \mathbf{A}$, and the inverse problem is solved using the first L eigenvectors (see Section A2), where we use Akaike's Information Criterion (AIC) to truncate the SVD expansion. We therefore have

$$\begin{aligned}\delta\boldsymbol{m}^{\text{est}} &= \left[\sum_{i=1}^{L}\frac{\boldsymbol{v}_i\boldsymbol{v}_i^T}{\lambda_i^2}\right]\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d}^{\text{obs}}\\ &= \left[\sum_{i=1}^{L}\frac{\boldsymbol{v}_i\boldsymbol{v}_i^T}{\lambda_i^2}\right]\mathbf{A}^{\mathbf{T}}\mathbf{A}\delta\boldsymbol{m}^{\text{true}}\\ &= \left[\sum_{i=1}^{L}\frac{\boldsymbol{v}_i\boldsymbol{v}_i^T}{\lambda_i^2}\right]\left[\sum_{j=1}^{M}\boldsymbol{v}_j\boldsymbol{v}_j^T\lambda_j^2\right]\delta\boldsymbol{m}^{\text{true}}\\ &= \mathbf{V}_L\mathbf{V}_L^{\mathbf{T}}\delta\boldsymbol{m}^{\text{true}}, \qquad (9)\end{aligned}$$

where we used the orthogonality of the eigenvectors in going from the next to last line of eq. (9) to the last line. By comparing eq. (9) to eq. (5) we can see that

$$\mathbf{R} = \mathbf{V}_L\mathbf{V}_L^{\mathbf{T}}. \qquad (10)$$

2.7.2 Model resolution matrix for CG

For the CG expansion using eqs (A24) and (A28), we have

$$\begin{aligned}\delta\boldsymbol{m}^{\text{est}} &= \mathbf{P}\mathbf{L}^{-1}\mathbf{P}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d}^{\text{obs}}\\ &= \mathbf{P}\mathbf{L}^{-1}\mathbf{P}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\delta\boldsymbol{m}^{\text{true}}. \qquad (11)\end{aligned}$$

By comparing eq. (11) to eq. (5) we can see that the model resolution matrix for the CG solution is

$$\mathbf{R} = \mathbf{P}\mathbf{L}^{-1}\mathbf{P}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}. \qquad (12)$$

3 APPLICATION TO OBSERVED DATA

3.1 Data

We choose the D″ layer beneath Central America (Fig. 3) as the target region for the 3-D inversion in this paper. The shear wave velocity structure here has been suggested to be laterally heterogeneous by previous studies based on ray theoretical approximations (e.g. Grand 2002; Hung *et al.* 2005; Hutko *et al.* 2006).

We carry out our inversion in two stages. In the first stage we use 412 records from the IRIS/USGS, SCSN, PNSN, BDSN and CNSN networks for the events listed in Table 1; ray paths are shown in Fig. 3. We use this data set to conduct a preliminary inversion and we then perform a series of tests to validate the robustness of the inversion results. In Section 4, we present results for the second-stage inversion, in which we use a much larger data set, consisting of the 412 records used in the first-stage inversion and an additional 1742 records from the US-Array for the 14 events shown in Table 2, for a total of 2154 records.

We use the transverse components of broad-band waveform data obtained by rotating the N–S and E–W components (Table 1 and Fig. 3). We deconvolve the instrument response and apply a band-pass filter to the data and construct data sets for the passband 0.005 to 0.08 Hz (i.e. for the period range, 12.5–200 s). We then select records which include *S*, *ScS* and the other phases that arrive between them.

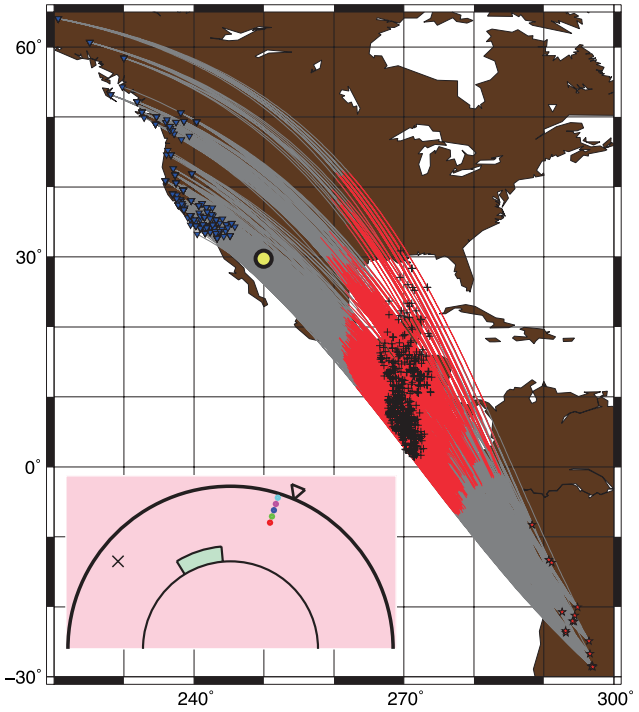

Figure 3. Event-receiver geometry for the first-stage data, with great circle ray paths. The portions of the great circles which sample D″ are shown in red, and plus signs indicate the turning points within D″. Blue reversed triangles show the sites of North American stations (data from IRIS/USGS, SCSN, PNSN, BDSN and CNSN) used in our study. Red stars show the 15 intermediate and deep earthquakes studied (Table 1). The yellow dot is the location for the target for the test of the effect of shallow structure shown in Fig. 18. Inset at lower left-hand side shows a schematic cross-section for the calculations in Fig. 18. The triangle shows a station and the cross shows a source. Coloured circles show the locations of the perturbations to the shallow structure and the green rectangle shows the target region in D″.

Table 1. Earthquakes for which we used IRIS/USGS, SCSN, PNSN, BDSN and CNSN data.

Event #	Date (Y/M/D)	Latitude	Longitude	Depth	M_w
1	1993/5/24	−23.45	−66.88	231.9	7.0
2	1993/10/19	−22.12	−65.69	278.9	6.0
3	1994/1/10	−13.28	−69.27	603.6	6.9
4	1994/4/29	−28.51	−63.22	565.9	6.9
5	1994/5/10	−28.62	−63.02	603.0	6.9
6	1994/8/19	−26.72	−63.42	562.6	6.4
7	1997/1/23	−22.04	−65.92	281.6	7.1
8	1997/11/28	−13.70	−68.90	600.5	6.6
9	1999/9/15	−20.73	−67.37	217.5	6.4
10	2000/4/23	−28.41	−63.04	607.9	6.9
11	2000/5/12	−23.72	−66.85	226.6	7.2
12	2002/10/12	−8.30	−71.66	539.4	6.9
13	2003/7/27	−20.05	−65.19	350.6	6.0
14	2004/3/17	−21.24	−65.60	297.0	6.1
15	2005/3/21	−24.86	−63.47	572.4	6.8

3.2 Synthetics and partials

We compute synthetics using the algorithm of Kawai *et al.* (2006) for the anisotropic PREM model (Dziewonski & Anderson 1981) for the Global CMT solutions using boxcar moment rate functions. We compute the ratio of the maximum amplitude of the data and the corresponding synthetic, and exclude records for which the ratio is greater than 2 or less than 0.5. The first-stage data set consists of 412 records that satisfy the criteria; 374 records that did not satisfy the criteria were rejected. The data are velocity seismograms (with units of m s^{-1}) with 1 Hz sampling. The reciprocal of the maximum amplitude of each record is used as the weighting factor in the inversion, so that all data have roughly the same importance.

Table 2. Earthquakes for which we used US-Array data.

Event #	Date (Y/M/D)	Depth	Latitude	Longitude	M_w
1	2005/06/02	193.2	−24.35°	−67.21°	6.0
2	2006/08/25	185.8	−24.44°	−67.18°	6.6
3	2006/09/22	602.4	−26.85°	−63.05°	6.0
4	2006/11/13	573.4	−26.10°	−63.47°	6.8
5	2007/05/25	188.5	−24.33°	−67.36°	5.9
6	2007/07/21	280.2	−22.31°	−66.00°	6.4
7	2007/11/18	262.4	−22.67°	−66.48°	6.0
8	2008/09/03	571.3	−26.85°	−63.30°	6.3
9	2008/10/12	361.5	−20.30°	−65.23°	6.2
10	2009/07/12	197.1	−15.25°	−70.75°	6.1
11	2009/09/05	209.8	−15.46°	−70.68°	5.8
12	2009/11/13	611.8	−17.96°	−64.22°	5.8
13	2009/11/14	221.2	−23.04°	−66.83°	6.2
14	2010/01/28	204.5	−23.64°	−66.96°	5.9

3.3 Static corrections

The source parameters (moment tensors, centroids and half durations) are fixed to the Global CMT solution. We convolve a boxcar moment rate function with the GCMT half duration with the synthetic seismograms and their partial derivatives. Since the inversion is only for the structure of D″ in the target region, other effects must be accounted for empirically. To correct for the effect of local structure near the stations and the sources, we make static corrections using the autopick method (Fuji *et al.* 2010).

3.4 First-stage inversion

The *S*-wave velocities at points 400 km or more above the CMB are fixed to PREM, while those within 400 km of the CMB are the unknown parameters. The study area is at latitudes between −10° and 30° and longitudes between 255° and 275°. We divide the studied volume into 5° × 5° × 50 km voxels. There are thus 336 unknown parameters. We conduct inversion using the first n basis vectors obtained by the CG method. We choose the value of n that minimizes AIC.

Model CG5, the model obtained for $n = 5$, is shown in Fig. 4. Defining the variance of the data to be 100 per cent, the variance (data minus synthetics) for the PREM synthetics is 88.4 per cent. A further variance reduction to 80.9 per cent is achieved by making the static corrections. The variance for model CG5, which minimizes AIC, is 63.0 per cent, as shown in Fig. 5 and Table 3.

Fig. 4 shows laterally heterogeneous structure mainly in the E–W direction for each depth; wiggling high-velocity anomalies are surrounded by low-velocity anomalies. The velocity perturbations are largest at the bottom of the mantle and second largest in the depth range from 350 to 400 km above the CMB. The velocity anomalies seem to continue from the CMB to 400 km above the CMB. The lateral scale of the high-velocity anomaly is about 250 km × 1500 km.

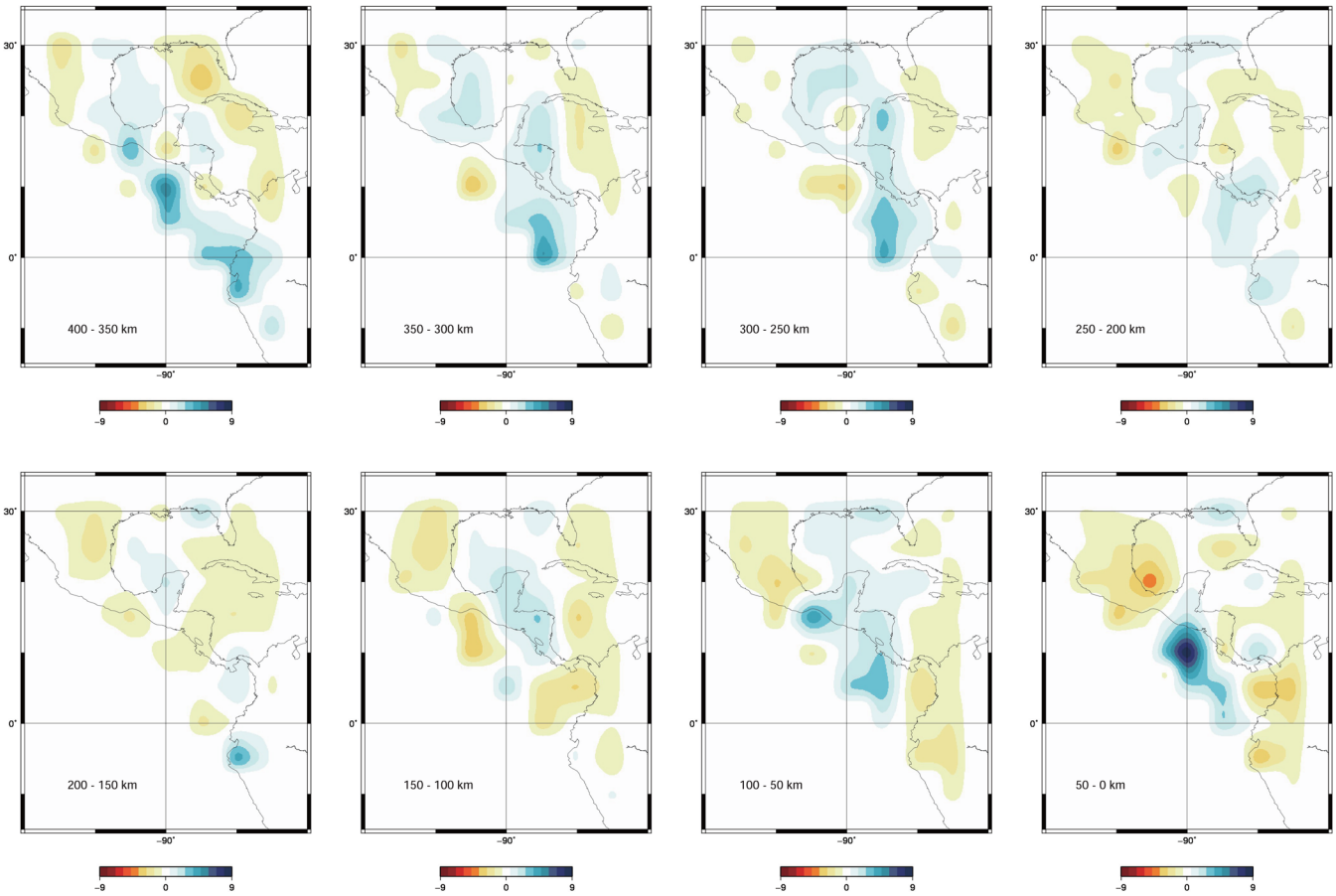

Figure 4. Results of the first-stage inversion (CG5) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. The reference for the velocity perturbation is the PREM model.

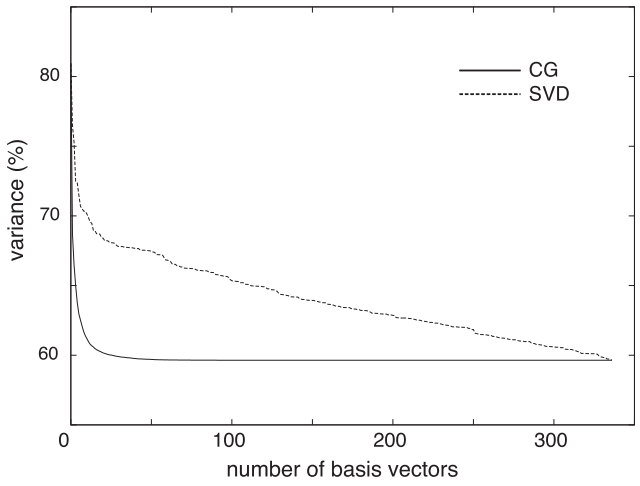

Figure 5. Variance reduction for stage 1 CG and SVD inversions. Model CG5 is shown in Fig. 4 and Models SVD50 and SVD200 are shown, respectively, in Figs 11 and 12.

3.5 Resolution check

We conduct synthetic resolution ('checkerboard') tests (Figs 6 and 7) to examine the ability of our methods to resolve various synthetic structure models. Since the input models are synthetic, static corrections are not needed. Synthetic seismograms were calculated using the Born approximation. We confirm that waveform inversion can resolve the lateral heterogeneity well in all depth ranges for this region, although the resolution in the depth range from 50 to 150 km above the CMB is relatively low. Waveform inversion can resolve laterally heterogeneous structure from waveforms propagating only in a relatively limited range of azimuths because of the large amount of information contained in the waveforms, which are linearly independent. This differs from waveform inversion for upper-mantle structure using surface waves, because our data set includes waveforms whose incidence angles differ from one another and whose most sensitive point, that is, the turning point, is different for each record. Fig. 7 shows that although the overall resolution of laterally heterogeneous structure is good, resolution where the number of bottoming points is large is better, for example, around profile B–B′. As we show in Section 4, better resolution in the northeast direction is obtained by also using data from the US-Array, which was deployed recently.

Table 3. Variance and AIC for each model obtained using the data for the events in Table 1. For calculation of AIC values, an empirical redundancy parameter α of 30 is used.

Model	Variance (per cent)	AIC
PREM	88.4	—
PREM with time-shift	80.9	304.2
CG5	63.0	285.1
SVD50	67.5	383.1
SVD200	62.9	676.9

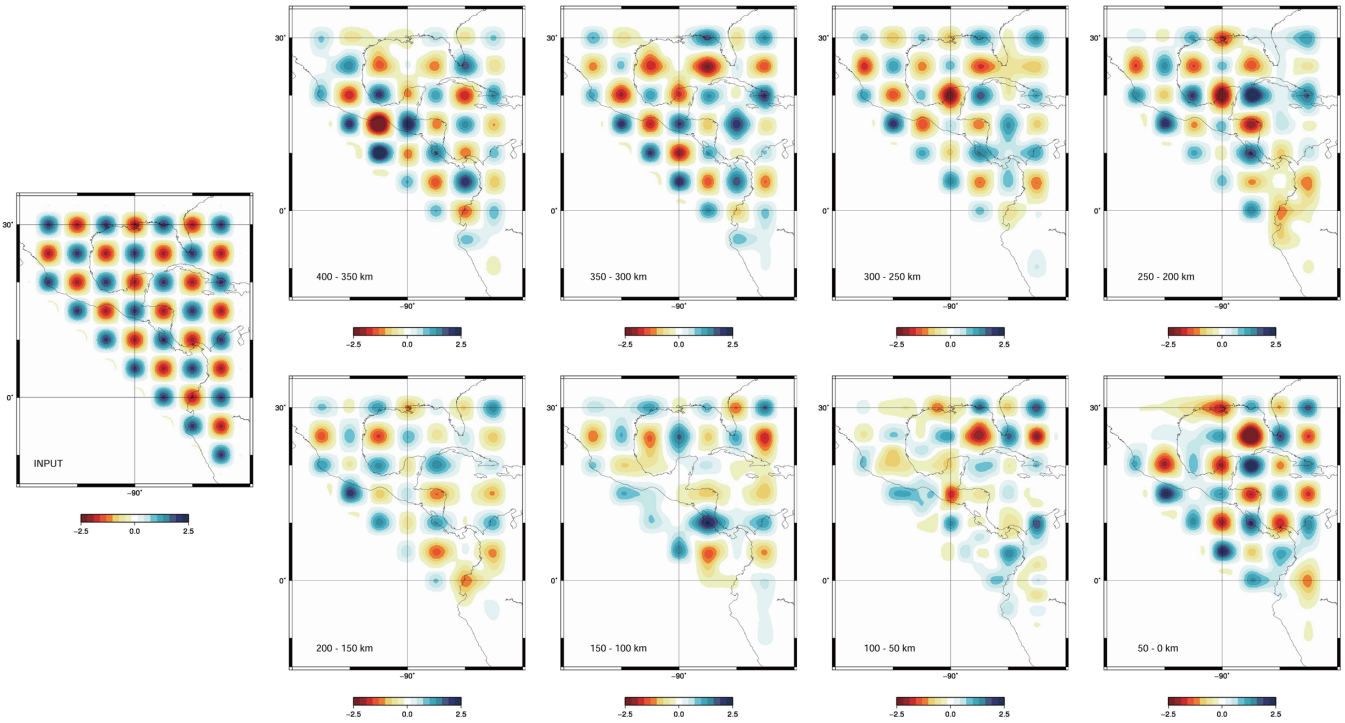

Figure 6. Recovered models for checkerboard patterns of the heterogeneities for each depth. The input model is shown in the left panel.

It should be noted that the above-mentioned checkerboard resolution test is a best-case estimate and that some studies have suggested that checkerboard tests can be misleading (e.g. Lévêque *et al.* 1993). Recent research is exploring other ways to estimate the uncertainty of waveform inversion results (e.g. Fichtner & Trampert 2011a,b).

We now discuss validation of our methods, the robustness of our obtained model and the resolution of the data set.

