AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1002/2015JB012432

Key Points:
- A new array-based method for seismic data is developed to detect small-scale heterogeneities
- The new method can also constrain the geometries of detected scatterers
- A total of 39 scatterers are identified in the vicinity of the Mariana subduction zone

Correspondence to:
J. Korenaga,
jun.korenaga@yale.edu

Citation:
Korenaga, J. (2015), Constraining the geometries of small-scale heterogeneities: A case study from the Mariana region, *J. Geophys. Res. Solid Earth*, *120*, 7830–7851, doi:10.1002/2015JB012432.

Received 6 AUG 2015
Accepted 7 OCT 2015
Accepted article online 11 OCT 2015
Published online 27 NOV 2015

Constraining the geometries of small-scale heterogeneities: A case study from the Mariana region

Jun Korenaga[1]

[1]Department of Geology and Geophysics, Yale University, New Haven, Connecticut, USA

Abstract A new array-based method for seismic waveform data is developed to detect small-scale heterogeneities in the mantle and constrain their possible locations and sizes. Detecting small-scale heterogeneities using scattered waves has been challenging because scattered phases usually suffer from low signal-to-noise ratios, and determining the shapes of such heterogeneities is even more difficult because of isochronal artifacts. In the proposed method, the problem of the low signal-to-noise ratio is addressed by adopting dual bootstrap stacking, which can detect weak signals even with a limited number of seismic traces, as well as by ensuring reproducibility among multiple events. The locations of detected scatterers and their sizes and shapes are then estimated using a new kind of hierarchical cluster analysis that could minimize the effect of isochronal artifacts. The new method is applied to a previously analyzed data set of Mariana earthquakes to facilitate a comparison with conventional methods. In the middle to lower mantle beneath the Mariana Trench region, a total of 39 scatterers are identified via *S*-to-*P* scattering, with their potential volumes ranging from $< 8 \times 10^4$ to $\sim 1 \times 10^6$ km^3.

1. Introduction

Plate tectonics constantly generates oceanic crust and depleted lithospheric mantle at mid-ocean ridges by partial melting of the upwelling mantle. These chemically differentiated materials return to the mantle at subduction zones and are expected to be eventually homogenized by mantle convection [e.g., *Olson et al.*, 1984; *Hoffman and McKenzie*, 1985; *Ferrachat and Ricard*, 1998]. Because chemical heterogeneities in the convecting mantle can affect magmatism [e.g., *Sleep*, 1984; *Cordery et al.*, 1997; *Korenaga*, 2004, 2005], it is important to understand how quickly this convective mixing takes place and how chemically homogeneous the mantle can be. Understanding the efficiency of convective mixing is, however, still challenging. Mixing efficiency depends on the details of rheology [e.g., *Manga*, 1996; *Karato*, 1997; *van Keken and Ballentine*, 1998], but our knowledge of mantle rheology is still subject to large uncertainty [e.g., *Korenaga and Karato*, 2008]. Even with the perfect understanding of rheology, it is a formidable task to predict theoretically the present state of chemical heterogeneity in the mantle, which could reflect the history of chemical differentiation in the last few billion years [e.g., *Davies*, 2002; *Xie and Tackley*, 2004; *Korenaga*, 2006]. Observational constraints are thus essential to distinguish between various theoretical possibilities.

Geochemical studies indicate that the mantle is chemically heterogeneous at a range of scales [e.g., *Allègre and Turcotte*, 1986; *Zindler and Hart*, 1986; *Hofmann*, 1997]. Some seismic tomographic studies point to chemical heterogeneities at a very large scale (a few 1000 km) [e.g., *Kennett et al.*, 1998; *Trampert et al.*, 2004], whereas the stochastic analyses of global travel time data [*Gudmundsson et al.*, 1990; *Nolet and Moser*, 1993] and the amplitude of *PKP* precursors [*Hedlin et al.*, 1997; *Margerin and Nolet*, 2003] suggest the presence of small-scale (10–100 km) heterogeneities throughout the mantle. Compared to large-scale anomalies, small-scale heterogeneities can more easily be incorporated into convective upwelling, so they generally have a higher likelihood to affect surface magmatism and could even explain the origin of some hot spots and large igneous provinces [e.g., *Hauri*, 1996; *Takahashi et al.*, 1998; *Korenaga and Kelemen*, 2000; *Ito and Mahoney*, 2005; *Sobolev et al.*, 2007]. Given the evolving configuration of plate tectonics, the distribution of such heterogeneities is likely to exhibit regional variations; the spatially averaged stochastic characterization of global data may thus be of limited value. In this regard, what appears most promising is the deterministic mapping of small-scale heterogeneities using scattered waves in teleseismic data [e.g., *Kaneshima and Helffrich*, 1998, 2003, 2010; *Castle and Creager*, 1999; *Hedlin and Shearer*, 2000; *Krüger et al.*, 2001; *Niu et al.*, 2003; *Shearer*, 2015].

