Geophysical Journal International

*Geophys. J. Int.* (2014) **196,** 1706–1723
Advance Access publication 2013 December 17

doi: 10.1093/gji/ggt475

# Teleseismic migration with dual bootstrap stack

Jun Korenaga

*Department of Geology and Geophysics, Yale University, New Haven,* 06520 *CT, USA. E-mail: jun.korenaga@yale.edu*

Accepted 2013 November 21. Received 2013 November 18; in original form 2013 September 19

## SUMMARY

A recently developed stacking scheme called dual bootstrap stack (DBS) is applied to teleseismic migration, using deep earthquakes in South America as sources and USArray as receivers. Built with statistical tests for signal significance and coherence, DBS can achieve a clean separation of signal and noise, and unlike conventional non-linear stacks such as *N*th-root stack (NRS) and phase-weighted stack (PWS), it can recover signals even when the signal-to-noise ratio is lower than unity. Migration results with DBS are compared with those with linear stack, NRS and PWS, and DBS-based migration is shown to be able to detect weak, small-scale scatterers that cannot be imaged by NRS and PWS and cannot be confidently identified by linear stack. Based on migration results reproducible among multiple data sets, a number of small-scale scatterers are suggested to exist in the top 400 km and the bottom 1500 km of the mantle in the study region. Though being time-consuming, teleseismic migration with DBS is suggested to be a promising tool to map out a variety of hitherto undetected small-scale heterogeneities in the mantle.

**Key words:** Time-series analysis; Body waves; Wave scattering and diffraction.



## 1 INTRODUCTION

Seismic migration is an imaging technique in which seismic energy recorded at surface receivers is projected back to its origin in the subsurface, using a background velocity structure. The technique was originally developed in exploration geophysics (e.g. Yilmaz 1987; Sheriff & Geldart 1995), and its adaptation to natural earthquake data is known as teleseismic migration (e.g. Revenaugh 1995; Freybourger *et al.* 2001; Rost & Thomas 2002; Thomas *et al.* 2004; Chambers & Woodhouse 2006; Hutko *et al.* 2008; Lay & Garnero 2011). As it directly uses the entire waveform of seismic data, migration has a potential to achieve the highest spatial resolution allowed by the given data.

The stacking of time-shifted seismic records constitutes the essential part of migration. Linear stack, which is equivalent to taking the arithmetic mean, is most commonly used in teleseismic migration, but it could suffer from high residual noise when the number of seismic traces to be stacked is not sufficiently large for a given signal-to-noise ratio (S/N). The noise level can be lowered by using non-linear stacks such as *N*th-root stack (NRS; Muirhead 1968; Kanasewich *et al.* 1973) or phase-weighted stack (PWS; Schimmel & Paulssen 1997). Kito & Krüger (2001), for example, used a variant of PWS when constructing a migration image for the lowermost mantle beneath the southwestern Pacific. These non-linear stacks, however, cannot detect signals when S/N is lower than unity. Recently, a new kind of non-linear stack, called dual bootstrap stack (DBS), has been developed to overcome these deficiencies of conventional stacks (Korenaga 2013). DBS is based on the hypothesis testing of signal significance and coherence, and it can work well even with a limited number of low S/N records.

The purpose of this paper is to demonstrate the performance of teleseismic migration with DBS, through an example with USArray data. The structure of the paper is as follows. After explaining a migration procedure with DBS, the nature of seismic data sets used in this study is described. Migration results with a single data set are presented first, and the advantage of DBS over conventional stacks is discussed. Though it is found that the probability of random noise to align as a signal is rather high in migration, the effect of such false signals can be minimized by focusing on reproducible signals among multiple data sets. Results based on 10 different data sets suggest the presence of numerous small-scale scatterers concentrated in the top 400 km and the bottom 1500 km of the mantle in the study region. The paper is closed by discussion on remaining issues on teleseismic migration with DBS.

## 2 METHOD

### 2.1 Basic migration procedure

Suppose that seismic energy radiated from an earthquake is received by $N_r$ seismic stations, whose data are denoted as $D_i(t)$ with $i = 1, 2, \ldots, N_r$. Here, $t$ denotes time. For each of $N_g$ nodes distributed in the subsurface, then, a corresponding migration energy may be defined as

$$E_j = \max\left[\left\{\mathbf{S}\left[A(i,j)W(t)D_i(t - t_i^{\mathrm{o}} + t_i^{\mathrm{t}} - t_{ij}^{\mathrm{t}})\right]\right\}^2\right],\qquad j = 1, 2, \ldots, N_g, \quad (1)$$

where $t_i^{\mathrm{o}}$ is the observed arrival time of a reference phase at the $i$th station, $t_i^{\mathrm{t}}$ is the theoretical arrival time of the same phase, $t_{ij}^{\mathrm{t}}$ is the theoretical traveltime from the hypocentre to the $i$th station through the $j$th node, $W(t)$ is a window function, $A(i, j)$ is a weighting factor to take into account the radiation pattern of forward scattering and the operator $\mathbf{S}$ denotes stacking of $N_r$ traces. Here each node is assumed to serve as a potential isotropic point scatterer. The window function may be defined as

$$W(t) = \begin{cases} 1 & \text{if } t \geq 0 \text{ and } t \leq \Delta t \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $\Delta t$ is a dominant period for seismic events under concern. For a regional array at a large epicentral distance, the weighting factor is nearly constant, and it may be set to unity for simplicity.

The calculation of theoretical traveltimes is done with a given reference earth model. In this study, I will focus on $P$-to-$P$ single scattering, so the reference phase is direct $P$, and the traveltime calculation is done solely with the compressional velocity information of a reference model. Migration for other types of scattering is of course possible with modifying the reference phase and/or the traveltime calculation. In eq. (1), the differential traveltime with respect to the reference phase, $t_{ij}^{\mathrm{t}} - t_i^{\mathrm{t}}$, is used to minimize the effect of 3-D velocity variations unaccounted for by an assumed reference model.

### 2.2 Dual bootstrap stack

In this study, the stacking operator $\mathbf{S}$ in eq. (1) can be any of the following: linear stack, NRS, PWS and DBS. Among them, DBS is the most recent, so how it works is briefly summarized here; for a full description, see Korenaga (2013). DBS is a doubly weighted stack defined as

$$\mathbf{S}_{\mathrm{DBS}}[X_i(t)] \equiv w_1(t) w_2(t) \overline{X}(t), \quad (3)$$

where $\overline{X}(t)$ denotes the arithmetic mean of $\{X_i(t); i = 1, 2, \ldots, N_r\}$, $w_1(t)$ is a weighting coefficient determined by a hypothesis test for signal significance and $w_2(t)$ is another coefficient determined by a test for signal coherence. Linear stack can be recovered by setting both coefficients to unity. The first weighting is calculated as

$$w_1 = \max(0, 1 - p_1/\alpha), \quad (4)$$

where $p_1$ is the probability for the null hypothesis that a given stack is indistinguishable from a noise stack, and $\alpha$ is the critical significance level below which the null hypothesis is rejected. The second weighting is calculated similarly as

$$w_2 = \max(0, 1 - p_2/\alpha), \quad (5)$$

where $p_2$ is the probability for the null hypothesis that a given data set is incoherent. In calculating these probabilities, bootstrap resampling is conducted in two different ways, first with the original data set, and second with trace scrambling to generate noise stacks. Based on a number of tests with synthetic data (Korenaga 2013), the critical significance level is set to 0.01, the number of bootstrap replicates is set to $10^3$ and the maximum period of trace scrambling is set to 20 s. That is, if either of the two null hypotheses has a probability higher than 1 per cent, a stack will be weighted down to zero, and time-shifts used for trace scrambling are randomly chosen from between −20 and 20 s.

As demonstrated by Korenaga (2013), DBS allows signal detection even when S/N is low, by exploiting the statistics of noise estimated by trace scrambling. Compared to linear stack, the number of traces required for clear signal identification is smaller by up to two orders of magnitude. Also, unlike linear stack, DBS can reject incoherent signals. Non-linear stacks such as NRS and PWS work well only when S/N is higher than unity, because they are based on the polarity statistics of data as a whole. When data are dominated by noise (i.e. S/N is lower than unity), these conventional non-linear stacks fail to identify a stacked signal as a true signal. This weakness of NRS and PWS remains even if the number of traces is large (see fig. 9 of Korenaga 2013).

The superior performance of DBS, however, comes with computational cost. The two weighting coefficients, $w_1$ and $w_2$, are a function of time, so at every sample in a given time-series, bootstrap-based hypothesis tests need to be conducted. As the number of nodes in a 3-D target volume is typically on the order of $10^6$, teleseismic migration with DBS would be extremely time-consuming if the original implementation of DBS is directly employed. Noting that only the maximum of stacked energy is important in eq. (1), however, some acceleration is possible. A linearly stacked waveform in a given window typically has only a few local maxima, and DBS may be considered only for those maxima. This local-maxima approach can reduce the computational cost by about one order of magnitude, and it is adopted in this study.

## 3 DATA

From four deep events in South America and the USArray network, 10 different data sets are prepared as shown in Table 1. The configuration of USArray is time-dependent, so restricting the spatial extent of each receiver array in a similar manner (Fig. 1) makes it more straightforward to compare results from different data sets. Also, for such subarrays with limited aperture, the migration procedure of eq. (1) can be simplified by setting the weighting factor $A(i, j)$ to unity.

A bandpass filter with the corner frequencies of 0.1 and 0.5 Hz was applied to broad-band seismic data, and data were resampled

**Table 1.** Data sets used in this study.

| ID | Source | USArray station range | $N_r$[a] |
|---|---|---|---|
| 1 | 2005 March 21/12:23:54.09, (24.98°S, 63.47°W), $z = 579$ km, $M$w 6.9 | 32°N–40°N, 130°W–113°W | 58 |
| 2 | 2006 September 22/02:32:25.64, (26.87°S, 63.15°W), $z = 598$ km, $M$w 6.0 | 32°N–40°N, 125°W–115°W | 114 |
| 3 | ” | 40°N–50°N, 125°W–115°W | 89 |
| 4 | 2006 November 13/01:26:35.87, (26.05°S, 63.28°W), $z = 572$ km, $M$w 6.8 | 30°N–38°N, 125°W–110°W | 92 |
| 5 | ” | 38°N–43°N, 125°W–110°W | 125 |
| 6 | ” | 43°N–50°N, 125°W–110°W | 99 |
| 7 | 2008 September 3/11:25:14.45, (26.74°S, 63.23°W), $z = 570$ km, $M$w 6.3 | 37°N–48°N, 125°W–115°W | 148 |
| 8 | ” | 42°N–49°N, 116°W–105°W | 205 |
| 9 | ” | 30°N–42°N, 120°W–110°W | 171 |
| 10 | ” | 28°N–42°N, 110°W–100°W | 335 |

[a]Number of seismic stations.

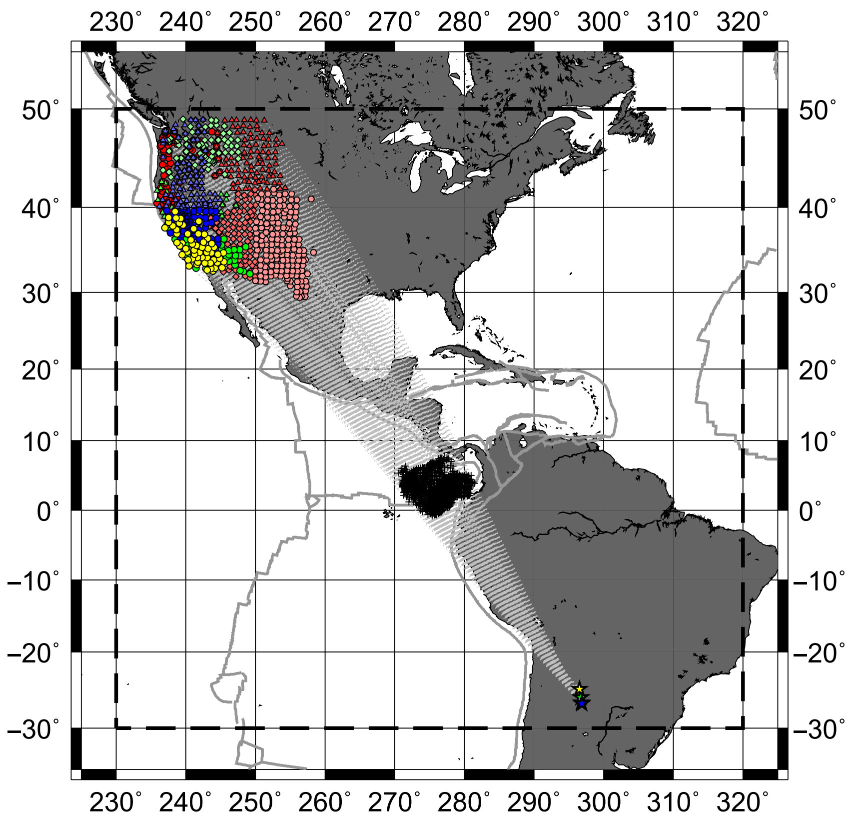


**Figure 1.** Locations of the events (stars) and stations (smaller symbols) for data used in this study. Symbols for stations are as follows: data set #1 (yellow circle), #2 (blue circle), #3 (blue triangle), #4 (green circle), #5 (green triangle), #6 (green diamond), #7 (red circle), #8 (red triangle), #9 (red diamond) and #10 (red hexagon). Ray paths of direct *P* waves are indicated by dashed grey curves, and *PP* bounce points are shown by crosses. Grey lines show plate boundaries, and thick dashed lines denote the extent of a target volume for migration.

at every 0.1 s. The data of each station were then normalized by the maximum amplitude of the direct *P* wavelet, and station data were aligned by cross-correlating direct *P* wavelets. A seismic station was rejected as too noisy if the root-mean-square amplitude before the direct *P* arrival was greater than 0.15. The number of good stations after this screening is listed in Table 1. The example of filtered data from the data set #5 is shown in Fig. 2. Based on the duration of the first main wavelet of direct *P*, the length of the time window for stacking is set to 6 s (eq. 2).