3.6 Inversion without static corrections

To address the question of the effect of the time-shifts (static corrections), we conduct inversions without making any time-shifts. A data set of 412 time windows with a variance reduction of 7.5 per cent (relative to synthetics for PREM when no time-shift was made) was used, and the resulting models for CG5 are shown in Fig. 8. The models obtained have essentially the same character as the models obtained by the actual inversions shown in Fig. 6, except that the absolute velocities are systematically shifted to lower values. Thus we conclude that the pattern of the 3-D velocity models is not an artefact of the time-shifting, but that hypothetical errors in the time-shifting procedure might cause small errors in the absolute velocities. The question of static corrections is complex, and further work is warranted. For example, it might be possible to use sensitivity kernels and a 3-D global structure model to compute correction terms and invert the corrected data for the velocity in D″. However, the results in the above-mentioned figures confirm the robustness of the overall spatial patterns found by our inversion.

3.7 Comparison with CG using 20 s periods waveforms

We conduct inversion using the first five basis functions obtained by the CG method in the same way as mentioned above, but using waveforms only up to periods of 20 s. The inversion for $n = 5$ is shown in Fig. 9. This model is very similar to that obtained using waveforms at periods up to 12.5 s in Fig. 4. We thus further confirm that our model (CG5 in Fig. 4) is robust.

We also conduct checkerboard tests to estimate the resolution of waveforms at periods up to 20 s and find that the resolution (Fig. 10) is a little worse than that up to 12.5 s (Fig. 6). This may imply that the scale of laterally heterogeneous velocity structure in the regions investigated in this study is large enough to be resolved by data sets at periods up to 20 s.

3.8 Comparison of CG inversion with inversion using SVD

We conduct inversion using the eigenvectors corresponding to the n largest eigenvalues of the SVD of the matrix of partial derivatives as the basis functions for the perturbation to the starting model. The results of our inversions for $n = 50$ and 200 are shown in Figs 11 and 12 (SVD50 and SVD200, respectively). Fig. 5 shows the variance for both SVD and CG inversions, which indicates that the SVD inversion converges more slowly than the CG inversion. Table 3 shows variance and AIC values for each model. The variance for SVD200 is almost equal to that for CG5 and a comparison of Fig. 4 and Fig. 12 shows that model SVD200 is almost equivalent to CG5. However, AIC for SVD200 is much larger than that for SVD50, while AIC for SVD50 is larger than that for PREM with time-shifts. This suggests that the data set does not contain sufficient information to justify inverting for even the first 50 basis vectors when the SVD basis is used, but the SVD inversion results are shown here for purposes of comparison.

Although the features of models SVD50 and SVD200 are similar in terms of pattern, the magnitude of the perturbations in SVD50 is smaller than SVD200. As well known for SVD inversion, eigenvectors for large eigenvalues tend to reflect long-wavelength structure in the model space. As discussed in Section 3.7, the similarity in pattern of inferred models between SVD50

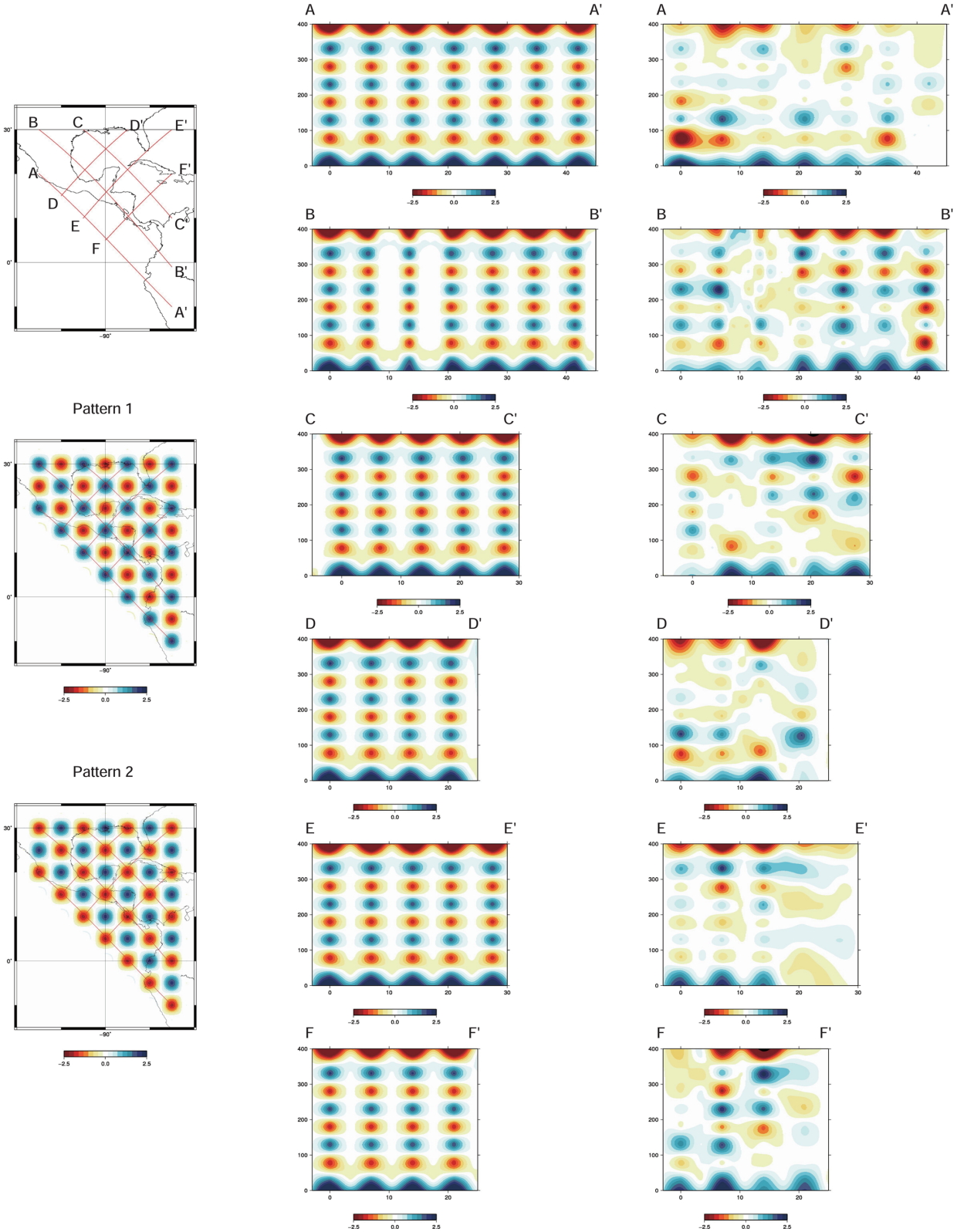

Figure 7. Recovered models for checkerboard patterns of the heterogeneities for six cross-sections (A–A′, B–B′, C–C′, D–D′, E–E′, F–F′ and G–G′). The input model is interlaminated with patterns 1 and 2 for each 50 km depth interval. The input models for each cross-section are shown in the centre column.

and SVD200 may imply that the long-wavelength components are dominant for this case. However, it is not necessarily true that the expansion coefficients for eigenvectors of large eigenvalues will be dominant.

We also conduct checkerboard tests to estimate the resolution of SVD50 (Fig. 13) and SVD200 (Fig. 14). We find that the resolution of SVD50 is, of course, worse than that of SVD200 and that the resolution of SVD200 is equal to that of CG5 (Fig. 6).

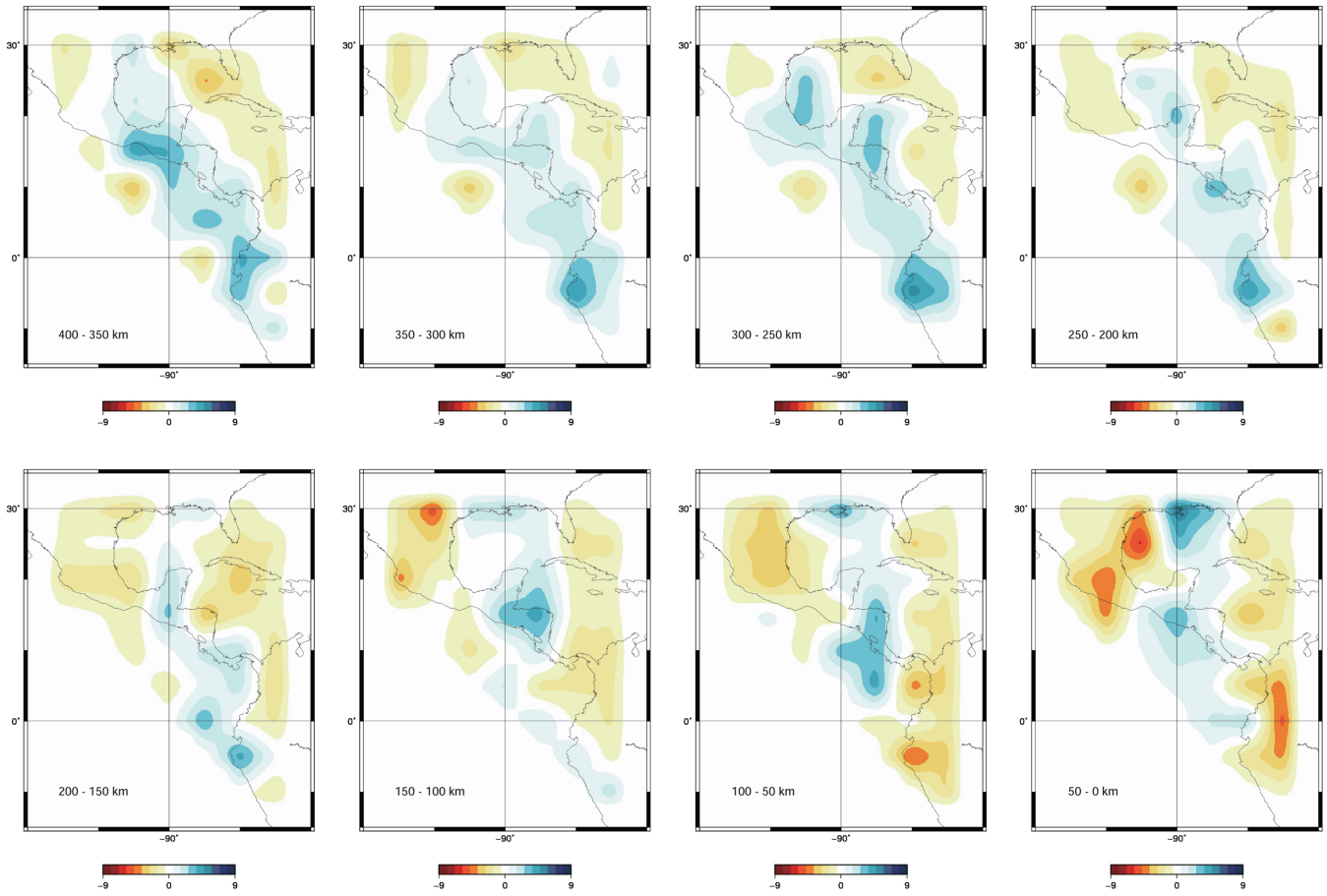

Figure 8. Results of the inversion (CG5) for the 3-D shear wave velocity structure without static corrections.

The contribution to the data space for SVD inversion can be described as the summation of contributions for model parameters up to the truncation index, q, as follows:

$$\boldsymbol{s}(q_{\rm svd}) = \mathbf{A}\sum_{i=1}^{q_{\rm svd}} \delta \boldsymbol{m}_i = \mathbf{A}\sum_{i=1}^{q_{\rm svd}} e_i \boldsymbol{v}_i = \sum_{i=1}^{q_{\rm svd}} \boldsymbol{s}_i, \tag{13}$$

where $\boldsymbol{s}_i = \mathbf{A}e_i\boldsymbol{v}_i$. When we describe the contribution to the data space for the SVD inversion in the checkerboard test as $\boldsymbol{s}^{\rm c}$, we know that $\boldsymbol{s}^{\rm c} = \boldsymbol{s}$. The contribution to the data space for the CG inversion can be written as:

$$\boldsymbol{c}(q_{\rm cg}) = \mathbf{A}\sum_{i=1}^{q_{\rm cg}} \delta \boldsymbol{m}_i = \mathbf{A}\sum_{i=1}^{q_{\rm cg}} a_i \boldsymbol{p}_i = \sum_{i=1}^{q_{\rm cg}} \boldsymbol{c}_i, \tag{14}$$

where $\boldsymbol{c}_i = \mathbf{A}a_i\boldsymbol{p}_i$. The contribution to the checkerboard test, $\boldsymbol{c}^{\rm c}$, is:

$$\boldsymbol{c}^{\rm c}(q_{\rm cg}) = \mathbf{A}\sum_{i=1}^{q_{\rm cg}} \delta \boldsymbol{m}_i^{\rm c} = \mathbf{A}\sum_{i=1}^{q_{\rm cg}} a_i \boldsymbol{p}_i^{\rm c} = \sum_{i=1}^{q_{\rm cg}} \boldsymbol{c}_i^{\rm c}. \tag{15}$$

As $\boldsymbol{c}^{\rm c}$ is generally not equal to $\boldsymbol{c}$, the question of whether the checkerboard test for the CG inversion reflects the resolution of the data set in the actual inversion remains open. As discussed later, our results qualitatively suggest that $\boldsymbol{c}^{\rm c} \approx \boldsymbol{c}$ using the relation $\boldsymbol{s}^{\rm c}(200) = \boldsymbol{s}(200)$. While $\boldsymbol{s}^{\rm c}(200) \approx \boldsymbol{c}^{\rm c}(5)$ from the similarity between Figs 6 and 14, $\boldsymbol{s}(200) \approx \boldsymbol{c}(5)$ between Figs 4 and 12. Hence, our results qualitatively suggest that $\boldsymbol{c}^{\rm c} \approx \boldsymbol{c}$. This should be quantitatively verified, but remains a topic for future work.

3.9 Solution errors

As noted in Section A2, the contributions to the data space of the eigenvectors of large eigenvalues include significant solution errors. Thus a sharp truncation of the SVD expansion can reduce solution errors (eq. A20). On the other hand, as noted in Section A3, while inversion using the CG method efficiently reduces the residuals between the observed data and synthetics, the truncation of the CG expansion will not necessarily contribute to reduction of solution errors. In this section, we discuss solution errors for both approaches for the data set used in this study.

The covariance matrix truncated at the k-th SVD eigenvector is given by eq. (A20):

$$\text{covar.}\,(\delta \boldsymbol{m}) = \sigma_d^2 \sum_{j=1}^{k} \frac{\boldsymbol{v}_j \boldsymbol{v}_j^{\rm T}}{\lambda_j^2}. \tag{16}$$

The covariance matrix truncated at the k-th CG iteration is given by eq. (A33):

$$\text{covar.}\,(\delta \boldsymbol{m}) = \sigma_d^2 \sum_{j=1}^{k} \frac{\boldsymbol{p}_j \boldsymbol{p}_j^{\rm T}}{\boldsymbol{p}_j^{\rm T}\mathbf{A}^{\rm T}\mathbf{A}\boldsymbol{p}_j}. \tag{17}$$

In order to estimate the solution errors, we compute the variance of the solution obtained using a statistical propagation-of-errors analysis. The definition of the covariance matrix for the data is as follows:

$$\text{covar.}\,(\delta \boldsymbol{d}) = \sum_{i.j}^{N} (\delta \boldsymbol{d}_i - \delta \bar{\boldsymbol{d}}_i)(\delta \boldsymbol{d}_j - \delta \bar{\boldsymbol{d}}_j). \tag{18}$$

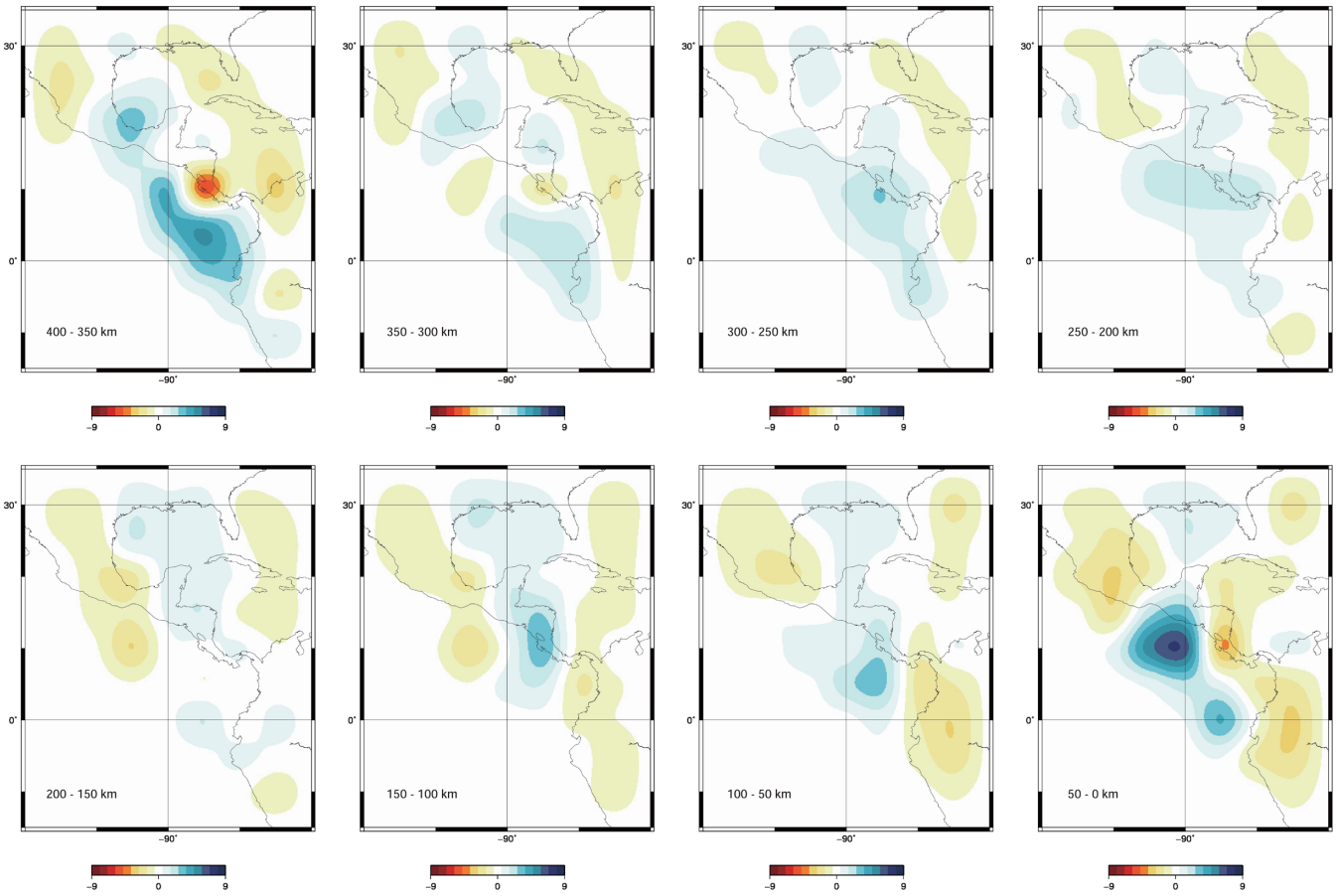

Figure 9. Results of the inversion (CG5) of low-pass filtered waveforms (at periods up to 20 s rather than up to 12.5 s) for each depth range. The reference for the velocity perturbation is the PREM model.

