Geophys. J. Int. (2004) **158,** 681–694 doi: 10.1111/j.1365-246X.2004.02334.x

Improvement of seismological earth models by using data weighting in waveform inversion

Nozomu Takeuchi[1,2] and Minoru Kobayashi[3]

[1] *Earthquake Research Institute, University of Tokyo, Yayoi* 1-1-1, *Bunkyo-ku, Tokyo* 113-0032, *Japan. E-mail: takeuchi@eri.u-tokyo.ac.jp*
[2] *Seismological Laboratory, University of California at Berkeley* 215 *McCone Hall, Berkeley,* CA 94720-4760, *USA. E-mail: takeuchi@seismo.berkeley.edu*
[3] *Institute of Space Technology and Aeronautics, Japan Aerospace Exploration Agency, Jindaiji Higashi-machi* 7-44-1, *Chofu-shi, Tokyo* 182-8522, *Japan. E-mail: mkoba@chofu.jaxa.jp*

Accepted 2004 April 16. Received 2004 April 9; in original form 2003 May 12

SUMMARY

We present a new method for data weighting in waveform inversion to improve the obtained seismological earth model. We define 'sensitivity' as an index to show the homogeneity of the waveform inversion data set. The sensitivity is directly evaluated from the partial derivatives of the synthetic seismograms and can be evaluated regardless of the method used to compute the synthetics. In this study we use the Direct Solution Method, which is a highly suitable method because of its accuracy and efficiency. If we weight the data so that the sensitivity is more or less homogeneous throughout the whole Earth, we can achieve homogeneous resolution for the entire region. The obtained model is an improvement over models obtained by previous inversion schemes, in the sense that the new model efficiently retrieves the robust information in the data. The inversion of a large data set using our methods can be expected to lead to further improvement of seismological earth models.

Key words: direct solution method, Earth's interior, inverse problem, synthetic seismograms.

1 INTRODUCTION

Seismological earth models have until now primarily been obtained by inversion of secondary data which are extracted from observed seismic waveform data, such as the traveltimes of body waves, the phase velocity of surface waves and splitting functions of modes of free oscillations (e.g. van der Hilst *et al.* 1997; Grand *et al.* 1997; Ritsema *et al.* 1999; Masters *et al.* 2000). Direct inversion of seismic waveforms, which utilizes all of the information contained in the data, has the potential to yield better earth models. In recent years waveform inversion for detailed 3-D Earth structure has become feasible (e.g. Megnin & Romanowicz 2000; Hara 2002), as has joint inversion of waveform data and secondary data (e.g. Gu *et al.* 2001). Further improvement can be expected by using more sophisticated forward modelling and inversion methods.

During the past decade our research group at the University of Tokyo has made significant progress in developing an accurate and efficient method for computing synthetic seismograms. Our basic approach, the Direct Solution Method (DSM) (Geller & Ohminato 1994), is a Galerkin weak form method that achieves both accuracy and computational efficiency when calculations are made using optimally accurate numerical operators (Geller & Takeuchi 1995). Our forward modelling software both for spherically symmetric earth models (Cummins *et al.* 1994; Takeuchi *et al.* 1996) and 3-D heterogeneous Earth models in spherical coordinates (Cummins *et al.* 1997; Takeuchi *et al.* 2000) can compute synthetic seismograms which are about 30 times more accurate than those computed by conventional algorithms using the same amount of CPU time. This allows rigorous computation of the partial derivatives of the synthetic seismograms without any approximations other than discretization. In contrast, most previous waveform inversion studies incorporate further approximations such as path-average approximation (Woodhouse & Dziewonski 1984) or non-linear asymptotic coupling theory (Li & Romanowicz 1995).

We have also developed efficient algorithms of waveform inversion for 3-D Earth structure and earthquake source parameters (Geller & Hara 1993; Hara 1997). These efficient algorithms make it feasible to perform iterative linearized waveform inversion for a 3-D heterogeneous initial model. The computational requirements for such iterative linearized inversion of surface wave data are relatively small, and our group has successfully carried out such inversions (Hara & Geller 2000). In contrast, waveform inversion of complete seismograms (including both surface waves and body waves) is much more computationally intensive, and is only now becoming feasible (see the Discussion in Takeuchi *et al.* 2000).

If it were possible to invert a completely homogeneous waveform data set, recorded by evenly distributed stations for evenly distributed events with evenly distributed magnitudes, we could obtain an ideal earth model. However, in practice, the actual data set is far from homogeneous, so the use of some data weighting is required in order for the inversion to simulate the results that would be obtained if an ideal data set were available. In this paper we develop a new weighting scheme by defining an index to show the

homogeneity of the waveform data set. Various studies (e.g. Pulliam & Stark 1993; Trampert & Snieder 1996) have pointed out that homogeneous sampling is one of the keys to obtaining a reliable earth model. For inversions of phase data such as delay time tomography, path coverage can be used as a weighting index to achieve homogeneity. However, for waveform inversion this is neither possible nor appropriate for the following reasons:

(1) Waveform inversion utilizes amplitude information as well as phase information, but path coverage cannot take amplitude effects (e.g. the effect of source magnitude and mechanism) into account.

(2) Waveform inversion rigorously considers finite wavelength (Fresnel zone) effects, but path coverage cannot take this into account.

(3) Waveform inversion utilizes all phases in the waveform data, but it is almost impossible to draw paths for all phases included in the data.

In this paper, we show that the 'sensitivity' of the waveform data set is a useful index that is easy to evaluate. Many studies, including Li & Tanimoto (1993), Zhao *et al.* (2000) and Dahlen *et al.* (2000), have computed sensitivity kernels for particular data. However, the sensitivity as defined in this paper is represented as the sum of the sensitivity kernels for the entire data set (or an arbitrary subset thereof). We will show that weighting the data to homogenize the sensitivity improves the robustness of the inversion. The reasons for this improvement are also discussed.

2 METHOD

2.1 Definition of sensitivity

Our definition of sensitivity is general and can be applied not only to waveform inversion but also to other types of geophysical inverse problems. In a geophysical inverse problem we first formulate the expected relation between the data set $\mathbf{d}$ and model parameters $\mathbf{m}$:

$$\mathbf{d} = f(\mathbf{m}), \tag{1}$$

where f is a function specifying the assumed dependence of the data set on the model.

If the model $\mathbf{m}$ is perturbed by a small amount $\delta\mathbf{m}$, the expected data will also be perturbed. The expected perturbation to the data $\delta\mathbf{d}$ can be linearly approximated as follows:

$$\delta\mathbf{d} = \frac{\partial f}{\partial \mathbf{m}}\delta\mathbf{m}. \tag{2}$$

The sensitivity for a particular perturbation to the model, $\delta\mathbf{p}$, $S(\delta\mathbf{p})$, is defined by the squared norm of $\delta\mathbf{d}$:

$$S(\delta\mathbf{p}) = \left|\frac{\partial f}{\partial \mathbf{m}}\delta\mathbf{p}\right|^2. \tag{3}$$

The sensitivity defined in eq. (3) can be readily evaluated from the partial derivatives $\partial f/\partial\mathbf{m}$ and is applicable to waveform inversion in a straightforward manner.

2.2 Explicit formulation of sensitivity

The definition of sensitivity, eq. (3), is general and formal. However, if we appropriately choose $\delta\mathbf{p}$ in eq. (3), we can define sensitivity with a physically clear meaning. We will show explicit formulations for block parametrizations and for global parametrizations. For simplicity, we consider model parameters only for the rigidity μ; however, the extension to other cases is straightforward. We represent the model parameters by basis function expansions such as:

$$\mu(\mathbf{x}) = \sum_l m_l \Phi^{(l)}(\mathbf{x}), \tag{4}$$

where the expansion coefficients m_l are the model parameters that are unknowns in the inversions, $\mathbf{x}$ is the position in the Earth and $\Phi^{(l)}$ are the basis functions for the model. (Note that the basis functions for the model and for the wavefield are in general distinct sets.)

2.2.1 *Sensitivity for block parametrizations*

If we use block-wise functions as the basis functions $\Phi^{(l)}(\mathbf{x})$ in eq. (4), we define the sensitivity for a perturbation in the rigidity of one particular block. The sensitivity of the kth block is defined for a perturbation in the rigidity

$$\delta\mu(\mathbf{x}) = \Phi^{(k)}(\mathbf{x}). \tag{5}$$

The corresponding $\delta\mathbf{p}$ in eq. (3) is

$$\delta p_{k'} = \delta_{k'k}, \tag{6}$$

where $\delta_{k'k}$ is a Kronecker delta. As we can see from eq. (5), the sensitivity depends on the block size. By homogenizing this sensitivity we will achieve higher (lower) spatial resolution for regions where smaller (larger) blocks are used. This definition can be straightforwardly extended to the case in which we use spline functions as the basis functions $\Phi^{(l)}(\mathbf{x})$ in eq. (4).

2.2.2 *Sensitivity for global parametrizations*

If we use global and orthogonal functions (e.g. spherical harmonics or Fourier basis) as the basis functions $\Phi^{(l)}(\mathbf{x})$ in eq. (4), we define the sensitivity for a point perturbation in the rigidity

$$\delta\mu(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_j), \tag{7}$$

where $\mathbf{x}_j$ is a particular point in the Earth. The corresponding $\delta\mathbf{p}$ in eq. (3) is

$$\delta p_{l'} = \left(\Phi^{(l')}(\mathbf{x}_j)\right)^*. \tag{8}$$

By homogenizing this sensitivity, we achieve a homogeneous spatial resolution which has no regional preference. In other words, these basis functions are not suitable if the goal is to achieve a higher spatial resolution for some particular region.