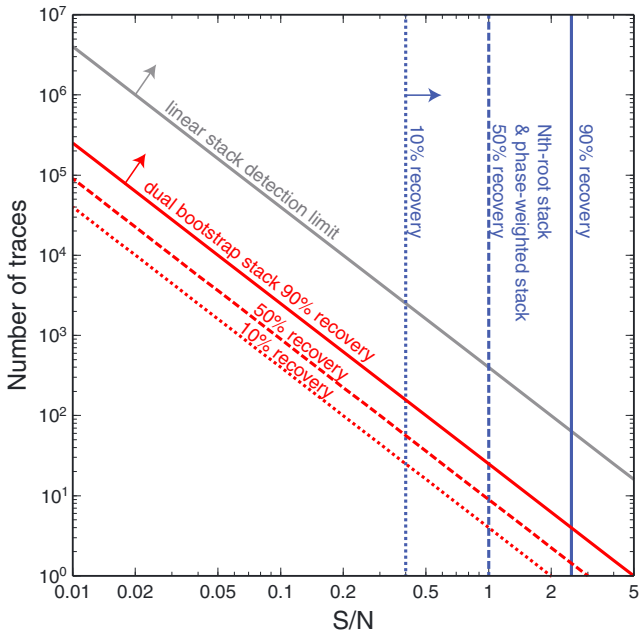

Figure 1. Comparison of four existing stacking schemes, in terms of the required number of traces for a given signal-to-noise ratio and a given rate of signal recovery (based on signal recovery experiments reported in *Korenaga* [2013]): linear (gray), NRS and PWS (blue), and DBS (red). There is no major difference between NRS and PWS in terms of signal recovery rate. The detection limit of linear stacking is based on the assumption that to detect signals unambiguously with linear stacking (i.e., with no statistical testing), the amplitude of stacked noise should be less than 10% of signal amplitude. Both NRS and PWS rely on the statistics of instantaneous polarity of data as a whole, and when S/N is low, noise dominates the statistics, resulting in lower signal recovery. Increasing the number of seismic traces does not rectify this situation.

The deterministic approach has been, however, hampered by two major difficulties. First, the amplitudes of scattered waves are usually very low. It is rare to observe a scattered phase with high signal-to-noise (S/N) ratio, and stacking a number of seismic traces is required to identify low S/N signals. Second, this stacking process leads to considerable uncertainty in the location of a scatterer. Scattered waves originating from different locations can produce similar results after stacking, and the spatial distribution of such potential scatterer locations is hereinafter referred to as an isochronal volume. In teleseismic migration, smearing appears in the isochronal volume. Small-scale heterogeneities are difficult to detect because of their limited spatial extent, and even if we are able to detect them by stacking, it is hard to estimate their locations and sizes because of the isochronal volume. Yet it would be vital to obtain such structural information, if we wish to constrain the efficiency of mantle mixing by seismological observations.

In this paper, I suggest a new array-based method to mitigate these difficulties. This is partly based on my previous work of teleseismic migration [*Korenaga*, 2014], and I continue to use dual bootstrap stacking (DBS) [*Korenaga*, 2013] to handle the low S/N issue. DBS is implemented with rigorous statistical testing, and compared with other conventional stacking schemes, it has a higher potential of signal detection (Figure 1). The isochronal volume issue is handled by developing a new kind of cluster analysis, which aims to extract the minimum set of scatterer locations from the result of teleseismic migration. I demonstrate how the new method works by exploring the Mariana region with seismic data analyzed previously [*Kaneshima and Helffrich*, 1998, 1999, 2003; *Kaneshima*, 2003]. Revisiting the same data set facilitates the comparison of methodological differences. Some important remaining issues are discussed at the end.

