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

Nobuaki Fuji[a], Kenji Kawai[b], Robert J. Geller[a]

[a] *Department of Earth and Planetary Science, Graduate School of Science, the University of Tokyo, Tokyo, Japan*
[b] *Department of Earth and Planetary Science, Graduate School of Science, Tokyo Institute of Technology, Tokyo, Japan*

Abstract

We develop a quantitative and objective methodology for inversion of body-wave waveform data for elastic and anelastic structure and apply it to the upper mantle and mantle transition zone (MTZ) beneath the Northwestern Pacific, using PREM as the starting model. We obtain an average positive S-wave velocity anomaly in the MTZ beneath the Philippine Sea Plate and the Japan Sea. We also find a negative S-wave velocity anomaly in the depth range 300-400 km beneath the subduction zone to the east of northern Japan and its southern extension beneath the Izu-Bonin-Marianas arc. Our inversions obtained a lower Q_μ, as compared to PREM, for all of the subregions.

Key words: mantle transition zone, seismic waveform inversion, anelastic attenuation, singular value decomposition

Email addresses: fuji@eps.s.u-tokyo.ac.jp (Nobuaki Fuji), kenji@geo.titech.ac.jp (Kenji Kawai), bob@eps.s.u-tokyo.ac.jp (Robert J. Geller)

1. Introduction

This paper has three aims: (1) to present techniques for automated inversion of broad-band body-wave waveform data for elastic and anelastic Earth structure; (2) to apply these techniques to invert for 1-D seismic structure for various subregions of the upper mantle and the mantle transition zone (MTZ) beneath the Northwestern Pacific; and (3) to estimate the uncertainty and study the robustness of the results.

There have already been relatively many waveform inversion studies using mainly long-period surface waves (e.g., Mégnin and Romanowicz, 2000; Hara, 2004; Takeuchi, 2007). Most such studies derived global models with a relatively low resolution. However, because of the efficient and accurate calculation of synthetic seismograms and partial derivatives, Hara (2004) was able to derive models with relatively better resolution.

Although there have been many studies which used body-wave waveforms to determine 1-D structure models for particular regions, almost all such studies have been conducted by forward modeling (e.g., Lay and Helmberger, 1983; Grand and Helmberger, 1983; Brudzinski et al., 1997; Tajima and Grand, 1998). One obstacle to performing waveform inversion of broad-band waveforms has been the need for accurate and efficient methods for computing synthetic seismograms and their partial derivatives. Our group has recently developed algorithms based on the Direct Solution Method (DSM: Geller and Ohminato, 1994) for computing highly accurate synthetics for spherically symmetric transversely isotropic Earth models (Kawai et al., 2006, and references cited). To date our applications of waveform inversion to Earth structure have only been for 1-D models, but our approach is formulated for

application to 3-D problems (Geller and Hara, 1993; Cummins et al., 1997; Takeuchi et al., 2000).

We have recently conducted waveform inversion studies of transverse component data for the localized 1-D S-wave velocity structure of D″ (Kawai et al., 2007ab; Konishi et al., 2009). Kawai et al. (2007a) inverted for the fine structure of the D″ layer beneath Central America and found higher S-velocities in the upper half of D″ and lower velocities in the lower half. Kawai et al. (2007b) and Kawai and Geller (2009) obtained similar results for D″ beneath the Arctic. Konishi et al. (2009) found a more complicated structure, which they called an "S-shaped" velocity model, for D″ beneath the Western Pacific. Their model suggests the existence of phase transitions at several depths within D″ and is consistent with the existence of depleted MORB in the lowermost mantle beneath the Western Pacific.

This study is our first attempt to simultaneously invert for elastic (S-wave velocity) and anelastic (Q_μ) structure. The target is the MTZ and upper mantle beneath the Northwestern Pacific. Accurate measurement of seismic attenuation can help to constrain the physical and chemical state of the interior of the Earth, as seismic attenuation reflects physical properties in the MTZ such as temperature, water content, and composition (e.g., Karato, 1993; Faul and Jackson, 2005). There have been several studies of attenuation structure, mainly using relatively long period surface wave (from 60 up to 1000 seconds) waveforms (e.g., Romanowicz, 1995, 1998; Gung and Romanowicz, 2004) or surface wave amplitudes (e.g., Dalton et al., 2008). Attenuation around Japan was also previously studied using forward modeling by Kato et al. (2001) and Revenaugh and Jordan (1991).

Extracting information on attenuation from recorded seismograms requires careful treatment of elastic structure because focusing and defocusing effects or scattering effects can also affect the amplitudes. This paper is the first attempt to utilize inversion techniques for relatively short period (20-200 s) body wave waveform data, in order to simultaneously invert for 1-D elastic and anelastic structure.

2. Methodology

Fig. 1 shows the overall flowchart of our inversion procedures. First we choose a target region, select events and stations, and calculate synthetic seismograms and partial derivatives for the initial model. After correcting for the effect of local structure near the stations and sources by making static corrections (time shifts), we check whether each of the observed seismograms satisfies criteria for inclusion in the dataset.

As shown in Fig. 1, we perform waveform inversion to minimize the variance between the observed seismograms and synthetic seismograms, subject to regularization. Since this inversion is non-linear, iterative linearized inversion is desirable. However, due to the computational requirements we are not presently routinely performing iterative inversion. In this paper we only perform one sample iterative inversion as an example.

We have developed software which automatically performs the various steps in Fig. 1. Some parts of this software are publicly available from our website (http://www-solid.eps.s.u-tokyo.ac.jp/~dsm).

2.1. Synthetics and partial derivatives

2.1.1. Synthetics

In this study we use the Direct Solution Method (Geller and Ohminato, 1994; Cummins et al., 1994ab; Geller and Takeuchi, 1995; Takeuchi et al., 1996; Takeuchi et al., 2000; Kawai et al., 2006) to compute full-wave synthetic seismograms. 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:

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

where $\mathbf{T}$ is the mass matrix, $\mathbf{H}$ is the stiffness matrix, $\mathbf{g}$ is the force vector, and $\mathbf{R}$ enforces continuity conditions at fluid-solid boundaries which exists only for the spheroidal (P-SV) case. Our methods do not use geometrical optics or Earth-flattening approximations in computing the synthetics and their partial derivatives. The DSM 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., Takeuchi et al., 2000).

2.1.2. Partial derivatives with respect to μ_0 and Q_μ^{-1}

When we compute the partial derivatives with respect to μ_0 (the real part of the elastic parameter at the reference frequency) and Q_μ (the anelastic attenuation parameter, which we assume to be frequency-independent within the frequency band being considered), we have to take anelastic attenuation into account. Azimi's law (Azimi et al., 1968) gives the frequency dependence of μ as follows:

$$\mu(\omega) \approx \mu_0 \left(1 + \frac{2q \ln(\omega/\omega_0)}{\pi}\right)(1 + iq) \tag{2}$$

where $i = \sqrt{-1}$, $q = Q_\mu^{-1}$, ω is the angular frequency, ω_0 is the reference angular frequency (a reference frequency of 1 Hz, i.e., $\omega_0 = 2\pi$, is used in this paper), and $\mu_0 = \mathrm{Re}(\mu(\omega_0))$. We thus have:

$$\frac{\partial \mu(\omega)}{\partial \mu_0} = \left(1 + \frac{2q \ln(\omega/\omega_0)}{\pi}\right)(1 + iq) \tag{3}$$

$$\frac{\partial \mu(\omega)}{\partial q} = \mu_0 \left[\frac{2 \ln(\omega/\omega_0)}{\pi} + i\left(1 + \frac{4q \ln(\omega/\omega_0)}{\pi}\right)\right]. \tag{4}$$

We use the above results to obtain the perturbation to the shear modulus, $\Delta\mu(\omega)$, for perturbations to the shear modulus at the reference frequency or to the anelastic parameter, $\Delta\mu_0$ or Δq, respectively. We obtain:

$$\Delta\mu(\omega) = \frac{\partial \mu(\omega)}{\partial \mu_0}\Delta\mu_0 + \frac{\partial \mu(\omega)}{\partial q}\Delta q. \tag{5}$$

We use eq. (5) to compute the perturbation to the synthetic seismogram, $\Delta u(\omega)$, for perturbations to μ_0 or q. We have:

$$\Delta u(\omega) = \frac{\partial u(\omega)}{\partial \mu}\Delta\mu(\omega), \tag{6}$$

where

$$\frac{\partial u(\omega)}{\partial \mu}$$

is the partial derivative of the synthetic seismogram $u(\omega)$ with respect to $\mu(\omega)$.

The partial derivatives for a "voxel" perturbation to a 3-D starting model are obtained from the following result, which is based on the formulation of Geller and Hara (1993):

$$\left\{ \frac{\partial u_i^{(k)}(\omega)[\mathbf{r}^{(p)}]}{\partial m_l} \right\} = \int_V (\omega^*)^2 [u_j^{(k)}][\rho^{(l)}]^* \eta_j^{i(p)} dV - \int_V [u_{j,q}^{(k)}]^* [C_{jqrs}^{(l)}] \eta_{r,s}^{i(p)} dV, \quad (7)$$

where u_i is the i-component of the displacement, $\eta_i^{j(p)}$ is the i-component of the back-propagated displacement excited by a point force in the j-direction at the p-th station, summation over j, q, r, and s is implied, and

$$u_{i,j} = \frac{\partial u_i}{\partial x_j}$$

is the locally Cartesian derivative.

In this paper we compute partial derivatives for a spherically symmetric perturbation to a spherically symmetric, anelastic, starting model. The perturbations to the density and the elastic constants at a particular depth $\mathbf{r}_l$ are written respectively as $\delta\rho^{(l)}\Psi_l(\mathbf{r})$ and $\delta C_{jqrs}^{(l)}\Psi_l(\mathbf{r})$, where $\Psi_l(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}_l)$ in this study. Note that, as discussed above, we include the effect of physical dispersion in $C_{jqrs}^{(l)}$. In this paper we invert only for SH-wave velocity and anelastic attenuation Q_μ^{-1}, but the above formulation can be easily extended to allow inversion for Q_κ^{-1} and P- and SV-wave velocities as well as an isotropic or 3-D structure.

2.2. Inverse problem

We define $\delta\mathbf{d}^{(kp)}$ to be the residual (the difference of the observed seismogram, $\delta\mathbf{d}_{\rm OBS}^{(kp)}$, and the synthetic seismogram for the initial model, $\delta\mathbf{d}_{\rm INIT}^{(kp)}$) for the k-th event and the p-th station:

$$\delta\mathbf{d}^{(kp)} = \mathbf{d}_{\rm OBS}^{(kp)} - \mathbf{d}_{\rm INIT}^{(kp)}. \quad (8)$$

For simplicity we assume here that we have K events and P stations with records available at all stations for all events, and we write the corresponding partial derivatives with respect to the l-th model parameter (in this study either μ_0 or Q_μ^{-1} for some depth) as

$$\frac{\partial \mathbf{u}^{(kp)}}{\partial m_l}$$

for all M model parameters. We also assume here that each time series has T points, and that data from all stations are available for all earthquakes; in this case the vector $\delta\mathbf{d}$ has dimension KPT and the matrix $\mathbf{A}$ has dimension $KPT \times M$. If, on the other hand, the various time series have different lengths and/or not all stations are available for all events, this can be handled by a straightforward extension of the above.

The matrix of partial derivatives $\mathbf{A}^{(kp)}$ for the k-th event and p-th station is written as follows:

$$\mathbf{A}^{(kp)} = \left[\frac{\partial \mathbf{u}^{(kp)}}{\partial m_1} \quad \frac{\partial \mathbf{u}^{(kp)}}{\partial m_2} \quad \cdots \frac{\partial \mathbf{u}^{(kp)}}{\partial m_l} \right]. \tag{9}$$

In order to conduct a simultaneous inversion for all event (k) and station (p) pairs, we combine the residuals $\delta\mathbf{d}^{(kp)}$ and partials $\mathbf{A}^{(kp)}$ into a single vector $\delta\mathbf{d}$ and a single matrix $\mathbf{A}$, respectively, as follows:

$$\delta\mathbf{d}^{\mathbf{T}} = \left[\delta\mathbf{d}^{\mathbf{T}(11)} \quad \cdots \delta\mathbf{d}^{\mathbf{T}(1P)} \delta\mathbf{d}^{\mathbf{T}(21)} \quad \cdots \delta\mathbf{d}^{\mathbf{T}(K1)} \quad \cdots \delta\mathbf{d}^{\mathbf{T}(KP)} \right] \tag{10}$$

$$\mathbf{A}^{\mathbf{T}} = \left[\mathbf{A}^{\mathbf{T}(11)} \quad \cdots \quad \mathbf{A}^{\mathbf{T}(1P)} \mathbf{A}^{\mathbf{T}(21)} \quad \cdots \quad \mathbf{A}^{\mathbf{T}(K1)} \quad \cdots \quad \mathbf{A}^{\mathbf{T}(KP)} \right]. \tag{11}$$

The inverse problem is usually written as:

$$\mathbf{A}\delta\mathbf{m} = \delta\mathbf{d} \tag{12}$$

where $\delta\mathbf{m}$ is the perturbation to the initial model. Since the number of unknowns (number of elements of $\delta\mathbf{m}$) is usually far smaller than the number of data points (number of elements of $\delta\mathbf{d}$), it is well known that eq. (12) implies the following minimization rather than strict equality:

$$|\mathbf{A}\delta\mathbf{m} - \delta\mathbf{d}|^2 = \text{minimum}. \tag{13}$$

Fig. 2 shows a schematic explanation of the inversion process.