2.3 Sensitivity for actual waveform inversion problems

As the basis functions $\Phi^{(l)}$ in the expansion of eq. (4), many recent waveform inversion studies have used spline functions in the vertically dependent part and spherical harmonics in the horizontally dependent part:

$$\Phi^{(l)}(r, \theta, \phi) = W^{(k)}(r)Y^{(s,t)}(\theta, \phi), \tag{9}$$

where $W^{(k)}(r)$ are spline functions and $Y^{(s,t)}(\theta, \phi)$ are spherical harmonics whose angular and azimuthal order are s and t respectively, and where the superscript (l) is an index for a set of (k, s, t). For this case, we define the sensitivity of Subsection 2.2.1 for the vertically dependent part and the sensitivity of Subsection 2.2.2 for the horizontally dependent part, i.e. the sensitivity for the rigidity perturbation

$$\delta\mu(r, \theta, \phi) = W^{(k)}(r)\frac{1}{\sin\theta}\delta(\theta - \theta_j)\delta(\phi - \phi_j), \tag{10}$$

where (θ_j, ϕ_j) is a particular horizontal position. The corresponding $\delta\mathbf{p}$ in eq. (3) is as follows:

$$\delta p_{l'} = \delta p_{k's't'} = \delta_{k'k} Y^{s't'*}(\theta_j, \phi_j). \tag{11}$$

If we evaluate the sensitivity at every horizontal position for a particular vertical model parameter we obtain a sensitivity map at that depth range, which indicates the distribution of the information included in the data set. Examples of computed sensitivity maps are presented in Section 3 (Figs 1, 2 and 5a).

2.4 Weighting algorithms to homogenize sensitivity

2.4.1 Homogenization for a particular depth range

We first formulate an algorithm to homogenize the sensitivity of eqs (3) and (11) for a particular (kth) vertical model parameter. We divide the whole data set $\mathbf{d}$ into several data bins $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N$ and determine the weighting factor for each bin. For example, when the data set consists of multiple waveform data traces one possibility is to treat each trace as a data bin and change the weighting factor for each trace. Another possible strategy is to define several time windows (e.g. define one window for each phase) in each trace and treat each time window as a data bin.

We define the weighting factor $\sqrt{w_i}$ for the ith data bin $\mathbf{d}_i$ as follows:

$$\mathbf{d}'_i = \sqrt{w_i}\,\mathbf{d}_i, \tag{12}$$

where $\mathbf{d}'_i$ is a weighted data bin. Note that w_i is the weighting factor for the square norm (i.e. $|\mathbf{d}'_i|^2 = w_i|\mathbf{d}_i|^2$), and the weighting factor in eq. (12) is defined by the square root of w_i. Our goal is to determine w_i so that the total sensitivity becomes homogeneous. Note that, when $w_i = 1$ for all i, all data are unweighted.

We denote the sensitivity for the ith data bin $\mathbf{d}_i$ at the jth horizontal position as S_{ij}. From the definition of the sensitivity (eqs 2 and 3) and from the definition of the weighting factor in eq. (12), the sensitivity S'_{ij} for the weighted data bin $\mathbf{d}'_i$ can be defined as follows:

$$S'_{ij} = w_i S_{ij}. \tag{13}$$

The total sensitivity T_j for the weighted data set is defined as follows:

$$T_j = \sum_i S'_{ij} = \sum_i w_i S_{ij}. \tag{14}$$

Our primary goal in choosing the weights is to homogenize T_j over j, i.e. to minimize the total variance of T_j. Here we define the total variance $\sigma^2(T)$ as

$$\sigma^2(T) = \sum_j (T_j - \bar{T})^2, \tag{15}$$

where $\bar{T}$ is the horizontal average of the total sensitivity for the kth vertical model parameter. The explicit form of $\bar{T}$ is

$$\bar{T} = \frac{1}{J}\sum_{j'} T_{j'}, \tag{16}$$

where J is the total number of (discretized) horizontal points. Here we assume evenly discretized horizontal points. If the horizontal points are unevenly discretized, we replace eq. (16) with the appropriate discretized equation for

$$\bar{T} = \frac{1}{\Omega}\int_S T(\mathbf{x})\, dS, \tag{17}$$

where $T(\mathbf{x})$ is the total sensitivity at horizontal position $\mathbf{x}$, S is the whole horizontal surface we are considering and Ω is the total area of S.

Weighting factors that minimize $\sigma^2(T)$ in eq. (15) can greatly homogenize the total sensitivity, but they are not the most appropriate weighting factors. If some of the weighting factors are very large, the noise in the data is greatly enhanced, which degrades the obtained model (see Section 4). Thus, in determining weighting factors we need to impose conditions not only to homogenize the resultant total sensitivity but also to avoid large perturbations from the weighting factors for the no-weighting method (i.e. $w_i = 1$ for all i). We define the penalty function as follows:

$$L = \sum_j (T_j - \bar{T})^2 + \gamma \sum_i (w_i - 1)^2, \tag{18}$$

where γ is an appropriate damping factor.

We impose the following conditions for the total norm of weighting factors w_i:

$$\sum_i w_i = N, \tag{19}$$

where N is the total number of data bins. This means that the total norm of the weighting factors should be equal to that for the no-weighting method (i.e. the case for $w_i = 1$ for all i). We also impose the condition

$$w_i \geq \epsilon, \tag{20}$$

where ϵ is a small positive number, to avoid negative or zero weighting which would diminish the physical interpretation of the weighting factors. The problem of minimizing eq. (18) under the condition of eqs (19) and (20) is a quadratic least-squares problem subject to linear equality/inequality constraints.

2.4.2 Homogenization for multiple depth ranges

Next we derive an algorithm to simultaneously homogenize sensitivity for multiple vertical model parameters. Seismological data have very different sensitivity for different depths. For example, if we invert surface wave waveform data they have great sensitivity in the upper mantle but almost no sensitivity in the deep lower mantle. It is critical to make an appropriate choice of spline functions $W^{(k)}(r)$ in eq. (9) so that the sensitivity for each vertical model parameter is more or less equal for each k. We must define narrow (broad) spline functions for depth ranges in which the data set is sensitive (insensitive), as has been done in most previous studies.

Nonetheless, we often have a great divergence of sensitivity for each vertical model parameter; we therefore introduce the following special method. For the case of multiple vertical model parameters, the penalty function which is a generalization of eq. (18) is

$$\begin{aligned} L &= \sum_k \sum_j \left(T_j^{(k)} - \bar{T}\right)^2 + \gamma \sum_i (w_i - 1)^2 \\ &= \sum_k \sum_j \left(T_j^{(k)} - \frac{1}{JK}\sum_{k'}\sum_{j'} T_{j'}^{(k')}\right)^2 \\ &\quad + \gamma \sum_i (w_i - 1)^2, \end{aligned} \tag{21}$$

where $T_j^{(k)}$ is the total sensitivity for the kth vertical model parameter at the jth horizontal position, $\bar{T}$ is the averaged total sensitivity for the whole volume, and K is the total number of vertical model parameters. If we determine the weighting factors by minimizing L in eq. (21) for an inhomogeneous data set, the data weighting

scheme essentially contributes only to homogenizing the sensitivity for different vertical parameters and does not contribute well to homogenizing the sensitivity for different horizontal positions. Each data bin is sensitive only to a specific depth range and specific azimuthal range, but weighting factors are tuned to enhance or reduce the depth sensitivity without reducing the variance of the horizontal sensitivity. We have confirmed this fact by numerical experiments not presented in this paper.

To avoid this problem, we replace eq. (21) with the following penalty function:

$$L' = \sum_k \sum_j \left(T_j^{(k)} - \bar{T}^{(k)} \right)^2 + \gamma \sum_i (w_i - 1)^2$$
$$= \sum_k \sum_j \left(T_j^{(k)} - \frac{1}{J} \sum_{j'} T_{j'}^{(k)} \right)^2$$
$$+ \gamma \sum_i (w_i - 1)^2 , \qquad (22)$$

where $\bar{T}^{(k)}$ is the horizontal average of the total sensitivity of eq. (16) for the kth vertical model parameter. The first term of eq. (22) is the summation of total variance around the averaged sensitivity for 'each' vertical parameter. Because we only consider variance over j, minimizing eq. (22) guarantees homogenization in the horizontal direction. In the next section, we will show that this penalty function works well in solving real problems.

3 NUMERICAL EXPERIMENTS

3.1 Model parameters used in this section

In the numerical experiments in this section we use the following initial model and model parameters in computing the sensitivity and inverting for 3-D Earth structure. The initial model is the modified isotropic PREM (Dziewonski & Anderson 1981). We replace the *S*-velocity model for Moho–220 km depth with the *SH* velocity model (propagating perpendicular to the axis of symmetry) of the anisotropic PREM. We determine only the rigidity μ; the other structural parameters (i.e. the density and Q structure) are all fixed. Because, in these numerical examples we invert only the transverse component of the waveform data and neglect toroidal–spheroidal coupling, the Lamé constant λ does not affect the data set. The source parameters are also fixed to the Harvard CMT solutions (with ramped source–time functions).

The model parameters are defined by the expansion eq. (4). The basis functions used are the spherical harmonics of degree 0–12 for the laterally dependent part and the six boxcar functions for the layers (Moho–310 km depth, 310–670 km depth, 670–1000 km depth, 1000–1500 km depth, 1500–2000 km depth and 2000 km depth–CMB (core–mantle boundary)) for the vertically dependent part. The total number of model parameters is 1014.