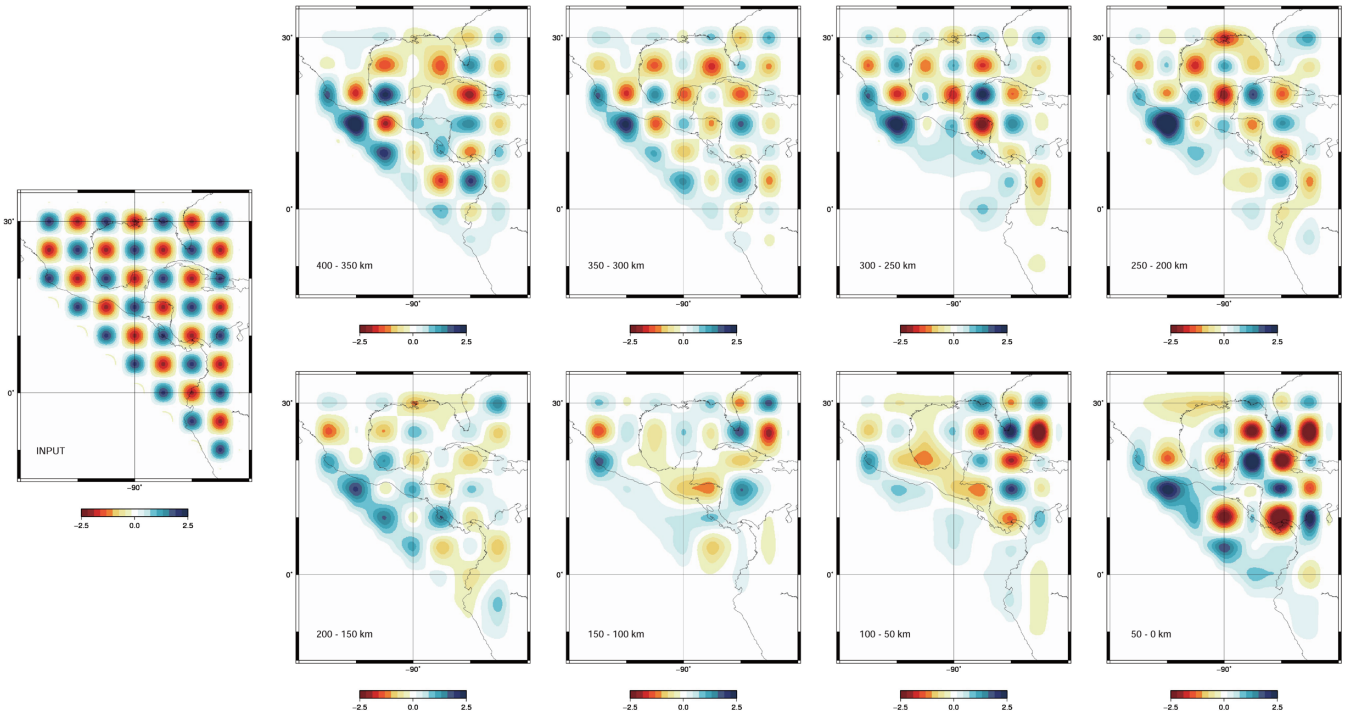

Figure 10. Recovered models for checkerboard patterns of the heterogeneities for each depth using low-pass filtered waveforms (at periods up to 20 s rather than up to 12.5 s). The input model is shown in the left-hand panel.

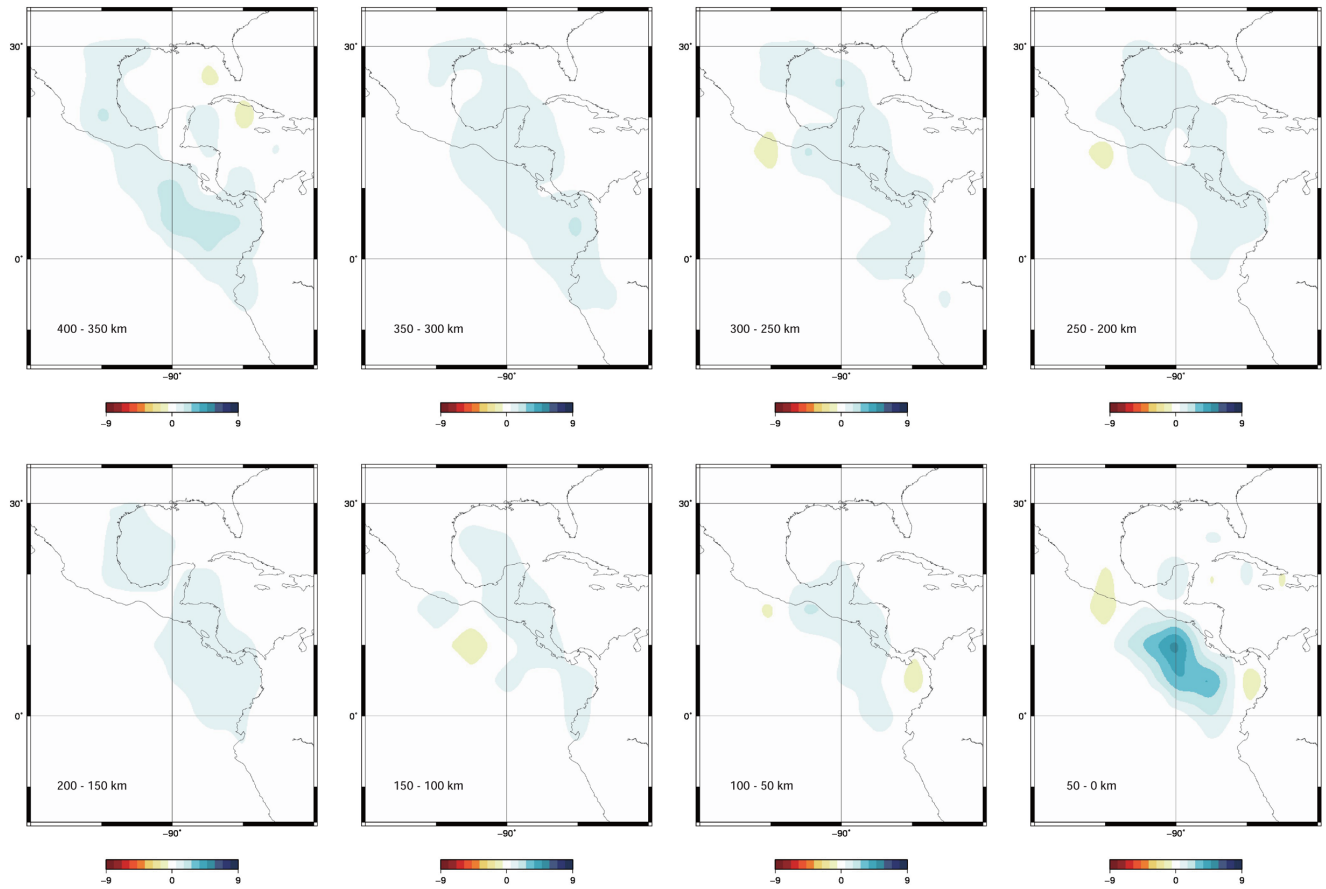

Figure 11. Results of the inversion (SVD50) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. The reference for the velocity perturbation is the PREM model.

In this study, we assume that the data errors are uncorrelated and equal, in which case the data covariance matrix becomes a constant times the identity matrix:

$$\text{covar.}\,(\delta \boldsymbol{d}) = \sigma_d^2 \mathbf{I}. \tag{19}$$

The above assumption is not necessarily valid, as the time windows include *S*, *ScS* and other phases arriving between them, which propagate along different paths and hence do not necessarily have the same data errors. Although more sophisticated estimation of solution errors may be required, this also is a topic for future work. We use the average of the noise amplitude before the arrival of the direct *S* phase as the data error. Fig. 15 shows the contribution to the variance of three typical voxels for each model parameters for both SVD and CG inversions. As shown by eq. (16), the SVD variance for small indices (large eigenvalues) is small, while the variance for large indices becomes large. On the other hand, the CG variance reduction improves rapidly for a small number of basis vectors as compared to the SVD inversion. Considering that the variance reduction of SVD200 is comparable to that of CG5, CG can achieve large variance reduction with relatively small solution errors. As noted, however, we assume that the data errors are uncorrelated and equal. This could also influence this discrepancy of the solution errors between CG5 and SVD200.

3.10 Quality control stacks

Ideally it would be possible to look at individual observed waveforms and compare them to synthetics for the various models to see visually as well as quantitatively the improvement in the fit of the final model as compared to the initial model. However, it is difficult to intuitively understand the effects of 3-D structure by inspection of individual waveforms. Also, unfortunately, the noise level is too high to allow meaningful visual comparison of data and synthetics for individual records. We therefore present 'quality control stacks' (QC stacks), some of which are shown in Fig. 16, for some of the events in this study. These stacks serve as an ancillary check to ensure that the inversion results are reasonable. The QC stacks are computed by aligning the records using the PREM arrival time and normalizing the maximum amplitude of each of the observed records to one. The synthetics are processed using the same weighting factors as the corresponding observed record. Model CG5 was used as the final model in constructing the QC stacks. The QC stacks shown in Fig. 16 are typical examples, and those for the other events are basically similar. The QC stacks in Fig. 16 show that the synthetics for the final model are, overall, an improvement over those for the initial model, thereby confirming that the inversion has reached a reasonable result. Comparing these QC stacks to those obtained by our previous studies on laterally homogeneous models (e.g. Kawai & Geller 2010a), the improvement is less clear. This is because this stacking method is most suitable for studies on laterally homogeneous layered structure but less suitable for a 3-D structure or for relatively smooth structure. It would be desirable to find some better methods to visualize the model improvement in the data space.

3.11 Record sections

We now show some examples of comparisons between observed data and synthetic seismograms for both the initial and final models.

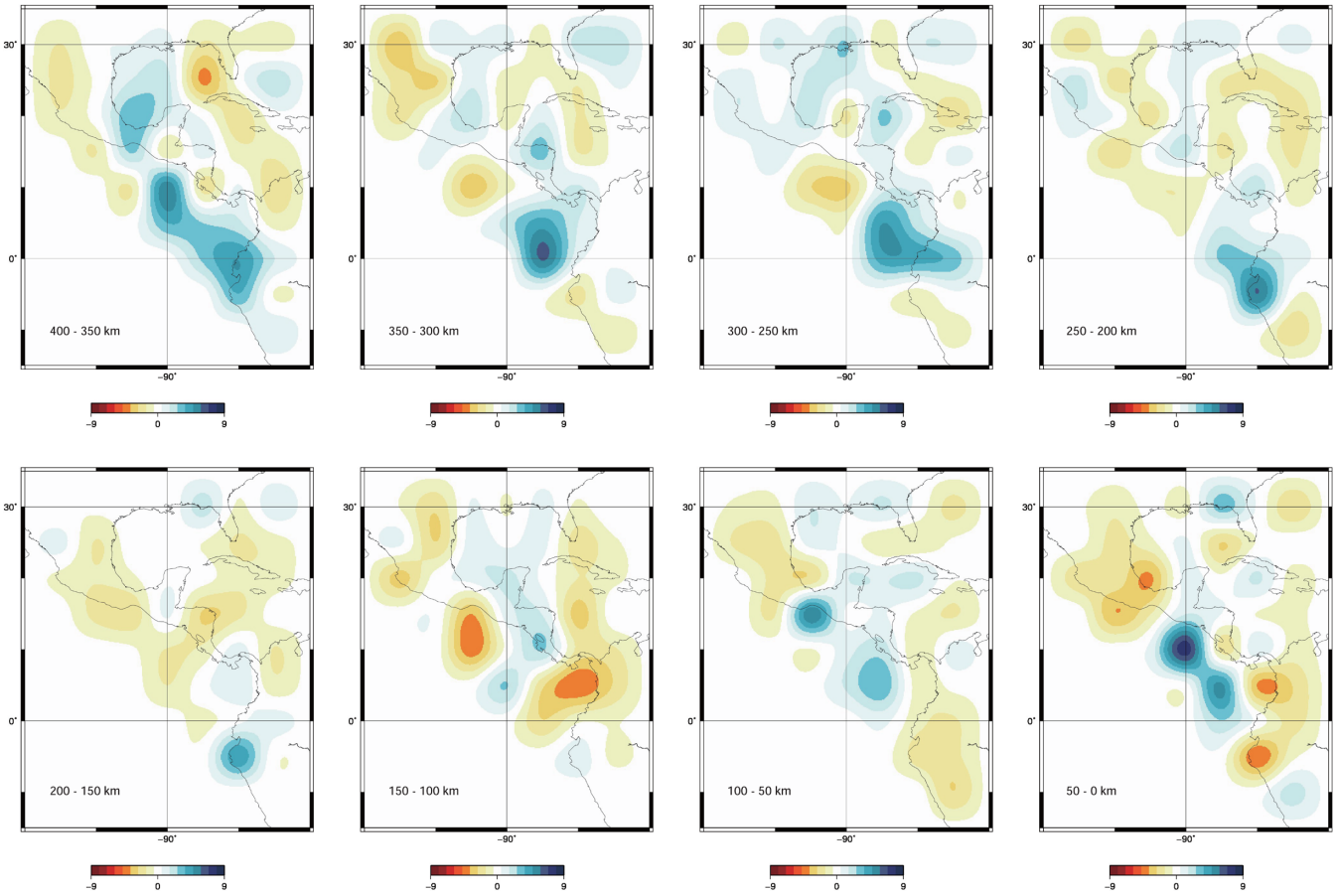

Figure 12. Results of the inversion (SVD200) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. The reference for the velocity perturbation is the PREM model.

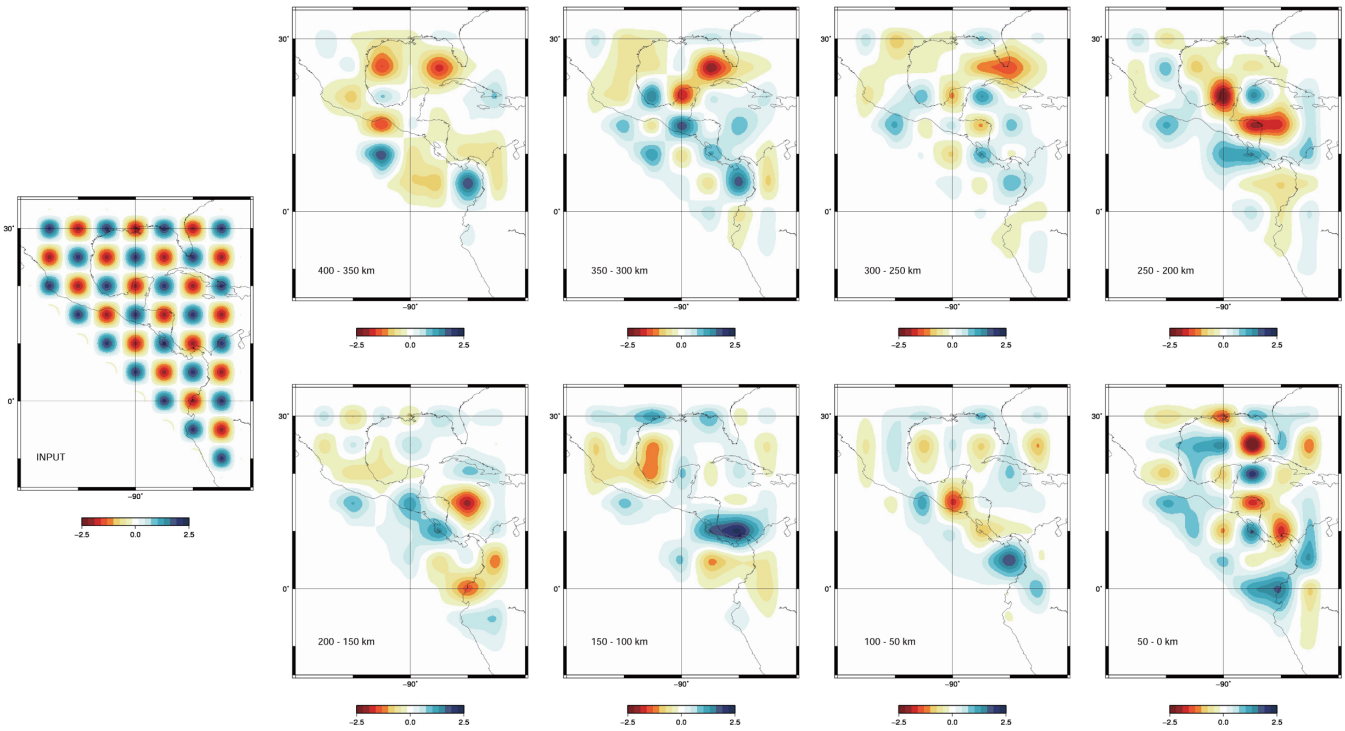

Figure 13. Recovered models for checkerboard patterns of the heterogeneities for each depth for SVD50. The input model is shown in the left-hand panel.

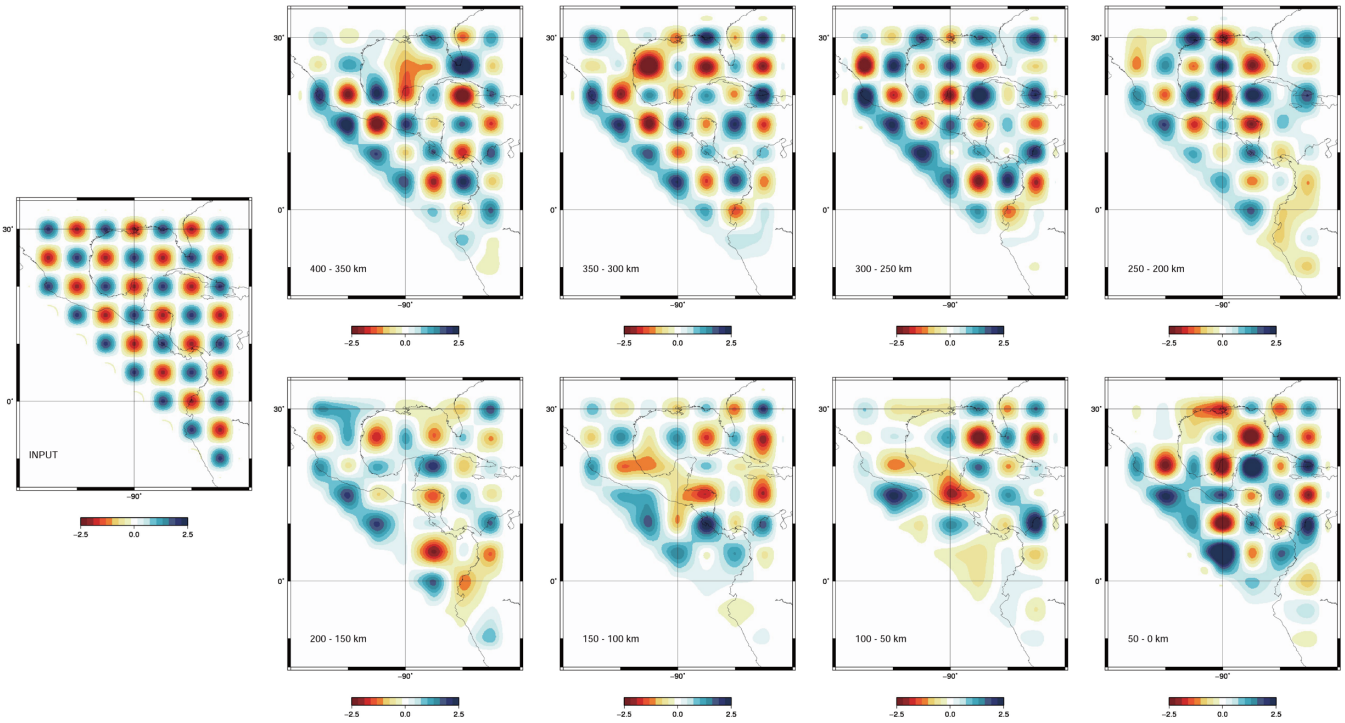

Figure 14. Recovered models for checkerboard patterns of the heterogeneities for each depth for SVD200. The input model is shown in the left-hand panel.

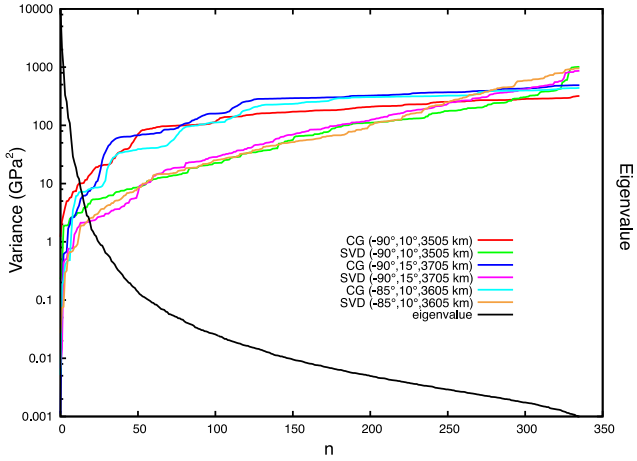

Figure 15. Contribution to variance of three typical voxels (longitude, latitude, radius) = (−90°, 10°, 3505 km), (−90°, 15°, 3705 km) and (−85°, 10°, 3605 km), for each model parameter for both SVD and CG inversions. These are obtained using a statistical propagation-of-errors analysis. SVD eigenvalues for each index normalized by maximum eigenvalue are also plotted.