2. Data

The data used in this study are short-period seismograms from the University of Washington (UW) network for intermediate to deep earthquakes that occurred at the Mariana subduction zone, from 1993 to 1998 (Figure 2). Table 1 lists 16 events, for which the UW network provides data for at least 160 s after the arrival of the direct

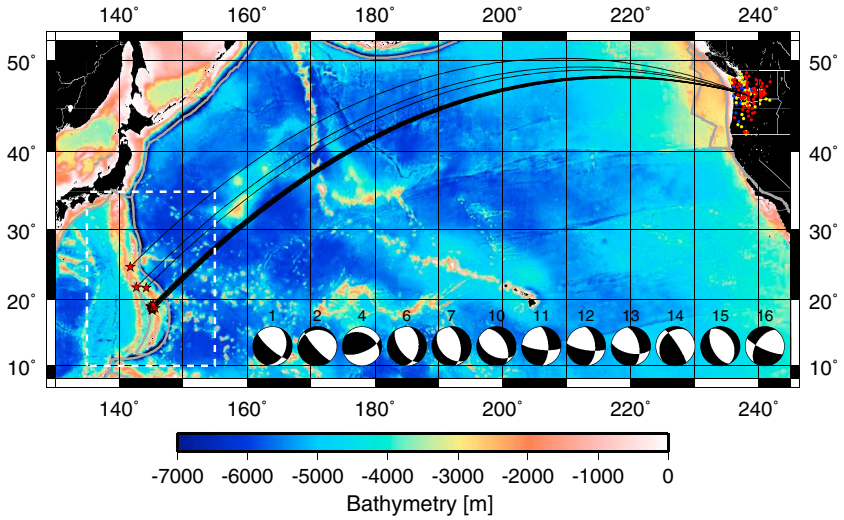

Figure 2. Locations of the events (stars) and stations (circles) for 16 data sets used in this study (Table 1), along with the predicted bathymetry of *Smith and Sandwell* [1997]. Stations used by at least 10 data sets are shown in red, those by at least 5 data sets in yellow, and those by less than 5 in blue. Raypaths of direct *P* waves are indicated by solid curves. Gray lines show plate boundaries, and dashed lines denote the extent of a target volume for migration. Source mechanisms are taken from the Global Centroid-Moment-Tensor Catalog [*Dziewonski et al.*, 1981].

P wave. Four other events analyzed by *Kaneshima and Helffrich* [2003] and *Kaneshima* [2003] are not considered because corresponding archived seismograms are of too short duration.

A band-pass filter with the corner frequencies of 0.2 and 2 Hz was applied to the seismic data, and data were resampled at every 0.05 s (the original sampling interval was 0.01 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 0.1. The number of good stations after this screening is listed in Table 1. The standard phases such as *P*, *PcP*, and *pP* in the seismic data are masked from 1 s before to 3 s after the theoretical travel time of each phase. Wider masking (from 1 s before to 8 s after) is applied for

Table 1. Events Used in This Study

ID	Date	Time (GMT)	Latitude (°N)	Longitude (°E)	Depth (km)	m_b	N_r^a
1	1993/04/02	14:32:19.00	18.423	145.221	500.8	5.2	88
2	1993/07/22	12:15:36.10	21.760	144.261	126.9	5.6	99
3	1995/02/25	14:47:59.60	18.952	145.091	602.0	4.7	45
4	1995/04/08	17:45:12.90	21.833	142.691	267.4	6.4	113
5	1995/04/23	04:30:48.80	18.497	145.169	514.9	4.9	95
6	1995/08/18	19:07:36.90	18.820	145.271	584.7	5.0	99
7	1995/08/23	07:06:02.70	18.856	145.218	594.9	6.3	128
8	1995/08/23	07:57:35.50	19.083	144.954	567.3	5.2	79
9	1995/08/23	09:44:01.70	18.898	145.192	592.3	4.7	40
10	1995/08/24	01:55:34.60	18.902	145.047	587.7	6.0	117
11	1995/08/24	06:28:54.90	18.847	145.123	602.2	5.7	105
12	1995/08/25	11:29:40.90	18.711	145.204	602.8	5.3	52
13	1995/11/14	15:14:03.60	18.792	145.247	600.0	5.3	34
14	1996/07/15	16:51:22.00	18.726	145.628	176.5	5.9	100
15	1998/02/07	01:13:36.80	24.787	141.752	525.9	5.4	71
16	1998/10/18	08:33:54.00	19.285	145.341	152.3	5.4	93

[a]Number of seismic stations after quality control.