3.2 Examples of computed sensitivity

In this subsection we provide some simple examples to show how the sensitivity is evaluated. In all examples below we use the sensitivity for $\delta\mu$ of eq. (10). Fig. 1 shows the sensitivity for the case in which the whole data set **d** in eq. (1) consists of a single trace. Fig. 1(a) shows the sensitivity map at Moho–310 km depth for a shallow event. The data set is a trace of the receiver at Ala-Archa in Kyrgyzstan (AAK; 42.639°N, 74.494°E) for the Aleutian earthquake ($M_w = 7.9$) of 1996 June 10. A Butterworth bandpass filter with corner frequencies of 0.025 and 0.02 Hz is applied. Because the centroid depth of this event is shallow (29 km depth), surface waves are dominant in the trace.

Fig. 1(a) intuitively shows the accuracy of the DSM. Because the DSM accurately computes the effect of the couplings between multiplets (although we do not explicitly compute the modes), we can accurately compute the sensitivity including the contribution from the off-great circle. In contrast, previous global waveform inversion studies using modal summation methods have assumed that sensitivity exists only on the great-circle path. However, Fig. 1(a) shows that the off-great circle makes a significant contribution to the sensitivity, and that the DSM can accurately and automatically evaluate the effects of finite wavelength. Also, because waveform inversion is a procedure for interpreting the residual of the waveform fit between synthetics and data based on the distribution of sensitivity, the ability to accurately compute the sensitivity means the ability to accurately invert for 3-D Earth structure. Thus, if we analyse a large data set using the DSM we can expect further improvement of the seismological earth model.

The sensitivity can be computed not only for a whole trace but also for a subset of a trace in an arbitrary time window. In Fig. 1(a) we also show the sensitivities for the time windows G1, G2 and G3 in a data set trace. The left three sensitivity maps for G1, G2 and G3 employ the same colour scale as that used for the whole trace. The right two sensitivity maps for G2 and G3 employ a colour scale that is normalized by the maximum sensitivity for each plot. We can see that the sensitivity for the whole trace is almost equivalent to the combined sensitivity of G1, G2 and G3, and that each contribution can be automatically evaluated with the amplitude effects included.

The sensitivity can be evaluated not only for surface waves but also for body waves. Fig. 1(b) shows the sensitivity at 1500–2000 km depth for a deep event. The data set is a trace of the receiver at Matsushiro in Japan (MAJO; 36.5425°N, 138.2073°E) for the Bolivian earthquake ($M_w = 8.2$) of 1994 June 9. A Butterworth bandpass filter with corner frequencies of 0.025 and 0.02 Hz is applied. Because the centroid depth of this event is deep (647.1 km depth), body waves are dominant in the trace.

Fig. 1(b) shows the automatic evaluation of the sensitivity of body waves. It is well known that body waves are highly sensitive to the regions around their bottoming points. For this distance ($\Delta = 147.89°$), the *SS* minor arc and the *SSS* major arc bottom around this depth. Although we do not specify the types of ray, a mechanical evaluation of the sensitivity using eqs (3) and (11) can reproduce this feature.

The above examples are all for cases in which the whole data set consists of a single trace, but our methods can also be applied to a data set constituting of multiple traces. Fig. 2 shows the sensitivity at Moho–310 km depth for the whole data set for the Bolivian earthquake of 1994 June 9. The results show that a higher sensitivity is automatically computed around the epicentre and its antipode. A higher sensitivity in the region in which the path coverage is dense is also automatically computed. Thus the use of our index 'sensitivity' allows us to automatically evaluate the coverage of the waveform inversion data set.

3.3 Improvement of the coverage

In this and the following subsections we demonstrate how, and the degree to which, the use of our data weighting method changes the obtained earth model. Fig. 3 shows the method used to choose the data set employed in this inversion. The basic idea is to include all useful data regardless of the type of phase. We apply three

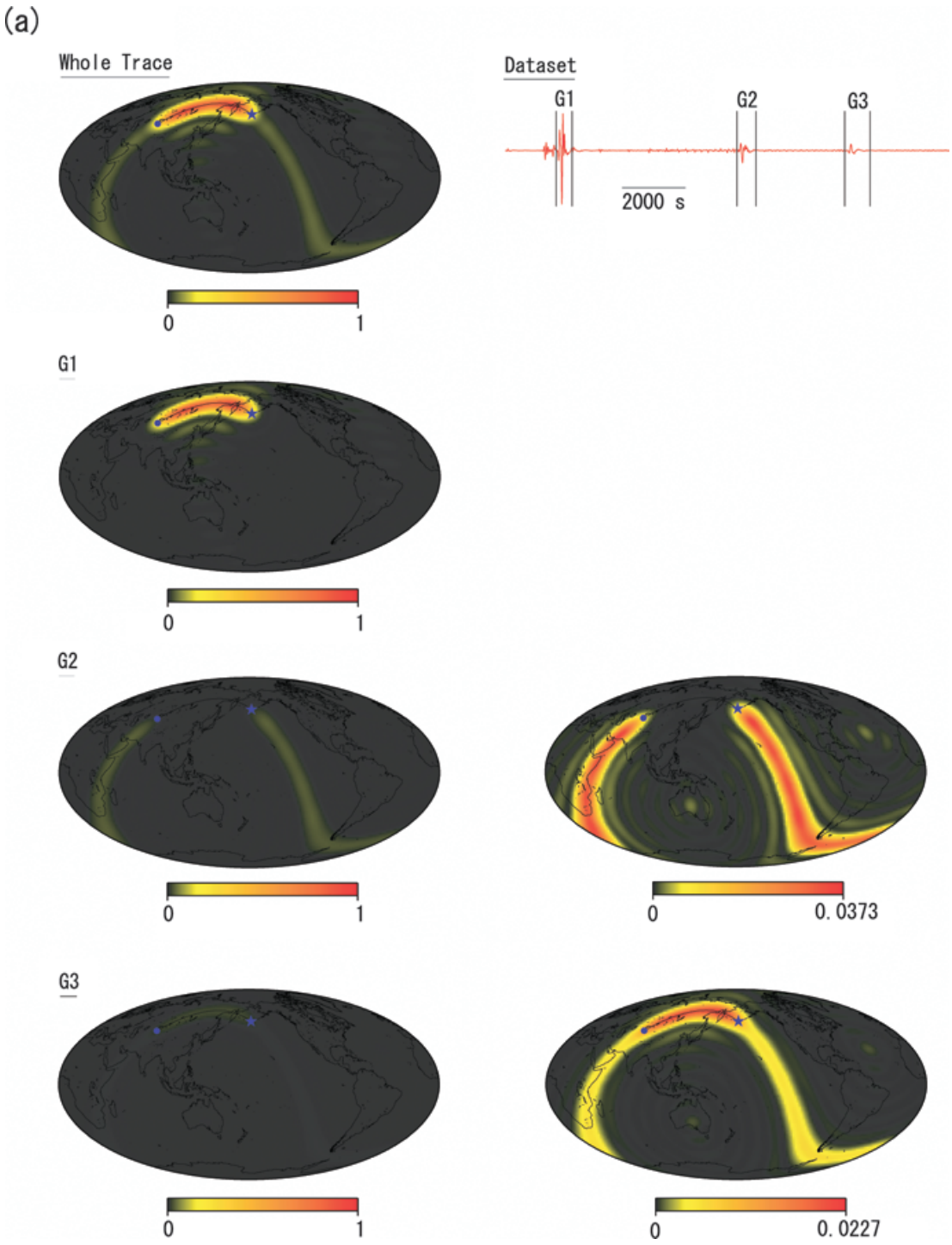

Figure 1. (a) Sensitivity map at Moho–310 km depth (uppermost left) for the whole waveform trace (shown uppermost right), and sensitivity maps for the waveform data within time windows in the trace for G1 (second uppermost), G2 (third uppermost left and right), and G3 (lowermost left and right). For each sensitivity map, the blue star and circle indicate the epicentre and the receiver, respectively, the solid line indicates the minor arc path, the red area indicates the high-sensitivity region and the black area indicates the low-sensitivity region. For the G2 and G3 maps we show sensitivity maps both in the same colour scale as the uppermost left figure (the two figures on the left) and in the normalized colour scale by the maximum value for each plot (the two figures on the right). The values of each colour palette show the relative amplitude compared with the maximum for the whole trace. (b) Sensitivity at 1500–2000 km depth (left) for the whole waveform trace (shown on the right). The blue star, blue circle, solid line and colour notations are the same as in (a). The values of the colour palette show the relative amplitude compared with the maximum of this plot.

different bandpass filters (200–400 s, 100–200 s and 50–100 s) to a single trace and compare them with corresponding synthetic seismograms. We extract time windows in which the residuals between the data and synthetics are reasonably small for each filtered trace and adopt data in those time windows as a data set. The width of the adopted time windows can be different for each filtered trace, and the time window for lower-frequency data is usually larger than that for higher-frequency data. The time window for each trace is not

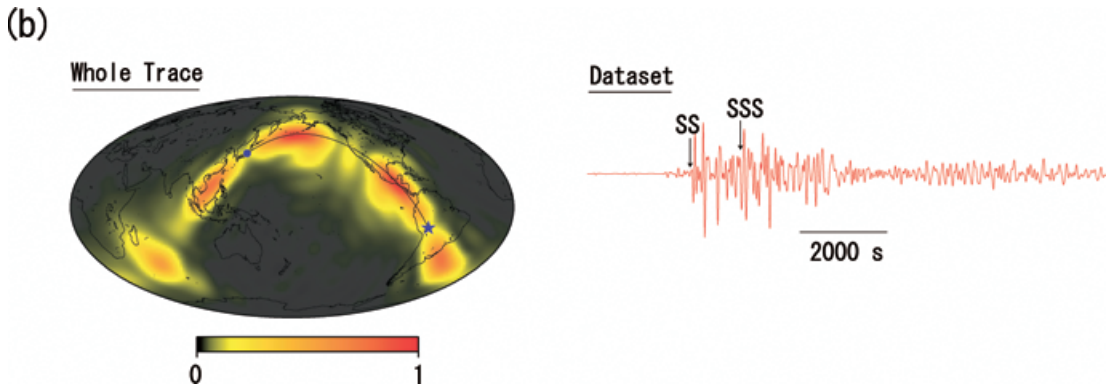

Figure 1. (*Continued.*)