For studies of 3-D structure, record sections are found to be more effective from the visual point of view than QC stacks. However, we emphasize that these should also be considered as ancillary, and that the primarily validation of the inversion results is the statistical data (variance reduction and AIC). A record section for six events is shown in Fig. 17. As relatively long-period (12.5–200 s) seismograms at epicentral distances around 90° are used in this study, it is difficult to identify phases such as *S*, *ScS* and other phases associated with complex structure in the lowermost mantle. Note that as the waveforms shown here are only a small portion of our data set, the fit for individual waveforms is not particularly meaningful. However, the fit of the synthetics for the inferred model (CG5) to the observed data is reasonably good for all epicentral distances. Even for overlapping phases at epicentral distances around 90° and diffracted waves at epicentral distances over 100°, significant improvements are confirmed. It is, however, difficult to intuitively understand which structure in CG5 contributes to the improved fit of the synthetics to the observed data.

Our models were all obtained by linearization with respect to a spherically symmetric starting model. Our methods for computing synthetics can also be used with respect to a 3-D starting model (e.g. Cummins *et al.* 1997) but the computational requirements for waveform inversion would be greatly increased. In the future it would be desirable to conduct inversions with respect to 3-D starting models. These and many other important technical issues, such as the effect of the choice of starting model on the results of the inversion, should be further explored, but as we have shown in this paper our present techniques are capable of obtaining robust 3-D earth models. It thus seems reasonable to press ahead with applications of waveform inversion to real data, and to incrementally improve the analytical techniques in parallel with obtaining new models of the Earth's interior.

3.12 Orthogonality

In this study, the target region for inversion is only in the lowermost mantle and the structure for shallow regions is assumed to be the anisotropic PREM model (Dziewonski & Anderson 1981). In order to examine the possible effects of shallow structure, such as the subducted Farallon slab, on the inversion results, we check the independence of partial derivatives for perturbations to structure in such regions and to structure in the lowermost mantle. We compute partial derivatives at intervals of 100 km for the depth range 100–500 km at the yellow dot in Fig. 3. The schematic image of the target region and shallow perturbation points is shown in the inset in the lower left-hand side of Fig. 3. We compute the direction cosines

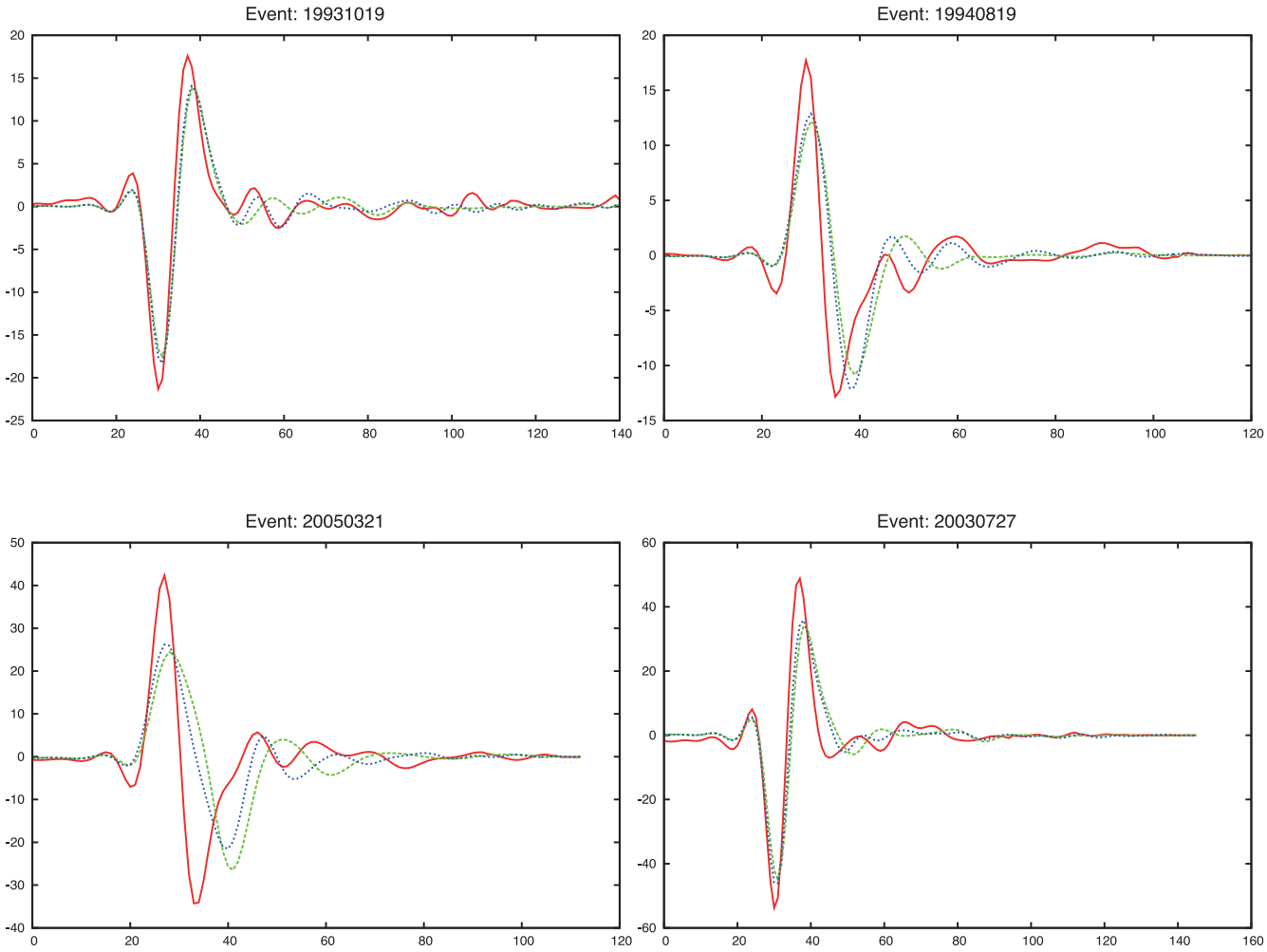

Figure 16. 'Quality control (QC)' stacks for events 19931019, 19940819, 20030727 and 20050321 computed as follows: The horizontal axis is in seconds, while the vertical axis indicates amplitude for stacked waveforms. First, all of the observed waveforms for each event were time-shifted using PREM, after making the same source and station corrections as in the inversion. These waveforms were then filtered in the passband 12.5–200 s using a bandpass filter. The maximum amplitude of each observed record was normalized to one, and the waveforms were then stacked (red curves). The synthetics for the initial model (PREM, green curves) and final model (blue curves) were stacked using the same weighting factors as for the corresponding observed record. Model CG5 was used as the final model. Note that since QC stacks are merely an ancillary presentation of visual data to confirm the absence of bugs and show the general state of some subset of the data, it is not significant whether we use PREM or the final model for the QC stacks.

between the *i*-th partial derivative in the target region and the *j*-th partial derivative in the shallower region (Fig. 18). We select a single event (19970123) and a single station (WDC) for comparing the extent of coupling for this data set. The deviation from 90° is very large for one event and one station (Fig. 18a). This indicates that it is difficult to separate the effects of shallow structure and structure in the target region from any individual seismogram. However, we can see that for larger numbers of waveforms the partials are close to orthogonal. Fig. 18(b) shows direction cosines for one station and all 27 events. Although the deviations from 90° become smaller, the partial derivatives with respect to the target region are not sufficiently independent of those of shallow structure. Fig. 18(c) shows direction cosines for one event and about 80 stations. It is noteworthy that although the number of waveforms in this case (80) is larger than in the former case (27), the deviation from 90° is sometimes larger than for Fig. 18(b). Also note that the deviation has periodic peaks, which indicates constructive interference between the shallow structure and structure in the target region. However, the data set in this study is large enough to effectively eliminate the trade-off between shallow structure and structure in the the target region when all seismograms and all stations are used (Fig. 18d) .

4 SECOND-STAGE INVERSION

4.1 Inversion

In this section, we perform waveform inversion using the second-stage data set, discussed in Section 3.1. The additional sources and stations used in the second-stage study are shown in Fig. 19 and Table 2. The studied region and the number of unknown parameters are the same as in the first-stage inversion in Section 3.4.

Map views of model CG5, the model obtained for $n = 5$, are shown in Fig. 20, and cross-sections are shown in Fig. 21. The variance reduction and AIC for several values of the empirical redundancy parameter α are shown in Fig. 22 and Table 4. Defining the variance of the data to be 100 per cent, the variance (data minus synthetics) for the PREM synthetics is 78.4 per cent. A further variance reduction to 58.1 per cent is achieved by making the static

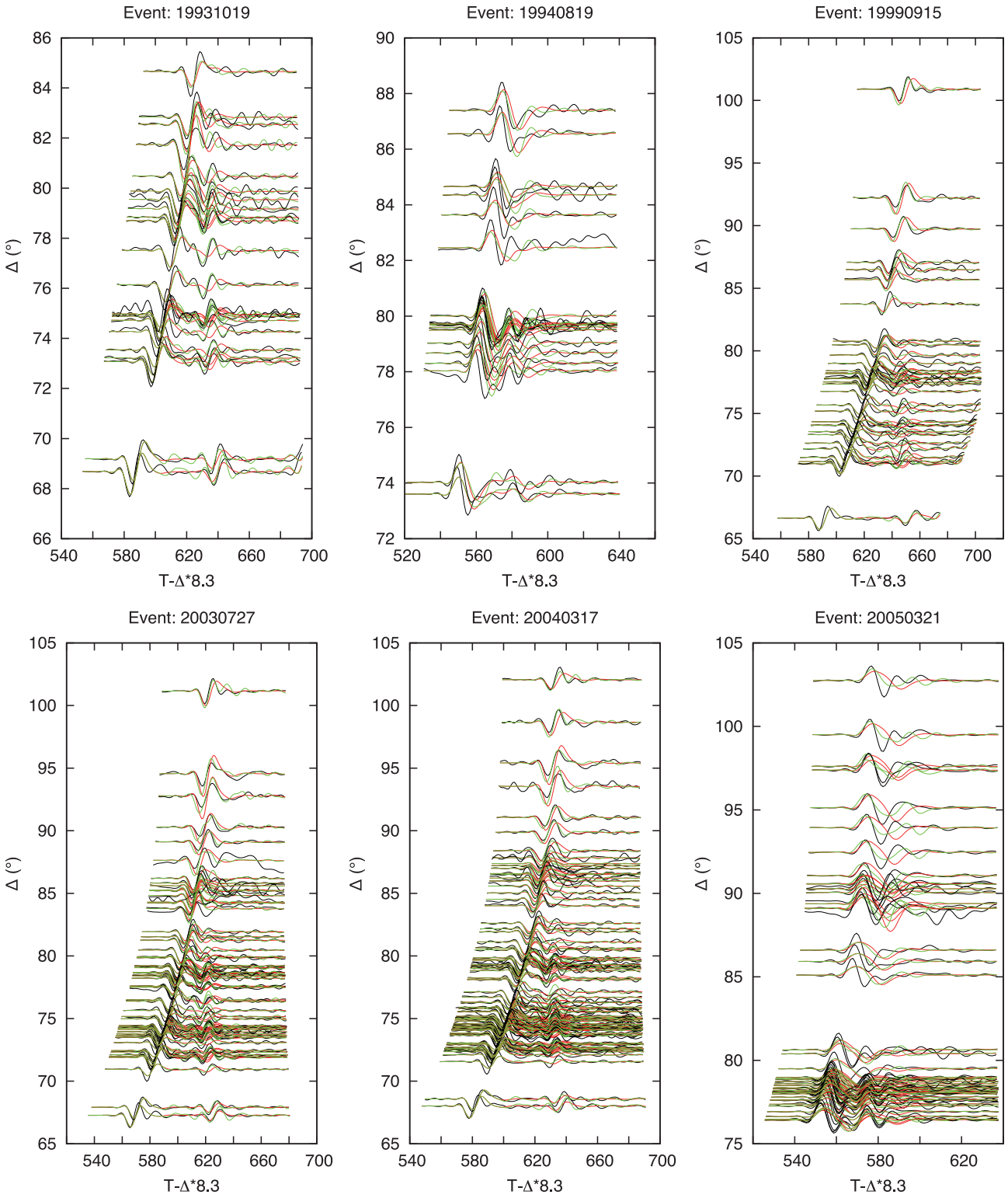

Figure 17. A record section of the transverse component for the six events. The traveltime is reduced at an apparent velocity of 8.3 km s^{-1}. A bandpass filter at periods between 12.5 and 200 s is applied. The black, red and blue traces show observed waveforms, synthetics for the initial model (PREM) and synthetics for the final model (CG5), respectively. *S*, *ScS* and other phases associated with complex structure such as the D″ discontinuity are almost entirely overlapped at the epicentral distances used in this study for this period band.

corrections. The variance for model CG5, which minimizes AIC, is 55.2 per cent, as shown in Table 4. While a variance reduction of 17.9 per cent was obtained in the first-stage inversion (Table 3), a reduction of 2.9 per cent is achieved here. The additional data used in the second-stage inversion were recorded by the US-Array. These waveforms have a relatively higher S/N ratio and a much lower level of short-wavelength noise than the waveforms used in the first-stage inversion. This is confirmed by the fact that the initial variance for

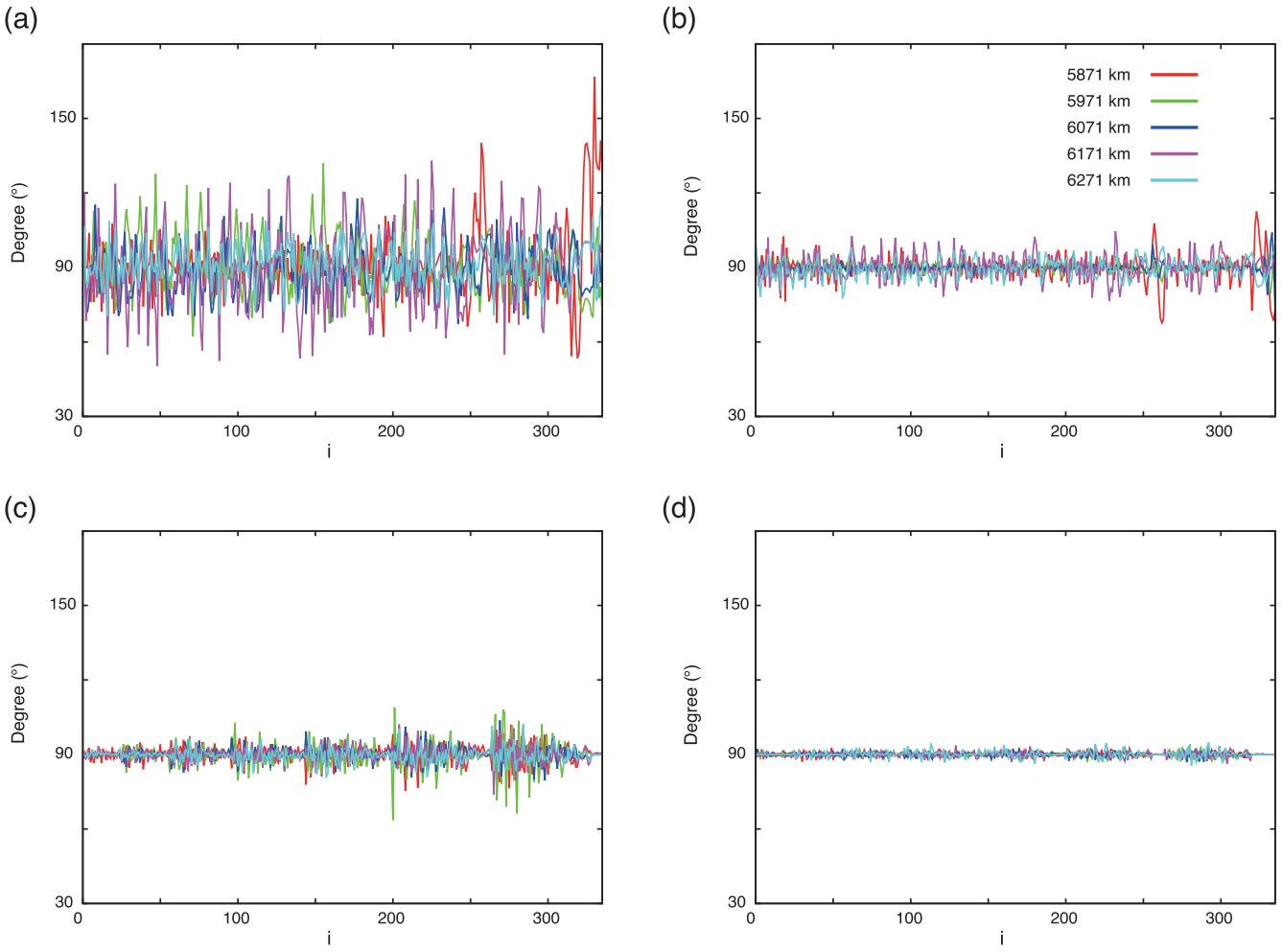

Figure 18. Panels show the value of the cosine of the angle between a vector of partial derivatives for the *i*-th point in the target region and the vector of partial derivatives for the depth shown by each coloured region at shallower depths. See inset panel at lower left of Fig. 3 for details. Panel (a) is for one waveform for one event, panel (b) is for all waveforms for one event, panel (c) is for all events for one station and panel (d) is for all events for all stations. The number of waveforms used in the four panels is 1, 29, 80 and 2154, respectively.

the second-stage inversion (78.4 per cent) is lower than that for the first-stage inversion (88.4 per cent). The static correction inherently tends to make time-shifts based on the long-wavelength component in the waveforms rather than short-wavelength noise. Since the data set in the second-stage inversion contains mainly of high-quality data observed by the US-Array, the variance reduction for the static correction is greater than that for the first-stage inversion.

Figs 20 and 21 show wiggling high-velocity anomalies surrounded by low-velocity anomalies. The velocity perturbations are largest at the bottom of the mantle. Although the pattern is almost the same for each depth, the amplitude of the velocity perturbation decreases as the distance from the CMB increases. The amplitude of the velocity perturbation in the lowermost 150 km of the mantle is notably large. Two strong low-velocity anomalies exist: one is to the northeast of the largest high-velocity anomaly in the depth range from 100 to 150 km, another is to the southwest of the largest high-velocity anomaly in the depth range from 50 to 100 km above the CMB.

The overall features of the model obtained by the second-stage inversion (Fig. 20) are almost the same as the results of the first-stage inversion (Fig. 4) but the details differ. Fig. 4 shows that the second largest velocity perturbation occurs in the depth range from 350 to 400 km above the CMB, and this is also the depth range with the largest velocity perturbation for the second-stage inversion. The lateral scale of the high-velocity anomaly just above the CMB obtained by the first-stage inversion (Fig. 4) is about 250 km × 1500 km, while it is about 250 km × 250 km for the second-stage inversion (Figs 20 and 21). The model obtained in the second-stage inversion is much sharper than that for the first-stage inversion. This appears to be due to the superior quality and quantity of the data set for the second-stage inversion.

4.2 Effects of noise

We conduct synthetic resolution tests (Fig. 23) to examine the ability to resolve various synthetic structure models for the second-stage inversion in the same manner as those presented for the first-stage inversion in Section 3.5. Compared to the results for the first-stage inversion (Fig. 6), Fig. 23 shows improved resolution.