With the resampling rate of 10 Hz, the number of samples per trace in the stacking window is 60. The direct application of DBS would thus require 60 calls of DBS per every node in the migration volume. In the acceleration based on local maxima (Section 2.2), DBS is calculated only at local maxima detected by linear stack. Because the number of local maxima in a 6-s-long stacked waveform rarely exceeds 10 at any given node, this local-maxima approach reduces the number of DBS calls by about one order of magnitude, thereby making DBS-based migration computationally more feasible.

The choice of these particular data sets is somewhat arbitrary, as the most important point to be addressed in this paper is the relative performance of DBS with respect to convectional stacking schemes, in the context of teleseismic migration. Which part of the mantle is to be studied is a secondary issue. Still, some may prefer to revisit a region with known heterogeneities. Previously detected small-scale heterogeneities in the mantle (e.g. Kaneshima & Helffrich 1999) are, however, based mostly on signals that can be identified even before stacking, and such high-S/N signals do not provide a critical test bed to compare different stacking schemes (*cf.* fig. 12 of Korenaga 2013). As shown in this paper, the strength and novelty of DBS-based migration is to be able to detect hitherto undetected signals. The adopted data sets are appropriate in this regard.

## 4 RESULTS WITH SINGLE DATA SET

In this section, migration results with the data set #5 are presented. The purpose of showing results with a single data set here is twofold; the simplicity of the procedure involved facilitates the comparison of different stacking schemes, and the results also highlight the difficulty of avoiding false signals, even with DBS, when using only one data set. The latter issue will be alleviated by using multiple data sets, but it is important to first understand the nature of the problem, which is particularly prominent in migration.

A target volume for migration, which is common to all cases considered in this paper, spans from the surface to the core–mantle boundary and is bounded between 30°S and 50°N and between 130°W and 40°W (Fig. 1). The volume is discretized horizontally with an interval of 0.5° and vertically with an interval of ∼50 km. The total number of nodes in the volume is $\sim 1.7\times 10^6$. For the theoretical calculation of traveltimes, the IASP91 earth model (Kennett & Engdahl 1991) is used.

In what follows, migration results are shown first with data up to 240 s after the direct *P* (Section 4.1). DBS-based results indicate the presence of a large number of blob-like small-scale scatterers, and to eliminate the possibility that they are some kind of artefacts introduced by the standard phases, migration results are repeated

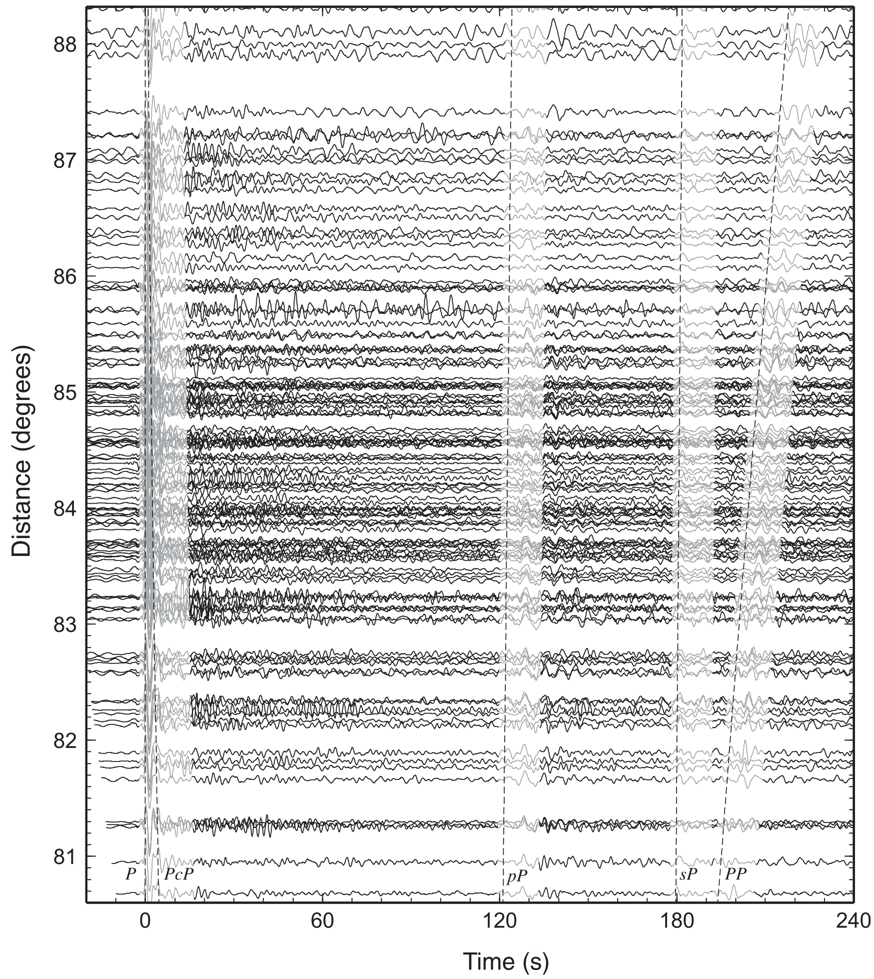


**Figure 2.** Vertical component data of data set #5, plotted in order of epicentral distance. For display, records have been aligned on the direct *P* arrival and normalized to the amplitude of the first maximum of the arrival. Theoretical arrivals of *PcP*, *pP*, *sP* and *PP* (based on IASP91, Kennett & Engdahl 1991) are shown as dashed lines. Masking of the standard phases (Section 4.2) is indicated by grey shading.

using only part of data bounded by *P* and *pP* (up to ∼120 s, Fig. 2; Section 4.2). The small-scale scatterers are still seen with DBS. To better understand the nature of these scatterers, two kinds of tests are considered. A test with synthetic point scatterers is conducted to analyse the extent of isochronal artefacts (Section 4.3), and this analysis leads to a cluster analysis to estimate the minimum number of point scatterers. Another test is conducted with randomized data to quantify the probability of false signal (Section 4.4), which is shown to be rather high. This test underscores the importance of verifying reproducibility using multiple data sets, which is the subject of Section 5.

### 4.1 Migration with unmasked data

Migration results using data up to 240 s after the direct *P* are shown at selected depths in Fig. 3, for the cases of linear stack, NRS, PWS and DBS. Throughout this paper, NRS is done with the power of 3, and PWS is done with the order of 2; results shown in this paper remain the same with different exponents, because different exponents affect waveforms but not the maximum energy. The data length of 240 s was chosen here to include the *PP* phase, and there is no fundamental restriction on data length. As expected, the strongest migration energy with near-zero dB is observed along the ray paths of direct *P*. The migration with linear stack is generally characterized by a much higher energy level than those with non-linear stacks, and numerous small-scale features seen in the former are mostly absent from the latter, suggesting that these small-scale features are artefacts. The most prominent artefact is the one caused by the *pP* phase, which occurs about 120 s after the direct *P* (Fig. 2) and is seen as an envelope around the direct *P* energy in the migration image (Fig. 3). Because the *pP* phase is of relatively large amplitude, its artefact can be recognized even with the non-linear stacks though it is much attenuated.

Grey shading in Fig. 3 denotes the region that cannot be probed by *P*-to-*P* single scattering with 240-s-long data. Around the edges of the migratable region, especially around the southern edge, high migration energy can be seen with linear stack, NRS and PWS. This high energy around the edges is another artefact caused by the stacking of a relatively small number (<∼20) of station data, because other stations are out of reach in terms of single scattering. When the number of traces is too small, noise reduction by stacking is not effective even with NRS and PWS. It is notable, therefore, that this kind of artefact is entirely absent with DBS. This attests to the efficacy of DBS in distinguishing clearly between signal and noise based on stringent statistical tests.

Excluding the direct *P* image, the *pP* artefact, and edge artefacts, migration results with conventional non-linear stacks (NRS and

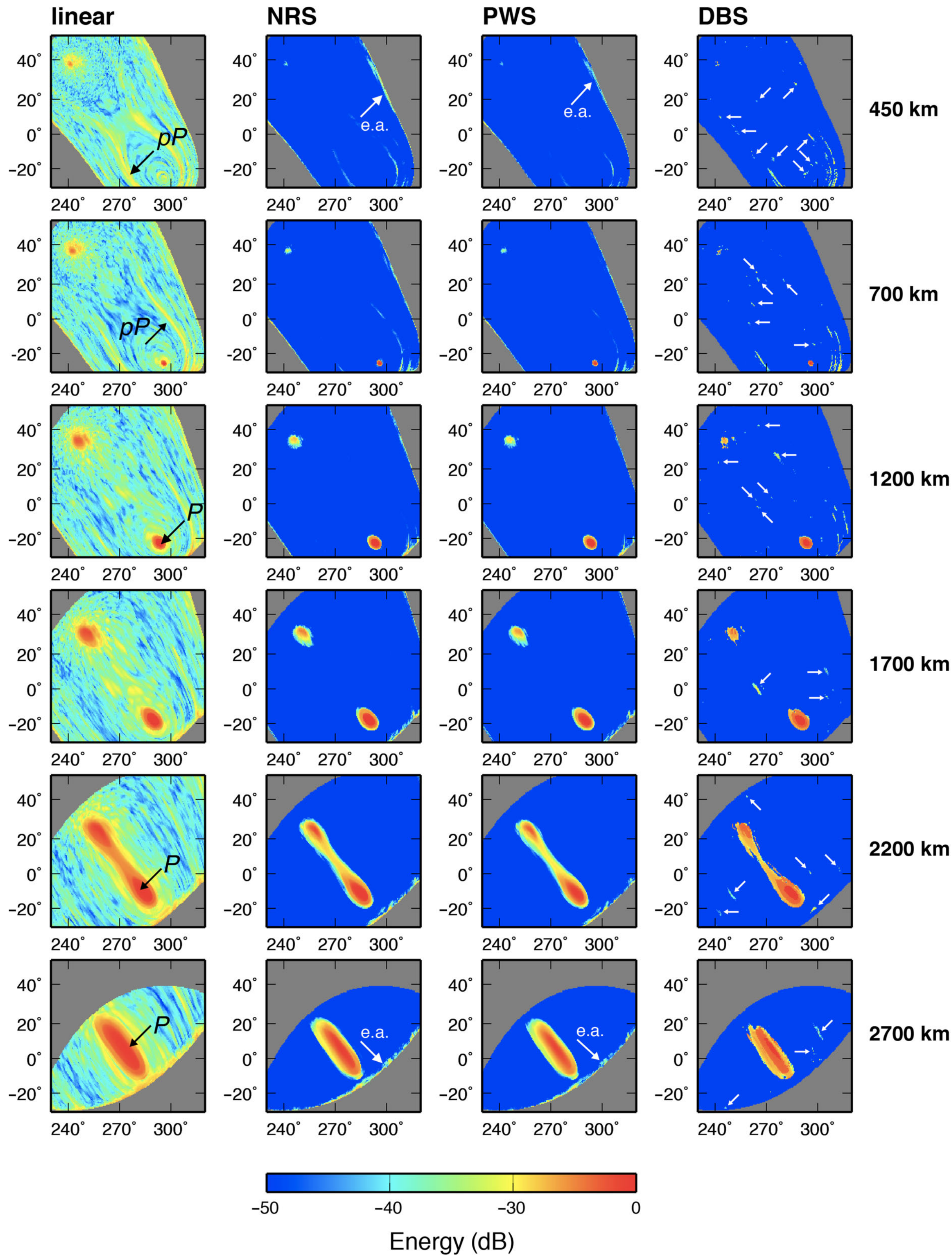


**Figure 3.** Migration energy for data set #5 at selected depths, with different stacking schemes. From left to right-hand panels: linear stack, *N*th-root stack (NRS), phase-weighted stack (PWS) and dual bootstrap stack (DBS). Grey shading denotes the region that cannot by sampled by *P*-to-*P* single scattering with data up to 240 s after direct *P*. 'e.a.' denotes edge artefacts. Arrows in DBS migration images point to some of small-scale features that are not detected by NRS and PWS (and are hard to identify by linear stack because of other numerous migration artefacts). Migration energy is shown in the unit of decibel (dB), that is, $10\log_{10}E$, where zero dB corresponds to the migration energy of direct *P*.