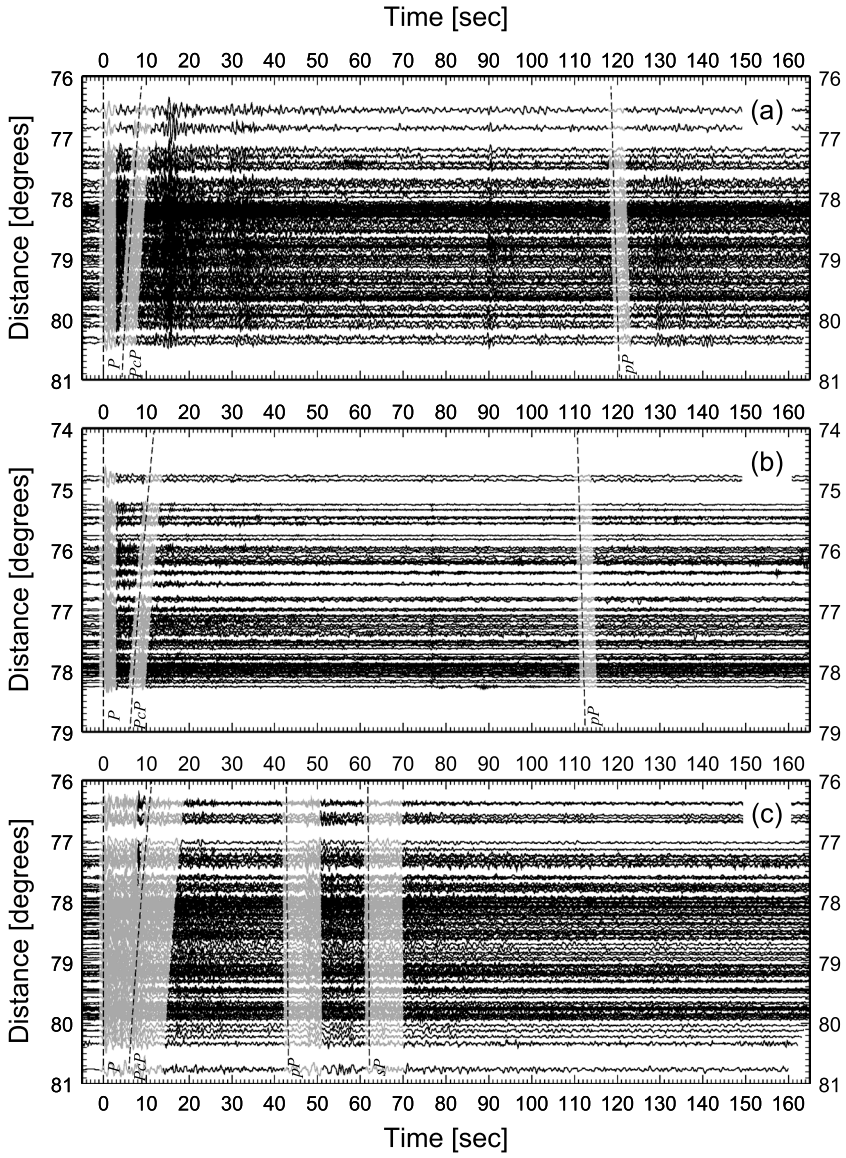

Figure 3. Vertical component data plotted in order of epicentral distance, for (a) event #8, (b) event #15, and (c) event #14 (Table 1). For display, records have been aligned on the direct *P* arrival and normalized to the amplitude of the first maximum of the arrival. Theoretical arrivals of *PcP*, *pP*, and *sP* (based on IASP91 [*Kennett and Engdahl*, 1991]) are shown as dashed lines. Masking of the standard phases is indicated by gray shading.

events #7 and #14 because the waveform of their direct *P* phase is much less impulsive than others. Masked data are simply excluded while stacking seismic traces in migration; i.e., seismic traces to be stacked need to have all of their data points unmasked within a chosen stacking window. Some of processed data are shown in Figure 3. Figure 3a (event #8) shows a few clear examples of scattered phases, e.g., at ~15 s, ~30 s, and ~90 s, and Figure 3b (event #15) shows a less prominent example at ~75 s, but these cases are not common. For most of events, scattered phases are not easily recognized in individual seismograms (Figure 3c).