We now test ‘checkerboard’ models with random noise added to the synthetics. A comparison of waveforms with/without noise is shown in Fig. 24. White noise is used with a fixed amplitude and randomly chosen phase in the frequency domain, and transformed to the time domain. The variance of the residuals is 9 and 70 per cent respectively. With a variance of the residuals of 70 per cent, it is difficult to visually identify signals of even direct *S* phases. We can

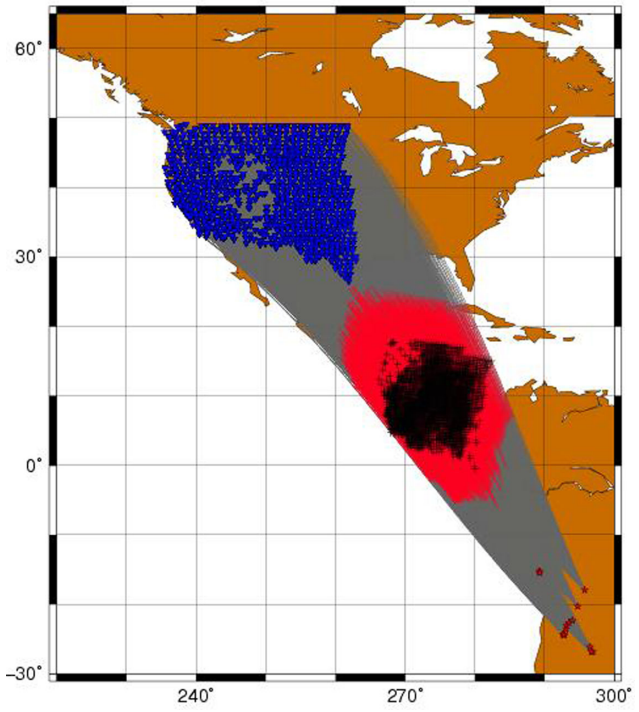

Figure 19. Event–receiver geometry, with great circle ray paths for the additional data used in the second-stage inversion. The portions of the great circles which sample D″ are shown in red, and plus signs indicate the turning points within D″. Blue reversed triangles show the sites of the US-Array stations used in our study. Red stars show the 14 intermediate and deep earthquakes studied (Table 2).

see a small deviation from the input mosaic pattern, but the data set still has sufficient resolving power to basically recover the input model (Fig. 25).

4.3 Model resolution matrix

For completeness we also conduct an SVD inversion using the eigenvectors of the 200 largest eigenvalues as the basis. The resulting model, SVD200, is shown in Fig. 26, and the variance reduction and AIC are shown in Table 4. The overall features of the respective models are generally similar. In contrast to the first-stage inversion, the AIC values in Table 4 show that the stage 2 data set appears to contain sufficient information to invert for the first 200 components of the SVD expansion, although the AIC value is slightly higher than that for CG5. (SVD200 is not necessarily the best SVD model in terms of minimizing AIC, but we are presenting this model primarily for purposes of comparison.)

We calculate the model resolution matrices for the SVD200 and CG5 expansions using eqs (10) and (12), respectively. Figs 27 and 28 show plots of selected rows of the respective model resolution matrices. As shown by the equations in Section 2.7, what the model resolution matrix tells us is not how much resolution we actually have, but just the extent to which our truncation of the solution of the inverse problem is smearing out the model parameters. Fig. 27 shows that the model resolution matrix for SVD200 is much less smeared out than that for CG5 (Fig. 28). As the AIC values for SVD200 are reasonable in Table 4, Fig. 28 may be a reasonable estimate of the actual resolution in the data set. On the other hand, CG requires only five basis vectors to arrive at essentially the same model, which requires 200 SVD basis vectors, and as a result the

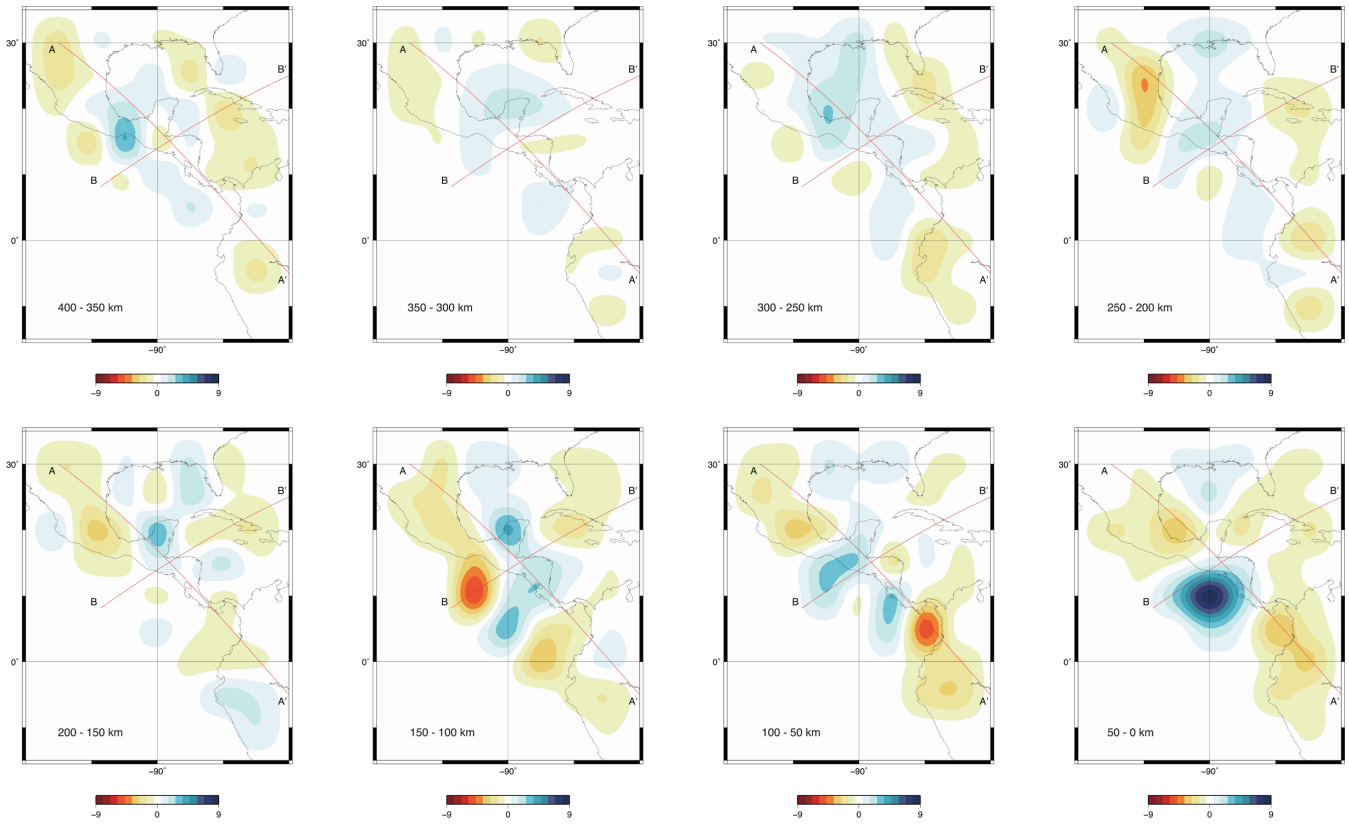

Figure 20. Results of the second-stage inversion (CG5) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. The reference for the velocity perturbation is the PREM model.

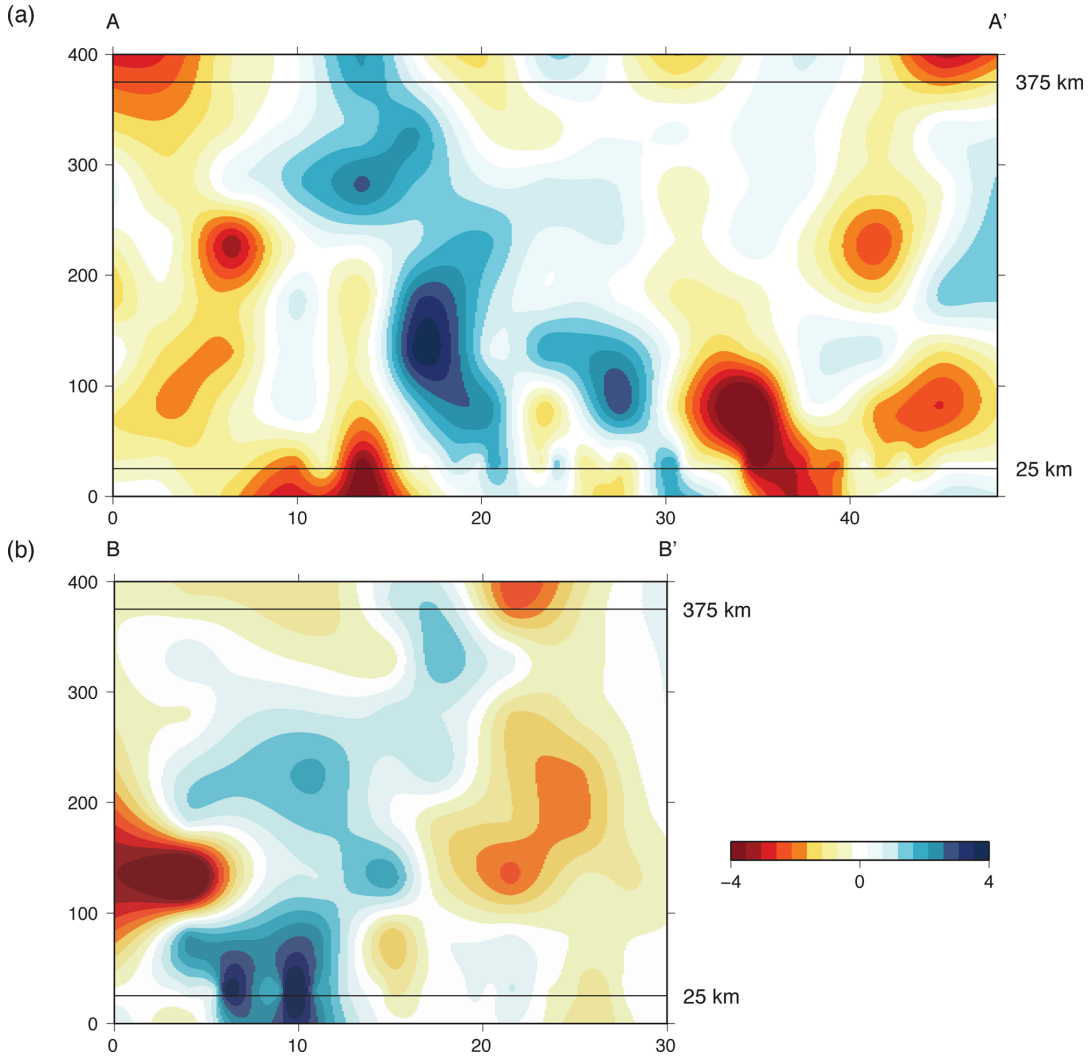

Figure 21. Results of the inversion for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for cross-sections A–A′ (panel a) and B–B′ (panel b). The orientation of the cross-sections is shown in Fig. 20.

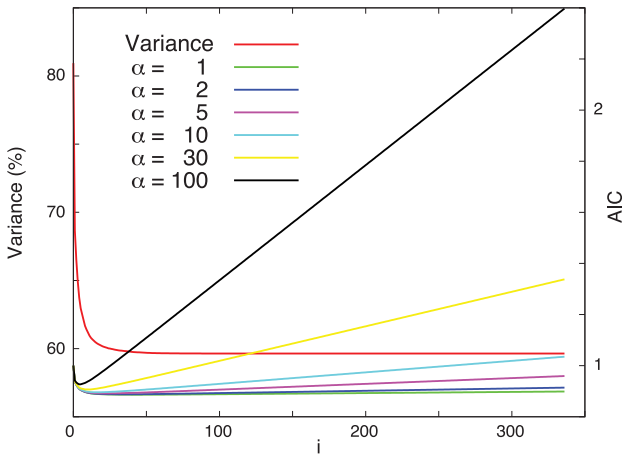

Figure 22. Variance and AIC values for CG inversions. AIC values are shown for several choices of the empirical redundancy parameter α.

nominal resolution matrix for CG5 (Fig. 28) appears very poor, whereas the resolution is probably really about the same as Fig. 27. The question of how to accurately quantify the resolution of CG inversions thus remains as an important subject for future work.

Table 4. Variance and AIC for each model obtained using the data for both the events in Table 1 and the events in Table 2. For calculation of AIC values, an empirical redundancy parameter α of 30 is used.

Model	Variance (per cent)	AIC
PREM	78.4	—
PREM with time-shift	58.1	3888.3
CG5	55.2	3861.4
SVD50	56.0	3871.5
SVD200	55.4	3867.5

4.4 Jack-knife test

We take random subsets of 50 per cent of our data set for the second-stage inversion (jack-knife) and invert for velocity structure. We conduct 10 such inversions. The results of three such inversions are shown in Fig. 29. The features of all three models are almost identical to those of CG5 (Fig. 20). While the overall features shown in Figs 20 and 29 match, they differ at some points. This suggests that the inference of models will become unstable using smaller data sets.

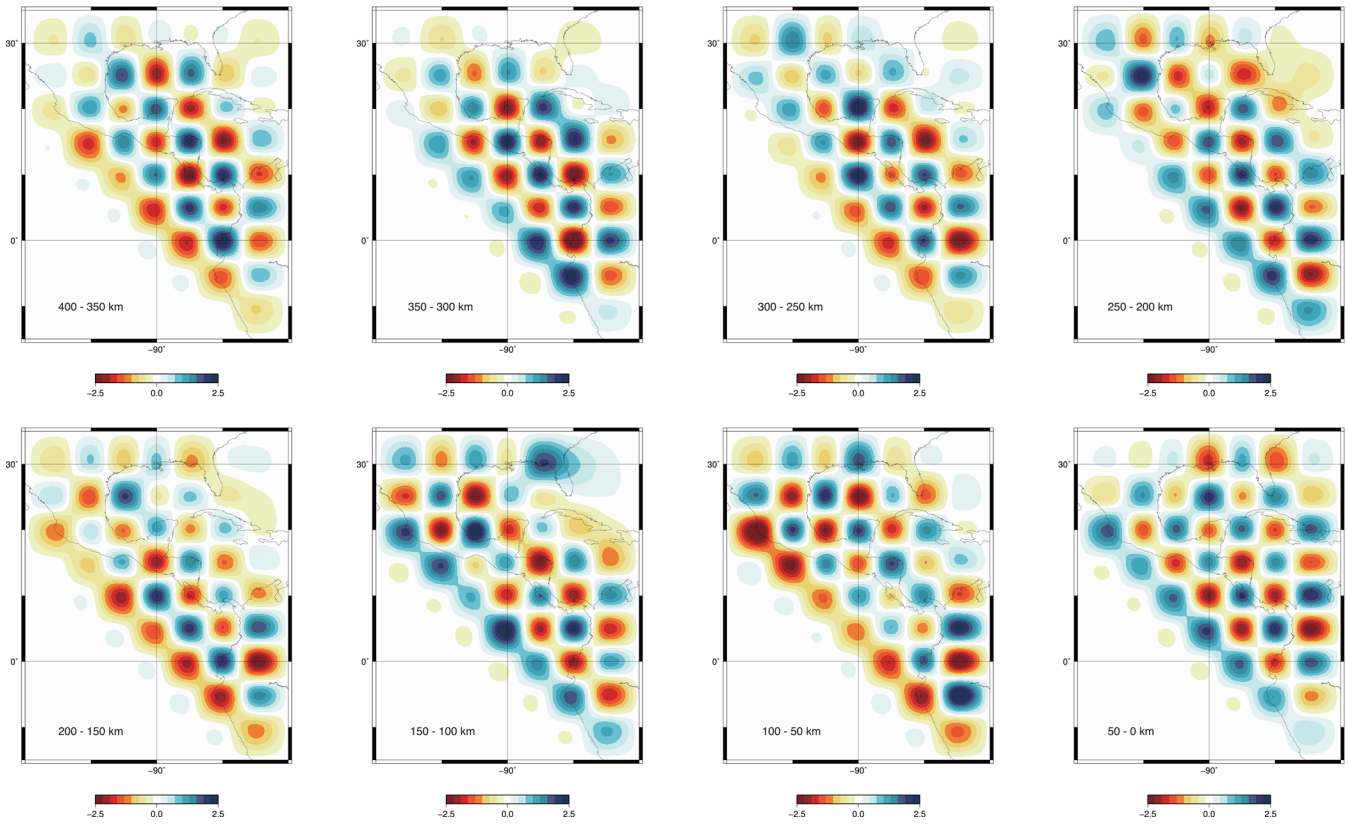

Figure 23. Recovered models for checkerboard heterogeneity input model for each depth. The input model is shown in Fig 14.

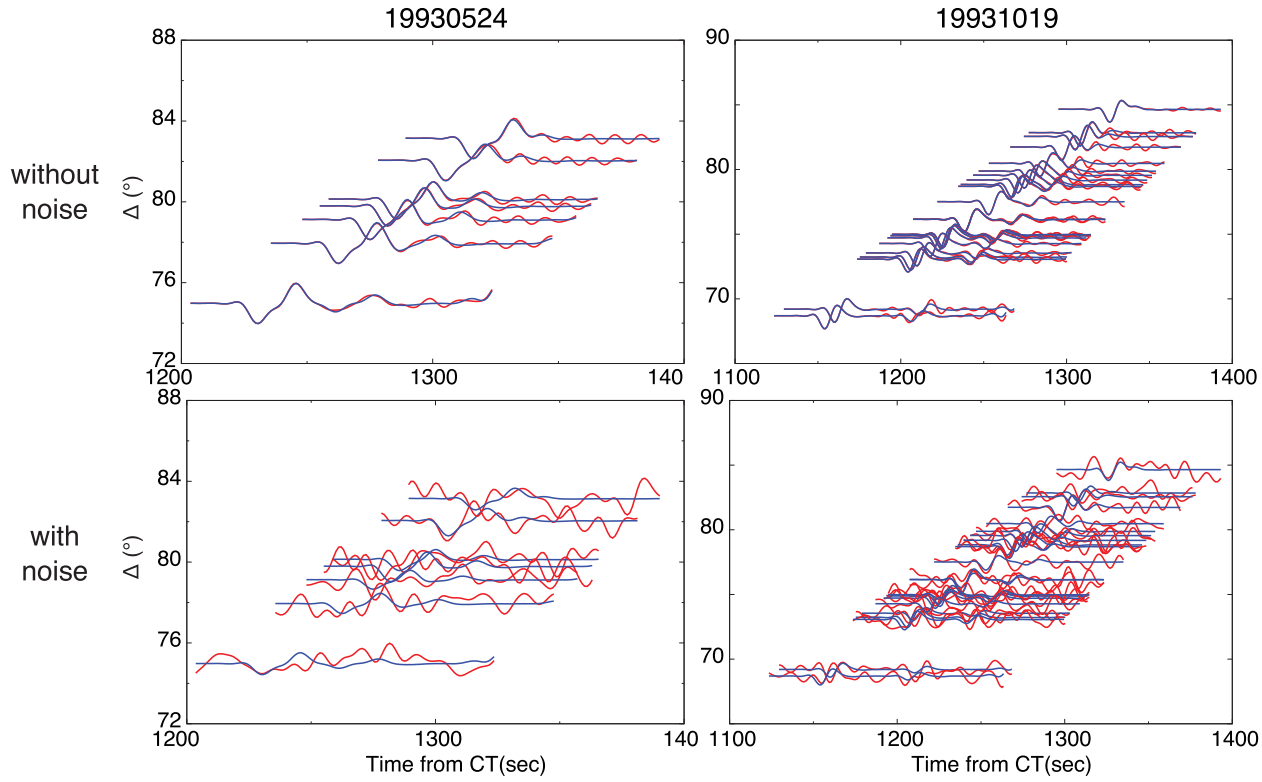

Figure 24. Waveforms with and without noise.