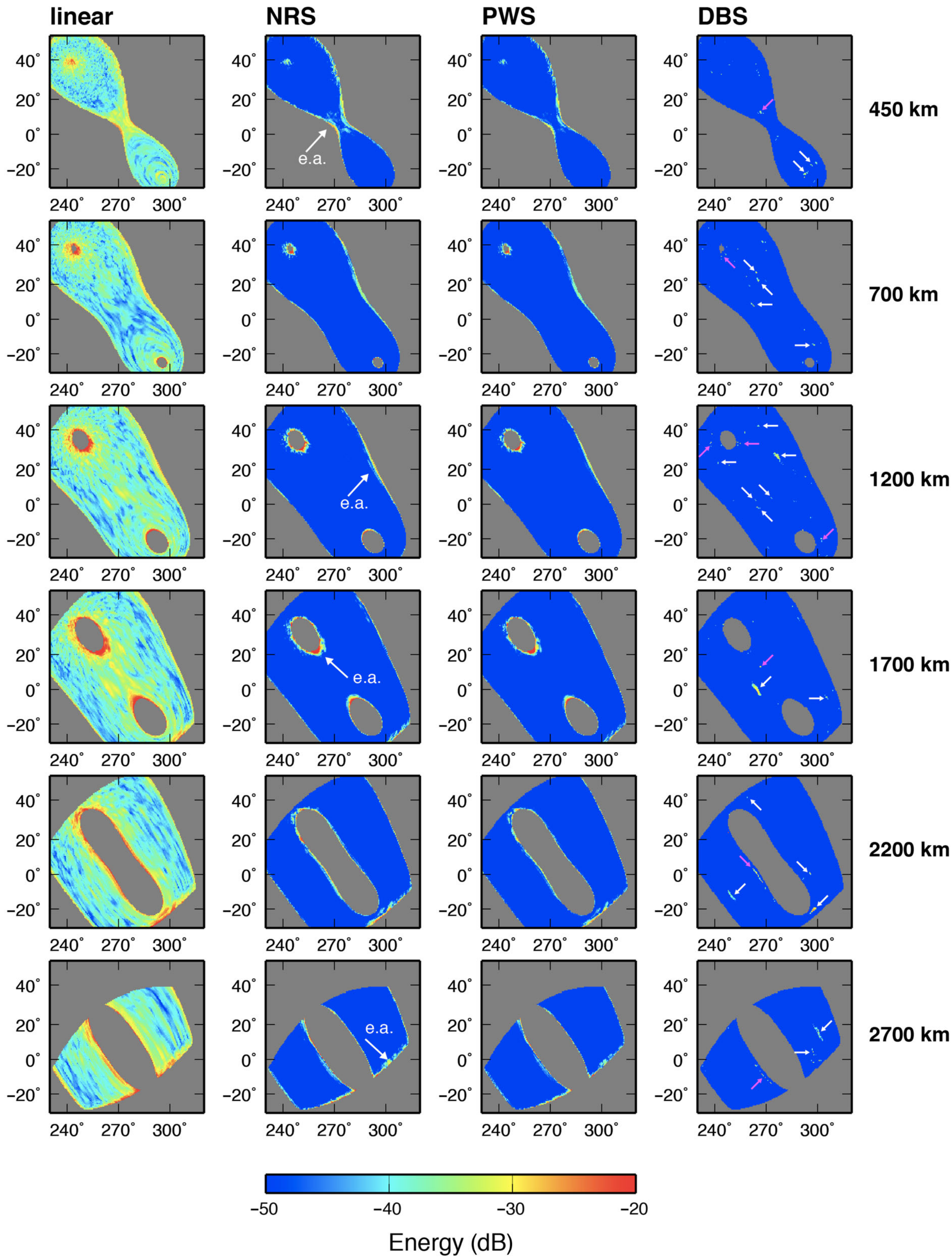


**Figure 4.** Same as Fig. 3 but with masking of the standard phases and with data up to 120 s after direct *P*. Pink arrows in DBS migration images are small-scale features not previously seen with unmasked data.

PWS) contain no signal. The DBS-based migration, however, is different, exhibiting quite a few small-scale features; that is, some of numerous local maxima seen with linear stack are judged as signal by DBS. NRS and PWS can detect signal only when S/N is greater than unity, whereas DBS can when (S/N) $n^{1/2}$, where $n$ is the number of traces, is greater than ∼2 (Korenaga 2013). In this case, the number of traces is 125, so DBS can detect signal if S/N is greater than ∼0.18. It is thus quite possible that these small-scale features imaged by DBS correspond to hitherto undetected seismic heterogeneities in the mantle. To scrutinize the reliability of these subtle signals, migration is repeated with more restricted data, as shown next.

### 4.2 Migration with masked data

The standard phases such as *P*, *PcP* and *pP* are masked in the seismic data by setting the amplitude to zero from 2 s before to 12 s after the theoretical traveltime of each phase (Fig. 2). The depth phases such as *pP* and *sP*, however, may not be entirely removed by this masking if their coda waves are too prolonged, so to minimize migration artefacts originating in the standard phases, the duration of data is further limited to the first 120 s after *P*. The spatial extent of the migratable region decreases accordingly.

Migration results with the masked data are shown in Fig. 4. Those with NRS and PWS exhibit only edge artefacts with virtually no signal, whereas that with DBS maintains most of small-scale features seen in Fig. 3. Indeed, a close comparison between DBS results with and without masking reveals that additional small-scale features are detected with the masked data; they were not detected as signal before because the stacking of subtle signals was disturbed by the presence of the large-amplitude standard phases.

To quantify the magnitude of a detected signal with respect to the background noise, an apparent S/N is calculated as

$$(\mathrm{S/N})_{\mathrm{ap}} = \left(\frac{E}{R^2 - E}\right)^{1/2}, \tag{6}$$

where $E$ is the migration energy and $R$ is the root-mean-square amplitude of corresponding data before stack. This is only a lower bound for a true S/N because signal recovery by stacking is not always perfect. For example, a true S/N of 1 would be reduced to an apparent value of ∼0.4 and ∼0.07, respectively, with a signal recovery rate of 0.5 and 0.1. For DBS results, the apparent S/N is mostly below unity (Fig. 5a). The true S/N of those signals should not much higher than unity, however, because such high S/N signals should have been detected by NRS or PWS. In terms of the apparent (S/N) $n^{1/2}$, about half of detected signals are below the detection limit of DBS (∼2, as determined by Korenaga 2013; Fig. 5b). Those signals below the threshold clearly have the signal recovery rate below unity. In other words, their true values of (S/N) $n^{1/2}$ must be higher than ∼2; otherwise they would not have been detected. The apparent (S/N) $n^{1/2}$ of 2 is thus a convenient threshold to identify significant signal with high recovery rate, which in tern indicates that true (S/N) $n^{1/2}$ would be sufficiently high (∼5 or higher). With this threshold, the migration energy of significant signals is greater than −40 dB (Fig. 5b); that is, the amplitude of those signals is greater than 1 per cent of the maximum amplitude of direct *P*.

### 4.3 Test with synthetic point scatterers

DBS can distinguish between signal and noise more reliably than conventional stacks, but this does not mean that DBS-based migration is free of artefacts. All kinds of migration suffer from isochronal artefacts to a varying degree, and DBS-based one is no exception. Scattered waves recorded at different receivers would add up most constructively when projected back to a true scattering point, but similarly good backprojection can also be achieved at neighbouring points having similar traveltimes to the receivers. In this study, the collection of such neighbouring points is referred to as an isochronal volume, and the smearing of a true point scatterer into its isochronal volume is called isochronal artefact.

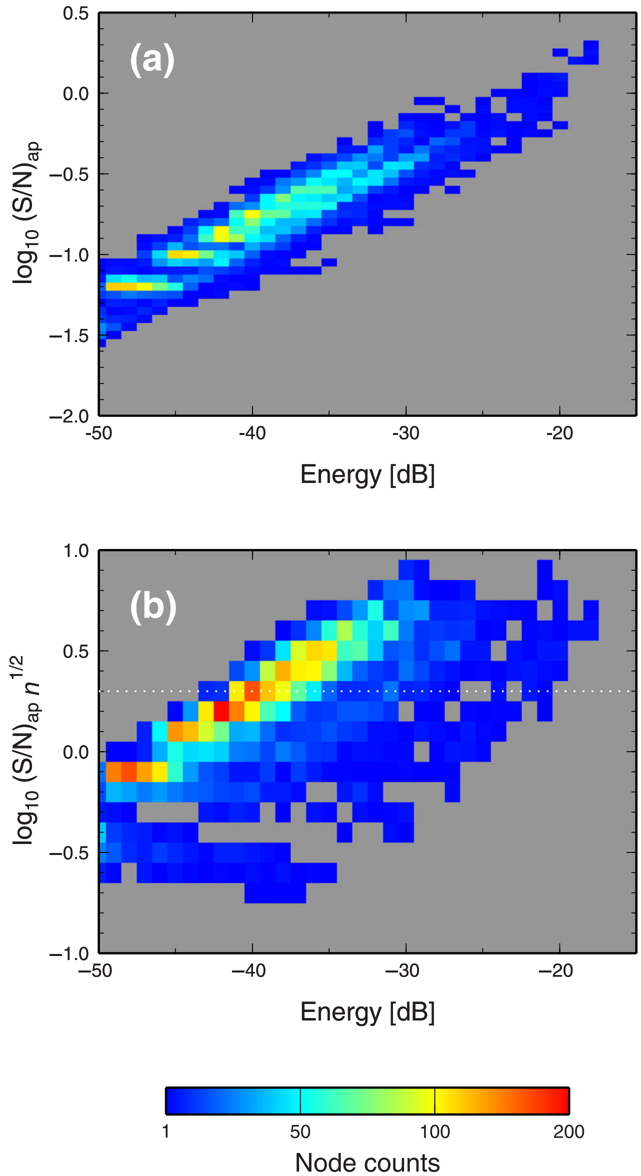


**Figure 5.** 2-D histograms for (a) apparent S/N and migration energy and (b) apparent (S/N) $n^{1/2}$ and migration energy, for DBS-based migration results shown in Fig. 4. Dashed line in (b) denotes the critical value of 2.0, above which signals are considered as significant; see text for details.