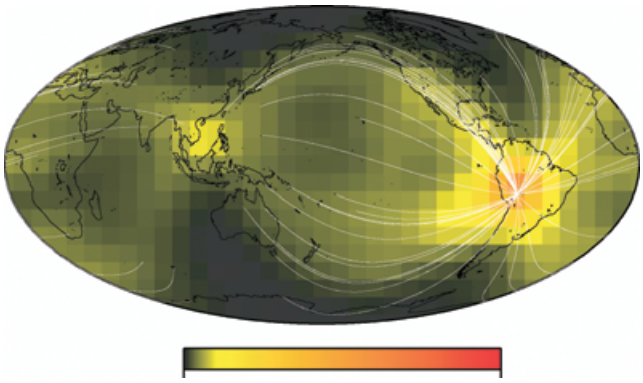

Figure 2. Total sensitivity at Moho–310 km depth for the whole data set for the Bolivian earthquake of 1994 June 9 used in the inversion in Section 3. Minor arc great-circle paths (shown by white lines) are also plotted.

necessarily continuous but sometimes has breaks (i.e. we can have multiple time windows). Because surface waves sample a shallower region than body waves, their fit to synthetic seismograms is sometimes poor. A typical example is shown in Fig. 3(b). In this case the fits for G2 and G3 are poor, and we exclude those surface wave parts from the data set. As a result, we have multiple time windows, and we can utilize the information from minor arc body waves as well as major arc body waves and multi-orbit phases.

The data used are Pacific21 (Pacific21/OHP and Pacific21/SPANET), IRIS (IRIS/IDA, IRIS/USGS and IRIS/CDSN) and GEOSCOPE data. We made a data set for the 20 events ($M_w \geq 7.0$ between 1991 and 1999) shown in Fig. 4(a). The resultant data set consists of 2592 time windows in 1161 independent traces. The data in each time window consist of velocity seismograms (in m s^{-1}) at each time step (1 Hz sampling). The total path coverage is shown in Fig. 4(b). Although the energy of the major arc and multi-orbit phases is smaller than that of the minor arc phases, these later phases sample the regions where minor arc path coverage is poor (e.g. beneath the Indian Ocean and the Central Pacific). It is thus important to utilize the later phases as much as possible.

We inverted the above data set using three different weighting methods: our weighting method, the moment normalization method and the no-weighting method. The moment normalization method is one of the weighting methods used in several previous waveform inversion studies. The weighting factors in eq. (12) are given as

$$\sqrt{w_i} = A/M_0^{(i)}, \tag{23}$$

where A is a constant and $M_0^{(i)}$ is the moment of the event for ith data. In the no-weighting method, the weighting factors in eq. (12) are constant:

$$\sqrt{w_i} = B, \tag{24}$$

where B is a constant. (For most cases $B = 1$, i.e. $w_i = 1$.) In this numerical example, constants A and B in eqs (23) and (24) are determined so that the sum of the total sensitivity $\sum_k \sum_j T_j^{(k)}$ is equal for each weighting method. This normalization allows direct comparison of the total sensitivity distribution and the resultant earth models.

Fig. 5(a) shows a comparison of the total sensitivity distribution at three layers of different depth. In defining weighting factors in eq. (12), we treat each time window as a data bin. We determine 2592 weighting factors. In determining the weighting factors for our method, we minimize eq. (22) using $\gamma = 10^{-5}$ as a damping factor. We can see that the total sensitivity distribution for the data set used in the no-weighting method is very heterogeneous, especially at greater depths. This is caused by the fact that energy from the great Bolivian earthquake of 1994 June 9 dominates. Using the moment normalization method, the heterogeneity of the total sensitivity distribution is slightly improved. However, as expected, the resultant total sensitivity distribution merely reflects the seismic ray path density, and thus this method cannot improve the sampling in regions where path coverage is poor (e.g. beneath the Indian Ocean and the Central Pacific). In contrast, using our weighting method the total sensitivity of the data set becomes more or less homogeneous over all regions.

3.4 Improvement of the obtained model

Other than weighting the data, the inversion scheme in our numerical experiments follows the standard damped least-squares procedure. We assume the following matrix equation between $\delta\mathbf{d}_{obs}$ (the residuals between the observed data and the synthetic data for the initial model) and $\delta\mathbf{m}_0$ (the residuals between the true model and the initial model):

$$\delta\mathbf{d}_{obs} = \mathbf{A}\delta\mathbf{m}_0, \tag{25}$$

where $\mathbf{A}$ is the assumed dependence of $\delta\mathbf{d}_{obs}$ on $\delta\mathbf{m}_0$. We apply the damped least square scheme to the following equation for the weighted data $\delta\mathbf{d}'_{obs}$:

$$\delta\mathbf{d}'_{obs} = \mathbf{A}'\delta\mathbf{m}_0, \tag{26}$$

where

$$\delta\mathbf{d}'_{obs} = \mathbf{W}^{1/2}\delta\mathbf{d}'_{obs} \tag{27}$$

$$\mathbf{A}' = \mathbf{W}^{1/2}\mathbf{A}, \tag{28}$$

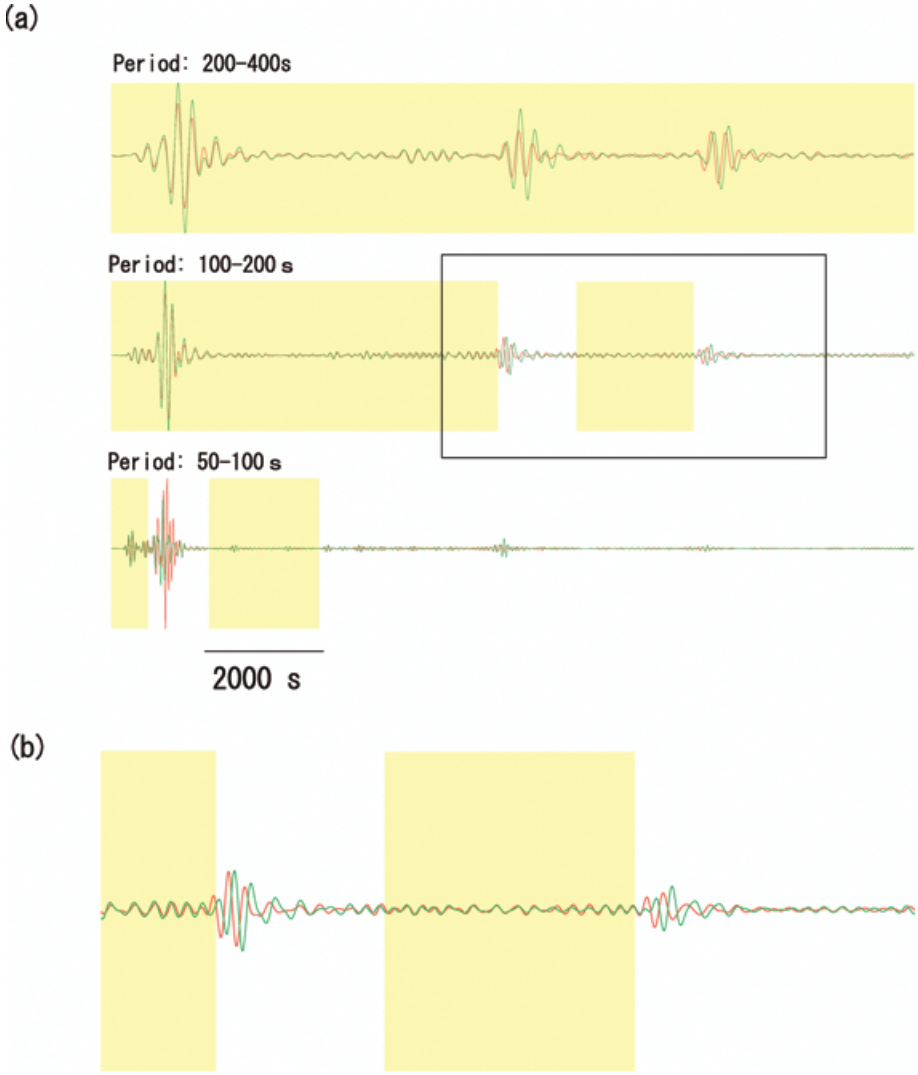

Figure 3. (a) Comparison of waveform data (red lines) and synthetic seismograms (green lines). The traces are bandpass filtered seismograms with corner periods of 200–400 s (uppermost), 100–200 s (middle) and 50–100 s (lowermost) of transverse component data of the HRV station for the Aleutian earthquake of 1996 June 10. Yellow boxes show time windows in which data are adopted as a data set. (b) Magnification of the black box in (a).

and $\mathbf{W}$ is a diagonal weighting matrix whose explicit elements are $\mathbf{W}_{ij} = w_i\,\delta_{ij}$. The inverted model $\delta\hat{\mathbf{m}}$ is obtained from the equation:

$$\begin{aligned}\delta\hat{\mathbf{m}} &= (\mathbf{A}'^{\mathrm{T}}\mathbf{A}' + \Gamma)^{-1}\mathbf{A}'^{\mathrm{T}}\delta\mathbf{d}'_{\mathrm{obs}} \\ &= (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \Gamma)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\delta\mathbf{d}_{\mathrm{obs}},\end{aligned} \tag{29}$$

where $\mathbf{\Gamma}$ is an appropriate diagonal matrix of damping factors.