We use the SVD to decompose the matrix $\mathbf{A}$ as follows. The singular value decomposition (SVD) has been widely discussed in the geophysical literature (e.g., Wiggins, 1972; Menke, 1984). However, as discussed below, their results can be considerably simplified for the case when $\mathbf{A^T A}$ is non-singular:

$$\mathbf{A} = \mathbf{U\Lambda V^T}, \tag{14}$$

where the non-zero eigenvalues are ordered as follows:

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_M^2 > 0. \tag{15}$$

Note that we are assuming that the problem is overdetermined (i.e., $M < KPT$), and that the first M eigenvalues of $\mathbf{A}$ (as defined in eq. 15) are non-zero, i.e., that $\mathbf{A^T A}$ is positive definite.

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

$$\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_M] \tag{16}$$

normalized to satisfy:

$$\mathbf{v}_j^{\mathbf{T}}\mathbf{v}_l = \delta_{jl}, \tag{17}$$

the unitary $KPT \times KPT$ matrix $\mathbf{U}$ contains the left eigenvectors of $\mathbf{A}$, normalized to satisfy:

$$\mathbf{u}_j^{\mathbf{T}}\mathbf{u}_l = \delta_{jl}, \tag{18}$$

and the $KPT \times M$ matrix $\mathbf{\Lambda}$ contains the eigenvalues:

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_M \\ & \mathbf{0} & \end{pmatrix}, \tag{19}$$

where δ_{jk} is a Kronecker delta. It follows from eqs. (14), (17) and (18) that:

$$\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{u}_j. \tag{20}$$

From eqs. (14) and (18) we have

$$\mathbf{A}^{\mathbf{T}}\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{\mathbf{T}}\mathbf{U}^{\mathbf{T}}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}} = \mathbf{V}\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}. \tag{21}$$

From eqs. (17) and (21) we have

$$\mathbf{v}_j^{\mathbf{T}}(\mathbf{A}^{\mathbf{T}}\mathbf{A})\mathbf{v}_k = \lambda_k^2 \delta_{jk}. \tag{22}$$

We now use the above results to obtain the least squares solution of eq. (13) by solving the normal equations:

$$\mathbf{A}^{\mathbf{T}}\mathbf{A}\delta\mathbf{m} = \mathbf{A}^{\mathbf{T}}\delta\mathbf{d}. \tag{23}$$

We write the solution of eq. (23) as an eigenfunction expansion:

$$\delta\mathbf{m} = \sum_{j=1}^{M} e_j \mathbf{v}_j = \mathbf{V}\mathbf{e}, \tag{24}$$

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

$$\mathbf{e}^{\mathbf{T}} = (e_1 \; e_2 \; \cdots \; e_M). \tag{25}$$

We substitute eq. (24) into eq. (23), multiply both sides of the resulting equation on the left by $\mathbf{v}_j^{\mathbf{T}}$ and use eq. (22) to solve for each expansion coefficient of eq. (24) as follows:

$$e_j \;\; = \;\; \frac{\mathbf{v}_j^{\mathbf{T}} \mathbf{A}^{\mathbf{T}} \delta\mathbf{d}}{\lambda_j^2}. \tag{26}$$

We can write eq. (26) in matrix form as follows:

$$\mathbf{e} = (\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1}\mathbf{V}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\delta\mathbf{d}. \tag{27}$$

The numerator of eq. (26) shows that the expansion coefficient of the j-th eigenvector will become larger as the vector $\mathbf{A}^{\mathbf{T}}\delta\mathbf{d}$ becomes more nearly parallel (in the multi-dimensional vector space) to the eigenvector $\mathbf{v}_j$, while the denominator of eq. (26) shows that as the eigenvalues become smaller the expansion coefficients will tend to be magnified, and the solution of the inverse problem will thus tend to become unstable.

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

$$\begin{aligned} \text{covar. } (\mathbf{e}) &= \sigma_d^2 \, (\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1}\mathbf{V}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{A}\mathbf{V}(\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1} \\ &= \sigma_d^2 \, (\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1}\mathbf{V}^{\mathbf{T}}(\mathbf{V}\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}})\mathbf{V}(\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1} \\ &= \sigma_d^2 \, (\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Lambda})^{-1}. \end{aligned} \tag{28}$$

If we transform eq. (28) from the eigenspace to the model space, the covariance matrix is written as:

$$\begin{aligned} \text{covar. } (\delta\mathbf{m}) &= \sigma_d^2 \, (\mathbf{A^T A})^{-1} \\ &= \sigma_d^2 \sum_{j=1}^{M} \frac{\mathbf{v}_j \mathbf{v}_j^\mathbf{T}}{\lambda_j^2} \\ &= \sigma_d^2 \, \mathbf{V}(\mathbf{\Lambda^T \Lambda})^{-1}\mathbf{V^T}. \end{aligned} \tag{29}$$

Note that as is pointed out in several inversion works (e.g., Okal and Geller, 1979), parameters not included in the inversion cannot be quantified. For example, these above formal covariance estimates cannot directly be taken as the "real" standard deviation, because the effect of our model does not include 3-D heterogeneity, anisotropy, CMT re-determination, etc. The real uncertainty can thus be several times the value of the nominal standard deviation.

2.3. SVD inversion

If we were to use eq. (24) as written, taking the sum over all eigenvectors, we would obtain the least squares solution of eq. (13) and eq. (23). However, in the actual inversions we sum over only the first n eigenvectors (which correspond to the n eigenvalues with the largest magnitudes — see eq. 15) as follows:

$$\delta\mathbf{m} = \sum_{j=1}^{n} e_j \mathbf{v}_j. \tag{30}$$

We discuss the choice of n below. For this case the covariance matrix in the model space is given by:

$$\text{covar. } (\delta\mathbf{m}) = \sigma_d^2 \sum_{j=1}^{n} \frac{\mathbf{v}_j \mathbf{v}_j^\mathbf{T}}{\lambda_j^2} \tag{31}$$

For each seismogram $\delta\mathbf{d}^{(kp)}$ we define the contribution of that seismogram to the perturbation of the j-th model parameter to be:

$$e_j^{(kp)} = \frac{\mathbf{v}_j^{\mathbf{T}} \mathbf{A}^{\mathbf{T}(kp)} \delta\mathbf{d}^{(kp)}}{\lambda_j^2}, \tag{32}$$

where the final model is obtained by summing the contributions from each record:

$$e_j = \sum_{k,p} e_j^{(kp)}. \tag{33}$$

In the same way, we define the contribution from the k-th event as follows:

$$e_j^k = \sum_{p} e_j^{(kp)} \tag{34}$$

and the contribution from the records at the p-th station as follows:

$$e_j^p = \sum_{k} e_j^{(kp)}. \tag{35}$$

By analyzing the values of $e_j^{(kp)}$, e_j^k, and e_j^p we can see which data affect which model components, and which eigenvectors of the model space $(\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\ \mathbf{v}_M)$ are most sensitive to the data.

2.4. Criteria for evaluating models

We truncate the eigenvector expansion (eq. 30), using as a guide the AIC (Akaike Information Criterion: Akaike, 1977). The variance VAR_n and AIC_n for the model SVD_n (the model obtained by using n in eq. 30) are written respectively as follows:

$$\mathrm{VAR}_n = \frac{|\mathbf{d}_{\mathrm{OBS}} - \mathbf{d}_{\mathrm{SYN}}|^2}{|\mathbf{d}_{\mathrm{OBS}}|^2} \tag{36}$$

$$\mathrm{AIC}_n = \mathrm{ND} \ln 2\pi + \mathrm{ND} \ln(\mathrm{VAR}_n) + \mathrm{ND} + 2(n+1), \tag{37}$$

where ND is the number of independent data points. We discuss the evaluation of ND below.

In this study the upper bound on the value of ND is $\mathrm{ND}^1 = KPT/20$ because we use 20-200 s bandpass-filtered waveforms, interpolated at a 1 Hz sampling frequency (the meaning of the superscript is explained below). 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 the following we introduce the empirical parameter α, which we define as follows:

$$\mathrm{ND}^\alpha = \frac{1}{\alpha} KPT/20. \tag{38}$$

The meaning of α is that the data are assumed to have α-fold redundancy. AIC_n for ND^α, as obtained from eq. (37) is hereafter written as AIC_n^α.

Eq. (37) shows that AIC decreases as the variance decreases but increases as the number of model parameters, n, is increased. The minimum value of AIC thus formally represents the optimum balance between model complexity and variance reduction. In evaluating eqs. (36) and (37) in this study we use the variance reduction computed using the partial derivatives rather than re-computing synthetics for the models obtained by each inversion.

3. Dataset and static corrections

3.1. Event and station selection

We select the upper mantle and MTZ beneath the Northwestern Pacific as the target for this study. In this paper we invert two datasets: (1) the entire dataset of waveforms which sample the region (Fig. 3, table of events in electronic supplement): (2) the respective datasets for six subregions: the

upper mantle and the transition zone beneath the Japan Sea; Philippines Sea West, Middle and East; Pacific West and East (Fig. 4). We collect data for events with $M_w \geq 5.5$ with a half source duration time less than 10.0 s and a source depth greater than 100 km, using records from stations at epicentral distances between 5° and 35° (based on Global CMT Solution parameters).

We use only transverse component SH data. The number of events from 1995 to 2007 is 161 and the number of waveforms is 6025. All of the seismograms are velocity seismograms and we deconvolve the instrumental response function. In this study we apply a Butterworth filter whose passband is between 20 s and 200 s to both the data and the synthetics; thus we can approximate the source time function (moment rate function) as a δ-function.

In this study we choose the direct S phase time window 60 seconds after the first arrival (which contains overlapped triplication arrivals which are not individually identifiable because of the relatively long-period passband) as the portion of the data to be analyzed. We weight each waveform by the reciprocal of the maximum amplitude of the observed seismogram.

The partial derivatives for the moment tensor and centroid are given by Hara (1997). These results can be used to quantitatively evaluate the effect of uncertainty of the source parameters on the Earth structure model and also to invert for CMT parameters. However, as we have been careful not to use stations near nodes of the radiation pattern we do not think uncertainly of the CMT solution has a significant effect on the results presented in this paper.

3.2. Static corrections

In order to correct for the effect of local structure near the stations and sources, we must make static corrections (Fig. 5). We align the onset times of the synthetic and the observed seismograms as follows. First we automatically pick the onset time of the synthetic seismogram and let the zero-to-peak time be Δt. We pick a time window from $4\Delta t$ before to Δt after the arrival time in the synthetic seismogram and also use this window for the observed seismogram. We then seek the time shift which gives the best cross-correlation coefficient in order to roughly determine the observed arrival time, which we denote as $t_{\rm tmp}$. $t_{\rm tmp}$ represents the power centroid of the incoming S wave; to align the respective onset times we apply a causal filter. We finally search the observed waveform for a local zero near $t_{\rm tmp}$ by comparing the noise level of the time windows Δt before and after $t_{\rm tmp}$. We find a good correlation between the autopicked arrival times and manually picked arrival times. The procedures for determination of the time shifts are empirical (see Konishi et al., 2009, for further details), and there is probably room for further improvements. However, our present procedures yield reasonable results and our forward modeling tests suggest that our correction procedures are reasonable for the frequency band used in this analysis (20-200 s).

3.3. Data selection criteria

We consider pairs consisting of an observed record and the corresponding synthetic seismogram (Fig. 6). The amplitude ratios in Fig. 6 are calculated by dividing the maximum absolute value of the synthetics in the time window by the maximum absolute value of the observed seismograms. There are cases

for which the observed and synthetic amplitudes differ greatly, and others for which the phases differ greatly even though the amplitudes are similar. For other waveforms the amplitudes differ greatly while the phases match quite well. For such waveforms the predominant period of the observed record is much shorter than that of the synthetic, perhaps because such seismograms contain waves that traveled through the slab, where there is a narrow high Q_μ zone. The above observations probably contain important information on 3-D structure, but we exclude them from the dataset for this paper.

4. "Initial Q_μ" model

The amplitude ratios in Fig. 6a suggest that the actual Q_μ values may be systematically lower than the PREM Q_μ model (Dziewonski and Anderson, 1981). We first investigate the behavior of our dataset by forward modeling with several Q_μ models (shown in Fig. 7). Figs. 6b and 6c illustrate the behavior for PREM1/2, which has half the Q_μ value of PREM, and Q19, which is modified from Romanowicz (1995) for this region. The first number above each panel of Fig. 6 shows the number of records which satisfy a criterion of an amplitude ratio between 0.3 and 3.0 and also have a cross correlation above 50%, which we regard as a good match. The second number is the total number of records in the dataset. "PREM1/2" improves the fit of the amplitude ratios but because of the broadening effect, the cross correlations drop significantly, while Q19 moderately changes the amplitude ratio but does not significantly degrade the cross-correlations. The above trade-offs suggest that we should simultaneously invert for elasticity and anelasticity.