The extent of an isochronal volume varies with the location of a scatterer and also with a source–receiver geometry. To understand isochronal artefacts expected for the source–receiver geometry of the data set #5, migration with synthetic data is conducted. Point scatterers are placed at 30° by 20° and at the depths of 250, 850, 1450, 2050 and 2650 km (Fig. 6), and corresponding scattered waves are simulated by embedding Ricker wavelets with the period of 5 s. The number of point scatterers in the migratable region is 29 in total. Gaussian noise is then added, and the synthetic data are processed by the same bandpass filter used for the real data. Two kinds of data sets are produced, one with S/N of 1 and the other with S/N of 0.5, and the maximum amplitude of Ricker wavelets (after the bandpass filtering) is set to 0.0333 and 0.0167, respectively (i.e. the corresponding migration energy of −30 and −36 dB). Data

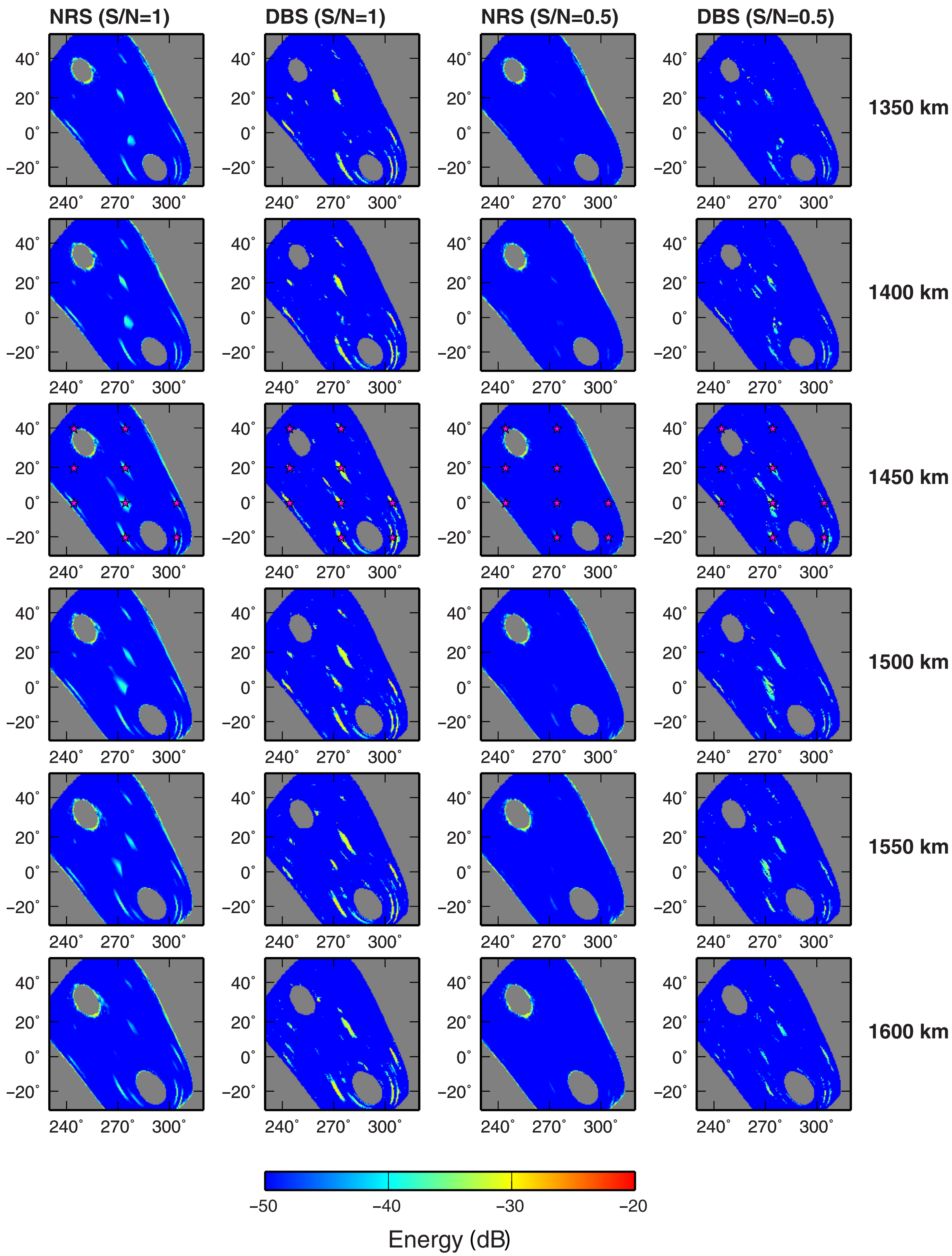


**Figure 6.** Results of signal recovery tests with two types of synthetic data: S/N of 1 (left half) and S/N of 0.5 (right half). Migration results are shown at depths ranging from 1350 to 1600 km, for the cases of NRS and DBS. The locations of true point scatterers at 1450 km are denoted by stars.

masking corresponding to the standard phases is applied as in the previous section.

Migration results with the synthetic data are shown in Fig. 6 for the depth of 1350–1600 km. Isochronal artefacts originating from the scatterers located at the depth of 1450 km are seen in all of these cross-sections. As noted earlier, migration with NRS suffers from edge artefacts as well. For the synthetic data with S/N of 1, both NRS and DBS show a similar distribution of isochronal artefacts,

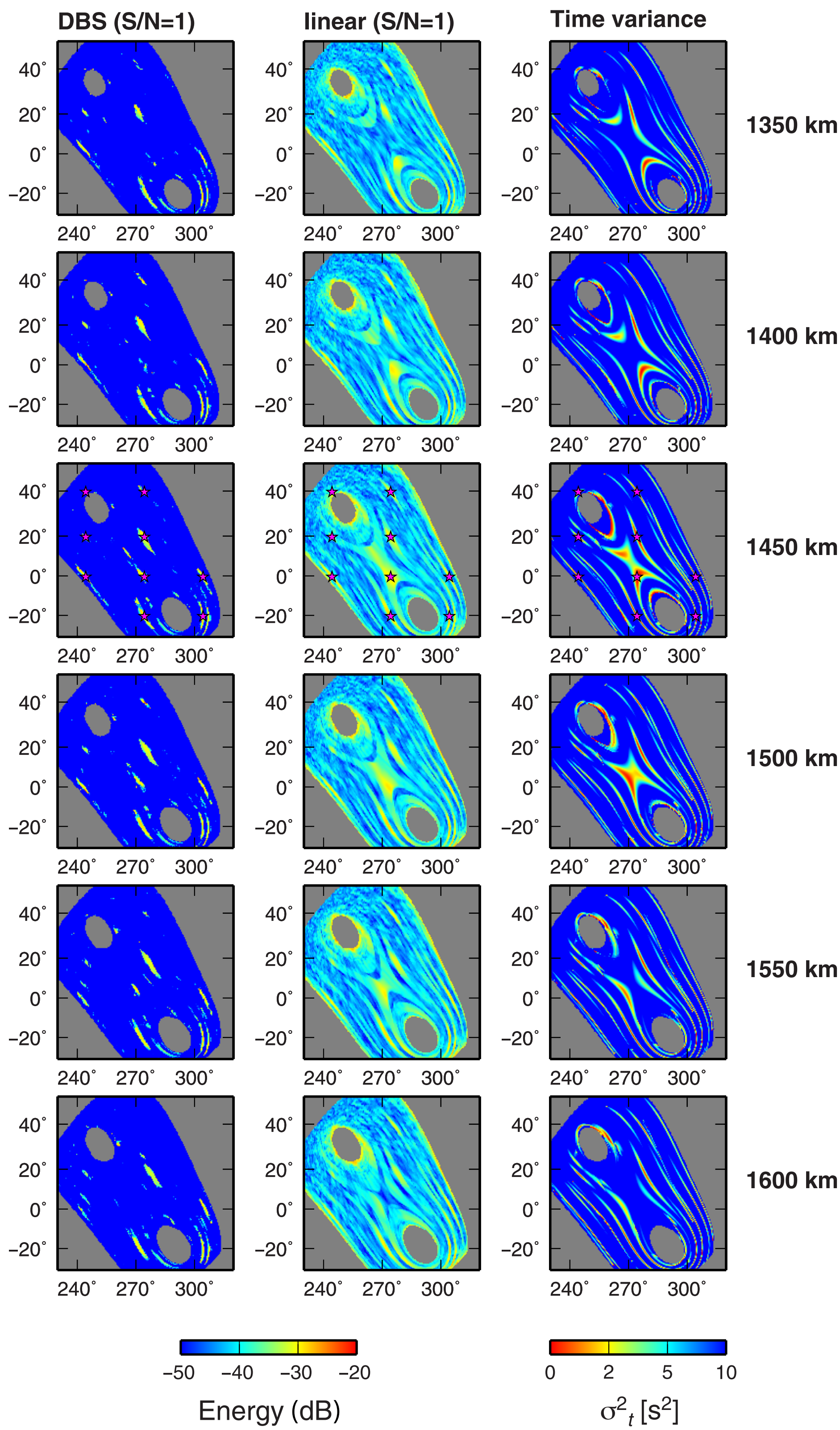


**Figure 7.** Isochronal artefacts and traveltime variance. The left-hand column is same as the second column of Fig. 6. The middle column shows migration results with linear stack for comparison. The right-hand column is traveltime variance with respect to the nearest point scatterer. As in Fig. 6, the locations of true point scatterers at 1450 km are denoted by stars.

though the latter is characterized by better signal recovery. When S/N is reduced to 0.5, NRS can barely detect embedded signals, whereas DBS can recover most of signals along with isochronal artefacts. These contrasting results between NRS and DBS are in accord with the scaling of signal recovery derived by Korenaga (2013). Results with PWS are virtually identical to those with NRS. Note that most of small-scale blobs seen only in the DBS-based migration images are part of the isochronal volumes associated with the point scatterers.

Different scatterers have different spreads of isochronal artefacts, and these migration images demonstrate that it is nearly impossible to constrain the size of a point scatterer in case of the source–receiver geometry used in this study. We may be able to constrain, however, the minimum number of point scatterers needed to explain a given migration image. Such information would still be useful when, for example, one is interested in characterizing the degree of chemical heterogeneity in the mantle.

To this end, isochronal volumes associated with the true scatterer positions are calculated by means of traveltime variance (*cf.* Kito & Korenaga 2010), which is defined at the *j*th node as

$$\sigma_t^2(j) = \frac{1}{N_r}\sum_{i=1}^{N_r}(t_{ij}^{\mathrm{t}} - t_{ij_r}^{\mathrm{t}})^2, \qquad (7)$$

where $t_{ij}^{\mathrm{t}}$ is the theoretical traveltime from the hypocentre to the *i*th station through the *j*th node (as in eq. 1), and the $j_r$th node corresponds to the nearest scatterer. The 'nearest' scatterer here means the scatterer that yields the smallest traveltime variance. The traveltime variance at all grid nodes (Fig. 7, right-hand column) is calculated using the shortest path method (Moser 1991), in which the scatterer nodes are treated as source and traveltime variance is used as distance. Visual comparison between the DBS-based migration image and the traveltime variance suggests that isochronal artefacts are seen where traveltime variance is small enough. This is quantified by the correlation between traveltime variance and migration energy (Fig. 8); isochronal artefacts of significant amplitude are characterized by $\sigma_t^2 < \sim 10$ s$^2$. This critical traveltime variance corresponds to the average traveltime difference of ∼3 s, which is half the dominant period of the signal being stacked.

The migration energy is not always high even when traveltime variance is small; quite a few nodes with $\sigma_t^2 < \sim 10$ s$^2$ have the energy lower than −40 dB (Fig. 8). This is because different combinations of time-shifting can have the same traveltime variance, but they can have different results for statistical tests on signal significance and coherence. In other words, strong isochronal artefacts always have small traveltime variance, but not vice versa; when there does exist a scatterer, only some fraction of its isochronal volume would show up in a migration image because of noise. The synthetic data here have the same scattered waveform for all traces, but in reality, scattered waveforms are expected to exhibit direction-dependent variations across traces, so an even smaller fraction of an isochronal volume would be visible by migration. This tendency would be enhanced with decreasing S/N. That is, as S/N decreases, a smaller fraction of an isochronal volume would show up in a migration image, and which part of the isochronal volume would be imaged becomes harder to predict, because it depends on how a particular realization of noise interferes with the signal detection of DBS.