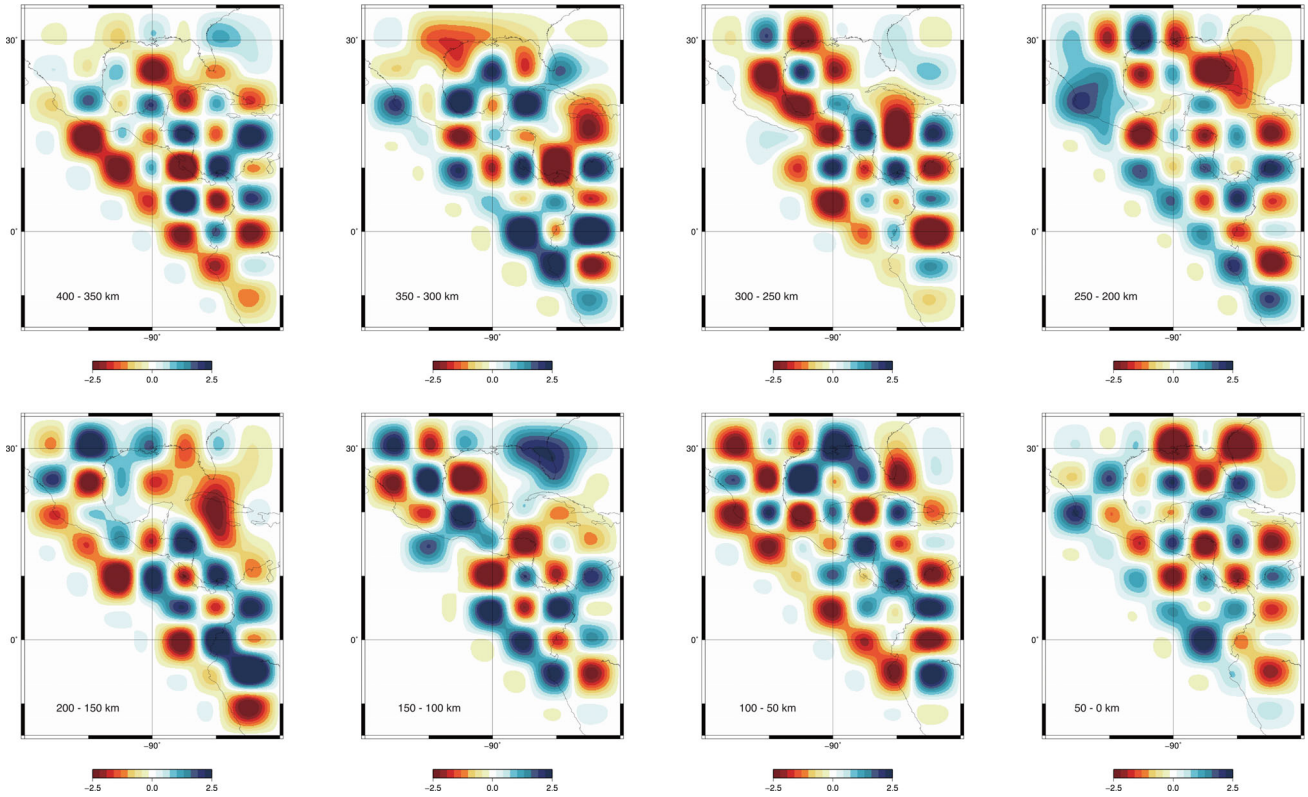

Figure 25. Recovered models for checkerboard heterogeneity input model for each depth. The input model is shown in Fig 14. Data set for the inversion contains white noise (see details in text).

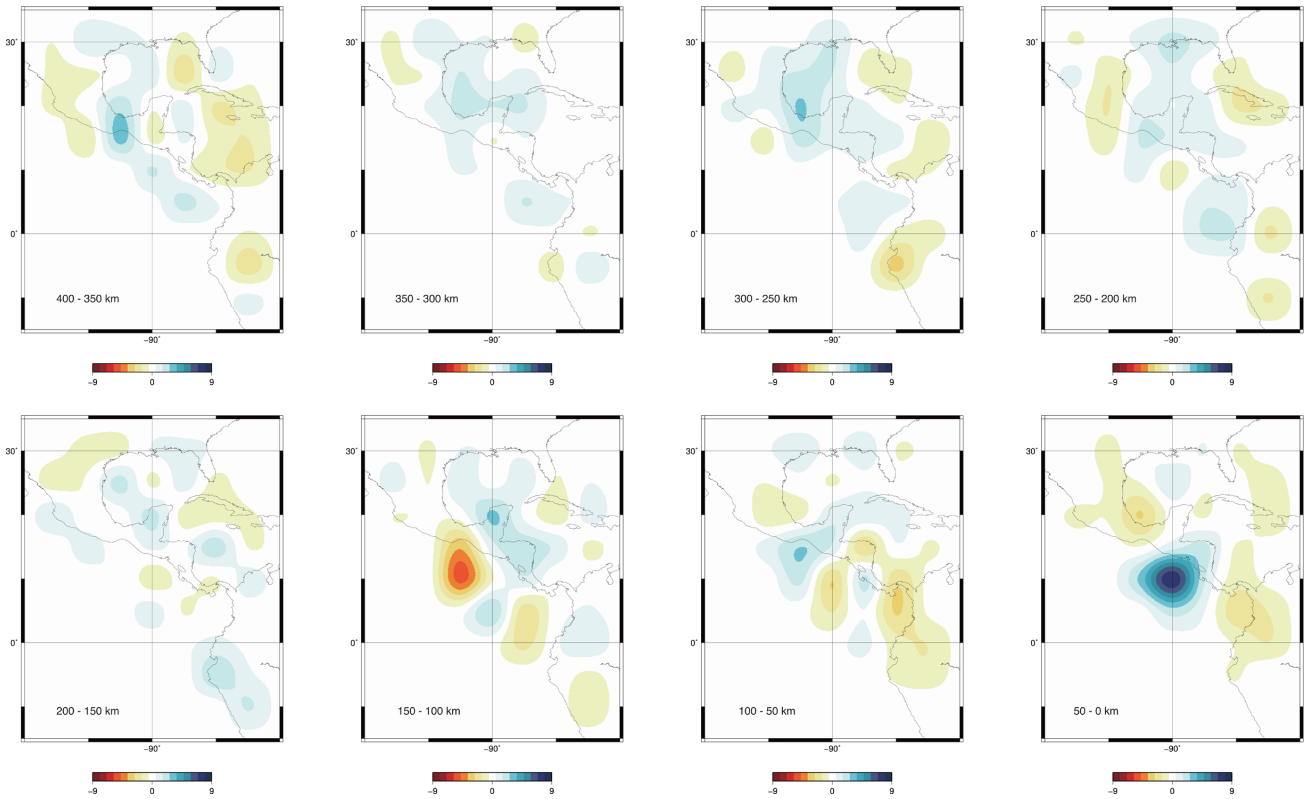

Figure 26. Results of the second-stage inversion (SVD200) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. The reference for the velocity perturbation is the PREM model.

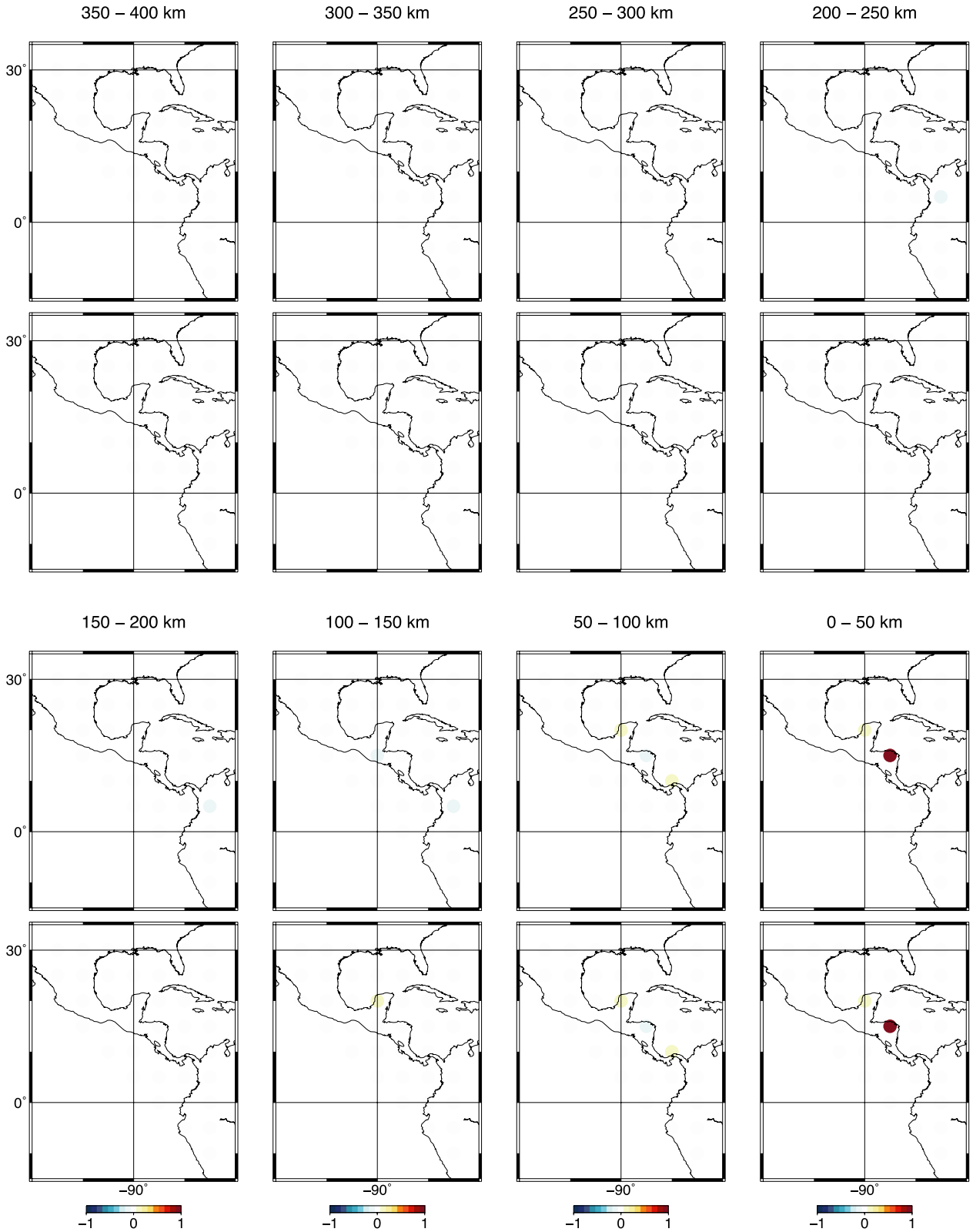

Figure 27. One selected row of the resolution matrix for the SVD200 model determined by our inversion with respect to a voxel at a latitude of 15° at a longitude of 275° in the depth range between 0 and 50 km (this page) and between 350 and 400 km above the CMB (next page). The upper panel for each depth range shows the resolution matrix for the first-stage inversion and the lower panel shows the resolution matrix for the second stage.

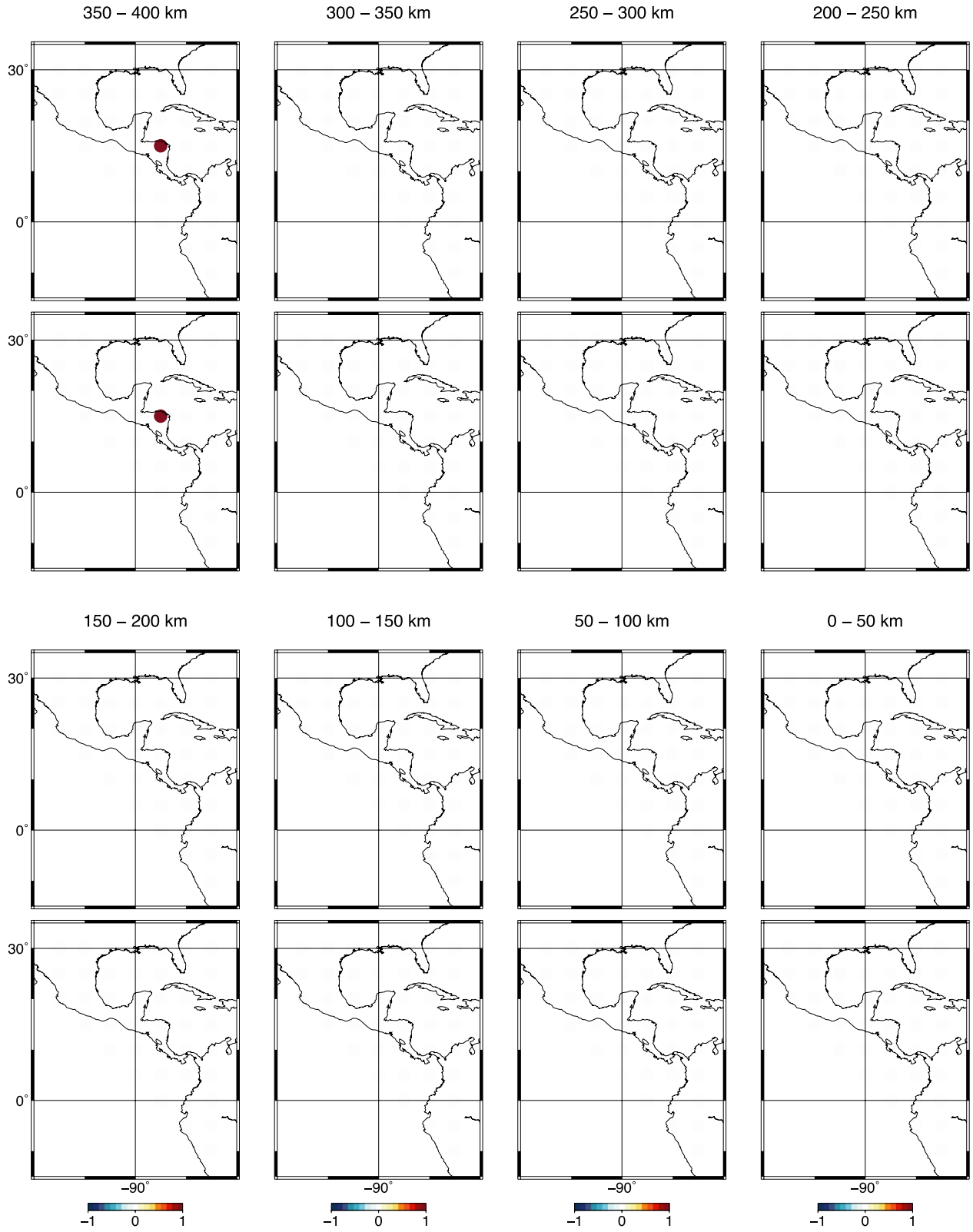

Figure 27 – (*Continued*)

5 GEOPHYSICAL INTERPRETATION

Fig. 20 shows lateral heterogeneous structure mainly in the E–W direction for each depth; wiggling high-velocity anomalies are surrounded by low-velocity anomalies. The velocity perturbations are largest at the bottom of the mantle and second largest in the depth range from 350 to 400 km above the CMB. As shown in Fig. 20, the velocity anomalies seem to continue from the CMB to 400 km above the CMB. The lateral scale of the high-velocity anomaly is about 250 km × 250 km.

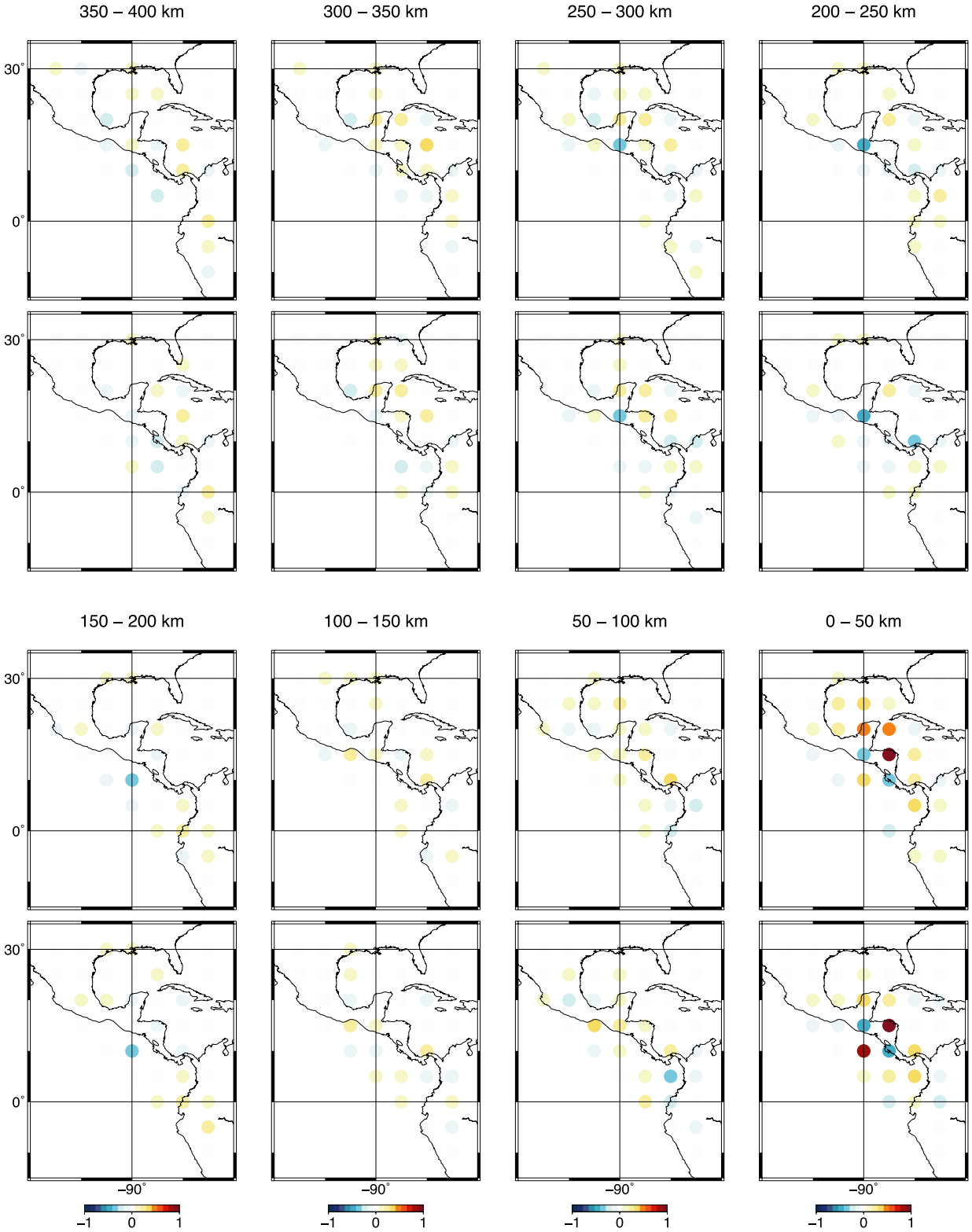

Figure 28. One selected row of the resolution matrix for the CG5 model determined by our inversion with respect to a voxel at a latitude of 15°, at a longitude of 275° in the depth range between 0 and 50 km (this page) and between 350 and 400 km above the CMB (next page). The upper panel for each depth range shows the resolution matrix for the first-stage inversion and the lower panel shows the resolution matrix for the second stage.