3. Migration Results

As mentioned in section 1, constraining the geometries of small-scale heterogeneities is not straightforward. Figure 4 depicts the entire workflow of the new approach proposed in this paper, but even with this rather involved data processing, the estimate of scatterer geometry is not guaranteed to be unique. Nevertheless, the proposed method is notable in identifying low S/N signals and reducing the effect of isochronal volume, so it is worth being tested to see how much can be achieved. Unlike migration in exploration seismology,

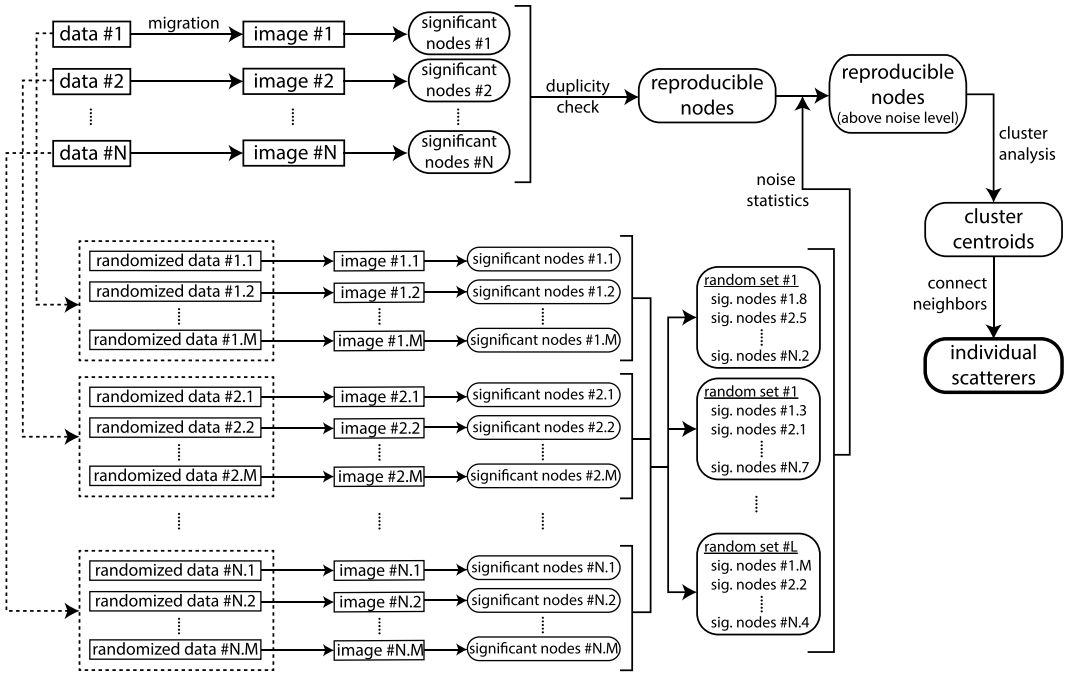

Figure 4. Flowchart for the new array-based method, from the migration of individual data sets to the final product of scatterers. For each of N data sets, M randomized data sets are generated, and from the total of $N \times M$ migration images for randomized data, L different random sets are created to estimate the background noise level for reproducibility. The use of DBS is assumed for migration.

where these difficulties can naturally be resolved by the sheer number of seismic traces and wide-aperture source and receiver arrays, teleseismic migration has to endure, in the foreseeable future, limitations both on fold and aperture. It is thus important to pursue how much can be extracted from teleseismic data, even under the simple framework of kinematic isotropic point scattering.

This section describes the workflow up to determining "reproducible" nodes above noise level (Figure 4). The remaining two steps are explained in the next section. Conducting teleseismic migration for a single data set, which corresponds to one earthquake event, is a starting point. From individual migration images, statistically significant nodes are identified, and their reproducibility among multiple data sets is then quantified. Only those nodes with sufficient reproducibility are forwarded to the subsequent cluster analysis.