We first invert for an initial Q_μ structure for our study region using the methods discussed in the previous section. The unknowns are μ_0 and Q_μ from the Earth's surface to a depth of 740 km at intervals of 10 km. We use 754 waveforms chosen randomly from those parts of our dataset which have amplitude ratios between 0.1 and 10.0 and cross correlation coefficients above 50%. The results obtained from SVD inversions (Fig. 7) using the n largest eigenvalues (up to $n = 20$) show a lower Q_μ model than PREM. We use a smoothed version of this model (Fig. 7c), which we call the "Initial Q_μ" model, as the Q_μ model in our subsequent inversions. Fig. 6 shows that when "Initial Q_μ" is used the number of waveforms satisfying both the amplitude ratio and cross-correlation criteria is 3629, as opposed to 3355, 3402, and 3077 for PREM, Q19, and PREM1/2 respectively.

Since the dataset used in this inversion is noisy and samples a complicated structure, a single S-wave velocity model cannot be meaningfully determined for the whole region. This is confirmed by the fluctuations of the S-wave velocity model in Fig. 7b. Fig. 7a shows AIC_n^1 and VAR_n for $n \leq 20$. AIC suggests that using a basis of 16-18 eigenvectors yields the most appropriate model.

5. Synthetic tests

We conduct several tests using synthetic data to study the resolution of our inversions and the potential that the static corrections could cause systematic errors. As shown in Fig. 8 we use an input model with alternating slow and fast layers of 100 km thickness, with an amplitude of $\pm$ 1.5 % relative to the initial model (PREM + "Initial Q_μ"). We compute synthetics for a

small number of stations and events (only 50 records), and conduct inversions varying n. The unknowns in this case are the perturbations to the elastic constant μ_0 from the Earth's surface to a depth of 700 km at intervals of 20 km. Perturbations to the anelastic model parameters are not included as unknowns in this case. As shown in Fig. 8, the inversions for reasonable ranges of n (e.g., 17-23) are successful in reproducing the main features of the input model.

We next conduct synthetic tests to study the effect that hypothetical errors in the static corrections might have on the inversion results. Figs. 9a through 9c show synthetic results for systematically shifted synthetic data with shifts of 2, 3, and 4 s later than the actual travel times, respectively. The models in Fig. 9a recover the input model reasonably well, but those in Figs. 9b and 9c do not, due to the non-linearity of the inversion. Fig. 9d shows that when the data are randomly shifted by up to ± 10 s the synthetic inversion can still reasonably recover the input model.

Fig. 10 examines synthetic inversion results for the case of synthetic seismograms systematically shifted to be 1 s late to simulate hypothetical errors in the static corrections. The left panel shows that the basic shape of the input model is recovered, but with a shift of about 0.2% towards lower velocities. The right panel shows $\theta(z)$, the angle between the partial derivatives with respect to the time shift and those with respect to μ_0 for each depth z, defined as follows:

$$\cos\theta(z) = \frac{\sum \frac{\partial u}{\partial t}\frac{\partial u}{\partial \mu_0(z)}}{\sqrt{\sum \left(\frac{\partial u}{\partial t}\right)^2}\sqrt{\sum \left(\frac{\partial u}{\partial \mu_0(z)}\right)^2}}. \tag{39}$$

On the basis of the results in Figs. 9 and 10 we conclude that hypothetical small systematic errors or reasonably large but random errors in the static corrections would cause only small errors in the absolute velocities while still preserving the basic character of the model, but that systematic errors of several seconds in most or all of the static corrections could be problematical. Such large systematic errors do not appear to be present in our actual data analyses. Nevertheless the subject of static corrections remains a subject for further investigation. Note that this applies not only to our work in particular, but to all studies of Earth structure.

5.1. "Checkerboard" test for Q_μ perturbation

In order to study the resolving power of a simultaneous inversion for elastic and anelastic parameters we used two input models for a synthetic inversion. The first has alternating slow and fast layers of 100 km thickness, with a perturbation of ± 2.0 % relative to the initial velocity model (PREM) but with no perturbation to the Q_μ model (Fig. 11). The second has the same elastic perturbation but also has a perturbation to the Q_μ model (Fig. 12). In our test inversions, the unknowns are μ_0 and Q_μ from the Earth's surface to a depth of 700 km at intervals of 20 km; we calculate 6000 synthetic "records" for each of the two perturbed models.

In Figs. 11a and 12a, we plot the variance (VAR) and the values of AIC_n^α (for $\alpha = 1, 10, 50, 100$ respectively) as a function of n and find AIC minima around SVD22 and SVD24 respectively if we take $\alpha \geq 50$. Fig. 11a shows that there is a drop in AIC at $n = 22$ regardless of the value of α. However this is only a local minimum for $\alpha = 1$ or $\alpha = 10$, but is a global minimum for $\alpha = 50$ or $\alpha = 100$. Since the absolute value of AIC_n^α is not important

(only the relative values for a given α as n varies are significant) we have adjusted all the AIC curves in Fig. 11a (as well as similar later figures) to have the same AIC values for $n = 0$ and normalized by multiplying by α.

Figs. 11b and 12b show that even for a small number of eigenvectors (15 or 13, respectively), the main S-wave velocity features of the input model can be resolved, except for the layer from 600 to 700 km depth, for which the resolving power of this basis set is low. From Fig. 11c we see that the Q_μ structure is not changed significantly for the input model with no Q_μ perturbation even if we use a larger basis ($n = 22$).

Fig. 11 shows that including the partial derivatives for anelasticity in the inversion for a purely elastic model perturbation does not result in a significant perturbation to Q_μ. On the other hand, Fig. 12 shows that a larger number of eigenvectors ($n = 24$) is necessary to reproduce the anelastic structure, whereas we cannot reproduce it with a smaller number of eigenvectors ($n = 13$).

5.2. Sensitivity of eigenvectors to elastic and anelastic structure

Fig. 13a shows the magnitudes of the eigenvalues, while Fig. 13b shows the relative contribution to the eigenvectors of the elastic and anelastic parts of the model, defined respectively as $P_j^{\rm elast}$, and $P_j^{\rm anel}$ as follows:

$$P_j^{\rm elast} = \frac{|\mathbf{v}_j^{\rm elast}|^2}{|\mathbf{v}_j|^2} \tag{40}$$

$$P_j^{\rm anel} = \frac{|\mathbf{v}_j^{\rm anel}|^2}{|\mathbf{v}_j|^2}, \tag{41}$$

where $\mathbf{v}_j^{\text{elast}}$ and $\mathbf{v}_j^{\text{anel}}$ are the elastic and anelastic components of each eigenvector $\mathbf{v}_j$. In Fig. 13b, we normalize the value of P_j^{elast} and P_j^{anel} so that:

$$\sum_{j=1}^{M} P_j^{\text{elast}} = \sum_{j=1}^{M} P_j^{\text{anel}}. \tag{42}$$

Fig. 13b shows that the eigenvectors for the modes with the largest eigenvalues reflect primarily elastic structure, while the eigenvectors for modes with smaller eigenvalues tend to be dominated by anelastic structure. This suggests that we cannot determine the Q_μ structure reliably and stably unless we first obtain a reliable elastic model, which is consistent with the results of previous surface wave attenuation tomographic studies (e.g., Romanowicz 1995, 1998; Gung and Romanowicz, 2004; Dalton et al., 2008).

As we will see in the next section, the results obtained from the real dataset require a large number of eigenvectors to describe the Q_μ structure with some fluctuations in the S-wave velocity model. This occurs because the actual anelastic structure is far from the starting model while the starting elastic model is much closer to the final model.

6. Inversion Results

On the basis of the success of these synthetic tests we now proceed to to conduct inversions of observed data for elastic and anelastic structure.

In Fig. 14 we show the inversion results for the Japan Sea region using a "small dataset" (104 records) with a criterion of amplitude ratio factors between 0.5 and 2.0 and cross correlations above 80 %, while in Fig. 15 we show results for the same region using a "large dataset" (284 records)

with a criterion of an amplitude ratio factor between 0.25 and 4.0 and cross correlations above 40 %.

In Fig. 14a the thick curve is the variance reduction and the thin curves (labeled 1, 2, 3, 4 respectively) are the AIC_n^α values (see eqs. 37 and 38) for $\alpha = 1, 2, 3, 4$ respectively. Regardless of the value of α, there is a local minimum at $n = 3$ and another at $n = 21$. However, for $\alpha = 1$ it is just barely clear in Fig. 14a that $n = 21$ is a global minimum, whereas for $\alpha = 4$, $n = 21$ is a local minimum but has a higher AIC value than for $n = 3$. If $\alpha = 2$ or $\alpha = 3$ is used, then $n = 21$ gives a clear global minimum.

In this paper, as the number of records in a dataset increases we use higher values of α. This seems reasonable, because as the size of the dataset increases the redundancy of the dataset will increase. Our methods for selecting α are still empirical, and we will investigate more objective methods for choosing α in our future work.

Using $\alpha \geq 3$ for the small dataset and $\alpha \geq 5$ for the large dataset, we conclude that SVD21 for the small dataset or SVD30 for the large dataset are the most preferable models for this region. All of the preferred models show a positive S-wave velocity anomaly in the MTZ and in the layer from 300 to 400 km. A kink around 400-500 km suggests the possibility of a velocity discontinuity. Because the large dataset has a larger number of scattered data, the results fluctuate more but still show the same feature (a positive velocity anomaly). As for Q_μ, we do not see any significant perturbation when we choose SVD3 for either dataset but we obtain a lower Q_μ if we take a larger number of eigenvectors.

6.1. Technical issues

6.1.1. Variance reduction: Born vs. re-linearized

In the above inversion we estimate the variance using the first order Born approximation to compute the variance (rather than re-linearizing with respect to each model obtained by the inversion, as was done by Kawai et al., 2007ab and Konishi et al., 2009). To estimate the errors due to using the first order Born approximation to estimate the variance, we re-compute the synthetics for each of models SVD1 through SVD30 for the small dataset. Fig. 14a shows the variance reduction and AIC values ($\alpha = 1, 2, 3, 4$) calculated using the first order Born approximation, while Fig. 14b shows the AIC values obtained using the re-computed synthetics up to $n = 30$. Due to the non-linearity of the inversion, the behavior of the re-computed variance reduction differs from that for the Born approximation and the absolute value of the variance reduction is smaller. However, both methods of calculation show that we have large variance reductions when go from $n = 2$ to $n = 3$ and from $n = 15$ to $n = 16$. The "real" variance has a global minimum (between $n = 0$ and $n = 30$) for $n = 20$. Because we do not use re-linearization in our inversions when we compute the variance, we obtained a global minimum of AIC_n^1 for $n = 3$. We also plot $\mathrm{AIC}_n^{0.5}$ for reference in order to find the value of α which gives a global AIC minimum around $n = 21$. We re-compute synthetics for each SVD_n model only for the examples in this subsection because it is computationally intensive, but we can expect that the "real" variance will have a minimum near the AIC minimum calculated using the first order Born approximation for the models presented below.

6.1.2. Iterative inversion

As the large difference between the variance reductions computed based on the Born approximation and re-computed without linear approximation indicates, the inversions are non-linear. One might thus question whether the inversions have sufficiently converged after a single inversion. We present one example of iterative inversion for the Japan Sea region. We use the model obtained by the initial inversion for the "small dataset" as the starting model for the second iteration (note that we fix the starting model to PREM and the "Initial Q_μ" above the 200 km depth, just for simplicity).

Fig. 16a shows the variance reduction and AIC values ($\alpha = 1, 2, 3, 4$) calculated using the first order Born approximation. Fig. 16b shows the SVD1 model which is suggested as the best by AIC analysis. For $\alpha = 1$ only, $n = 13$ seems to be a global minimum, whereas for $\alpha \geq 2$, $n = 1$ is a global minimum. As we can see, the second iteration suggests just a small S-velocity shift (in a positive sense) in the shallower part and no significant change in the anelastic model. If we take $n = 13$, we get lower Q_μ and some fluctuations in the S-velocity model. Still, the overall features, compared to the starting model of the first iteration, seems stable in the sense of positive/negative anomaly for both elastic and anelastic structure.

We now consider why the second iteration produces such a small change. The inversion process is non-linear, but the data themselves have a large degree of noise due to scattering. Due to this noise, the first iteration seems to have resolved almost all observable information. Thus the second iteration does not significantly change to model obtained by the first iteration. Thus the non-linearity affects the estimates of the variance reductions, but not the

models themselves.

6.1.3. Statistical estimation

From eq. (28), we can rewrite the variance for each SVD expansion coefficient e_j as follows:

$$\text{var. } (e_j) = \frac{\sigma_d^2}{\lambda_j^2}. \tag{43}$$

If we assume that σ_d is the variance of the data itself, we can estimate the nominal error bars δe_j for each of the 70 coefficients:

$$\delta e_j = \frac{\sigma_d}{\lambda_j}. \tag{44}$$

Fig. 17 shows that the nominal error bars are small compared to the expansion coefficients. Of course in any inversion of this type, it is well known that the real errors may be many times greater than the nominal error bars due to systematic errors (e.g., the use of a 1-D parametrization). Thus the nominal error bars should not be taken at face value.