Fig. 5(b) shows a comparison of the obtained models. The obtained model is significantly improved in the sense that it reproduces the robust feature of previous results. Because, in this numerical example, we use a relatively small data set containing only 20 events, even the combined use of both the DSM and our weighting method might yield a model with worse accuracy and resolution than the models from previous studies. However, only the model obtained by our weighting method agrees well with the models from previous studies, which intuitively shows that our weighting method is effective. Note that this does not mean that we cannot obtain a better model using our weighting method. If we use a large data set, the resultant model should be improved, especially for small-scale structures.

We summarize the notably improved features below. At Moho–310 km depth, only from the model using our data weighting method we can confirm that the distribution of mid-ocean ridges agrees well with very low-velocity anomalies, and that the distribution of the old continent and old ocean agrees well with very high-velocity anomalies. In contrast, for the models by the previous methods, the overall high- and low-velocity patterns are more or less similar, but their detailed distribution does not agree with the surface tectonics.

At the depth range of 670–1000 km, velocity structures at the subduction zones are greatly improved. Most of the previous studies have suggested the existence of high-velocity anomalies beneath Indonesia and South America, which are considered to be related to a slab penetration into the lower mantle. The model obtained by our method clearly confirms these features.

Looking at the lower mantle, we see that the model obtained by our method avoids spotty and extremely large anomalies. For example, in the model obtained by the no-weighting method we can see extremely large high-velocity anomalies beneath Panama and the Gulf of Guinea in the depth range of 1000–1500 km, whose

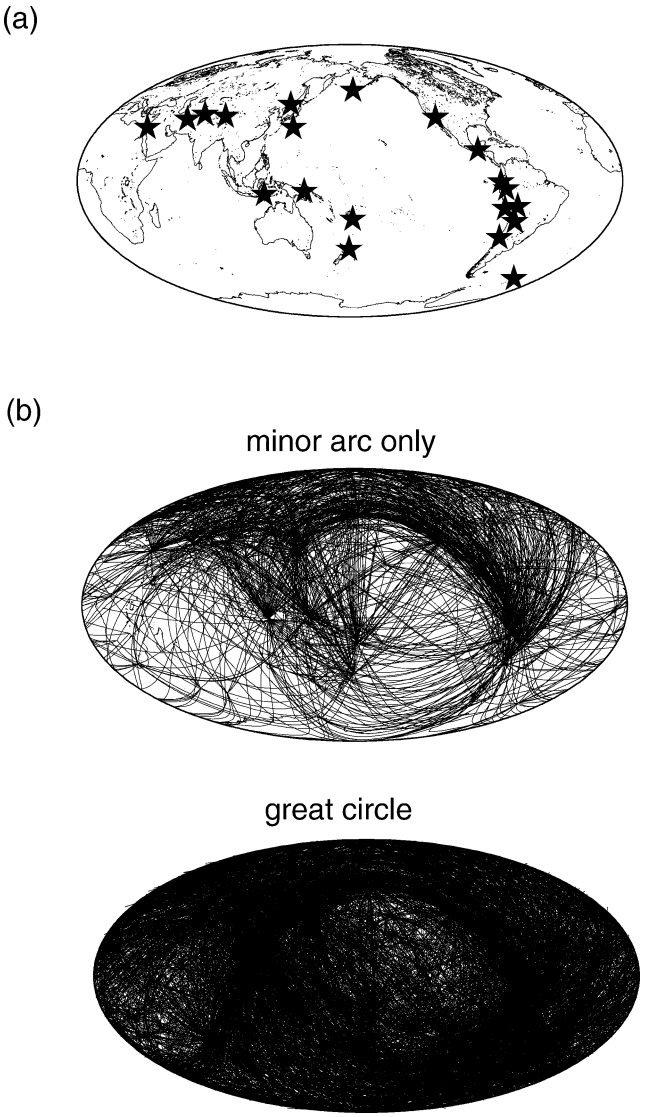

Figure 4. (a) Distribution of the epicentre used in the inversion in Section 3. (b) Minor arc (upper) and great-circle (lower) path coverage of data set used in the inversion.

amplitude (about 3 per cent) seems to be too large because the mid-mantle is considered to be weakly heterogeneous. Also, at a depth range of 2000 km CMB, we can see 2 per cent high-velocity anomalies beneath Africa, which were not reported in previous studies. However, most such spotty features disappear using our method. We can see that deep colours are dominant in the model obtained by our weighting method compared with the models obtained using previous methods. Because the colour scale is normalized by the maximum amplitude of heterogeneity, this dominance of deep colours indicates the disappearance of spotty features.

4 REASONS FOR THE IMPROVEMENT

In this section we will discuss the reasons for the improvement in the model obtained by our weighting method. We show that our method is approximately equivalent to a homogenization of the resolution of the obtained model. We also show that this procedure improves the shape of the resolution kernel because the model is constrained by various independent data, whereas using the previous weighting methods the model is essentially constrained by only some particular subset of the data. We discuss how this improvement in the resolution kernel enhances the robustness of the inversion.

Fig. 6 compares the resolution between the model obtained by our method and those obtained by the previous weighting methods at three different depth ranges. Although our model parameters are represented by spherical harmonic expansions for the laterally dependent part, we can transform the basis of the resolution matrix from a spherical harmonic basis to a block basis. Fig. 6 shows the distribution of diagonal elements of the transformed resolution matrix. Although the absolute values of the diagonal elements depend on the block size, we can show that the difference is only a scaling factor as long as the number of blocks is sufficiently large. Obviously, the resolution of the model obtained by our method is much more nearly homogeneous than that using the previous methods. Comparing Fig. 6 with Fig. 5(a) we can see that the resolution is more or less proportional to the sensitivity, and that homogenization of the resolution is achieved by the homogenization of total sensitivity. This fact is also shown theoretically if we substitute the definition of sensitivity of eq. (3) into the definition of the resolution matrix and ignore contributions from off-diagonal elements.

The homogenization of resolution, in general, improves the shape of the resolution kernel. Fig. 7(a) shows examples of computed resolution kernels at particular target points (20°N–100°W and 0°N–10°E both at 670–1000 km depth) for the models obtained using our weighting method and previous weighting methods. The path coverage in these regions is also shown. If we use the no-weighting method, the contribution of the path travelling in the vicinity of the target point from the Bolivian earthquake (shown by white lines) is large, and the resolution kernel is prolonged in that direction. Even if we normalize the moment magnitude, the resolution kernel is ill-shaped (has a large azimuthal dependence) because of the inhomogeneity of the azimuthal coverage. Red arrows in the figure indicate the regions where the azimuthal dependence is prominent. On the other hand, if we use our weighting method the azimuthal dependence is suppressed and we achieve an horizontally isotropic resolution kernel. This is because the data travelling in the azimuth where the path coverage is poor are enhanced. Because this should be a general feature of our weighting scheme, we can generally expect such an improvement of the resolution kernels.

Actually, if we compare the resolution kernels for the global model parameter at each depth range, we can observe the improvement in each of the ranges. Fig. 7(b) shows a comparison of the resolution kernel for a particular degree-12 pattern (cosine part of $s = 12, t = 4$) at 670–1000 km depth. This comparison is equivalent to a comparison of the recovery of models for this degree-12 pattern using noise-free synthetic data. The correlation between the input pattern and the output pattern is 0.88, 0.79 and 0.73 for our weighting method, the moment normalization method and no-weighting method respectively. While the recovered models obtained by the previous methods are seriously distorted, especially in the southern Hemisphere, due to the contributions of off-diagonal elements of the resolution matrix, the model obtained by our weighting method shows coherence with the input model. Considering that the obtained model is the convolution of the resolution kernel and the true earth model with some addition of noise, this improvement in coherence should enhance the robustness of the inversion.

One intuitive illustration of the improvement in the robustness is the reduction of the variance for a data set which was *not* used in the inversions. Fig. 8 compares the variance for the model obtained using our weighting method with that for the model obtained using one of the previous weighting methods. We compute the variances