3.1. Migration of Single Data Sets

As in the previous studies on the Mariana region, I focus on *S*-to-*P* scattering, which could be more pronounced than *P*-to-*P* scattering because the direct *S* wave has greater energy than the direct *P* wave. At the same time, the *S* wave travels much slower than the *P* wave, so the volume that can be probed by *S*-to-*P* scattering, using the first 100–200 s after the direct *P* arrival, is restricted to the vicinity of the seismic sources. The later section of seismograms is populated with a variety of standard phases and is not suitable for a search for weak scattered phases.

To detect scattered phases and locate their possible origins, I use an imaging technique called teleseismic migration [e.g., *Rost and Thomas*, 2002]. A target volume for migration, which is common to all cases considered in this paper, spans from 200 km to 2000 km in depth and is bounded between 10°N and 35°N and between 135°E and 155°E (Figure 2). 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 $\sim 8 \times 10^4$. For each of these nodes, a travel time for a hypothetical scattered wave is calculated for all stations, and seismic energy corresponding to the arrival of the scattered wave is stacked to see how seismic waveforms from different stations add up. The nodes at which waveforms add up coherently can be considered as potential scatterer locations. For details on the migration procedure, see *Korenaga* [2014]. For the theoretical calculation of travel times, the IASP91 Earth model [*Kennett and Engdahl*, 1991] is used. The migration uses the differential travel time with respect to the direct *P* arrival to minimize the effect of three-dimensional (3-D) velocity variations