In order to know how the residual for each record $\delta \mathbf{d}^{(kp)}$ contributes to the model $\delta \mathbf{m}$, we examine the distribution of angles (in the multi-dimensional vector space) between the SVD coefficient vector $\mathbf{e}_n$ and the expansion coefficients for each record $\mathbf{e}_n^{(kp)}$ truncated above n:

$$\cos \theta_n^{(kp)} = \frac{\mathbf{e}_n \cdot \mathbf{e}_n^{(kp)}}{|\mathbf{e}_n| \; |\mathbf{e}_n^{(kp)}|}, \tag{45}$$

where

$$\mathbf{e}_n = (e_1 \; e_2 \;\; \cdots \;\; e_n \;\; 0 \;\; \cdots \; 0)^{\mathbf{T}} \tag{46}$$

and

$$\mathbf{e}_n^{(kp)} = (e_1^{(kp)} \; e_2^{(kp)} \;\; \cdots \;\; e_n^{(kp)} \;\; 0 \;\; \cdots \; 0)^{\mathbf{T}}. \tag{47}$$

As shown by eq. (33), the final model vector $\delta\mathbf{m}$ in eq. (30) can be expressed as the sum of the vector for each station for each event. If all of these individual vectors were nearly parallel, each would yield the same model for each individual waveform for each station for each event. However, this is not the case for the actual inversions. Figs. 18a-c show histograms for $\theta_3^{(kp)}$, $\theta_{21}^{(kp)}$, $\theta_{70}^{(kp)}$ for the small dataset, while Figs. 19a-c show histograms for $\theta_3^{(kp)}$, $\theta_{30}^{(kp)}$, $\theta_{70}^{(kp)}$ for the large dataset. We find that even for $\theta_3^{(kp)}$, the angles between the individual coefficients and the coefficients determined from eq. (27), are much closer to 90° than to 0° and that as we increase the value of n, the average value of $\theta_n^{(kp)}$ increases. This tells us that no single individual datum $\delta\mathbf{d}^{(kp)}$ has sufficient power to determine the structure by itself.

For the expansion coefficients for k-th event, we calculate angles between the SVD coefficient vector $\mathbf{e}_n$ and the expansion coefficients $\mathbf{e}_n^k$ truncated above n:

$$\cos\theta_n^k = \frac{\mathbf{e}_n \cdot \mathbf{e}_n^k}{|\mathbf{e}_n|\ |\mathbf{e}_n^k|}, \tag{48}$$

where

$$\mathbf{e}_n^k = (e_1^k\ e_2^k\ \ \cdots\ e_n^k\ \ 0\ \ \cdots\ 0)^{\mathbf{T}}. \tag{49}$$

Figs. 18d-f show histograms with black bars for θ_3^k, θ_{21}^k, θ_{70}^k for the small dataset, while Figs. 19d-f show histograms with black bars for θ_3^k, θ_{30}^k, θ_{70}^k for the large dataset. We find that the angles are closer to 0° than those for individual waveforms. This means that the combined set of waveforms for each particular event has stronger information than any individual waveform for that event. Figs. 18d-f and Figs. 19d-f also show, with gray bars, the

angles between $\mathbf{e}_n^{k_1}$ and $\mathbf{e}_n^{k_2}$ for all pairs of events ($k_1 \neq k_2$):

$$\cos\theta_n^{(k_1,k_2)} = \frac{\mathbf{e}_n^{k_1} \cdot \mathbf{e}_n^{k_2}}{|\mathbf{e}_n^{k_1}|\ |\mathbf{e}_n^{k_2}|}. \tag{50}$$

We find that there are some events which have similar information on the structure (angles close to 0°). This fact supports the above discussion concerning the need for the factor α in AIC_n^α, because the data are redundant to some extent. However, we also find that if we take larger n, there exist pairs of events which have very different information (angles larger than 90°, i.e., negative correlation of the respective model contributions). This shows that there is a discrepancy between some events for large n and that we therefore should not use all 70 coefficients in our models, which is confirmed by the increase in the AIC values for larger values of n. This discrepancy is probably due to 3-D structure.

Similarly we calculate angles between the SVD coefficient vector $\mathbf{e}_n$ and the expansion coefficients $\mathbf{e}_n^p$ for the p-th station truncated above n:

$$\cos\theta_n^p = \frac{\mathbf{e}_n \cdot \mathbf{e}_n^p}{|\mathbf{e}_n|\ |\mathbf{e}_n^p|}, \tag{51}$$

where

$$\mathbf{e}_n^p = (e_1^p\ e_2^p\ \ \cdots\ e_n^p\ \ 0\ \ \cdots\ 0)^{\mathbf{T}}. \tag{52}$$

Figs. 18g-i show histograms with black bars for θ_3^p, θ_{21}^p, θ_{70}^p for the small dataset, while Figs. 19g-i show histograms with black bars for θ_3^p, θ_{30}^p, θ_{70}^p for the large dataset. We find that the angles are slightly closer to 0° than for the individual waveforms, but not as much as is observed for θ_n^k. This is because the independent information on the depth dependence of the structure is very limited in the waveforms for some stations, as the sources are very close to

each other and the epicentral distances are similar to those for other nearby for each of individual stations. Figs. 18g-i and Figs. 19g-i also show, with gray bars, the angles between $\mathbf{e}_n^{p_1}$ and $\mathbf{e}_n^{p_2}$ for all pairs of stations ($p_1 \neq p_2$):

$$\cos\theta_n^{(p_1,p_2)} = \frac{\mathbf{e}_n^{p_1} \cdot \mathbf{e}_n^{p_2}}{|\mathbf{e}_n^{p_1}|\ |\mathbf{e}_n^{p_2}|}. \tag{53}$$

We find (gray bars in Figs. 18d and 18g) that $\theta_3^{p_1,p_2}$ has one peak in the range 0-20° and another is around 60-80° for the small dataset, and that $\theta_3^{p_1,p_2}$ for the large dataset are widely distributed from 0° to 180°. The smaller angles $\theta_3^{p_1,p_2}$ are primarily for stations or events near to one another, suggesting the partial redundancy of the dataset. As n increases, θ_n^p is normally distributed around 90° because of the lack of resolving power for the larger basis set, which includes eigenvectors for smaller eigenvalues. We conclude that in general each record contributes weakly to the inversion results, and that a large number of waveforms must be analyzed systematically as a single dataset in order to extract reliable information on the dataset. Thus visual inspection of individual seismogram pairs (synthetic and observed) is unlikely to be edifying.

6.1.4. Quality control stacks

As waveform inversion is a "black box" statistical procedure, we desire some way to visually confirm the reasonableness of the inversion results. We therefore calculate "quality control stacks" of both the data and synthetics. We align the waveforms using the onset time after applying the same time shift (static correction) as used in the inversion. Note that these stacks are merely ancillary presentations for purposes of visualization of the results of the inversions, and are not used in the inversions. Also note that both the

data and synthetics have been stacked in the same way. From Fig. 20, we see that for SVD3 the onset-to-peak time of the arrival time has been modified because this model has a positive S-wave velocity anomaly in the MTZ. For SVD21 which has a lower Q_μ, the fit of the amplitudes is improved.

6.2. Philippines Sea Region

The sky-blue lines in Fig. 22 show the inversion results for the Philippines Sea West region using the "small dataset" (136 records) and the sky-blue lines in Fig. 22 show the inversion results for the same region using the "large dataset" (400 records). For the small dataset SVD15 is the preferred model for $\alpha = 2, 3, 4$, while for the large dataset SVD17 is the preferred model for $\alpha = 2, 5, 10$ (see Fig. 21a,f). Although the SH model obtained using the large dataset shows fluctuating features, both models have a positive anomaly in the MTZ; Q_μ is around 50 throughout the MTZ for both models.

The blue lines in Fig. 22 show the inversion results for the Philippines Sea Middle region for the "small dataset" (233 records) and the blue lines in Fig. 22 show the inversion results for the same region for the "large dataset" (471 records). For the small dataset SVD14 is the preferred model for $\alpha = 4$, while for the large dataset SVD13 is the preferred model for $\alpha = 10$ (see Fig. 21b,g). Although the S-wave velocity model shows fluctuating features which may be caused by laterally heterogeneous structure in the target region, the average velocity model is higher than PREM in the MTZ. The Q_μ model shows very low value (around 50) throughout the MTZ.

The purple lines in Fig. 22 show the inversion results for the Philippines Sea East region using the "small dataset" (156 records) and the purple lines in Fig. 22 show the inversion results for the same region using the "large

dataset" (419 records). For the small dataset SVD14 is the preferred model for $\alpha = 3, 4$, while for the large dataset SVD17 is the preferred model for $\alpha = 5, 10$ (see Fig. 21c,h). The S-wave velocity in the MTZ shows a positive anomaly in the MTZ and in the upper mantle. Q_μ is around 50 throughout the MTZ.

6.3. Pacific Sea Region

The red lines in Fig. 22 show the inversion results for the Philippines Sea West region for the "small dataset" (163 records) and the red lines in Fig. 22 show the inversion results for the same region using the "large dataset" (645 records). For the small dataset SVD32 is the preferred model for $\alpha = 1, 2, 3, 4$, while for the large dataset SVD33 is the preferred model for $\alpha = 5, 10$ (see Fig. 21d,i). From both inversions we can consistently see a positive SH anomaly in the MTZ with a kink above the 400 km discontinuity probably connected with a discontinuity depth change. Q_μ is again very low but in the MTZ is relatively larger (50-100) than in other subregions.

The orange lines in Fig. 22 show the inversion results for the Philippines Sea East region for the "small dataset" (87 records) and the orange lines in Figs. 22 show the inversion results for the same region for the "large dataset" (348 records). For the small dataset SVD20 is the preferred model for $\alpha = 2, 3, 4$, while for the large dataset SVD25 is the preferred model for $\alpha = 5, 10$ (see Fig. 21e,j). The inversion results show a fluctuating S-wave velocity model in the MTZ but the average velocity is not far from PREM. However, around 400 km depth we can see a slight negative S-wave velocity anomaly. Q_μ is very low but for the inversion using the large dataset, there might be a relatively large Q_μ (50-100) in the MTZ. In this subregion,

the difference between the attenuation models obtained using the small and large datasets is the strongest. The reason for this discrepancy might be that the target subregion includes a strong lateral heterogeneity because of the complicated geometry beneath the Izu-Bonin-Mariana Arc.

7. Discussion and conclusions

We have developed and applied methods for localized waveform inversion. As we have automated the data-handling procedures, our software is sufficiently robust to carry out inversion for 3-D structure in the near future.

In this study we performed quasi-iterative inversion for anelasticity by first determining an "Initial Q_μ" model and then performing inversions for 6 subregions. The fluctuations in our S-wave velocity models are due to the complicated 3-D structure and relatively poorer resolving power for the depth dependence of the actual inversions as compared to the synthetic resolution tests. However, we can obtain sufficient information on the average velocity structure in general to suggest the existence of a positive velocity anomaly in the MTZ beneath all of the 6 subregions. Of course this feature should not be regarded as having been definitely established to the omission of 3-D structure in our inversions.

This is consistent with other long-period wave tomographic models (e.g., Mégnin and Romanowicz, 2000; Takeuchi, 2007). The magnitude of the positive anomaly is stronger beneath the Philippines Sea than the Pacific. The negative velocity anomaly found in the layer from 300 to 400 km depth beneath the Pacific East region might be related to the negative P-wave velocity anomaly proposed by Zhao (2004) and Obayashi et al. (2006).

A lower Q_μ (50-100) than PREM, even lower than the "Initial Q_μ" model itself, is obtained from the inversions for the 6 subregions. There are two possible reasons for caution about these low Q_μ values: (1) the fact that we assumed the source time function is a δ-function; (2) the fact that we ignored the effects of 3-D elastic structure, which could also lessen the amplitude of the waveforms due to scattering and de-focusing effects, which are likely to occur during travel along a slab. $Q_\mu = 50$ is thus probably a lower limit for the average value in the MTZ beneath these region (for comparison, previous studies suggests a Q_μ of around 100-200 for the upper mantle or the MTZ: but they also found lower Q_μ for this region than other regions). If $Q_\mu = 50$, the partial derivative of the seismic velocity with respect to temperature is doubled from that for PREM ($Q_\mu = 143$) in the MTZ (Karato, 1993).

We also find a weak sub-regional dependence of Q_μ structure although further work is required to confirm this. Beneath the Philippines Sea region, we find the $Q_\mu \sim 50$ throughout the MTZ down to 660 km, while we find an increasing Q_μ in the deep part (500-660 km) beneath the Pacific. This may reflect the water rich content beneath the mantle wedge driven down to the lowermost part of MTZ in the Philippines Sea.

Heretofore almost all studies of body-wave waveforms for regional upper mantle or transition zone structure have been based on forward modeling of individual profiles. Such studies have made many important contributions, but there are limits because the forward modeling approach cannot systematically analyze large amounts of data at one time. In contrast, statistical approaches based on waveform inversion, as shown in this paper, can systematically analyze large volumes of data to invert not only for seismic velocities

but also for anelastic structure. Even when there is considerable misfit between observed data and synthetics, the use of statistically-based waveform inversion makes it possible to obtain robust models. Further information can be expected through application of waveform inversion to 3-D Earth models.

8. Data and Resources

All seismograms used in this study come from the National Research Institute for Earth Science and Disaster Prevention F-net at http://www.hinet.bosai.go.jp/fnet/ (last accessed December 2008).