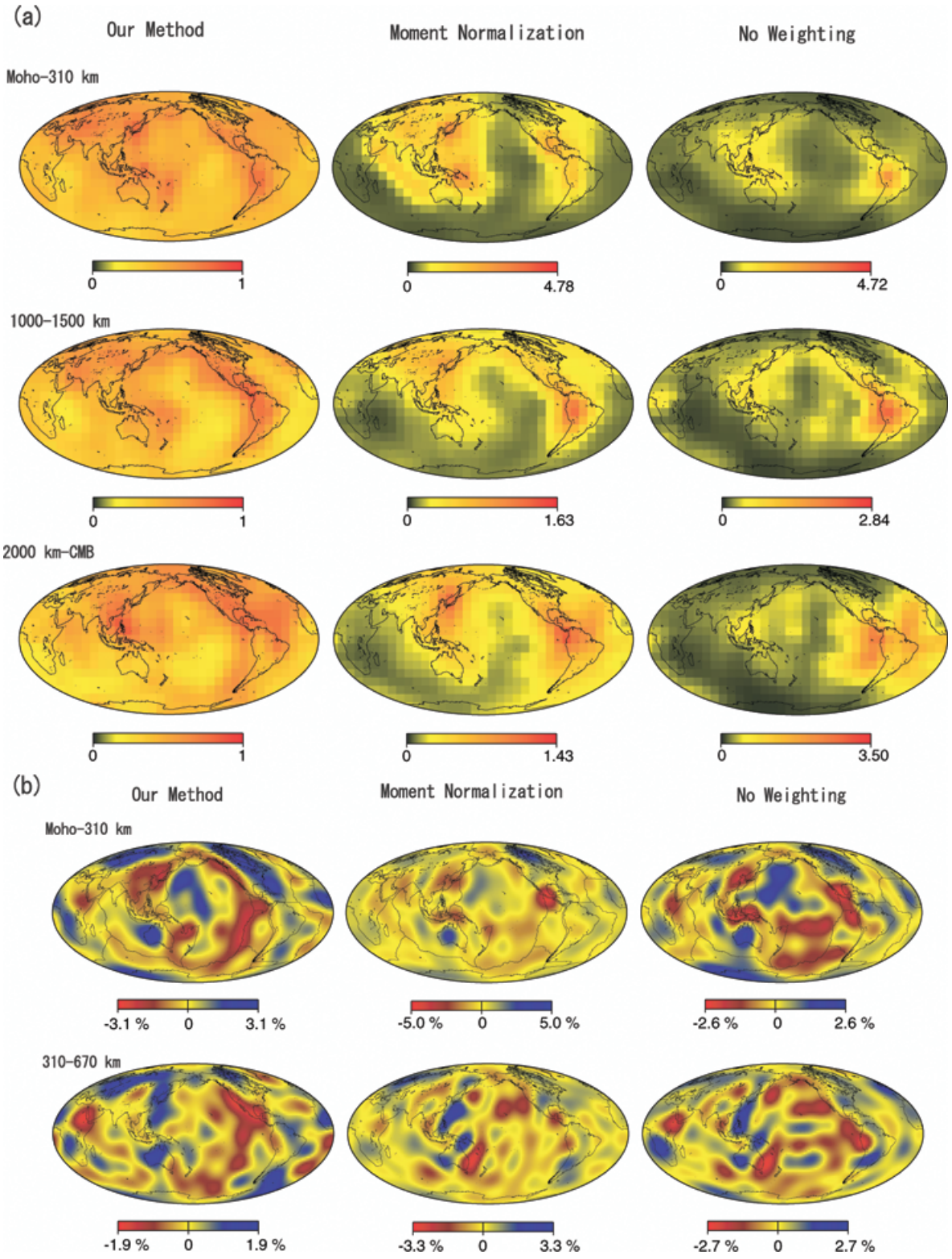

Figure 5. (a) Comparison of the total sensitivity map for our weighting method (left), for the moment normalization method (middle) and for the no-weighting method (right) at Moho–310 km depth (upper), 1000–1500 km depth (middle) and 2000 km depth–CMB (lower). The red (black) area indicates the high- (low-) sensitivity region. The colour scale is normalized by the maximum value of each plot. The values of each colour palette show the relative amplitude compared with the maximum for our method at that depth. (b) Comparison of the model obtained by our weighting method (left), by the moment normalization method (middle) and by the no-weighting method (right) at Moho–310 km depth (upper) and at 310–670 km depth (lower). The colour scale is normalized by the maximum absolute value of each plot. In the plots for Moho–310 km depth, we also show the plate boundaries by solid lines. (b) Comparison of the model obtained by our weighting method (left), by the moment normalization method (middle) and by the no-weighting method (right) at 670–1000 km depth (uppermost), 1000–1500 km depth (second uppermost), 1500–2000 km depth (second lowermost) and 2000 km depth–CMB (lowermost). The colour scale is normalized by the maximum absolute value of each plot.

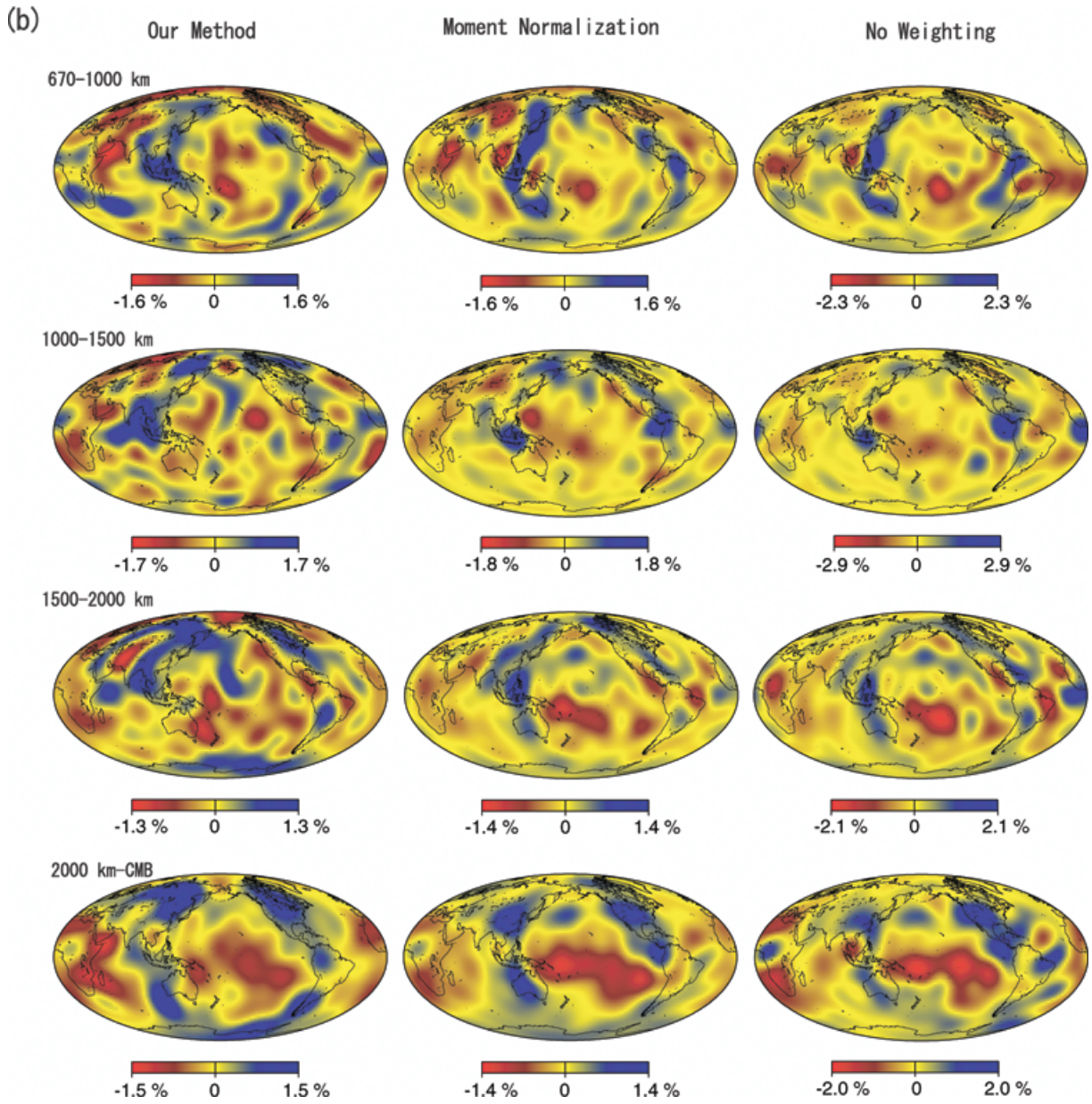

Figure 5. (*Continued*.)

for a data set consisting of the whole data set for one particular event. The definition of the variance is as follows:

$$\sum_i \left| \mathbf{d}_i - \mathbf{d}_i^{(0)} \right|^2, \tag{30}$$

where $\mathbf{d}_i^{(0)}$ is the ith synthetic data predicted by the obtained model. In Fig. 8 we plot the ratio of the variance for the model obtained by our weighting method against that for the model obtained by a previous weighting method (either the no-weighting method or the moment normalization method) at the epicentre of the corresponding event. Blue circles indicate that the ratio is smaller than 1, which means that the model obtained by our method better fits the data, and red circles indicate the opposite. We see that our weighting method is superior for most of the events. Note that the fitting for the original data set used in the inversion is worst using our weighting method (the variance reduction for the unweighted data set is 73.1 per cent using our weighting method, 76.6 per cent using the moment normalization method and 78.6 per cent using the no-weighting method). Better predictability for new data means the better retrieval of the information included in both the original data and new data, which means improvement in the robustness.

However, the robustness will not necessarily be improved as the shape of resolution kernels is improved. This is because extreme weighting will amplify the noise in data, thereby degrading the accuracy of the obtained model. This is the reason why we incorporate the damping factor γ in the penalty function when determining weighting factors (eq. 22). Actually, if we estimate the total variance of the obtained models (given by the trace of $\mathbf{C}$ in eq. 20 of Jackson 1979), it increases as γ in eq. (22) becomes small and the weighting becomes extreme. Fig. 9 shows the comparison of model error (the square root of total variance) of the model obtained by various γ in eq. (22). We assume that the covariances of the (unweighted) data set $\delta\mathbf{d}_{\rm obs}$ in eq. (25) are given by $\sigma_d{}^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. We also assume that the