Fig. 21 shows that a sheet-like high-velocity anomaly reaches to the CMB and that the highest velocity (about 8 per cent faster than PREM) is observed in the deepest depth range. This can be interpreted as a remnant of the ancient Farallon Plate, which is thought to have been subducted continuously over the past 100 Myr, and which recently sunk down to the CMB. Such a remnant should be about 1500 K colder than the ambient mantle. Considering a CMB temperature of 3800 K (Kawai & Tsuchiya 2009), 2300 K for

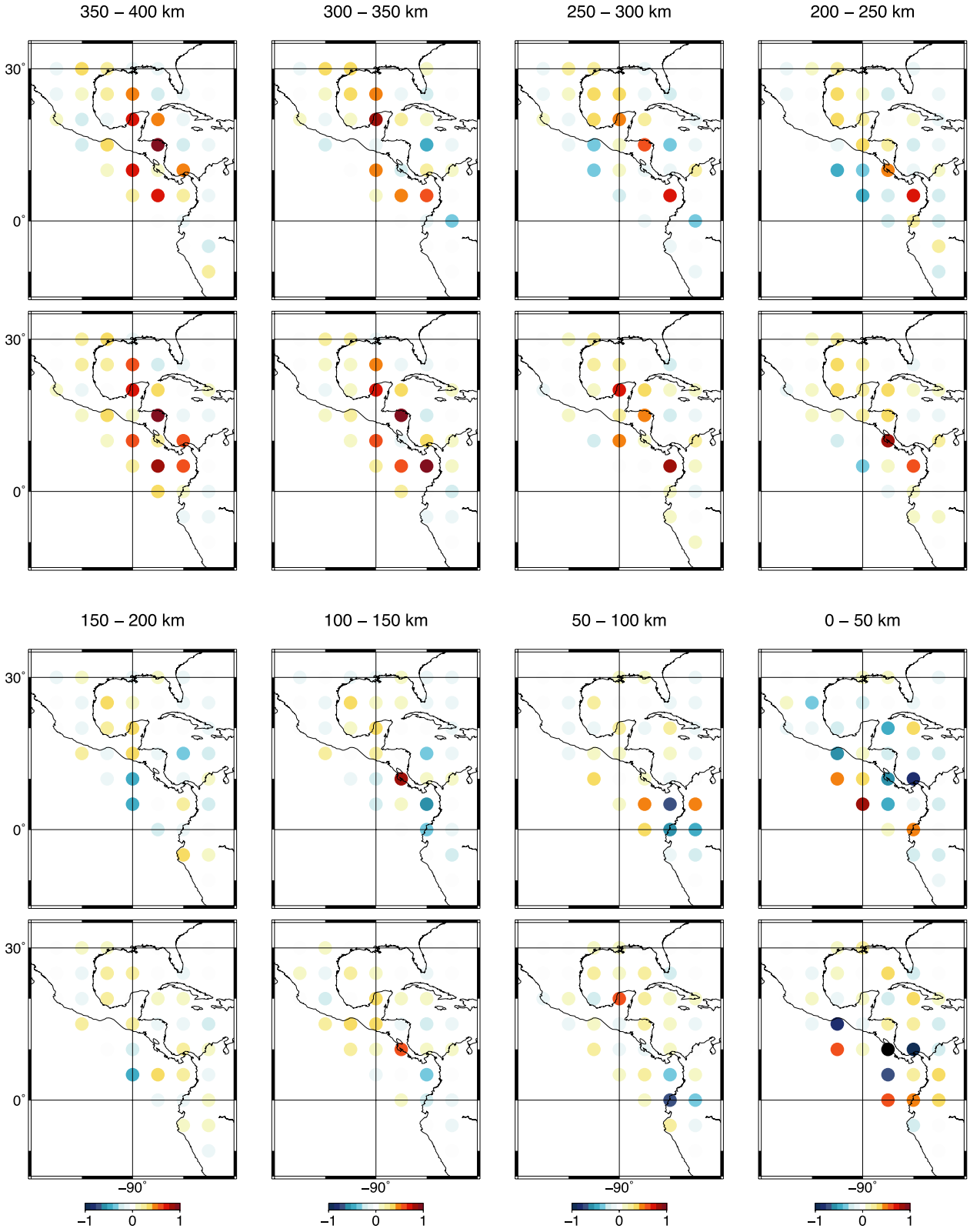

Figure 28 – (*Continued*)

the subducted slab is plausible, as it is just 200 K colder than the mantle adiabat of about 2500 K (Brown & Shankland 1981).

It is difficult to detect obvious evidence for the phase transition from Mg-perovskite (*pv*) to post-perovskite (*ppv*) in our model, perhaps because our inversion method is parametrized to infer a relatively smooth model rather than a model with sharp discontinuities. However, in the depth range from 200–400 km above the CMB, the fastest anomaly is found in the depth range 350–400 km above the CMB and the magnitude of the anomaly decreases as the depth increases to 300 km above the CMB. This suggests that

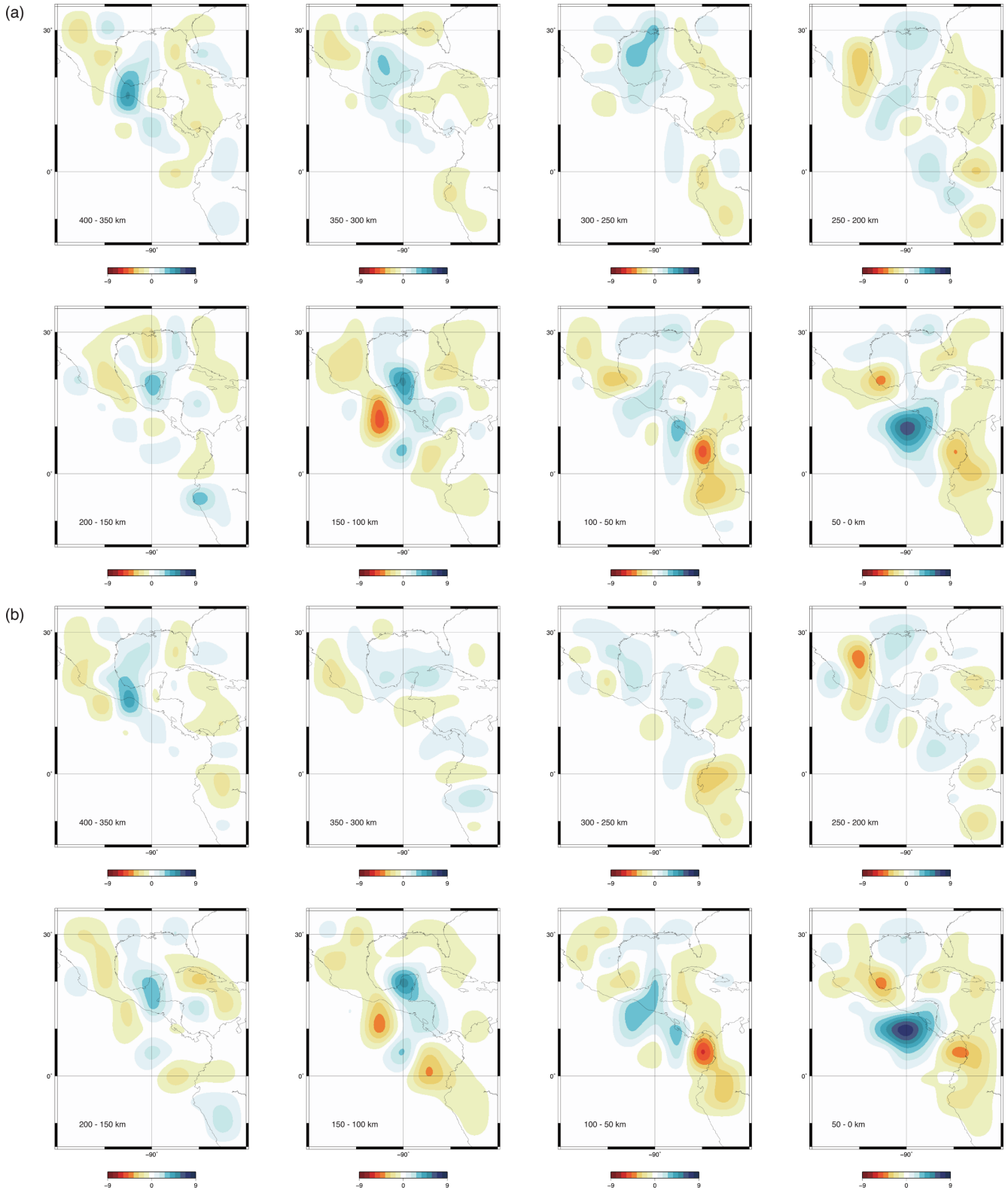

Figure 29. Results of the three jack-knife inversions (CG5) for the 3-D shear wave velocity structure in the lowermost mantle beneath Central America for each depth range. Each model (a and b, this page, and c, next page) was inferred from a randomly chosen data set consisting of 1000 waveforms.

the *pv–ppv* phase transition occurs over a depth range of at most about 100 km. Also the fact that the fast anomaly is concentrated at depths of 250–400 km above the CMB may also be indicative of the *pv–ppv* phase transition. This sheet-like high-velocity region is 250 km wide (E–W) and 250 km long (N–S).

As previous tomographic studies (Grand 2002) reported that the fast shear velocity anomaly which is interpreted as being the Farallon slab seems to penetrate the 660 km discontinuity, the length of the ancient subduction region has been thought to be 1500 km. This is comparable to the present trench-parallel Farallon slab width,

(c)

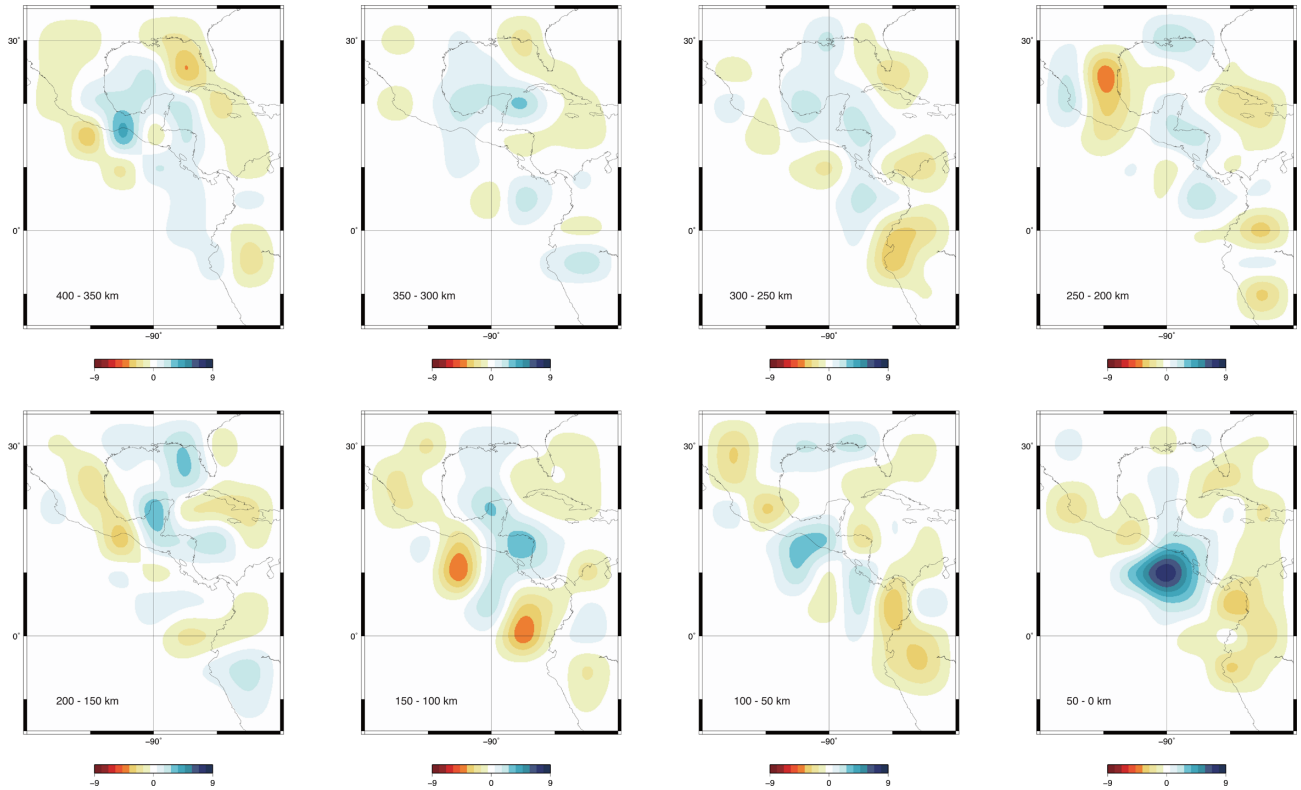

Figure 29 – (*Continued*)

although the width before 30 Ma is estimated to have been much larger than at present (Schellart *et al.* 2009). The width (250 km) seems to be two or three times greater than that of typical lithosphere. This could be because the sinking slab bulged out due to a slowing of its sinking speed in the more viscous lower mantle (Bunge & Grand 2000). The sheet-like fast velocity anomaly reported in this study seems to be isolated, while many previous tomographic studies (e.g. Grand 2002) reported a broader fast anomaly beneath North America and South America, which might imply the existence of other fast anomalies due to subduction of older slabs over the 100 Myr subduction history of the Farallon Plate. Such anomalies were not detected by this study, which might reflect events inducing subduction direction changes such as the birth of the Cocos Plate and the splitting of the Farallon and Nazca plates (Bunge & Grand 2000; Schellart *et al.* 2009).

6 CONCLUSION

Localized waveform inversion is well suited for studying fine structure using data from recently deployed arrays such as US-Array and Hi-net. In this study, we used data from a dense station network along the Pacific in western North America. As the US-Array is currently being extended eastwards, the resolution of future studies will be higher than that of this study.

ACKNOWLEDGEMENTS

Data were obtained from IRIS and CNSN data servers. This research was supported by grants from the JSPS (Nos. 24840020 and 25400442) and the Korean government (MEST, No. 2009-0092790).

REFERENCES

Akaike, H., 1977. Extension of method of maximum likelihood and the Stein's problem, *Ann. Inst. Stat. Math.,* **29,** 153–164.

Beckman, F.S., 1960. The solution of linear equations by the conjugate gradient method, in *Mathematical Methods for Digital Computers,* Vol. 1, pp. 62–72, eds Ralston, A. & Wilf, H.S., Wiley.

Brown, J.M. & Shankland, T.J., 1981. Thermodynamic parameters in the Earth as determined from seismic profiles, *Geophys. J. R. astr. Soc.,* **66,** 579–596.

Bunge, H.-P. & Grand, S.P., 2000. Mesozoic plate-motion history below the northeast Pacific Ocean from seismic images of the subducted Farallon slab, *Nature,* **405,** 337–340.

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

Cummins, P.R., Takeuchi, N. & Geller, R.J., 1997. Computation of complete synthetic seismograms for laterally heterogeneous models using the Direct Solution Method, *Geophys. J. Int.,* **130,** 1–16.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton Univ. Press.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187,** 1604–1624.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental and global scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.,* **175,** 665–685.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2009. Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179,** 1703–1725.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle, *Earth planet. Sci. Lett.,* **290,** 270–280.

Fuji, N., 2010. A methodology for inversion of seismic waveforms for elastic and anelastic structure and its preliminary application to the mantle transition zone beneath the Northwestern Pacific, *PhD thesis,* University of Tokyo.

Fuji, N., Kawai, K. & Geller, R.J., 2010. A methodology for inversion of broadband seismic waveforms for elastic and anelastic structure and its application to the mantle transition zone beneath the Northwestern Pacific, *Phys. Earth planet. Inter.,* **180,** 118–137.

Fuji, N., Chevrot, S., Zhao, L., Geller, R.J. & Kawai, K., 2012. Finite-frequency structural sensitivities of short-period compressional body waves, *Geophys. J. Int.,* **190,** 522–540.

Geller, R.J. & Hara, T., 1993. Two efficient algorithms for iterative linearized inversion of seismic waveform data, *Geophys. J. Int.,* **115,** 699–710.

Geller, R.J. & Ohminato, T., 1994. Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the Direct Solution Method, *Geophys. J. Int.,* **116,** 421–446.

Grand, S.P., 2002. Mantle shear-wave tomography and the fate of subducted slabs, *Phil. Trans. R. Soc. A,* **360,** 2475–2491.

Hara, T., Tsuboi, S. & Geller, R.J., 1993. Inversion for laterally heterogeneous upper mantle S-wave velocity structure using iterative waveform inversion, *Geophys. J. Int.,* **115,** 667–698.

Hung, S.-H., Garnero, E.J., Chiao, L.-Y., Kuo, B.-Y. & Lay, T., 2005. Finite frequency tomography of D′′ shear velocity heterogeneity beneath the Caribbean, *J. geophys. Res.,* **110,** B07305, doi:10.1029/2004JB003373.

Hutko, A.R., Lay, T., Garnero, E.J. & Revenaugh, J., 2006. Seismic detection of folded, subducted lithosphere at the core-mantle boundary, *Nature,* **441,** 333–336.

Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y., 1990. Whole mantle P-wave travel time tomography, *Phys. Earth planet. Inter.,* **59,** 294–328.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. astr. Soc.,* **28,** 97–109.

Kawai, K. & Geller, R.J., 2010a. Waveform inversion for localized seismic structure and an application to D′′ structure beneath the Pacific, *J. geophys. Res.,* **115,** B01305, doi:10.1029/2009JB006503.

Kawai, K. & Geller, R.J., 2010b. The vertical flow in the lowermost mantle beneath the Pacific from inversion of seismic waveforms for anisotropic structure, *Earth planet. Sci. Lett.,* **297,** 190–198.

Kawai, K. & Tsuchiya, T., 2009. Temperature profile in the lowermost mantle from seismological and mineral physics joint modeling, *Proc. Natl. Acad. Sci. U.S.A.* **106,** 22 119–22 123.

Kawai, K., Takeuchi, N. & Geller, R.J., 2006. Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media, *Geophys. J. Int.,* **164,** 411–424.

Kawai, K., Takeuchi, N., Geller, R.J. & Fuji, N., 2007. Possible evidence for a double crossing phase transition in D′′ beneath Central America from inversion of seismic waveforms, *Geophys. Res. Lett.,* **34,** L09314, doi:10.1029/2007GL029642.

Konishi, K., Kawai, K., Geller, R.J. & Fuji, N., 2012. Waveform inversion of broad-band body-wave data for the S-velocity structure in the lowermost mantle beneath India, *Geophys. J. Int.,* **191,** 305–316.

Lanczos, C., 1961. *Linear Differential Operators,* Van Nostrand.

Lawson, C.L. & Hanson, R.J., 1974. *Solving Least Squares Problems,* Prentice-Hall.

Lévêque, J.-J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.,* **115,** 313–318.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: theory and verification in a physical scale model, *Geophysics,* **64,** 888–901.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O.I. & Kommedal, J.H., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophys. J. Int.,* **187,** 1495–1515.

Reeves, C.M., 1964. Algorithm-238: Conjugate gradient method, *Comm. ACM,* **7,** 481.

Saito, M., 1978. An automatic design algorithm for band selective recursive digital filters, *Butsuri Tanko,* **31,** 112–135 (in Japanese).

Schellart, W.P., Stegman, D.R., Farrington, R.J., Freeman, J. & Moresi, L., 2009. Cenozoic tectonics of western north America controlled by evolving width of Farallon slab, *Nature,* **329,** 316–319.

Shanks, J.L., 1967. Recursion filters for digital processing, *Geophysics,* **32,** 33–51.

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

Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U. & Kommedal, J.H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall, *First Break,* **28**(4), 65–70.

Takeuchi, N., 2012. Detection of ridge-like structures in the Pacific large low-shear-velocity province, *Earth planet. Sci. Lett.,* **319–320,** 55–64.

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

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

Tarantola, A., 1984. Inversion of seismic-reflection data in the acoustic approximation, *Geophysics,* **49,** 1259–1266.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics,* **51,** 1893–1903.

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

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

Wiggins, R.A., 1972. The general linear inverse problem implication of surface waves and free oscillation for Earth structure, *Rev. Geophys. Space Phys.,* **10,** 251–285.