9. Acknowledgments

We thank the staff of the National Research Institute for Earth Science and Disaster Prevention for providing seismograms. We thank K. Konishi for his cooperation. We thank S. Karato, H. Iwamori, F. Cammarano, and H. Mizutani for fruitful discussions. NF is supported by a JSPS Fellowship for Young Scientists.

References

Akaike, H., 1977. Extension of method of maximum likelihood and the Stein's problem Annals of the Institute of Statistical Mathematics 29, 153-164.

Azimi, S.A., Kalinin, A.V., Kalinin, V.V., Pivovarov, B.B., 1968. Impulse and transient charactereristics of media with linear and quadratic absorption laws. Izv. Earth Physics., (English Translation), 2, 88 93.

Brudzinski, M.R., Chen, W.P., Nowack, R.L., Huang, B.S., 1997. Variations of P wave speeds in the mantle transition zone beneath the northern Philippine Sea. J. Geophys. Res. 102, 11815-11827.

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

Cummins, P.R., Geller, R.J., Takeuchi, N., 1994b. DSM complete synthetic seismograms: P-SV, spherically symmetric, case. Geophys. Res, Lett. 21, 1663-1666.

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.

Dalton, C.A., Ekström, G., Dziewonski, A.M., 2008. The global attenuation structure of the upper mantle. J. Geophys. Res. 113, B09303, doi:10.1029/2007JB005429.

Dziewonski, A.M., Anderson, D.L., 1981. Preliminary reference Earth model. Phys. Earth Planet. Inter. 25, 297-356.

Faul, U.H., Jackson, I., 2005. The seismological signature of temperature and grain size variations in the upper mantle. Earth Planet. Sci. Lett. 234, 119-134.

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., Helmberger, D.V., 1984. Upper mantle shear structure beneath the Northwest Atlantic-ocean. J. Geophys. Res. 89, 1465-1475.

Gung, Y., Romanowicz, B., 2004. Q tomography of the upper mantle using three component long-period waveforms. Geophys. J. Int. 157, 813-830.

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., 2004. Waveform inversion for 3-D earth structure using the Direct Solution Method implemented on vector-parallel supercomputer. Phys. Earth Planet. Inter. 146, 65-74.

Karato, S., 1993. Importance of Anelasticity in the Interpretation of Seismic Tomography. Geophys. Res. Lett. 20, 1623-1626.

Kato, M., Misawa, M., Kawakatsu, H., 2001. Small subsidence of the 660- discontinuity beneath Japan probed by ScS reverberations. Geophys. Res. Lett. 28, 447-450.

Kawai, K., Geller, R.J., 2009. Impurities in Mg-perovskite in the lowermost mantle beneath the Arctic: Evidence from waveform inversion. Phys. Earth Planet. Inter. submitted (this issue).

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., 2007a. 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.

Kawai, K., Geller, R.J., Fuji, N., 2007b. D″ beneath the Arctic from inversion of shear waveforms. Geophys. Res. Lett. 34, L21305, doi:10.1029/2007GL031517.

Konishi, K., Kawai, K., Geller, R.J., Fuji, N., 2009. MORB in the lowermost mantle beneath the western Pacific: Evidence from waveform inversion. Earth Planet. Sci. Lett. 278, 219-225.

Lay, T., Helmberger, D.V., 1983. A lower mantle S-wave triplication and the shear velocity structure of D″. Geophys. J. R. Astr. Soc. 75, 799-837.

Mégnin, 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.

Menke, W., 1984. Geophysical data analysis: discrete inverse theory. Academic Press, Inc., Orlando, 260 pp.

Obayashi, M., Sugioka, H., Yoshimitsu, J., Fukao, Y., 2006. High temperature anomalies oceanward of subducting slabs at the 410-km discontinuity. Earth Planet. Sci. Lett. 243, 149-158.

Okal, E.A., Geller, R.J., 1979. On the observability of isotropic seismic sources: The July 31, 1970 Colombian earthquake. Phys. Earth Planet. Inter. 18, 176-196.

Revenaugh, J., Jordan, T.H., 1991. Mantle layering from ScS reverberations, 3. The upper mantle. J. Geohpys. Res. 96, 19781-19810.

Romanowicz, B., 1994. On the measurement of anelastic attenuation using amplitudes of low-frequency surface-waves. Phys. Earth Planet. Inter. 84, 179-191.

Romanowicz, B., 1995. A global tomographic model of shear attenuation in the upper-mantle. J. Geophys. Res. 100, 12375-12394.

Romanowicz, B., 1998. Attenuation tomography of the earth's mantle: A review of current status. Pure Appl. Geophys. 153, 257-272.

Tajima, F., Grand, S.P., 1998. Variation of transition zone high-velocity anomalies and depression of 660 km discontinuity associated with subduction zones from the southern Kuriles to Izu-Bonin and Ryukyu. J. Geophys. Res. 103, 15015-15036.

Takeuchi, N., Geller, R.J., 2000. Optimally accurate second order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media. Phys. Earth Planet. Int. 119, 99-131.

Takeuchi, N., 2007. Whole mantle SH velocity model constrained by waveform inversion based on three-dimensional Born kernels. Geophys. J. Int. 169, 1153-1163.

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. Inter. 119, 25-36.

Wiggins, R.A., 1972. General linear inverse problem - implication of surface-waves and free oscillations for earth structure. Rev. Geophys. Space Phys. 10, 251-285.

Zhao, D., 2004. Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics. Phys. Earth Planet. Inter. 146, 3-34.

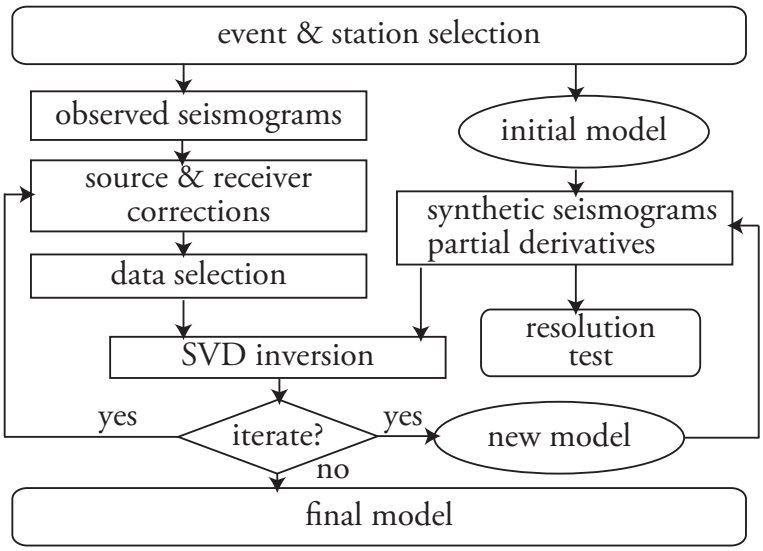

Figure 1: Our procedures for waveform inversion. We have developed software which automatically performs the above steps.

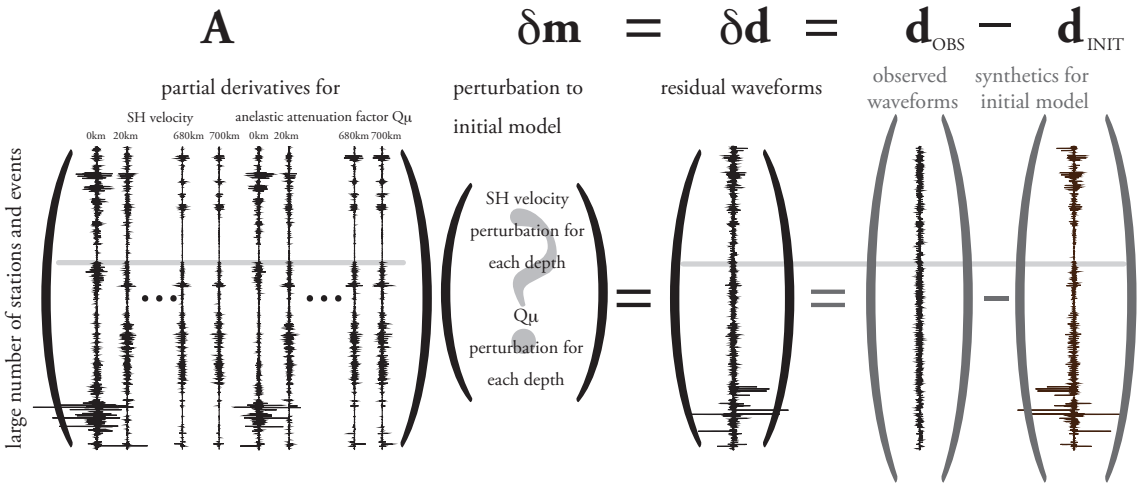

Figure 2: Schematic interpretation of the equation (eq. 12) being solved by waveform inversion. The residuals and partial derivatives for the various events are gathered in a single vector $\delta\mathbf{d}$ and a single matrix ($\mathbf{A}$) respectively. We then perform a simultaneous inversion for the perturbation to the model parameters ($\delta\mathbf{m}$) using the singular value decomposition (SVD).

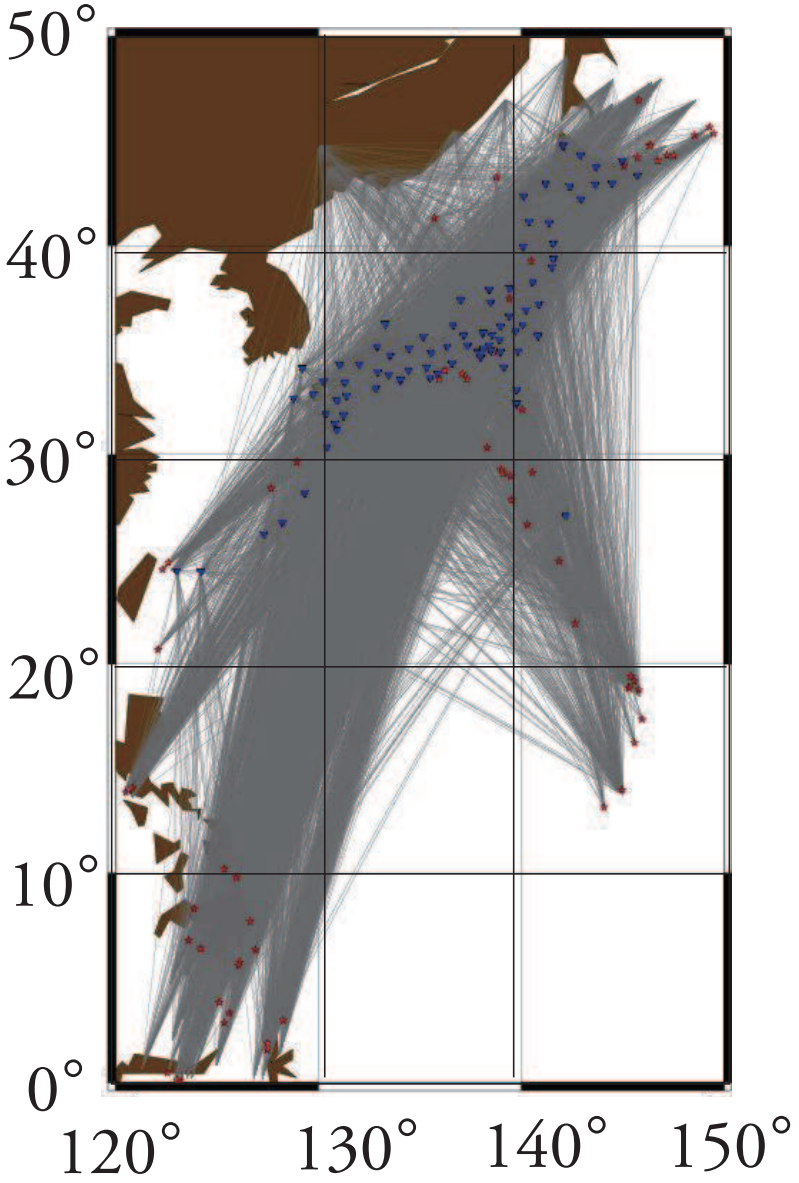

Figure 3: Dataset for this study. Blue dots denote F-net stations and red dots denote events with an epicentral distance between 5° and 40°. We use the entire dataset in the inversion for the "Initial 1-D Q_μ" model. We then divide the dataset into several subsets in order to invert for more localized structure.