These characteristics of isochronal artefacts suggest that hierarchical clustering (e.g. Defays 1977) may be suited to locate point scatterers. Hierarchical clustering is a method of cluster analysis that defines clusters in a hierarchical way based on a certain metric,

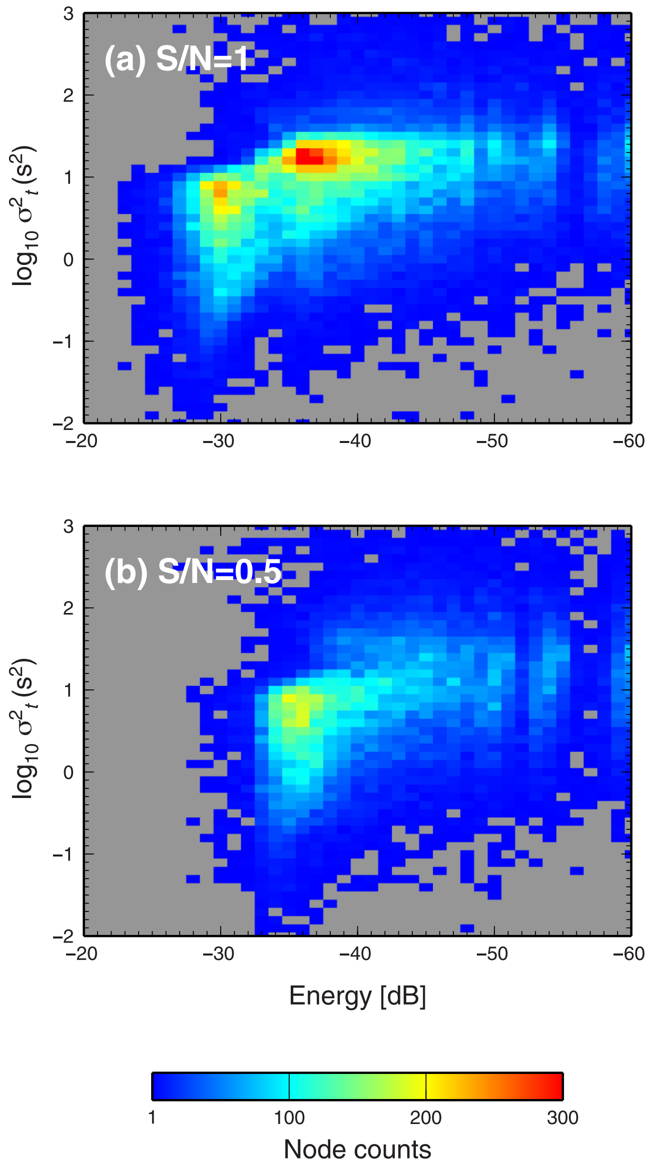


**Figure 8.** Covariation of traveltime variance and migration energy for DBS-based migration with synthetic data for the case of (a) S/N of 1 and (b) S/N of 0.5. These 2-D histograms are based on migration results at all depths (i.e. including all of point scatters), not just for the depth range shown in Figs 6 and 7.

which is a measure of distance between pairs of points. Traveltime variance is a natural choice for such a metric, and traveltime variance between two nodes, $j_1$ and $j_2$, is defined as

$$\sigma_t^2(j_1, j_2) = \frac{1}{N_r}\sum_{i=1}^{N_r}(t_{ij_1}^{\mathrm{t}} - t_{ij_2}^{\mathrm{t}})^2. \qquad (8)$$

In the agglomerative (bottom–up) hierarchical clustering, which is adopted here, each significant node starts in its own cluster, and pairs of clusters are merged with an increasing metric. At each step, the two clusters separated by the smallest traveltime variance are combined. The distance between clusters is based on complete linkage, that is, the maximum traveltime variance between all possible node pairs between two clusters.

The result of hierarchical clustering, using the DBS-based migration result for the case of S/N of 1, is shown in Fig. 9, and at $\sigma_t^2$ of 10 s$^2$, the number of clusters is reduced down to ∼40, which

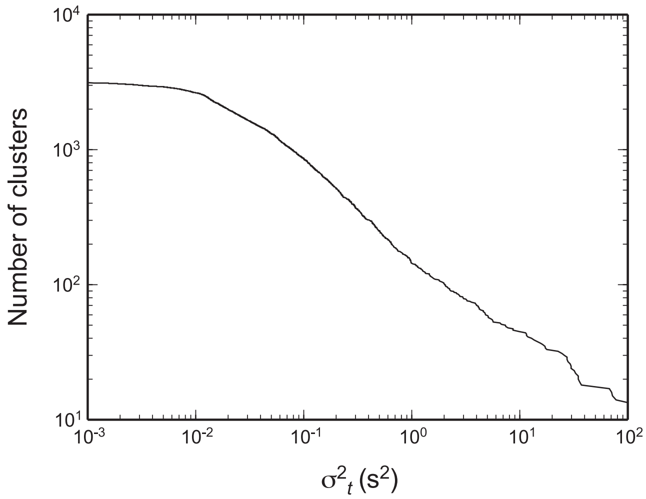


**Figure 9.** Progression of hierarchical clustering with increasing traveltime variance. At the beginning, all of ∼3000 significant nodes are their own clusters, and at each step, the two clusters separated by the smallest traveltime variance are combined, and the number of clusters decreases by one. The result of DBS-based migration using the synthetic data with S/N of 1 (Fig. 6, second column) is considered, and because of the high S/N, significant nodes are defined here as the nodes with (S/N) $n^{1/2} \geq 10$.

is reasonably close to 29, the number of true scatterers. For each cluster, a centre node is defined as the average position of all nodes contained in the cluster, and centre nodes for clusters at $\sigma_t^2$ of 10 s$^2$ are shown in Fig. 10 for the depth of 1350–1600 km. Scatterers a, d, e and h are correctly recovered by this procedure, whereas other scatterers are estimated to be at other depths; b, f and g are off by 50 km and i by 100 km. More than one centre node is sometimes assigned to one scatterer (a, d, f and h), and some scatterers are entirely missed (b and c). Given the complexity of the traveltime variance space (Fig. 7), it would be unrealistic to expect a perfect recovery of true scatterer locations. What is more important is that, because of the complete linkage clustering, all nodes in each cluster are guaranteed to be separated less than the critical traveltime variance, which is consistent with the notion of isochronal volume. In the example considered here, the number of significant nodes is over 3000, but the number of clusters is only ∼40 at $\sigma_t^2$ of 10 s$^2$. This clustering approach thus offers a convenient way to obtain approximate scatterer locations and estimate the minimum number of point scatterers. Note that the critical traveltime variance, which is set here as 10 s$^2$, should be lowered for shorter period seismograms and be raised for longer period ones.

### 4.4 Test with randomized data

The DBS-based migration image with the masked data (Fig. 4) contains about 2000 nodes with apparent (S/N) $n^{1/2} \geq 2$, and the cluster analysis yields ∼280 clusters at $\sigma_t^2$ of 10 s$^2$. The depth distributions of such significant nodes and clusters are shown in Fig. 11 (solid line). Both distributions have a broad peak at the mid-mantle level, but the cluster distribution is more skewed towards shallow depths. This is because isochronal volumes are generally larger in the lower mantle, that is, a large number of significant nodes can be contained in just one cluster. When the volume of a cluster is large, it could potentially contain more than one scatterer, but determining whether it is the case or not is beyond the resolution of the source–receiver geometry under consideration.

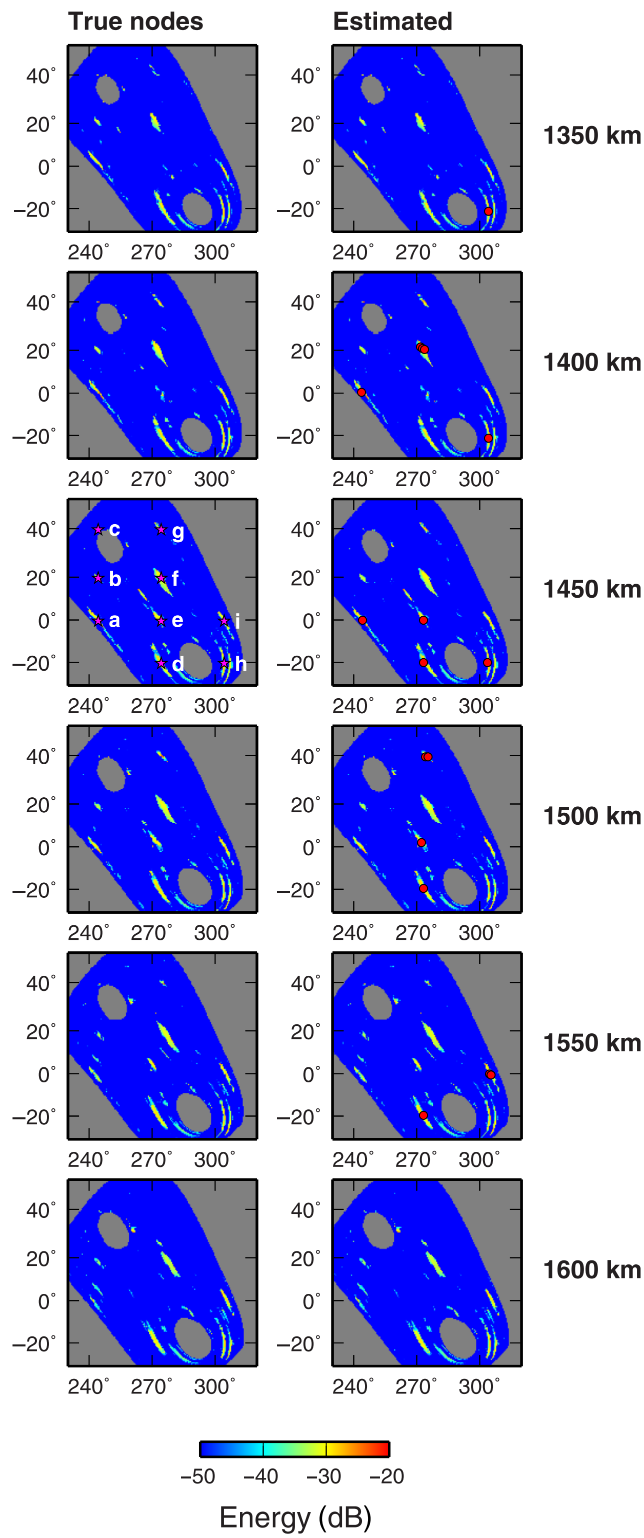


**Figure 10.** Estimating true node locations from DBS-based migration images through hierarchical clustering. The left-hand column is same as the second column of Fig. 6. Circles in the right-hand column denote the locations of the centre nodes of clusters at $\sigma_t^2$ of 10 s$^2$. See text for details.

The above estimate on the spatial distribution of point scatterers involves (1) statistical tests on signal significance and coherence in DBS, (2) screening with apparent (S/N) $n^{1/2}$ and (3) minimizing isochronal artefacts through cluster analysis. Even after these

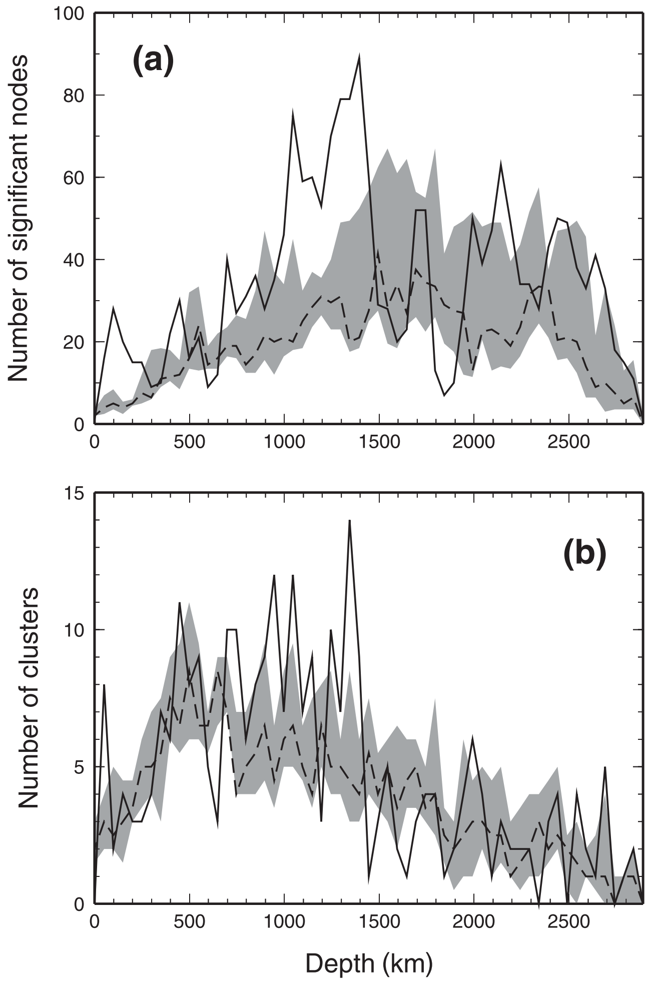


**Figure 11.** The depth distributions of (a) significant nodes and (b) clusters at $\sigma_t^2$ of 10 s$^2$, for DBS-based migration of data set #5. Solid line corresponds to results from the original data set, while dashed line and grey shading denote, respectively, the median and interquartile range of results from ten randomized data sets.

procedures, however, false signals can still remain at large owing to the very nature of migration. In migration, seismic records at different stations are time-shifted and stacked in a variety of ways because each of a large number of nodes in a migration volume corresponds to one unique way of time-shifting. The number of different time-shifting in migration is far greater than that in vespagram, so even when input data are pure random noise, it is still possible for random noise to be aligned as a coherent signal at some nodes. When S/N is low, such a false signal is indistinguishable from a true signal, at least by statistical tests implemented in DBS.

To evaluate the significance of false signals, DBS-based migration is repeated with randomized data sets. Each randomized data set is prepared by trace-scrambling the original data set with the maximum period of 20 s, in the same manner as noise stacks are generated in DBS. Results based on 10 randomized data sets are shown in terms of the depth distributions of significant nodes and clusters (Fig. 11). At most depths, the occurrence of false signals is high enough to cast a doubt on the reliability of the alleged signals from the original data.