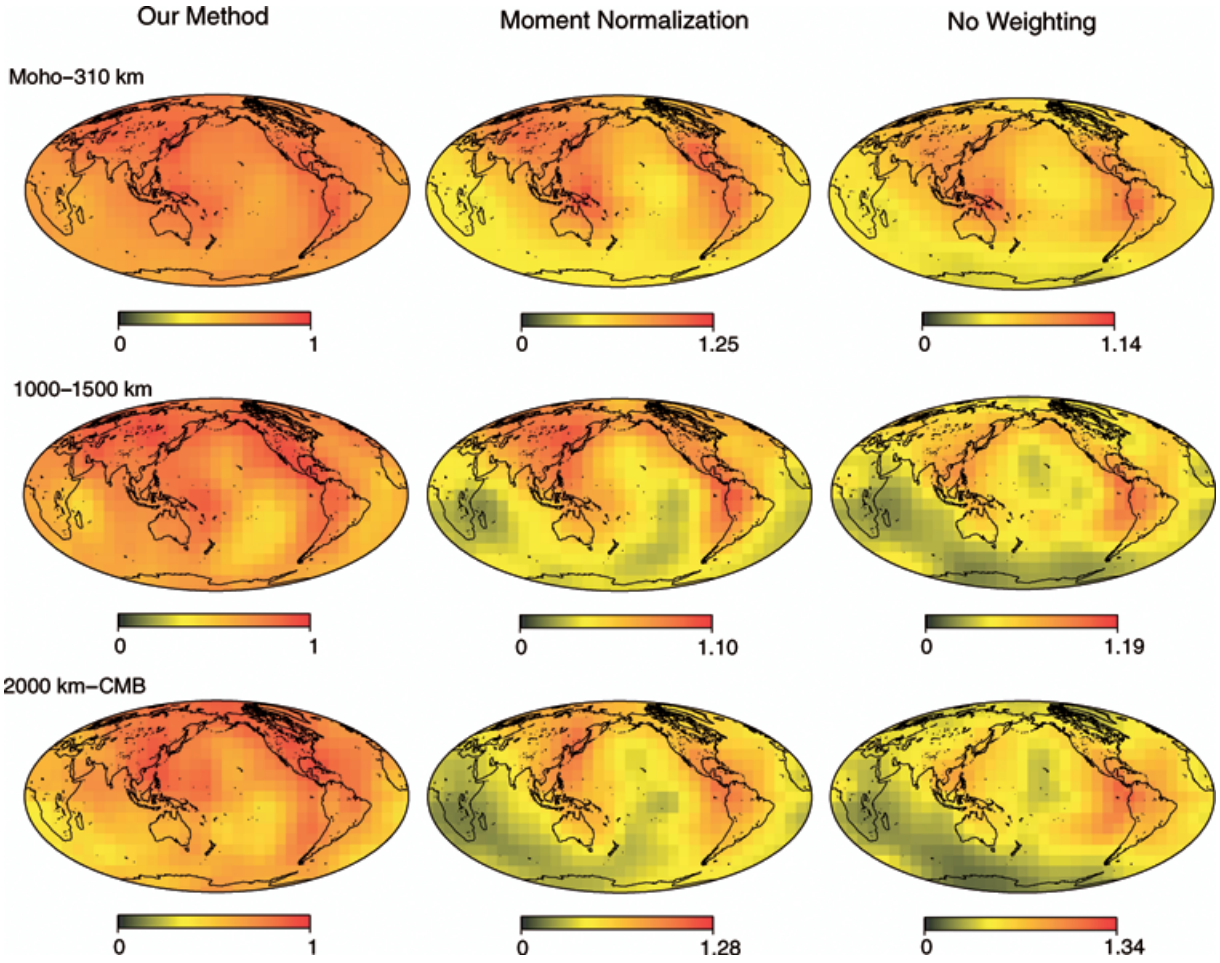

Figure 6. Comparison of the resolution of the model obtained by our weighting method (left), by the moment normalization method (middle) and by the no-weighting method (right) at Moho–310 km depth (upper), 1000–1500 km depth (middle) and at 2000 km depth–CMB (lower). The colour scale is normalized by the maximum value of each plot. The values of each colour palette show the relative amplitude compared with the maximum for our method at that depth.

damping factor $\mathbf{\Gamma}$ in eq. (29) is chosen to be $\Gamma_{ij} = \delta_{ij}\ {\sigma_d}^2/\sigma_m^{(i)2}$, where $\sigma_m^{(i)2}$ is the variance of the ith model parameter of the initial model $\delta\mathbf{m}_0$. The total variance of the inverted model is independent of σ_d other than a scaling factor, and the model errors in the plot show the ratio compared with the case when γ in eq. (22) is infinity (i.e. γ^{-1} is zero). When γ is infinity, the weighting factors w_i all become 1, which is the no-weighting method. Because the obtained model for this case is what we call a 'minimum variance solution' (Jackson 1979), the total variance is the smallest. When γ is zero (i.e. γ^{-1} is infinity), weighting occurs such that the total sensitivity is as homogeneous as possible. The total variance exponentially increases as γ^{-1} becomes larger and the weighting method approaches the extreme case. Thus we should choose an appropriate γ so as to compromise between the improvement of the resolution kernels and the degradation of the total variance. Development of a quantitative method for choosing this parameter remains a topic for future research.

5 DISCUSSION

We derived a weighting scheme that allows automatic and more or less objective determination of weighting factors in waveform inversion. Although this is not the first study to weight the data in waveform inversion, our method has significant differences and merits compared with previous studies. For example, Li & Romanowicz (1996) incorporate data weighting factors for each major phase in waveform traces. This method enhances phases, such as $S_{\rm dif}$, which sample regions where the information included in the whole data set is poor. However, the goal of their weighting is to estimate the covariance matrix of data error accurately, including the effect of off-diagonal elements. Their method is basically the inversion to obtain the minimum variance solution. Also, as explicitly stated in their paper, their method for determining weighting factors is somewhat *ad hoc*. Thus the basic concept is completely different from that of our method, and the latter is clearly preferable from the standpoint of objectivity. Other studies have used the moment normalization method, but this method is clearly not as effective as that presented here, as shown in Fig. 5.

Because the key feature of our method is homogenization of the resolution, the model obtained by our method is especially useful to discuss the scale of heterogeneity in the Earth. The difficulty in discussing the scale of heterogeneity from the models obtained by previous methods is that it is not easy to distinguish whether a difference of scale between two regions reflects a difference of the actual scale of the Earth's structure or simply a difference in resolution. However, if we homogenize the resolution, we can avoid this problem. In the near future, we hope to detect the scale of heterogeneity in detail by obtaining a detailed earth model using a

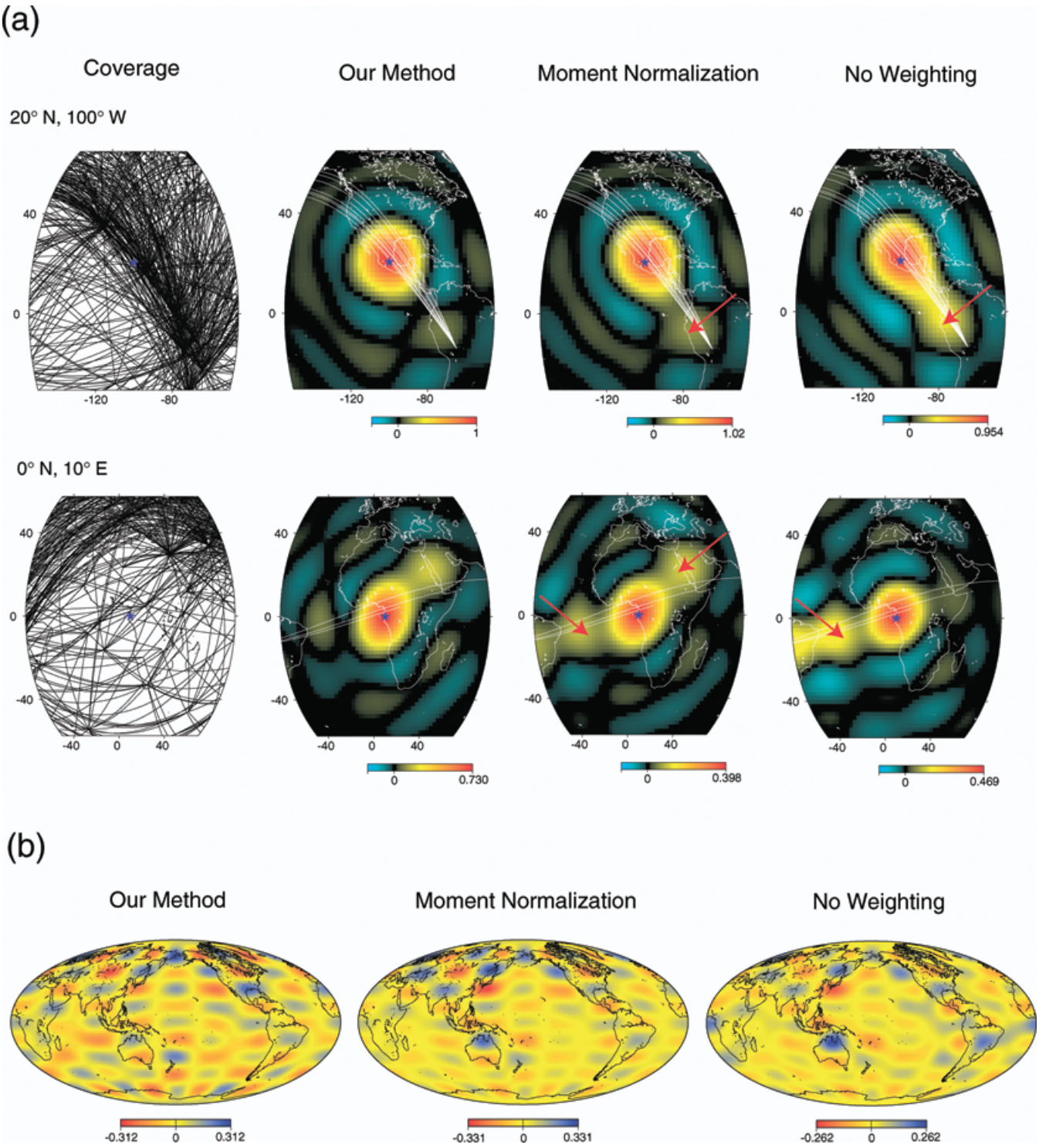

Figure 7. (a) Comparison of the resolution kernels for target points at 20°N–100°W (upper) and at 0°–10°E (lower) at 670–1000 km depth for the models obtained by our weighting method (second left), the moment normalization method (second right) and the no-weighting method (right). Minor arc path coverage at these regions is also shown (left). The target point is shown by a blue star. The red area indicates the region where the value of the resolution kernel is large, the black area the region where the value is around zero and the blue area the region where the value is negative. The colour scale is normalized by the maximum value of each plot. The values of each colour palette show the relative amplitude compared with the maximum for our method at 20°N–100°W. White lines are minor arc paths from the Bolivian earthquake travelling in the vicinity of the target point (within 10° radius from the target point). Red arrows show the regions where notable azimuthal dependence of resolution kernels is observed in the previous weighting methods. (b) Comparison of the resolution kernels for a degree-12 pattern (cosine part of $s = 12$, $t = 4$) at 670–1000 km depth for our weighting method (left), the moment normalization method (centre) and the no-weighting method (right). The values of each colour palette show the relative amplitude compared with the maximum of the input model.