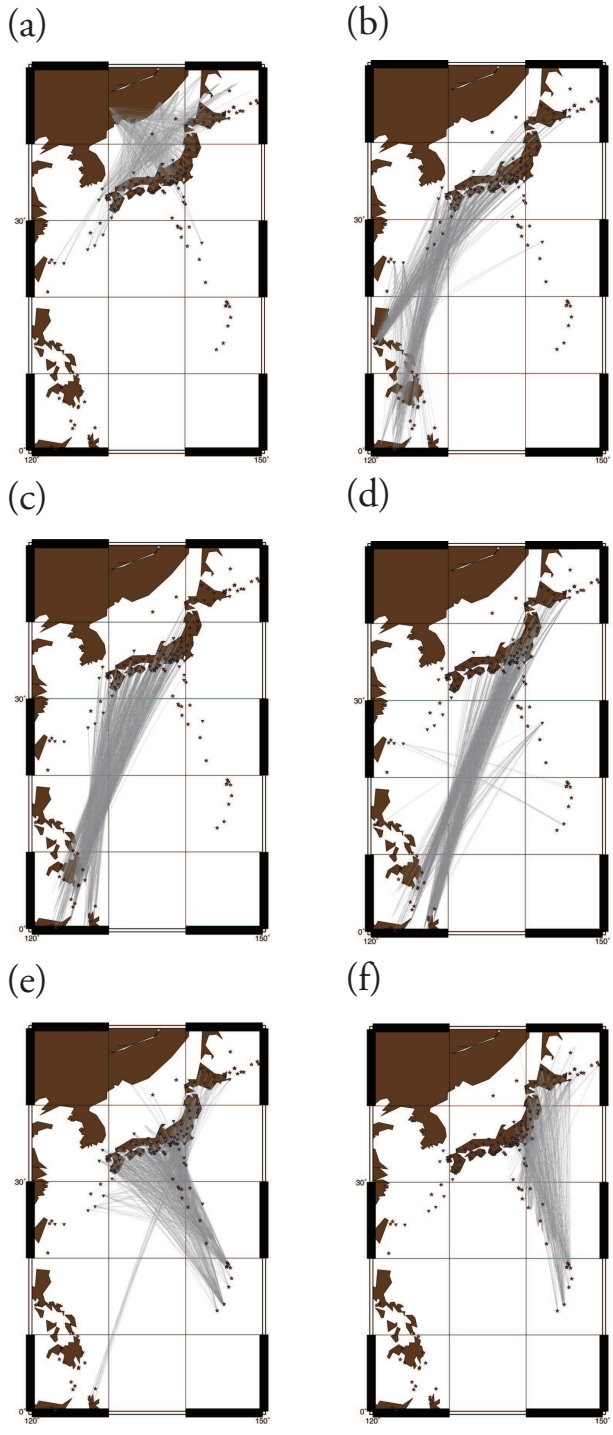

Figure 4: The six subregions for this study, and the number of waveforms used in the inversion for each subregion. (a) Japan Sea: 478 waveforms; (b) Philippines Sea West: 644; (c) Philippines Sea Middle: 437; (d) Philippines Sea East: 688; (e) Pacific West: 1176; (f) Pacific East: 525.

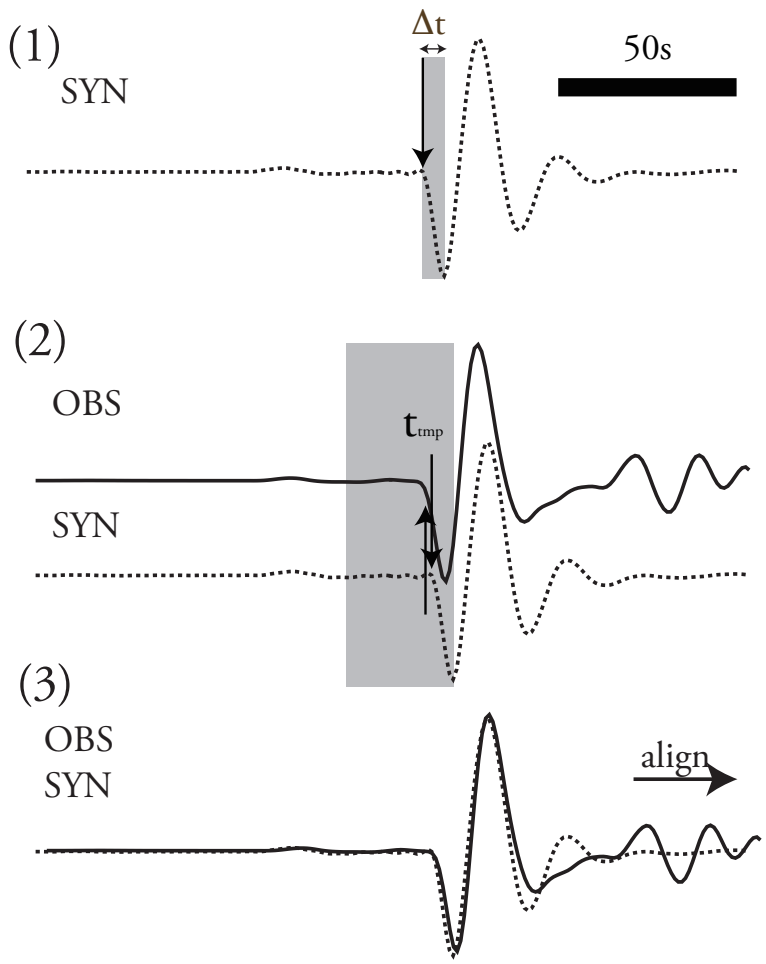

Figure 5: Procedure for static corrections. Step 1: Autopick the first arrival on the synthetics (the up arrow) and let the zero-to-peak time be Δt. Step 2: Use the time window from $4\Delta t$ before to Δt after the arrival) to seek the time shift $t_{\rm tmp}$ (the down arrow) which gives the best correlation coefficients. Step 3: Find the local zero around $t_{\rm tmp}$ in the observed seismogram and let this be the arrival time with which we align the synthetic.

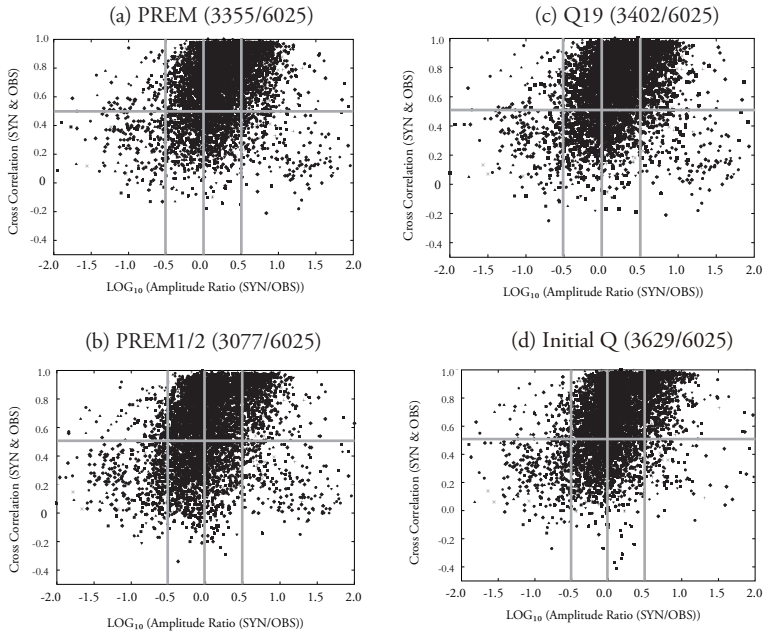

Figure 6: Plots of the amplitude ratios and cross correlations (the maximum values with a time lag in the range -10.0 s to 10.0 s) for pairs of observed and synthetic seismograms determined from 4 sets of synthetics, each calculated with a different Q_μ model. The different symbols (triangles, squares, etc.) denote individual events used in this study. The grey lines denote where the amplitude ratio is 3.0 ($\log_{10} 3.0 = 0.48$), 1.0, ($\log_{10} 1.0 = 0.0$), 0.33 ($\log_{10} 0.33 = -0.48$) and where the cross correlation coefficient is 50 %. The number in the parentheses above each panel show the number of waveforms which satisfy a criterion of 0.33 to 3.0 maximum amplitude ratio between the observed seismogram and the corresponding synthetic seismogram calculated for the the model, and also have a cross-correlation coefficient above 50 %. The second number is the total number of records in the dataset. As the first numbers in the parentheses show, the "Initial Q_μ" model is the suitable than the other 3 models in terms of the amplitude ratios and cross correlation coefficients.

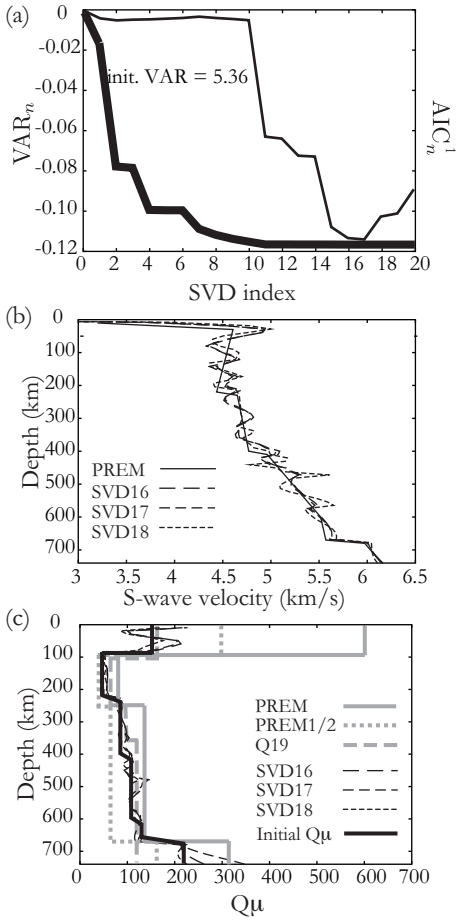

Figure 7: Inversion results for the dataset of all records with amplitude ratios (OBS/INIT) between 0.1 and 10.0 and cross correlation coefficients above 50 %. (a) Variance reduction (thick curve) and AIC_n^1 value (thin curve). (b) S-velocity for PREM (the starting elastic model for the inversion) and the inversion results for $n = 16$, $n = 17$ and $n = 18$. (c) Q_μ model for PREM (the starting anelastic model for the inversion) as well as "PREM1/2" and "Q19" for reference, and the inversion results obtained using $n = 16$, $n = 17$ and $n = 18$. The smoothed model "Initial Q_μ" obtained from the inversion is used as a starting anelastic model for subsequent inversions.

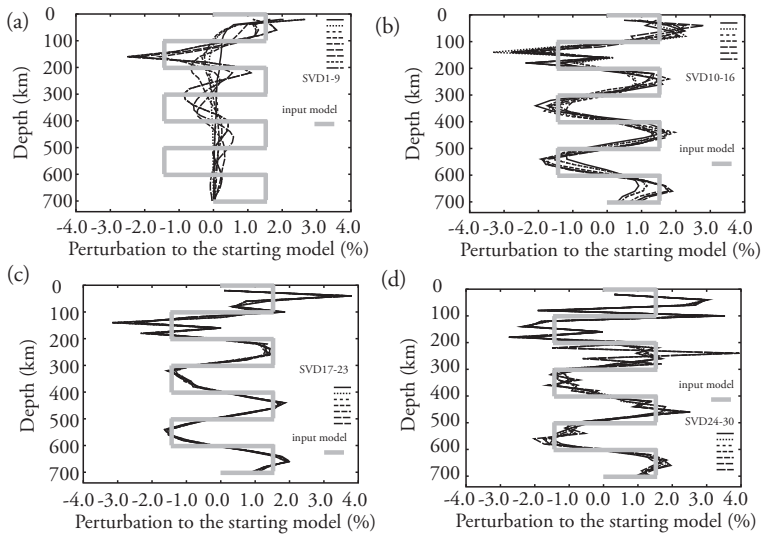

Figure 8: Synthetic inversion results for a dataset perturbed only with respect to the velocity structure (with no perturbation to Q_μ for this figure). Gray line denotes the input model and black lines are inversion results. Panel (a) shows models for $n = 1$ through 9, panel (b) for $n = 10$ through 16, panel (c) for $n = 17$ through 23, and panel (d) for $n = 24$ through 30. Panel (b) appears best, and panel (c) appears reasonable, while panel (a) cannot satisfactorily reproduce the input model and panel (d) fluctuates too much.

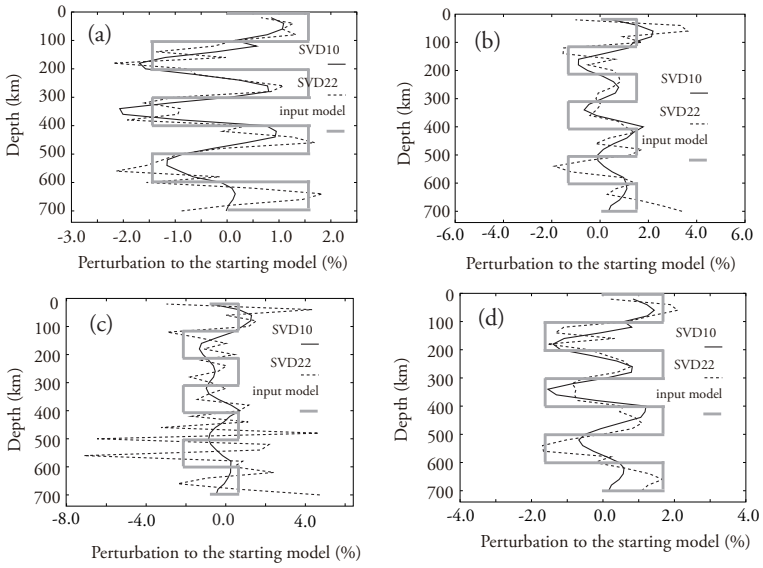

Figure 9: Synthetic inversion results for several perturbed datasets. Details are as same as Fig. 8 (only SVD10 and SVD22 are shown) except: (a) the input synthetic data are shifted 2 seconds; (b) shifted 3 seconds; (c) shifted 4 seconds; (d) randomly shifted (up to 10 seconds back and forward)

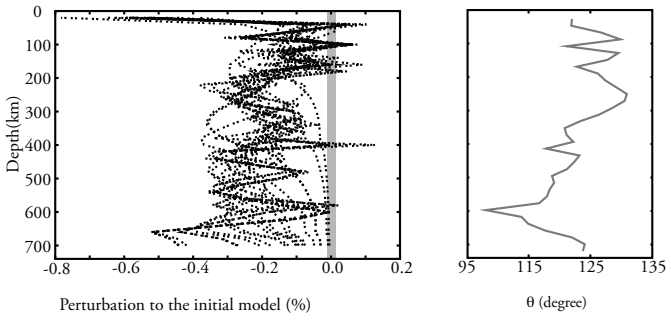

Figure 10: Synthetic inversion results for the systematically 1 s shifted seismograms (left). The time shift does not significantly affect the inversion results because the inner product of the partial derivatives with respect to the depth dependence of SH velocity and the partial derivatives for the time shift is small, albeit non-zero (right).