At this point, one may wonder why signals can show up from the randomized data, which are prepared by trace-scrambling, that is, the very method used in DBS to prepare noise stacks. This is because it is easy to generate a noise stack locally, but not globally. In DBS, the significance of a given stack is tested by comparing with the stack of scrambled traces. If there is a coherent signal in a given set of traces, it is easy to destroy them by trace-scrambling, and this is the basis for DBS. However, if we take the scrambled traces as a whole and consider migration with such randomized data, there is still a possibility of having random noise to be aligned at a certain node and look like a true signal.

The high probability of false signal is hard to avoid when trying to image low S/N signals with migration. Significant nodes from the original data may be just all false signals, or many of them could still be real. It is difficult to prove either way with a single data set. If most of significant nodes are false signals, however, they would not be reproduced with other data sets. If weak signals keep showing up in different data sets, then, it would lend more credence to their significance. A test for reproducibility is thus described next.

## 5 RESULTS WITH MULTIPLE DATA SETS

DBS-based migration is conducted for other nine data sets as well, using the same masking of standard phases and the first 120 s after the direct $P$. Results are shown in Fig. 12 in terms of the depth distributions of significant nodes and clusters at $\sigma_t^2$ of 10 s$^2$. The number of significant nodes varies from ∼1000 (data set #6) to ∼5000 (data set #8), and the average of 10 data sets is ∼3000. The number of clusters shows a similar variation, with the average of ∼380. The depth distribution of significant nodes varies considerably among different data sets, but that of clusters exhibits a common pattern, with a broad peak at ∼700 km depth. The probability of false signal is also estimated by migration with randomizing data, as done in Section 4.4, and for most data sets, the original data yield a much larger number of significant nodes than the randomized data for a substantial depth range (Fig. 12). This indicates that the original data sets contain a certain number of true signals, which are destroyed by trace-scrambling. This tendency becomes, however, less pronounced in terms of the number of cluster nodes.

The reproducibility of these significant nodes is tested in the following way. Each of significant nodes in a certain data set is checked against all other data sets, and the number of supporting data sets are counted. For example, a significant node with the duplicity of 5 means that the presence of a signal at the particular node is confirmed by five different data sets. Given the highly incomplete illumination of an isochronal volume by one source–receiver geometry (as discussed in Section 4.3), nodes in different data sets are considered to coincide if they are within the traveltime variance of 1 s$^2$. Though the extent of an isochronal volume is better quantified with $\sigma_t^2$ of 10 s$^2$, a tighter volume is used here for a conservative measure of overlapping. Also, when checking for this overlapping, the amplitude of the original node is replaced by that of a supporting node, if the latter is greater, to compensate for incomplete signal recovery due to low S/N. For example, if a node in data set #1 is supported by a node in data set #2, and if the latter happens to have a larger amplitude than the former, the amplitude of the former is set to that of the latter. From the total of 10 data sets, about 2200 significant nodes are found to have the minimum duplicity of 5, that is, supported by at least half of the data sets. The number of nodes decreases with increasing duplicity: ∼1100, ∼380 and ∼80 for the minimum duplicity of 6, 7 and 8, respectively. All data sets contribute to these reproducible nodes in a similar manner (Fig. 13,

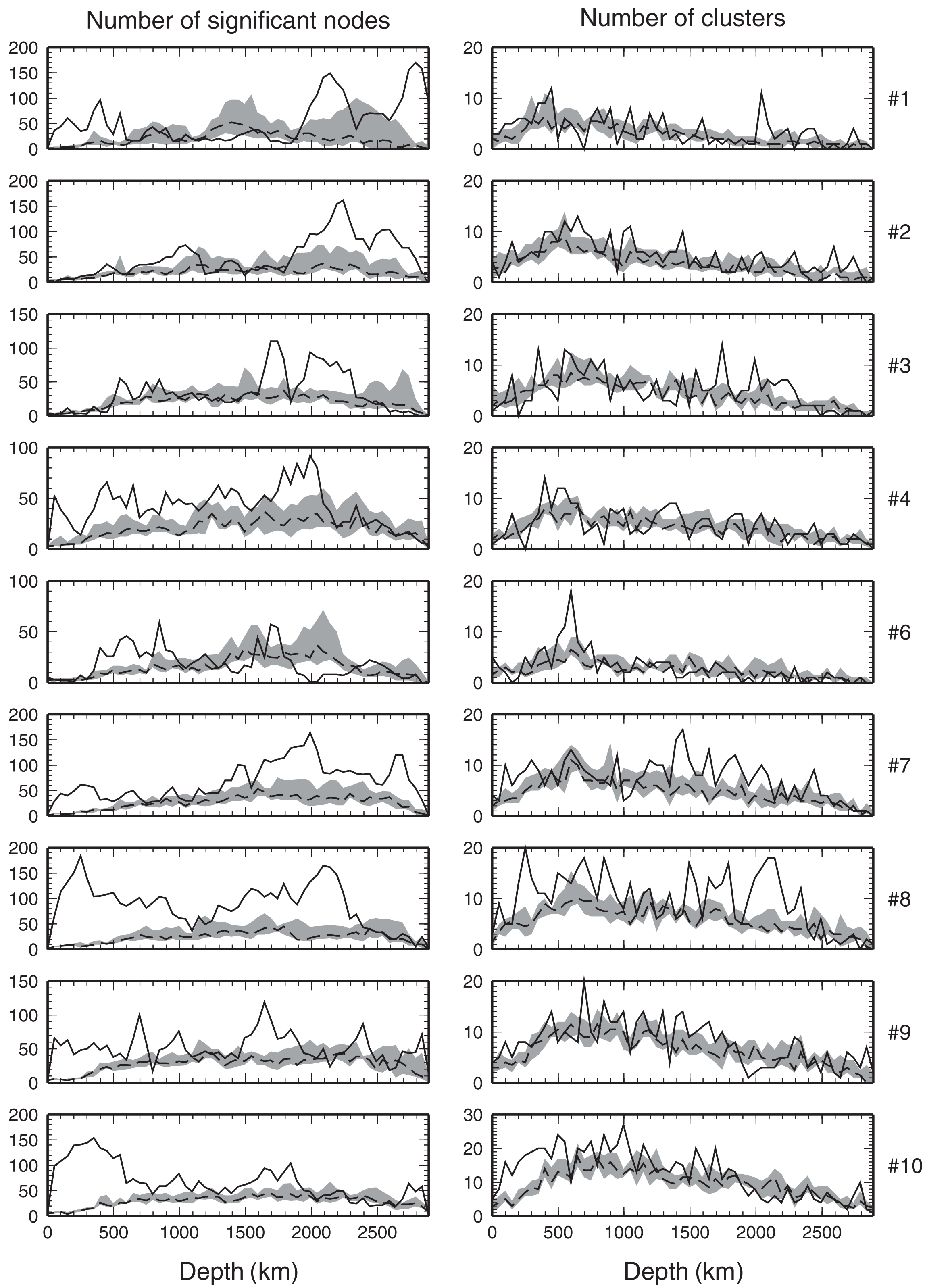


**Figure 12.** The depth distributions of (right-hand side) significant nodes and (left-hand side) clusters at $\sigma_t^2$ of 10 s$^2$, for DBS-based migration of all data sets except #5 (which is shown in Fig. 11).

top row). Data set #5, which is considered in the previous section, in fact contributes the largest number of nodes, suggesting that the high probability of false signal (Fig. 11) alone does not readily mean the low reliability of detected signals.

The depth distribution of the reproducible nodes is contrasting to that of significant nodes in individual data sets. Instead of a broad peak in the mid-mantle level, significant nodes are concentrated in the top 400 km and the bottom 1500 km of the mantle (Fig. 13). This

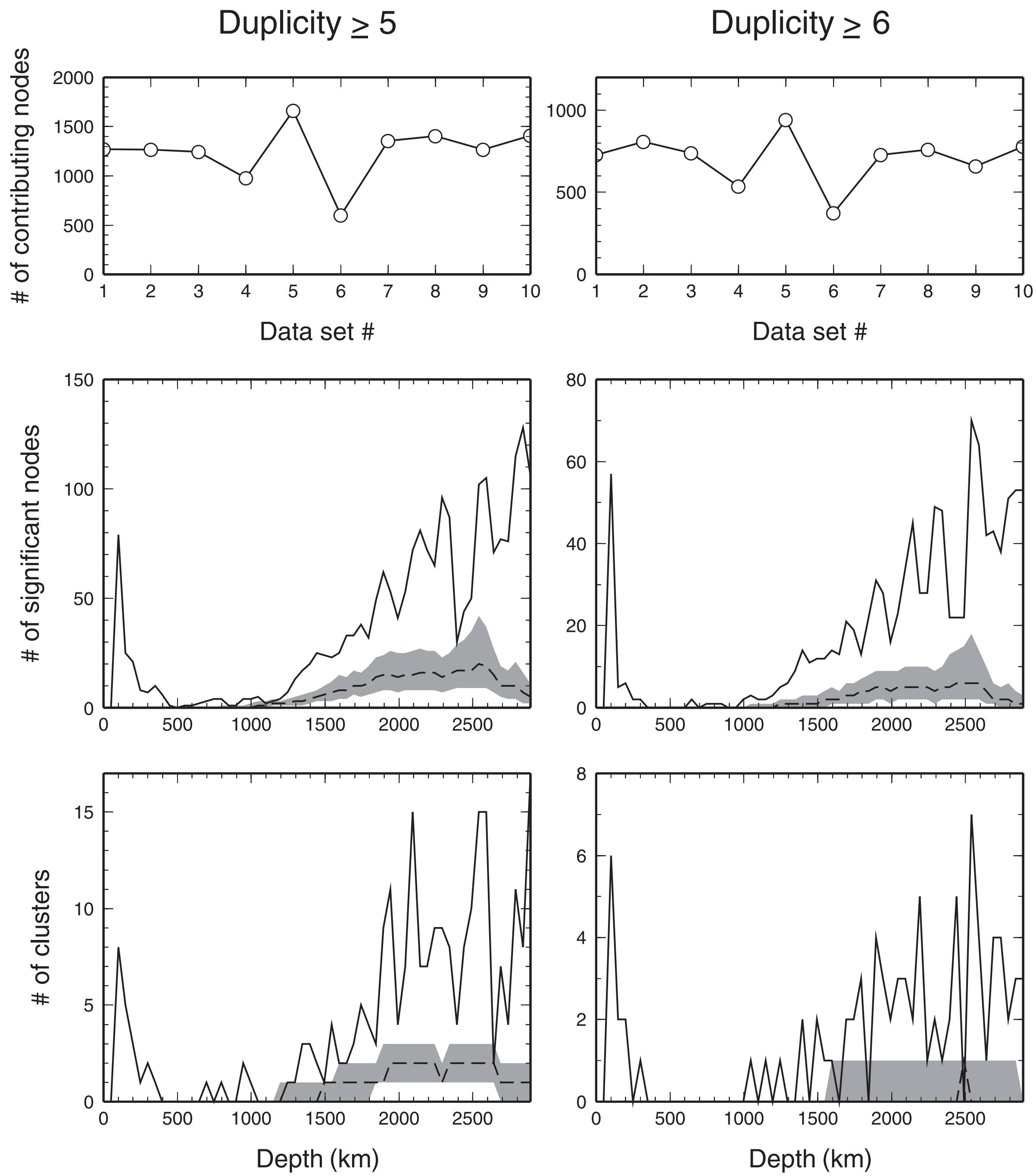


**Figure 13.** Results of reproducibility test, for the minimum duplicity of 5 (left-hand side) and 6 (right-hand side). The top row shows the number of nodes supported by each data set. The middle and bottom rows show the depth distributions of reproducible significant nodes and clusters at $\sigma_t^2$ of 10 s$^2$, respectively. Dashed line and grey shading denote, respectively, the median and interquartile range of 1000 reproducibility test results using randomized data sets.

suggests that an impression from a single data set can be misleading, and that checking for reproducibility is a vital step to distinguish between true and false signals. Also, because the topology of isochronal volumes can be complex, it is not easy to guess the distribution of reproducible nodes from individual depth distributions (Fig. 12). For example, the high concentration of reproducible nodes near the core–mantle boundary (Fig. 13, middle row) may seem to be supported only by three data sets #1, #9 and #10 (Fig. 12, left-hand column), but they are actually supported by other data sets as well, but by significant nodes located at shallower depths. The cluster analysis is conducted on these reproducible nodes, and the number of clusters at $\sigma_t^2$ of 10 s$^2$ is ∼250, ∼90, ∼30 and ∼10, respectively, for the minimum duplicity of 5, 6, 7 and 8. The depth distribution of these clusters is similar to that of reproducible nodes (Fig. 13). Calculating traveltime variance with multiple data sets is more involved than that with a single data set. Consider the traveltime variance between nodes $j_1$ and $j_2$. If the node $j_1$ is supported by data sets #1, #2, #3, #4 and #5, and the node $j_2$ is supported by data sets #2, #3, #4, #5 and #6, traveltime variance is calculated for each of the common data sets (i.e. #2, #3, #4 and #5), and the largest variance is adopted as $\sigma_t^2(j_1, j_2)$. This is consistent with the use of the complete linkage in hierarchical clustering.