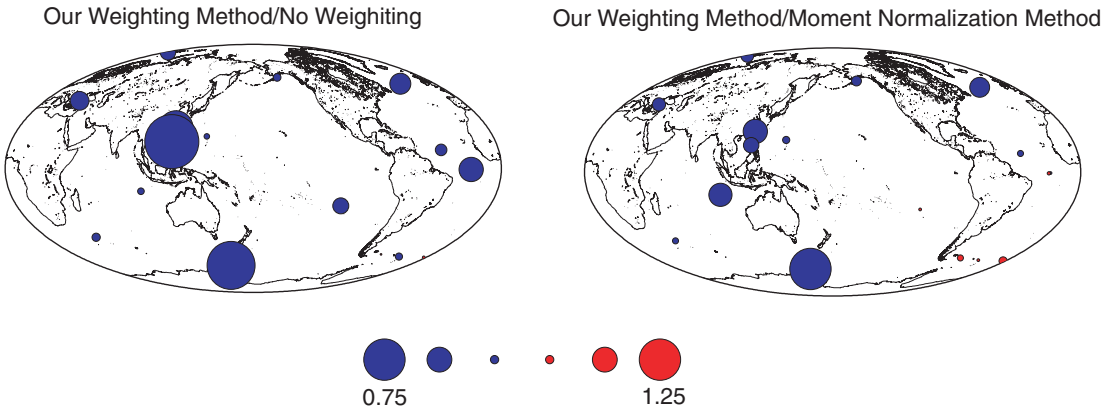

Figure 8. Ratios of the variance for the model obtained by our weighting method to that for the model obtained by the no-weighting method (left), and ratios of the variance for the model obtained by our weighting method to that for the model obtained by the moment normalization method (right). The variance (defined by eq. 30) is computed for the whole data set for one particular event which was *not* used in the inversion. The variance ratios are plotted at the epicentre of the corresponding event. Blue circles indicate that the variance ratio is smaller than 1 (the variance for the model obtained by our weighting method is smaller), and red circles indicate the opposite. The radius of the circles shows the value of the ratio.

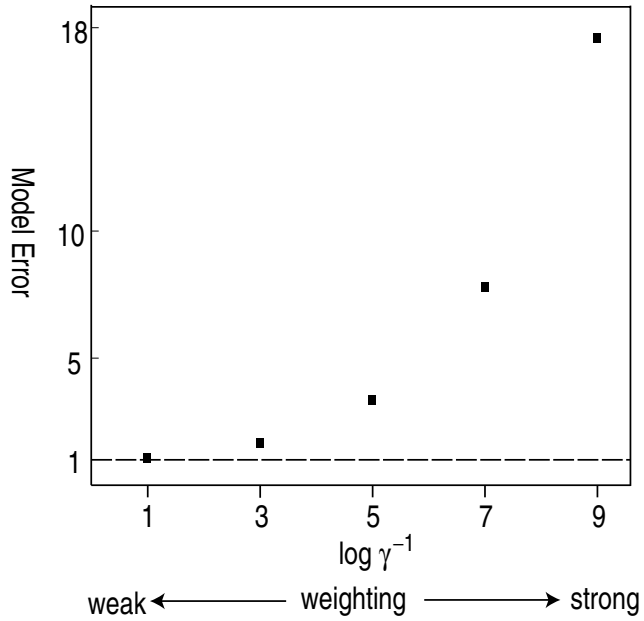

Figure 9. Comparison of the error (square root of total variance) of the models obtained by using various γ in eq. (22). The model error is shown by the ratio compared to that for the minimum variance solution (the model when γ^{-1} is zero).

large data set, the DSM, and the weighting method presented in this paper.

An apparent drawback of our weighting method is that it suppresses information for regions where higher resolution can be expected. One straightforward solution for this problem is to use non-uniform local functions (e.g. spherical splines, Wang & Dahlen 1995) to define model parameters for the horizontally dependent part as well as the vertically dependent part of the physical parameters of the Earth. In phase inversions, such parametrizations have already been applied both in delay time tomography (e.g. Fukao *et al.* 1992; Bijwaard *et al.* 1998) and surface wave tomography (e.g. Wang *et al.* 1998). Besides *a priori* parametrizations, adaptive parametrizations have been applied (e.g. Sambridge & Faletič 2003). They usually define model parameters so that the density of parameters is proportional to the density of optical rays. These approaches seem to be very successful. If we use our sensitivity instead of ray density (note that the ray density is equivalent to the sensitivity for the case of phase inversions), a similar approach should be possible for the waveform inversion. This is also an important future research topic.

ACKNOWLEDGMENTS

We are greatly indebted Yoshio Fukao for his valuable input and suggestions. We also thank Robert J. Geller, Shingo Watada, Malcom Sambridge, Jeannot Trampert and two anonymous reviewers for their critical reading of the manuscript. We are grateful for the use of SGI Origin 2000 and SGI LX3700 at the Earthquake Research Institute, University of Tokyo, the SR8000/MPP at the Information Technology Center, University of Tokyo, and the Earth Simulator at the Earth Simulator Center of the Japan Marine Science and Technology Center. This research was partly supported by grant from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (no 14540393).

REFERENCES

Bijwaard, H., Spakman, W. & Engdhal, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.,* **103,** 30 055–30 078.

Cummins, P.R., Geller, R.J., Hatori, T. & Takeuchi, N., 1994. DSM complete synthetic seismograms: SH, spherically symmetric, case, *Geophys. Res. Lett.,* **21,** 533–536.

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., Huang, A.-H. & Nolet G., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.,* **141,** 157–174.

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

Fukao, Y., Obayashi, M., Inoue, H. & Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone, *J. geophys. Res.,* **97,** 4809–4822.

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.

Geller, R.J. & Takeuchi, N., 1995. A new method for computing highly accurate DSM synthetic seismograms, *Geophys. J. Int.,* **123,** 449–470.

Grand, S.P., van der Hilst R.D. & Widiyantro S., 1997. Global seismic tomography; a snapshot of convection in the Earth, *GSA Today,* **7,** 1–7.

Gu, Y.J., Dziewonski, A.M., Su, W.-J. & Ekström, G., 2001. Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities, *J. geophys. Res.,* **106,** 11 169–11 199.

Hara, T., 1997. Centroid moment tensor inversion of low-frequency seismic spectra using Green's functions for aspherical earth models, *Geophys. J. Int.,* **130,** 251–256.

Hara, T., 2002. Degree 16 model of S-wave heterogeneity in the upper and uppermost mantle determined by the Direct Solution Method, *Abstracts of the Superplume International Workshop,* pp. 40–43, Superplume International Workshop, Tokyo.

Hara, T. & Geller, R.J., 2000. Simultaneous waveform inversion for three-dimensional Earth structure and earthquake source parameters considering a wide range of modal coupling, *Geophys. J. Int.,* **142,** 539–550.

Jackson, D.D., 1979. The use of *a priori* data to resolve nonuniqueness in linear inversion, *Geophys. J. R. astr. Soc.,* **57,** 137–157.

Li, X.D. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.,* **121,** 695–709.

Li, X.D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101,** 22 245–22 272.

Li, X.D. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. Int.,* **112,** 92–102.

Masters, G., Laske, G., Bolton, H. & Dziewonski, A., 2000. The relative behaviour of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale,* pp. 63–87, eds Karato, S., Forte, A.M., Liebermann, R.C., Masters, G. & Stixrude, L., AGU, Washington, DC.

Megnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.,* **143,** 709–728.

Pulliam, R. & Stark, P.B., 1993. Bumps on the core-mantle boundary: are they facts or artifacts?, *J. geophys. Res.,* **98,** 1943–1955.

Ritsema, J., van Heijst, H.J. & Woodhouse, H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science,* **286,** 1925–1928.

Sambridge, M. & Faletič, R., 2003. Adaptive whole Earth tomography, *Geochem. Geophys. Geosyst.,* **4,** doi:10.1029/2001GC000213.

Takeuchi, N., Geller, R.J. & Cummins, P.R., 1996. Highly accurate P-SV complete synthetic seismograms using modified DSM operators, *Geophys. Res. Lett.,* **23,** 1175–1178.

Takeuchi, N., Geller, R.J. & Cummins, P.R., 2000. Complete synthetic seismograms for 3-D heterogeneous Earth models computed using modified DSM operators and their applicability to inversion for Earth structure, *Phys. Earth planet. Int.,* **119,** 25–36.

Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: Possible artifacts in seismic tomography, *Science,* **271,** 1257–1260.

van der Hilst, R.D., Widiyantro, S. & Engdahl, E.R., 1997. Evidence of deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parameterization of three dimensional Earth models, *Geophys. Res. Lett.,* **22,** 3099–3102.

Wang, Z., Tromp, J. & Ekström, G., 1998. Global and regional surface-wave inversions: a spherical-spline parameterization, *Geophys. Res. Lett.,* **25,** 207–210.

Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.,* **89,** 5953–5986.

Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.,* **141,** 558–576.