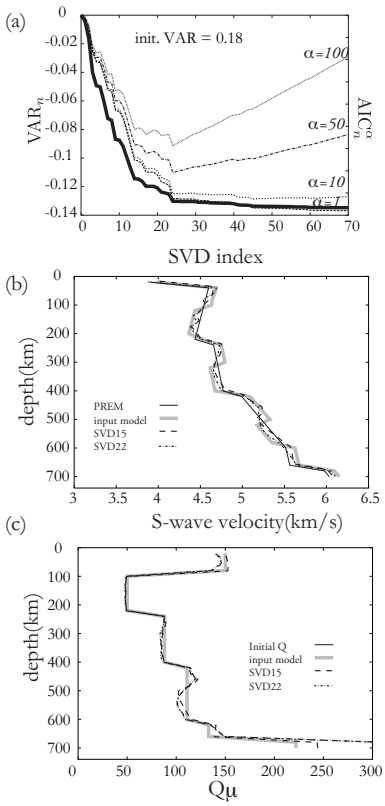

Figure 11: Results of inversion of synthetic data for a perturbation to only the elastic part of the model. (a) Variance reduction (thick curve) and AIC_n^α values (thin curves). Note that the AIC_n^α curves have been adjusted to have the same value for $n = 1$, as only the relative values of AIC_n^α for a given α are meaningful. Note that the scale for the variance reduction and AIC_n^α are respectively on the left and right hand sides of the box. (b) S-velocity for PREM (the starting elastic model for the inversion), the perturbed ("checkerboard test") input model used to computed the synthetic "observed data" for the test inversion, and the inversion results for $n = 15$ and $n = 22$. (c) "Initial Q" model (starting point for the inversion as well as the Q model used to compute the synthetic "observed data" (as Q was not perturbed in the input model), and the Q models obtained by the inversions for $n = 15$ and $n = 22$ (basically unchanged from the starting model for the synthetic inversion).

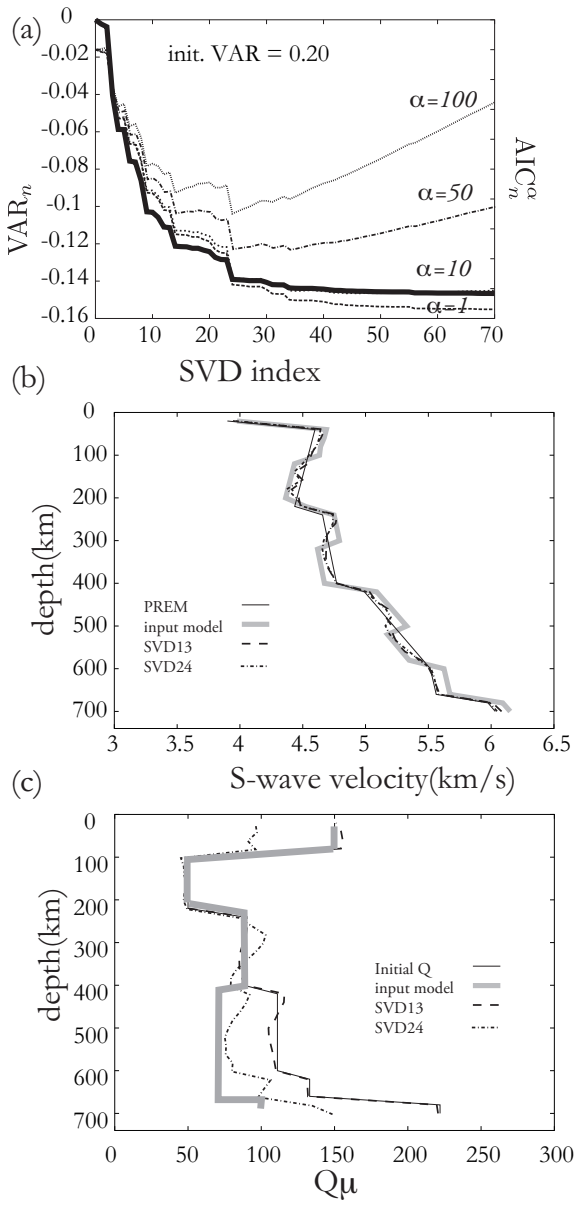

Figure 12: Same as Fig. 11, except that the model used to compute the synthetic "observed data" included perturbations to both the anelastic and elastic parts of the starting model for the synthetic inversion (PREM and "Initial Q" respectively).

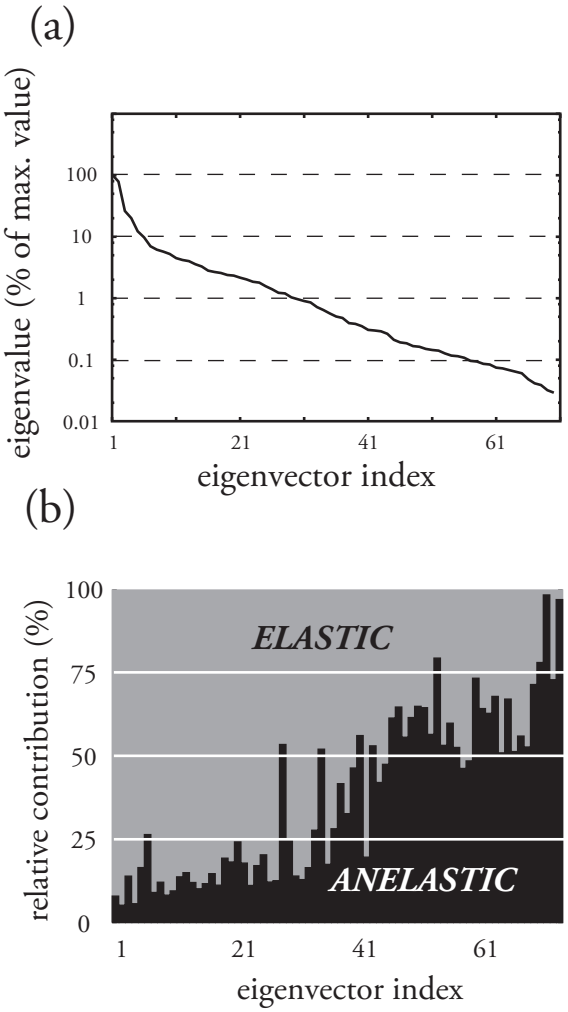

Figure 13: (a) the absolute value of the eigenvalues for the SVD inversion (plotted on a logarithmic scale); (b) the relative contribution of the elastic and anelastic components of the eigenvectors.

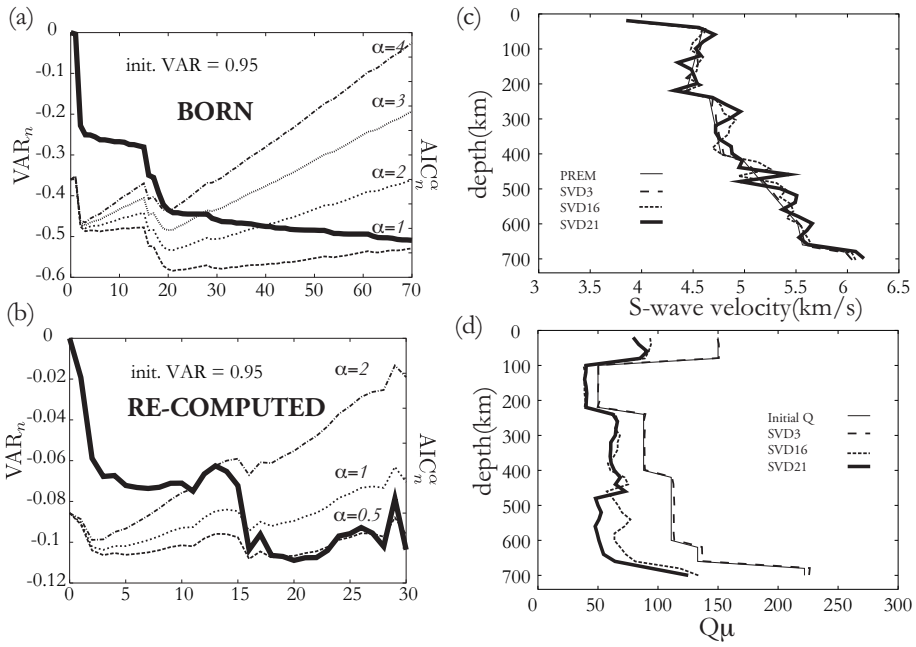

Figure 14: Results of inversion of the "small dataset" for the Japan Sea region. (a) Variance reduction (thick curve) (computed using the 1st order Born approximation) and AIC_n^α values (thin curves). (b) Re-computed variance reduction (thick curve) and AIC_n^α values (thin curves) up to $n \leq 30$. Note that we plot the re-computed variance reduction and AICs only in this figure. $\mathrm{AIC}_n^{0.5}$ is also shown as a reference. (c) S-velocity for PREM (the starting elastic model for the inversion), and the inversion results for $n = 3$, $n = 16$ and $n = 21$. (d) "Initial Q_μ" model (starting point for the inversion) and the inversion results for $n = 3$, $n = 16$ and $n = 21$.

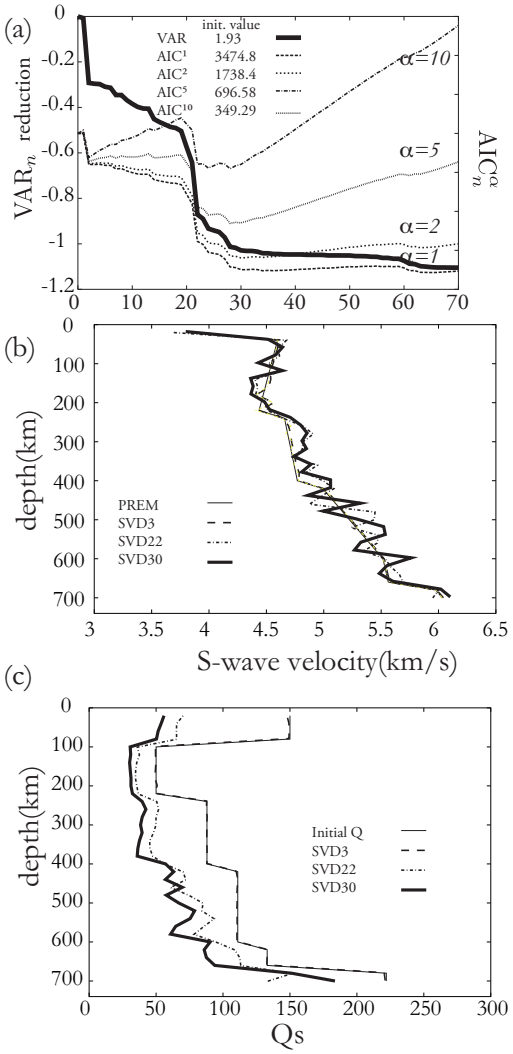

Figure 15: Inversion result for Japan Sea Region for the "large dataset."

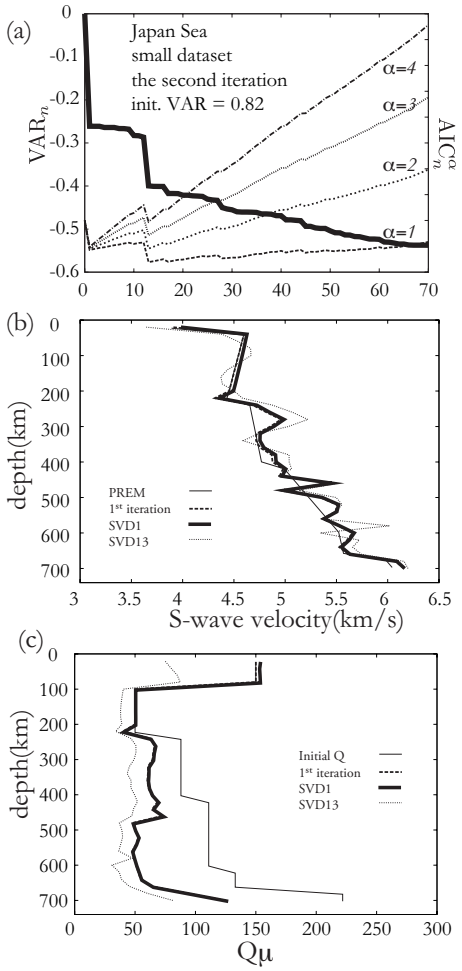

Figure 16: Results of the second iterative inversion of the "small dataset" for the Japan Sea region. (a) Variance reduction (thick curve) (computed using the 1st order Born approximation) and AIC_n^α values (thin curves). (b) S-velocity for PREM (the starting elastic model for the first iteration), the starting elastic model for the second iteration which represents the SVD21 model in the first iterative inversion below 200 km depth (above 200 km we fix to PREM), and the second inversion results for $n = 1$, $n = 13$. (c) "Initial Q_μ" model (the starting anelastic model for the first iteration), the starting anelastic model for the second iteration which represents the SVD21 model in the first iterative inversion below 200 km depth (above 200 km we fix to PREM), and the second inversion results for $n = 1$, $n = 13$.