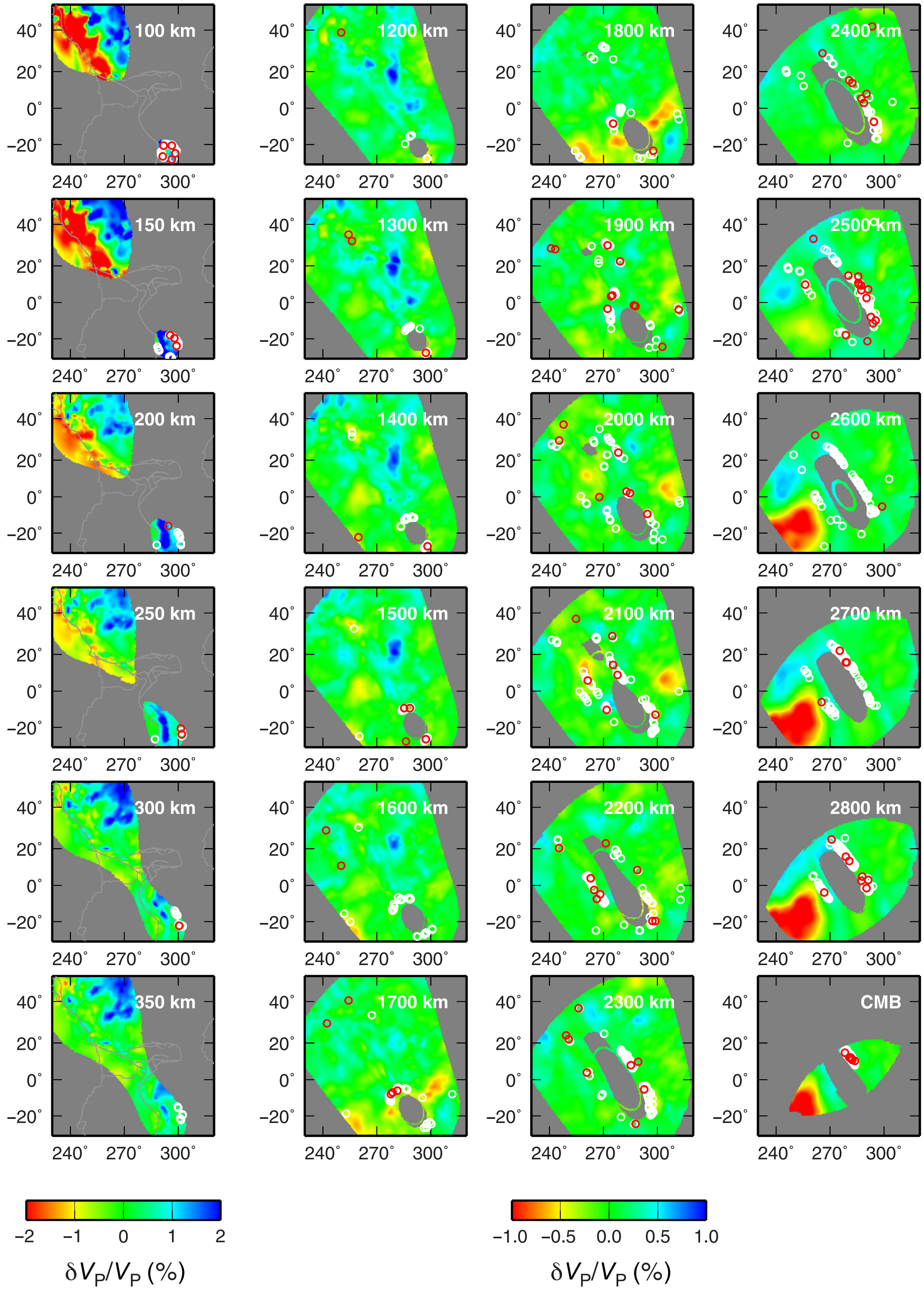


**Figure 14.** Spatial distributions of reproducible nodes (white circles) and the centre nodes of clusters (red circles) with the minimum duplicity of 5, shown on the *P*-wave tomography model of Simmons *et al.* (2012). Grey regions are out of reach by any of masked data sets used in this study. Depth sections are shown at 100 km interval for the lower mantle, so only half of relevant sections are seen here.

The same test for reproducibility is applied to the migration results with randomized data. Because migration with a randomized data set was repeated 10 times for each data set, there are in total 100 migration results with randomized data. For each data set, one migration result is chosen randomly out of 10, and the reproducibility of significant nodes is tested with the collection of 10 randomly chosen data sets. The total number of different combinations of such data sets is $10^{10}$, but testing all of them is of course impractical. This procedure is thus repeated up to 1000 times to obtain reasonably robust statistics, and results are shown in Fig. 13. Unlike the case for an individual data set, the numbers of significant nodes and clusters with the minimum duplicity of 5 are both considerably smaller than those for the original data sets. Migration with randomized data has zero reproducibility for depths less than 1000 km. Large isochronal volumes at greater depths allow even false signals to sometimes overlap among different data sets, but the level of reproducibility is only ∼25 per cent of the case of the real data sets. These results suggest that requiring the minimum duplicity of 5 or 6 can reduce the possibility of false signal to the extent that the majority of the reproducible nodes may be rendered for physical interpretation.

The spatial distribution of the significant nodes with the minimum duplicity of 5 is shown in Fig. 14, together with the tomographic *P*-wave velocity model of Simmons *et al.* (2012). There is no obvious correlation between those nodes and velocity anomalies, though there is no reason to expect such correlation if, for example, these scatterers represent crustal fragments subducted by ancient plate tectonics. Nonetheless, the correlation issue needs to be assessed more carefully by considering the extent of isochronal volumes associated with these nodes. The reproducible nodes in the upper mantle are all located near the sources, in the vicinity of the subducting plate. The lack of such upper-mantle nodes beneath the receivers may be because the variation of the scattering coefficient in eq. (1) is neglected in this study. For scatterers in the upper mantle beneath receivers, the aperture of the subarrays considered in this study would be too wide to be treated with such simplified approach. Still, it is remarkable is that the minimum number of point scatterers needed to explain the DBS-based migration results is as many as ∼100–200; the migratable volume for the data sets used in this study is only $\sim 1.5 \times 10^{10}$ km$^3$, which is about 1.7 per cent of Earth's mantle. The average migration energy of the detected signals is on the order of −20 to −30 dB with respect to the direct *P* (Fig. 15). These signals are of course visible in linear stack as well (Fig. 4), but they are surrounded by other numerous migration artefacts, and given that they are not imaged by NRS and PWS (Fig. 4), these subtle signals could easily be dismissed as noise in the framework of conventional stacks. It is the signal detection capability of the new stack that allows to bring out the presence of weak, small-scale scatterers in the mantle. When dealing with such subtle signals, the probability of false signal is particularly high for migration (Figs 11 and 12), but without detecting signals first, it is not even possible to test them for reproducibility. Migration with NRS or PWS yields almost no signal, so there is nothing to be tested for reproducibility. Migration with linear stack is characterized by a large number of high-amplitude artefacts almost everywhere, leading to numerous seemingly 'reproducible' nodes; such high background noise level would overwhelm the presence of true scatterers.

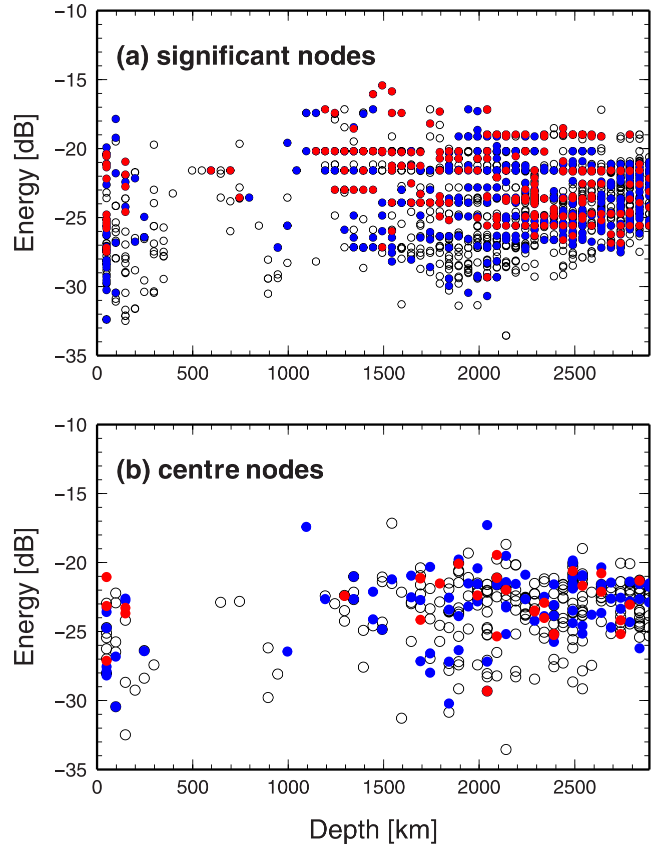


**Figure 15.** Covariation of the depth and migration energy of (a) reproducible nodes and (b) the centre nodes of clusters, for the minimum duplicity of 5 (open circle), 6 (blue circle) and 7 (red circle). The energy of a centre node is calculated by taking the logarithmic average of the energy of all nodes in a cluster.

## 6 DISCUSSION AND OUTLOOK

### 6.1 Comparison with weighted migration schemes

PWS and DBS both apply weighting on linear stack results. In PWS, weighting is calculated based on the similarity of the phase components of individual traces, and in DBS, weighting is based on statistical tests on signal significance and coherence. There are other kinds of weighting specifically designed for teleseismic migration. Slowness-backazimuth-weighted migration (SBWM; Kito *et al.* 2007) is one of them, and it may look similar to DBS-based migration. These two migration schemes are, however, fundamentally different, so some clarification is due.

In SBWM, a migration image constructed by linear stack is weighted according to the difference between the theoretical and observed values of slowness and backazimuth. Traveltime calculation used in migration provides the theoretical values of slowness and backazimuth for all nodes. The observed values of slowness and backazimuth are estimated by computing a large number ($\sim 10^3$) of stacks with a variety of slowness-backazimuth pairs around their theoretical values and choosing the one with the highest migration energy. Weighting is computed based on the following notion: The smaller the difference between the theoretical and observed values of slowness and backazimuth is, the more likely the node corresponds to a true scatterer. Because the observed values of slowness and backazimuth need to be estimated by conducting a large number of stacks at each node, SBWM is computationally expensive; it is in fact as costly as migration with the original DBS implementation (i.e. without the local maxima acceleration).

Both SBWM and DBS-based migration conduct a large number of stacks at each node, but the meanings of these stacks are different. In DBS, traces are time-shifted 'randomly' to generate noise stacks, against which the significance of the original stack is tested. In SBWM, traces are time-shifted 'gradually and systematically' as slowness and backazimuth used for stacking are varied from their theoretical values. The migration energy of these stacks thus varies gradually from the that of an original linear stack. If the maximum of the migration energy takes place at the theoretical slowness and backazimuth, this is interpreted that the original linear stack contains a reliable signal. Though this approach may appear sound and seems to work in certain applications (e.g. Kito *et al.* 2007, 2008), it would lead to interpreting all of well-isolated local maxima seen in a linear migration image as true signals. This approach is thus not suited for detecting subtle signals from low S/N data; being local maxima is not equivalent to being statistically significant (compare, for example, linear stack and DBS in Fig. 3).

DBS has three control parameters, whereas SBWM has only two. The control parameters of DBS are, however, not of arbitrary nature. The critical significance level should be based on one's need for statistical rigor (smaller value corresponds to more stringent testing), the number of bootstrap replicates should be high enough to achieve accurate statistical estimates and the maximum period of trace scrambling should reflect the dominant period seen in a given data set. In contrast, two control parameters in SBWM, one for slowness weighting and the other for backazimuth weighting, are rather *ad hoc*, and one needs to vary them to arrive at a 'reasonable-looking' result.