APPENDIX A: INVERSE PROBLEM FORMULATION

We define $\delta \boldsymbol{d}$ to be the residual (the difference of the observed seismogram, $\boldsymbol{d}_{\rm OBS}$, and the synthetic seismogram for the initial model, $\boldsymbol{d}_{\rm INIT}$):

$$\delta \boldsymbol{d} = \boldsymbol{d}_{\rm OBS} - \boldsymbol{d}_{\rm INIT}. \tag{A1}$$

The inverse problem is usually written as:

$$\mathbf{A}\delta \boldsymbol{m} = \delta \boldsymbol{d}, \tag{A2}$$

where $\mathbf{A}$ is the $N \times M$ matrix of partial derivatives, N is the number of observations, M is the number of model parameters and $\delta \boldsymbol{m}$ is the perturbation to the initial model. Since the number of unknowns (the number of elements of $\delta \boldsymbol{m}$) is usually far smaller than the number of data points (number of elements of $\delta \boldsymbol{d}$) in geophysics ($M < N$), it is well known that eq. (A2) implies the following minimization rather than strict equality:

$$|\mathbf{A}\delta \boldsymbol{m} - \delta \boldsymbol{d}|^2 = \text{minimum}. \tag{A3}$$

We now use the above results to obtain the least-squares solution of eq. (A3) by solving the normal equations:

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\delta \boldsymbol{m} = \mathbf{A}^{\mathrm{T}}\delta \boldsymbol{d}. \tag{A4}$$

A1 DLS

The normal equations using DLS can be written as:

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varepsilon^2\mathbf{I}\right)\delta \boldsymbol{m} = \mathbf{A}^{\mathrm{T}}\delta \boldsymbol{d}, \tag{A5}$$

where ε is a damping factor and $\mathbf{I}$ is the identity matrix. When weighted inversion is conducted, this equation can be written using a diagonal weighting matrix whose explicit elements are $W_{ij} = w_i \delta_{ij}$, where $\sqrt{w_i}$ is the weighting factor for the i-th data bin $\boldsymbol{d}_i$,

$$\left(\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A} + \varepsilon^2\mathbf{I}\right)\delta\boldsymbol{m} = \mathbf{A}^{\mathbf{T}}\mathbf{W}\delta\boldsymbol{d}. \tag{A6}$$

By replacing the identity matrix by another matrix, one can impose other constraints such as smoothness (e.g. Inoue *et al.* 1990). Kawai *et al.* (2007) inverted for seismic structure using eq. (A5).

A2 SVD

We can decompose an $M \times N$ matrix $\mathbf{A}$ using the SVD (e.g. Lanczos 1961):

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}^{\mathbf{T}}, \tag{A7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary $N \times N$ and $M \times M$ matrices, and the non-zero eigenvalues are ordered as follows:

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_M^2 > 0. \tag{A8}$$

The unitary $M \times M$ matrix $\mathbf{V}$ contains the right eigenvectors of $\mathbf{A}$:

$$\mathbf{V} = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_M] \tag{A9}$$

normalized to satisfy:

$$\boldsymbol{v}_j^{\mathbf{T}}\boldsymbol{v}_l = \delta_{jl}, \tag{A10}$$

and the $N \times M$ matrix ($M < N$) $\mathbf{\Lambda}$ contains the eigenvalues:

$$\Lambda = \begin{pmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_M \\ & & \\ & \mathbf{0} & \end{pmatrix}. \tag{A11}$$

For later use we define the matrix consisting of the first L right eigenvectors as follows:

$$\mathbf{V}_L = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_L]. \tag{A12}$$

In this section, we assume that the problem is overdetermined, that is, that $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is positive definite and that all of the λ_i^2 in eq. (A7) are thus positive. From eq. (A7) we have

$$\mathbf{A}^{\mathbf{T}}\mathbf{A} = \mathbf{V}\Lambda^{\mathbf{T}}\mathbf{U}^{\mathbf{T}}\mathbf{U}\Lambda\mathbf{V}^{\mathbf{T}} = \mathbf{V}\Lambda^{\mathbf{T}}\Lambda\mathbf{V}^{\mathbf{T}}, \tag{A13}$$

because $\mathbf{U}$ is unitary.

From eqs (A10) and (A13) we have

$$\boldsymbol{v}_j^{\mathbf{T}}(\mathbf{A}^{\mathbf{T}}\mathbf{A})\boldsymbol{v}_k = \lambda_k^2\delta_{jk}. \tag{A14}$$

We write the solution of eq. (A4) as an eigenfunction expansion:

$$\delta\boldsymbol{m} = \sum_{j=1}^{M} e_j\boldsymbol{v}_j = \mathbf{V}\boldsymbol{e}, \tag{A15}$$

where the expansion coefficients e_j are the unknowns, and the vector $\boldsymbol{e}$ is defined to be:

$$\boldsymbol{e}^{\mathbf{T}} = (e_1\ e_2\ \cdots\ e_M). \tag{A16}$$

We substitute eq. (A15) into eq. (A4), multiply both sides of the resulting equation on the left by $\boldsymbol{v}_j^{\mathbf{T}}$, and use eq. (A14) to solve for each expansion coefficient of eq. (A15) as follows:

$$e_j = \frac{\boldsymbol{v}_j^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\mathbf{d}}{\lambda_j^2}. \tag{A17}$$

We can write the contribution of the i-th eigenvector to the data space as:

$$\delta\boldsymbol{s}_i = \mathbf{A}\delta\boldsymbol{m}_i = \mathbf{A}e_i\boldsymbol{v}_i = e_i\lambda_i\boldsymbol{u}_i. \tag{A18}$$

We write eq. (A17) in matrix form as follows:

$$\boldsymbol{e} = (\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{V}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d}. \tag{A19}$$

As we assume that the data errors are uncorrelated and equal, the data covariance matrix is a constant times the identity matrix. Hence, the covariance matrix is written as:

$$\begin{aligned}\text{covar.}\,(\delta\boldsymbol{m}) &= \sigma_d^2\,(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1} \\ &= \sigma_d^2\sum_{j=1}^{M}\frac{\boldsymbol{v}_j\boldsymbol{v}_j^{\mathbf{T}}}{\lambda_j^2} \\ &= \sigma_d^2\,\mathbf{V}(\mathbf{\Lambda}^T\mathbf{\Lambda})^{-1}\mathbf{V}^{\mathbf{T}}.\end{aligned} \tag{A20}$$

$|\lambda_i|$ becomes smaller as i increases in eq. (A20). When $|\lambda_i|$ is close to zero, the variance of the solution increases. On the other hand, when the eigenvectors of only the largest L eigenvalues are used and other eigenvectors are omitted, the solution error can be suppressed (Wiggins 1972; Jackson 1972; Menke 1984). As discussed in section 3.4, we use AIC (Akaike 1977) as the criterion for the cut-off of the SVD expansion. AIC is a measure that rewards variance reduction and penalizes increases in the number of model parameters. The model with the minimum AIC is the model which nominally gives the optimum trade-off between variance reduction and model complexity.

A3 CG

Suppose that $\{\boldsymbol{p}_j = \boldsymbol{p}_1, \boldsymbol{p}_2, \cdots\}$ is a sequence of M mutually orthogonal conjugate vectors with respect to $\mathbf{A}^{\mathbf{T}}\mathbf{A}$; that is,

$$\boldsymbol{p}_j^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\boldsymbol{p}_k = 0 \qquad \text{for} \qquad j \neq k. \tag{A21}$$

We write the solution of eq. (A4) as an expansion in terms of the conjugate vectors:

$$\delta\boldsymbol{m} = \sum_{j=1}^{M} a_j\boldsymbol{p}_j, \tag{A22}$$

where the expansion coefficients a_j are the unknowns. We define the vector $\boldsymbol{a}$ to be:

$$\boldsymbol{a}^{\mathbf{T}} = (a_1\ a_2\ \cdots\ a_M). \tag{A23}$$

Eq. (A22) can be written as follows in matrix form:

$$\delta\boldsymbol{m} = \mathbf{P}\boldsymbol{a}, \tag{A24}$$

where we define the matrix $\mathbf{P}$ to be:

$$\mathbf{P} = (\boldsymbol{p}_1\ \boldsymbol{p}_2\ \cdots\ \boldsymbol{p}_M). \tag{A25}$$

Substituting eq. (A22) into eq. (A4), multiplying both sides on the left-hand side by the transpose of some particular $\boldsymbol{p}_k$, and using eq. (A21), we obtain:

$$\boldsymbol{p}_k^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d} = a_k\boldsymbol{p}_k^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\boldsymbol{p}_k. \tag{A26}$$

The k-th expansion coefficient is thus given by:

$$a_k = \frac{\boldsymbol{p}_k^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d}}{\boldsymbol{p}_k^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\boldsymbol{p}_k}. \tag{A27}$$

Using eq. (A23) we can write this in matrix form as follows:

$$\boldsymbol{a} = \mathbf{L}^{-1}\mathbf{P}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\boldsymbol{d}, \tag{A28}$$

where we define the $M \times M$ diagonal matrix $\mathbf{L}$ as follows:

$$\mathbf{L} = \begin{pmatrix} \boldsymbol{p}_1^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{p}_M^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_M \end{pmatrix}. \tag{A29}$$

Assuming that the data are uncorrelated and all have equal variance σ_d^2, we can write the covariance matrix for a as follows in the CG representation of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$:

$$\text{covar.}\,(\boldsymbol{a}) = \sigma_d^2 \mathbf{L}^{-1}, \tag{A30}$$

while in the model space, the covariance matrix is written as:

$$\text{covar.}\,(\delta\mathbf{m}) = \sigma_d^2(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} = \sigma_d^2 \mathbf{P}\mathbf{L}^{-1}\mathbf{P}^{\mathrm{T}}. \tag{A31}$$

In the CG inversions we sum over only the first n (where $n \leq M$) conjugate vectors as follows:

$$\delta\boldsymbol{m}_n = \sum_{j=1}^{n} a_j \boldsymbol{p}_j. \tag{A32}$$

Thus the covariance matrix in the model space is given by:

$$\text{covar.}\,(\delta\boldsymbol{m}_n) = \sigma_d^2 \sum_{j=1}^{n} \frac{\boldsymbol{p}_j \boldsymbol{p}_j^{\mathrm{T}}}{\boldsymbol{p}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_j}. \tag{A33}$$

While eq. (A20) clearly shows that the solution error of the contribution to the data space of the i-th eigenvector becomes larger as i increases, this is not clear in eq. (A33). We discuss this in section 3.9 using the actual values.

In order to minimize the data variance as rapidly as possible, the first conjugate vector is defined to be the gradient vector:

$$\boldsymbol{p}_1 = -\mathbf{A}^{\mathrm{T}}\delta\boldsymbol{d}. \tag{A34}$$

Each successive conjugate vector is then chosen to be as close as possible to the gradient vector while still being conjugate. We follow Reeves (1964) below. Let $\boldsymbol{r}_k$ be the residual (in the model space) at the k-th iteration:

$$\boldsymbol{r}_k = \mathbf{A}^{\mathrm{T}}\delta\boldsymbol{d} - \mathbf{A}^{\mathrm{T}}\mathbf{A}\delta\boldsymbol{m}_k. \tag{A35}$$

We then choose the next basis vector $\boldsymbol{p}_{k+1}$ to be as close as possible to $\boldsymbol{r}_k$, while also being conjugate to all previous basis vectors:

$$\boldsymbol{p}_{k+1} = \boldsymbol{r}_k - \sum_{i=1}^{k} \frac{\boldsymbol{p}_i^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{r}_k}{\boldsymbol{p}_i^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_i}\boldsymbol{p}_i. \tag{A36}$$

We then calculate the next expansion coefficient as follows:

$$a_{k+1} = \frac{\boldsymbol{p}_{k+1}^{\mathrm{T}}\mathbf{r}_k}{\boldsymbol{p}_{k+1}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_{k+1}}. \tag{A37}$$

As the CG basis is problem-specific, it converges using a smaller number of independent parameters (expansion coefficients of the basis vectors) than is required for SVD inversion.

We truncate the conjugate vector expansion (eq. A32), using AIC. The variance VAR_n and AIC_n for the model CG_n (the model obtained using n in eq. A32) are written, respectively, as follows:

$$\mathrm{VAR}_n = \frac{|\boldsymbol{d}^n_{\mathrm{OBS}} - \boldsymbol{d}_{\mathrm{SYN}}|^2}{|\boldsymbol{d}^n_{\mathrm{OBS}}|^2}, \tag{A38}$$

$$\mathrm{AIC}_n = \mathrm{ND}\ln 2\pi + \mathrm{ND}\ln(\mathrm{VAR}_n) + \mathrm{ND} + 2(n+1), \tag{A39}$$

where ND is the number of independent data, and $\boldsymbol{d}^n_{\mathrm{SYN}}$ is the vector of synthetic seismograms including the contribution of the first n basis vectors of the CG basis. We assume here that we have K events and P stations with records available at all stations for all events. We also assume here that each time-series has T points. In this study the upper bound on the value of ND is $\mathrm{ND}^1 = \mathrm{KPT}/12.5$ because we use 12.5–200 s bandpass-filtered waveforms, interpolated at a 1 Hz sampling frequency (the meaning of the superscript is explained later). It is reasonable to consider that the data are, to some extent, redundant, but at present we have no rigorous way to quantify the extent of the redundancy. In this study we use the empirical parameter α which we define as follows:

$$\mathrm{ND}^{\alpha} = \frac{1}{\alpha}\mathrm{KPT}/12.5. \tag{A40}$$

From eq. (A33) we can rewrite the variance for each CG expansion coefficient a_j as follows:

$$\text{var.}\,(a_j) = \frac{\sigma_d^2}{\boldsymbol{p}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_j}. \tag{A41}$$

If we assume that σ_d is the variance of the data itself, we can estimate the nominal error bars δa_j for each model parameter as follows:

$$\delta a_j = \frac{\sigma_d}{\sqrt{\boldsymbol{p}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_j}}. \tag{A42}$$

APPENDIX B: RECURSIVE FILTER

As we design the filter following Saito (1978), which is written in Japanese, we summarize this method here. The bandpass filter is characterized by five parameters (Fig. B1): A_P, A_S, f_H, f_L and f_S, which are parameters related to amplitudes for the passband and stop band, higher and lower frequencies for passband and higher frequency for stop band, respectively. The frequency response for the Butterworth filter is:

$$|B_n(\sigma)|^2 = \frac{1}{1+\sigma^{2n}}, \tag{B1}$$

where σ is non-dimensional angular frequency. The passband is defined as: $|\sigma| \leq \sigma_P$ and the stop band is defined as $|\sigma| \geq \sigma_S$. For a finite number of poles n, we design the response function as:

$$\frac{1}{1+A_P^2} \leq |B_n(\sigma)|^2 \leq 1 \quad |\sigma| \leq \sigma_P < \sigma_S$$

$$|B_n(\sigma)|^2 \leq \frac{1}{1+A_S^2} \qquad \sigma_S \leq |\sigma|. \tag{B2}$$

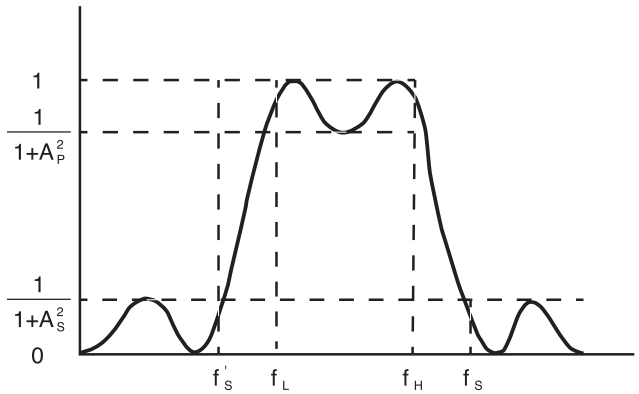

Figure B1. Schematic image drawn after Saito (1978) to explain the Butterworth bandpass filter, which is characterized by the five independent parameters: A_P, A_S, f_H, f_L, and, f_S.

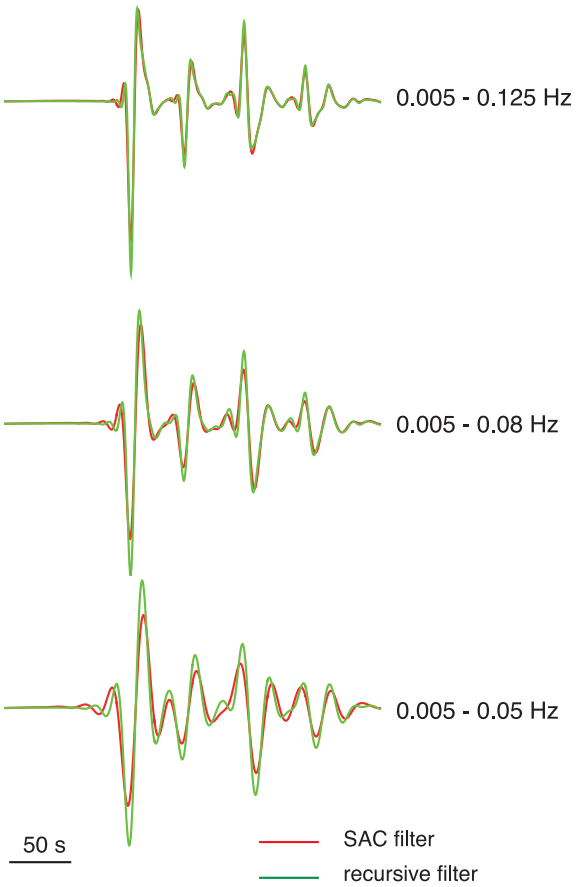

Figure B2. Comparisons between waveforms filtered by SAC (Seismic Analysis Code) and those filtered by our filter (which follows Saito 1978) for the three passbands (0.005–0.125, 0.005–0.08, 0.005–0.05 Hz).

In order to satisfy eq. (B2),

$$n \geq \frac{\log(A_S/A_P)}{\log(\sigma_S/\sigma_P)}. \tag{B3}$$

The minimum integer n satisfying eq. (B3) can be obtained. In order to design a bandpass filter, we transform σ as:

$$\sigma = \frac{\lambda^2 - \lambda_0^2}{\lambda}, \tag{B4}$$

where λ is dimensionless angular frequency. When the roots of λ for $\sigma = \sigma_P$ are defined as λ_H and $-\lambda_L$, $|\sigma| \leq \sigma_P$ is transformed to $\lambda_L \leq |\lambda| \leq \lambda_H$. Also, when the roots of λ for $\sigma = \sigma_S$ are defined as λ_S and $-\lambda'_S$, $|\sigma| \geq \sigma_S$ is transformed to $|\lambda| \geq \lambda_S$ and $|\lambda| \leq \lambda'_S$. We can also obtain the following relations:

$$\sigma_P = \lambda_H - \lambda_L,$$

$$\sigma_S = \lambda_S - \lambda'_S,$$

$$\lambda_0^2 = \lambda_H \lambda_L = \lambda_S \lambda'_S\,. \tag{B5}$$

We then transform eq. (B2) using the following bilinear transformation:

$$\lambda = \frac{c}{i}\frac{1-z}{1+z} \quad c > 0\,, \tag{B6}$$

where $\lambda = c\,\tan\omega/2$. We can write the quantities in eq. (B5) in terms of ω as:

$$\sigma_P = c(\tan\omega_H/2 - \tan\omega_L/2),$$

$$\sigma_S = c(\tan\omega_S/2 - \tan\omega'_S/2),$$

$$\lambda_0^2 = c^2 \tan\omega_H/2 \tan\omega_L/2 = c^2 \tan\omega_S/2 \tan\omega'_S/2\,. \tag{B7}$$

c satisfies the following relation:

$$c^2 = (A_P A_S)^{1/n}(\tan\omega_H/2 - \tan\omega_L/2)^{-1} \cdot|\tan\omega_S/2 - \tan\omega_H/2 \tan\omega_L/2 \cot\omega_S/2|^{-1}\,. \tag{B8}$$

Using angular frequencies of f_H, f_L, and f_S for the bandpass filter, we write, after Saito (1978):

$$\frac{\sigma_S}{\sigma_P} = \frac{|\tan(\omega_S/2) - \tan(\omega_H/2)\tan(\omega_L/2)\cot(\omega_S/2)|}{\tan(\omega_H/2) - \tan(\omega_L/2)}. \tag{B9}$$

While Saito (1978) uses A_P, A_S, f_H, f_L and f_S as five free parameters, we use A_P, A_S, f_H, f_L and n. Also, we set $1/(1+A_P^2)$ and $1/(1+A_S^2)$ to 0.9 and 0.1, respectively. Hence, in our software three parameters (f_H, f_L, n) should be given. In this study, for example, with f_H, f_L and n of 0.08, 0.005 Hz and 4, we obtain f_S of 0.133 and f'_S of 0.003 Hz.

We compare waveforms filtered by our bandpass filter to those filtered by SAC (Fig. B2). Since the SAC filters are proprietary, there may be some small difference between the SAC filter and our filter. This cannot be confirmed but, if so, it would probably be due to differences in the definition of n.