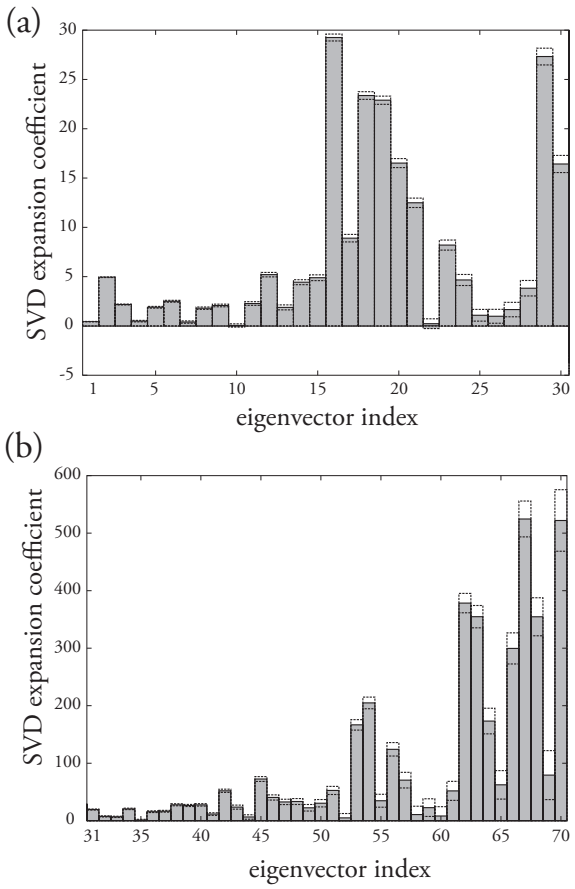

Figure 17: SVD expansion coefficients e_j for the inversion for the small dataset. Gray bars show the estimated values and two dotted bars show the error bars for each coefficient. Eq. (26) shows that e_j tends to increase with the eigenvector index j. We thus show this histogram in two columns: (a) from the first to thirtieth modes; (b) from the thirty-first to the seventieth modes.

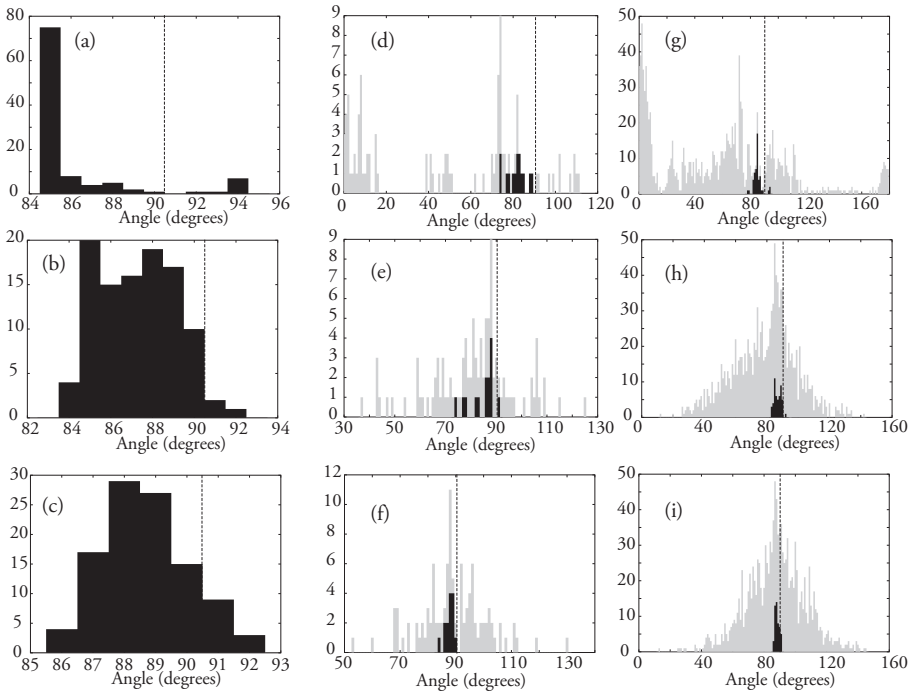

Figure 18: For the small dataset we plot histograms for the angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^{(kp)}$ for (a) $\theta_3^{(kp)}$, (b) $\theta_{21}^{(kp)}$, and (c) $\theta_{70}^{(kp)}$; angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^k$ for each event (black bars), and angle between the vector $\mathbf{e}_j^{k_1}$ and $\mathbf{e}_j^{k_2}$ (gray bars) for (e) $j = 3$, (e) $j = 21$ and (f) $j = 70$; angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^p$ for each station (black bars), and angle between the vector $\mathbf{e}_j^{p_1}$ and $\mathbf{e}_j^{p_2}$ (gray bars) for (g) $j = 3$, (h) $j = 21$ and (i) $j = 70$. Note that in each panel, the dotted line denotes an angle of 90°.

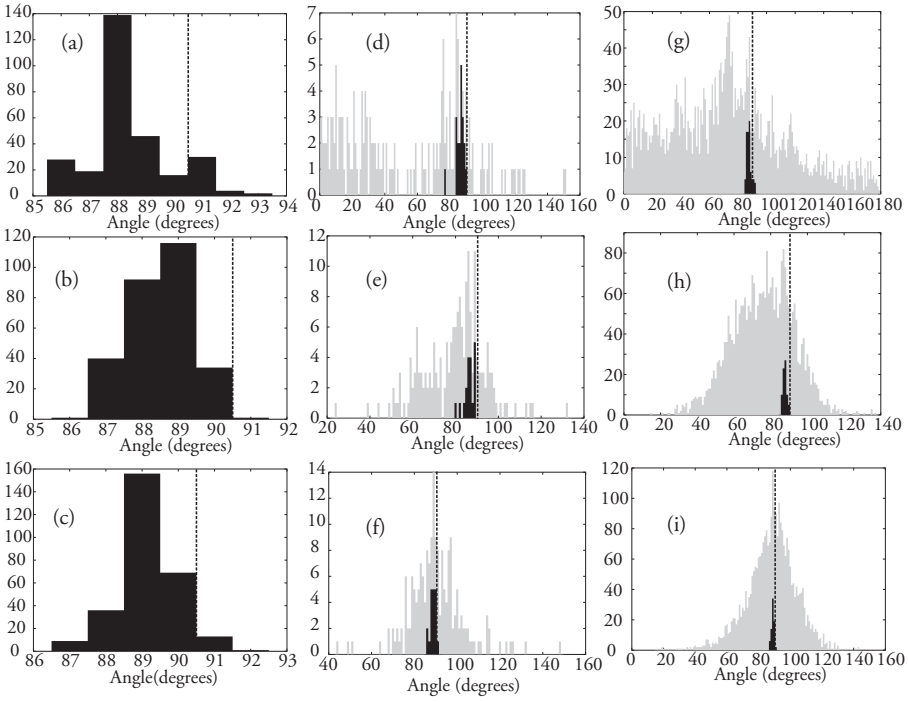

Figure 19: For the large dataset we plot histograms for the angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^{(kp)}$ for (a) $\theta_3^{(kp)}$, (b) $\theta_{30}^{(kp)}$, and (c) $\theta_{70}^{(kp)}$; angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^k$ for each event (black bars), and angle between the vector $\mathbf{e}_j^{k_1}$ and $\mathbf{e}_j^{k_2}$ (gray bars) for (e) $j = 3$, (e) $j = 30$ and (f) $j = 70$; angle between the vector $\mathbf{e}_j$ and $\mathbf{e}_j^p$ for each station (black bars), and angle between the vector $\mathbf{e}_j^{p_1}$ and $\mathbf{e}_j^{p_2}$ (gray bars) for (g) $j = 3$, (h) $j = 30$ and (i) $j = 70$. Note that in each panel, the dotted line denotes an angle of 90°.

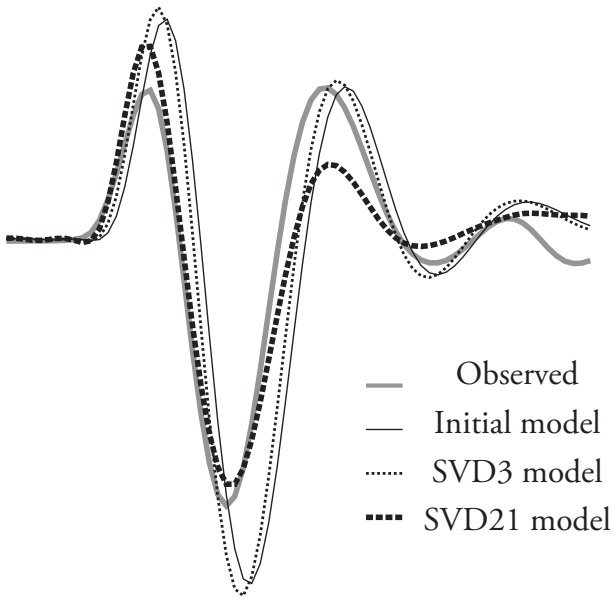

Figure 20: Stacked waveforms for the event 20041107. Synthetic waveforms are calculated using the 1st order Born approximation.

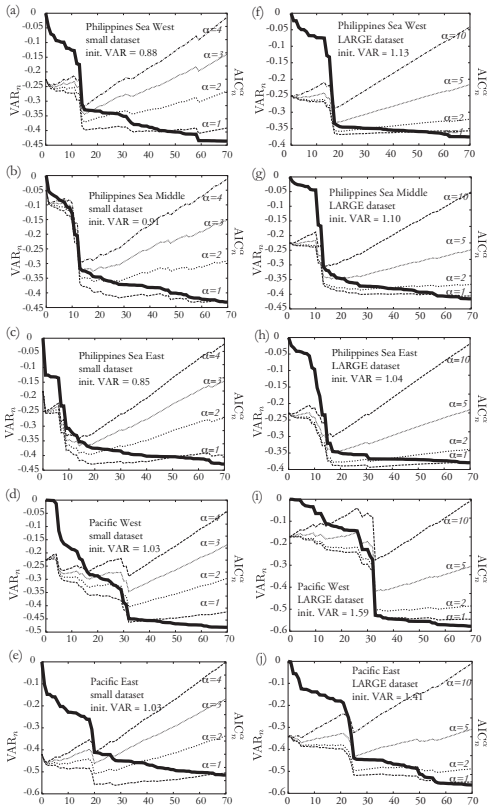

Figure 21: Variance reduction (thick curve) and AIC_n^α values (thin curves) for inversion of the "small dataset" (a-e) and "large dataset" (f-j) for five sub-regions: (a,f) show the results of the Philippines Sea West Region, (b,g) show the results of the Philippines Sea Middle Region, (c, h) show the results of the Philippines Sea East Region, (d,i) show the results of the Pacific West Region, (e, j) show the results of the Pacific East Region.

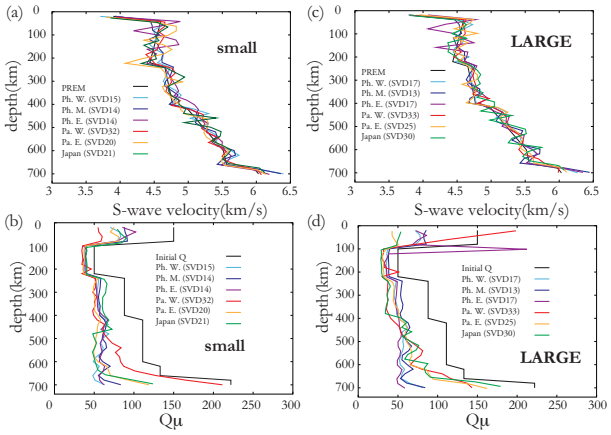

Figure 22: Results of inversion of the "small dataset" (a, b) and "large dataset" (c, d) for the six sub-regions. (a, c) S-velocity for PREM (black curve: the starting elastic model for the inversion), and the preferred inversion result (colored curve). (b, d) "Initial Q" model (thin curve: starting point for the inversion) and the preferred inversion result (colored curve). Sky blue curves are for the Philippines Sea West Region, blue curves for the Philippines Sea Middle Region, purple curves for the Philippines Sea East Region, red curves for the Pacific West Region, orange curves for the Pacific East Region, green curves for the Japan Sea Region.