Kito & Korenaga (2010) developed another scheme called cross-correlation-weighted migration (CCWM), to imitate the performance of SBWM with less computational cost by computing the cross-correlation of traces. This scheme does not work well for low S/N data either, because it is nearly impossible to extract meaningful cross-correlation when S/N is low. It introduced, however, the concept of using traveltime variance to quantify the extent of an isochronal volume, which is shown in this paper to be useful to minimize the effect of isochronal artefacts.

### 6.2 Origin of small-scale scatterers

When isochronal volumes are large, which is usually the case for teleseismic migration with a regional array, it is difficult to discuss the shape of a scatterer. Nonetheless, the clusters of the reproducible nodes are seen to congregate at some depths, for example, 2500 and 2800 km (Fig. 14), and because the effect of isochronal artefacts is mostly removed through the cluster analysis, these assemblies of clusters may indicate the presence of relatively coherent structures. At the same time, the majority of the clusters are more dispersed, and for them, their migration energy is the only clue for their physical origin.

As shown in Fig. 15, most of the reproducible nodes have the migration energy in the range of −20 to −30 dB; the amplitude of scattered waves is 10–3 per cent of the direct $P$ wave. Perhaps the simplest approach to interpret such relative amplitude is to assume scattering by a spherical heterogeneity. In general, the amplitude of a scattered wave is direction-dependent, so for the sake of simplicity, I focus on the maximum amplitude here. For $P$-to-$P$ scattering, it may be expressed as (e.g. Miles 1960; Gubernatis *et al.* 1977)

$$\max|\Phi_1| = \frac{V}{4\pi r}\frac{\omega^2}{\alpha_0^2}\left|\frac{\delta\rho}{\rho_0} - \frac{\delta\lambda + 2\delta\mu}{\lambda_0 + 2\mu_0}\right| \sim \frac{V}{2\pi r}\frac{\omega^2|\delta\alpha|}{\alpha_0^3}, \tag{9}$$

where $\Phi_1$ is the amplitude of a scattered wave with respect to a unit-amplitude primary wave, $V$ is the volume of a spherical heterogeneity, $r$ is the distance between the heterogeneity and a receiver, $\omega$ is the angular frequency of a seismic wave, $\alpha_0$, $\rho_0$, $\lambda_0$ and $\mu_0$ are, respectively, the $P$-wave velocity, density and two Lamé parameters of the ambient medium and $\delta\alpha$, $\delta\rho$, $\delta\lambda$ and $\delta\mu$ denote corresponding perturbations due to the heterogeneity. For a plausible range of velocity perturbation, the radius of the spherical scatterer corresponding to the observed range of the migration energy is found to be on the order of 100 km (Fig. 16). As eq. (9) assumes Rayleigh scattering (the size of a scatterer being much smaller than wavelength), the estimate shown in Fig. 16 must be regarded as a lower bound. When the size of a scatterer is comparable with wavelength (the Mie scattering regime), the predicted amplitude would become smaller than indicated by eq. (9), because of the interference between the wave fields from different parts of a heterogeneity (e.g. Wu & Aki 1985). Taking into account such a finite volume effect would not be worthwhile, because the assumption of a spherical heterogeneity is already too simplistic and actual scatterers are expected to have more complicated shapes. This order-of-magnitude estimate on the likely size of scatterers would still be useful, however, when considering the implication of such velocity heterogeneities for mantle dynamics and terrestrial magmatism (e.g. Korenaga 2008).

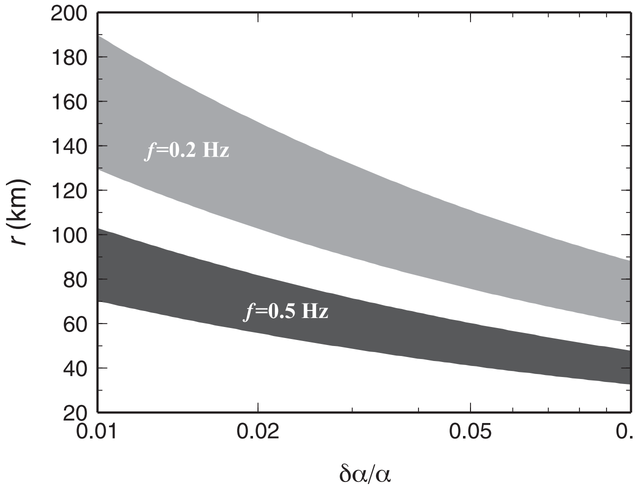


**Figure 16.** Radius of a spherical scatterer as a function of $P$-wave velocity perturbation, expected from the migration energy at the level of −20 to −30 dB. Assuming the mid-mantle condition, $r$ and $\alpha_0$ are set to 5000 km and 12 km s$^{-1}$, respectively. The cases of two frequencies ($f$ of 0.5 and 0.2 Hz) are shown.

### 6.3 Future directions

Judging from the example given in this study, DBS-based migration appears promising for detecting hitherto undetected (or disregarded) weak signals. The disadvantage of being much more time-consuming than conventional stacks is probably compensated by its superior performance in signal detection. The deterministic detection of small-scale scatterers in the mantle has so far relied on scattered waves with reasonably high amplitudes, which can already be identified in seismic traces before stacking, and this limitation on signal strength has restricted us mostly to spatially coherent features such as lateral discontinuities and dipping reflectors (e.g. Kawakatsu & Niu 1994; Kaneshima & Helffrich 1999; Vanacore *et al.* 2006; Kaneshima & Helffrich 2010; Niu 2013). By applying DBS to a range of array methods such as vespagram and migration, we can now start mining the vast wealth of teleseismic data for a variety of small-scale features in the mantle.

In this study, the procedure of teleseismic migration was kept simple to highlight the effects introduced by using DBS in place of conventional stacks. With the source–receiver geometry used in this study, the neglect of radiation pattern is reasonable for imaging the deep mantle, but properly calculating $A(i, j)$ in eq. (1) would become important when imaging the upper mantle beneath the receiver array or when using a wider receiver array. Similar efforts would be essential when handling a source array. Assuming a 1-D reference earth model for traveltime calculation may be justifiable because of the use of differential traveltime, but how migration results would differ by using a 3-D velocity model needs to be seen. The data used are also highly limited; only a 120-s-long section bounded by *P* and *pP* was searched for *P*-to-*P* single scattering. This is to minimize the possibility of a false alarm by reducing the input data to a bare minimum, but with adequate care, more aggressive approaches should become possible. In this study, multiple data sets were processed separately, and the individual results were used to test the reproducibility of detected signals. Alternatively, one may combine all data sets into one and process it by double-beam forming (e.g. Krüger *et al.* 1996; Scherbam *et al.* 1997). This may reduce the probability of false signal, but at the same time, using only one data set would not allow a reproducibility test. Which is better is yet to be seen, and there may be other possibilities. In any case, it seems warranted to accumulate more worked examples with DBS by exploring and experimenting further.

## ACKNOWLEDGEMENTS

This work was sponsored by the U.S. National Science Foundation under grant EAR-0842753. Figures were generated with GMT (Wessel & Smith 1995). Data used in this study were obtained from the IRIS Data Management Centre. This work was also supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. The author thanks Peter Shearer for constructive comments on the earlier version of the manuscript. Reviews from two anonymous referees were helpful to improve the clarify of the manuscript.

## REFERENCES

Chambers, K. & Woodhouse, J.H., 2006. Investigating the lowermost mantle using migration of long-period S-ScS data, *Geophys. J. Int.,* **166,** 667–678.

Defays, D., 1977. An efficient algorithm for a complete link method, *Comp. J.,* **20,** 364–366.

Freybourger, M., Chevrot, S., Krüger, F. & Achauer, U., 2001. A waveform migration for the investigation of P wave structure at the top of D" beneath northern Siberia, *J. geophys. Res.,* **106,** 4129–4140.

Gubernatis, J.E., Domany, E. & Krumhansl, J.A., 1977. Formal aspects of the theory of the scattering of ultrasound by flaws in elastic materials, *J. appl. Phys.,* **48,** 2804–2811.

Hutko, A., Lay, T., Revenaugh, J. & Garnero, E.J., 2008. Anticorrelated seismic velocity anomalies from post-perovskite in the lowermost mantle, *Science,* **320,** 1070–1074.

Kanasewich, E.R., Hemmings, C.D. & Alpaslan, T., 1973. *N*th-root stack nonlinear multichannel filter, *Geophysics,* **38,** 327–338.

Kaneshima, S. & Helffrich, G., 1999. Dipping low-velocity layer in the mid-lower mantle: evidence for geochemical heterogeneity, *Science,* **283,** 1888–1891.

Kaneshima, S. & Helffrich, G., 2010. Small scale heterogeneity in the mid-lower mantle beneath the circum-Pacific area, *Phys. Earth planet. Inter.,* **183,** 91–103.

Kawakatsu, H. & Niu, F., 1994. Seismic evidence for a 920-km discontinuity in the mantle, *Nature,* **371,** 301–305.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105,** 429–465.

Kito, T. & Korenaga, J., 2010. Cross-correlation weighted migration: toward high-resolution mapping of mantle heterogeneities, *Geophys. J. Int.,* **181,** 1109–1127.

Kito, T. & Krüger, F., 2001. Heterogeneities in D" beneath the southwestern Pacific inferred from scattered and reflected P waves, *Geophys. Res. Lett.,* **28,** 2545–2548.

Kito, T., Rietbrock, A. & Thomas, C., 2007. Slowness backazimuth weighted migration: a new array approach to a high-resolution image, *Geophys. J. Int.,* **169,** 1201–1209.

Kito, T., Thomas, C., Rietbrock, A., Garnero, E.J., Nippress, S. & Heath, A.E., 2008. Seismic evidence for a sharp lithospheric base persisting to the lowermost mantle beneath the Caribbean, *Geophys. J. Int.,* **174,** 1019–1028.

Korenaga, J., 2008. Urey ratio and the structure and evolution of Earth's mantle, *Rev. Geophys.,* **46,** RG2007, doi:10.1029/2007RG000241.

Korenaga, J., 2013. Stacking with dual bootstrap resampling, *Geophys. J. Int.,* **195,** 2023–2036.

Krüger, F., Scherbaum, F., Weber, M. & Schlittenhardt, J., 1996. Analysis of asymmetric multipathing with a generalization of the double-beam method, *Bull. seism. Soc. Am.,* **86,** 737–749.

Lay, T. & Garnero, E.J., 2011. Deep mantle seismic modeling and imaging, *Annu. Rev. Earth planet. Sci.,* **39,** 91–123.

Miles, J.W., 1960. Scattering of elastic waves by small inhomogeneities, *Geophysics,* **25,** 642–648.

Moser, T.J., 1991. Shortest path calculation of seismic rays, *Geophysics,* **56,** 59–67.

Muirhead, K.J., 1968. Eliminating false alarms when detecting seismic events automatically, *Nature,* **217,** 533–534.

Niu, F., 2013. Distinct compositional thin layers at mid-mantle depths beneath northeast China revealed by the USArray, *Earth planet. Sci. Lett.,* in press. Available at: http://dx.doi.org/10.1016/j.epsl.2013.02.015.

Revenaugh, J., 1995. A scattered-wave image of subduction beneath the transverse ranges, *Science,* **268,** 1888–1892.

Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.,* **40,** 2-1–2-27.

Scherbam, F., Krüger, F. & Weber, M., 1997. Double beam imaging: mapping lower mantle heterogeneities using combinations of source and receiver arrays, *J. geophys. Res.,* **102,** 507–522.

Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.,* **130,** 487–505.

Sheriff, R.E. & Geldart, L.P., 1995. *Exploration Seismology,* 2nd edn, Cambridge Univ. Press.

Simmons, N.A., Myers, S.C., Johannesson, G. & Matzel, E., 2012. LLNL-G3Dv3: global P wave tomography model for improved regional and teleseismic travel time prediction, *J. geophys. Res.,* **117,** B10302, doi:10.1029/2012JB009525.

Thomas, C., Kendall, J.-M. & Lowmann, J., 2004. Lower mantle seismic discontinuities and the thermal morphology of subducted slabs, *Earth planet. Sci. Lett.,* **225,** 105–113.

Vanacore, E., Niu, F. & Kawakatsu, H., 2006. Observations of the mid-mantle discontinuity beneath Indonesia from S to P converted waveforms, *Geophys. Res. Lett.,* **33,** L04302, doi:10.1029/2005GL025106.

Wessel, P. & Smith, W.H.F., 1995. New version of the generic mapping tools released, *EOS, Trans. Am. geophys. Un.,* **76,** 329.

Wu, R. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics,* **50,** 582–595.

Yilmaz, O., 1987. *Seismic Data Processing,* Soc. of Exp. Geophys.