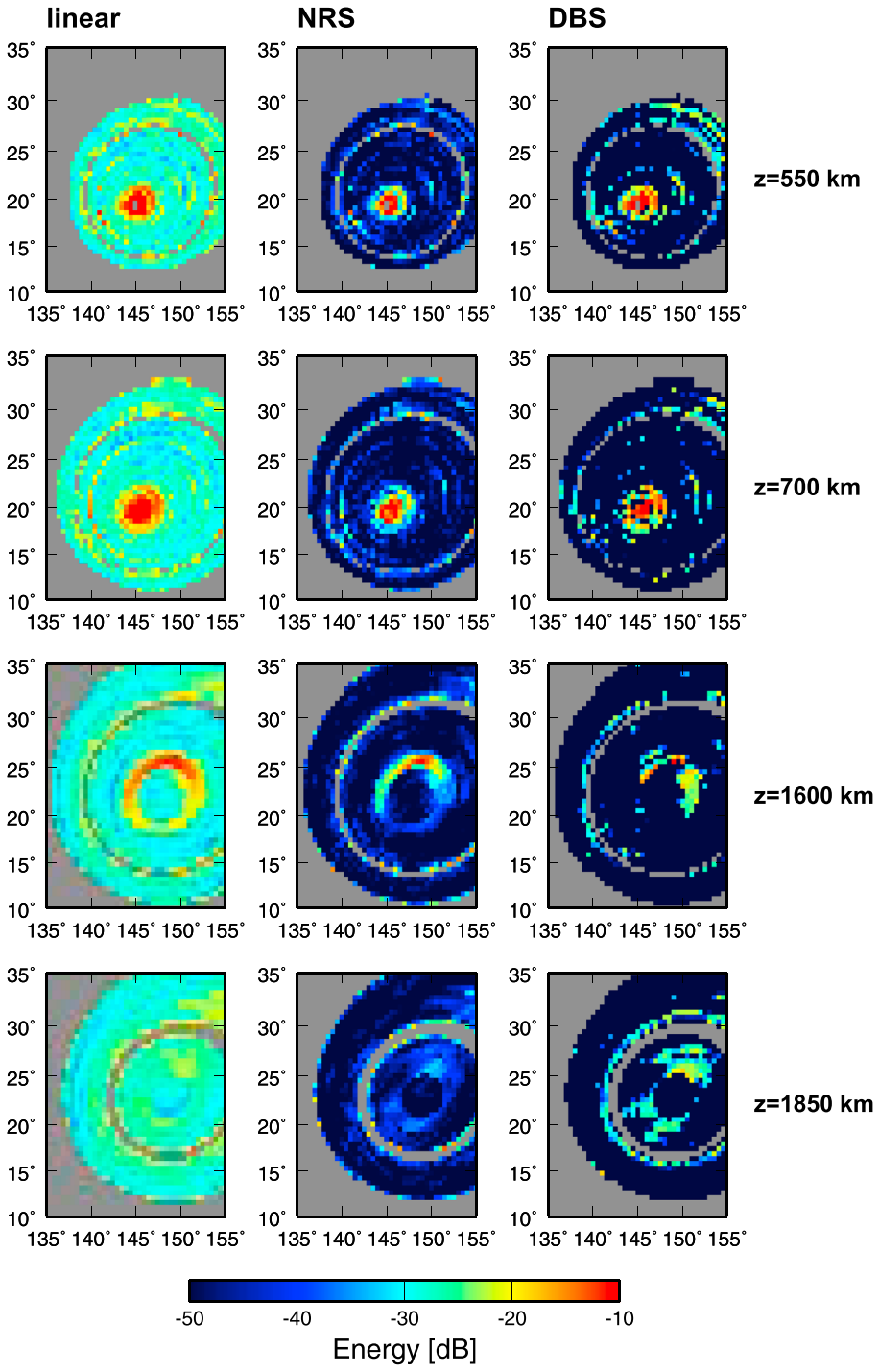

Figure 5. Migration energy for event #8 at selected depths, with different stacking schemes. From left to right: linear stack, *N*th-root stack (NRS), and dual bootstrap stack (DBS). Gray shading denotes the region that cannot by sampled by *S*-to-*P* single scattering with data up to 160 s after direct *P*; it also reflects the masking of standard phases. Migration energy is shown in the unit of decibel (dB), i.e., $10 \log_{10} E$, where zero dB corresponds to the migration energy of direct *P*.

unaccounted for by the reference model. The use of differential travel time would not reduce the effect of velocity variations near sources (e.g., within the migration volume), where the raypaths for the direct *P* and scattered waves can be widely apart, and travel time calculations based on a 1-D velocity model could lead to some errors in scatterer locations. Given the typical velocity perturbations of 1–2% seen in tomographic models, however, such errors are much smaller than the adopted grid resolution. Data up to 160 s after the direct *P* are used, and the length of the time window for stacking is set to 4 s based on the duration of the first main wavelet of direct *P*.

Some fraction of the migration volume is inaccessible by the data of this duration, and stacking is conducted only when a hypothetical scattered phase can be expected at the minimum of 10 stations. The number of seismic traces used for stacking is close to this minimum near the edges of the migratable region.

Migration results for event #8 are shown in Figure 5 for selected depths, using three different stacking schemes: linear stacking, *N*th-root stacking (NRS) [*Muirhead*, 1968; *Kanasewich et al.*, 1973], and DBS. NRS is done with the power of 3 (i.e., cube-root stacking). Phase-weighted stacking (PWS) [*Schimmel and Paulssen*, 1997] is another popular nonlinear stacking, and its performance is similar to that of NRS because both schemes are based on the same kind of statistical consideration [*Korenaga*, 2013, 2014]. For DBS, 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. As in *Korenaga* [2014], I use the local maxima acceleration for computational expediency; for a few data sets, migration was done also with the original DBS, which is more time consuming, and the results were virtually the same.

Linear stacking is simple averaging, so it suffers most from migration artifacts. There are two kinds of artifacts, one created by isochronal smearing and the other by low folds at boundaries. For the migration volume considered in this study, the isochronal artifact appears mostly annular, and ring-shaped features seen in the linear stack image could be generated by only one scatterer per each ring. The low-fold artifact appears at the edges of the migratable region; the number of waveforms to be stacked is too low to reduce noise sufficiently. In addition, the linear stack image is generally quite bright, with a background energy level of around −30 dB.

Both the NRS and DBS images have a lower background level (< −50 dB), with the isochronal and low-fold artifacts substantially reduced (Figure 5). At first look, they appear similar to each other, but except for the strong peaks seen at the depths of 550 and 700 km, they are different in many details. The DBS image exhibits a larger number of nodes with high migration energy, which is expected because the recovery rate of low S/N signals is lower in NRS (Figure 1). There are some nodes where NRS shows high energy but DBS shows nothing, and most of such nodes are located at the edges of the migratable region. This means that NRS does not eliminate low-fold artifacts completely, which is also expected because the residual noise level of NRS is higher than that of DBS when the number of traces is smaller than ∼50 [*Korenaga*, 2013].

The magnitude of a detected signal with respect to the background noise may be quantified as an apparent S/N as *Korenaga* [2014]

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

where E is the migration energy and R is the root-mean-square amplitude of corresponding data before stacking. This serves as a lower bound for a true S/N because signal recovery by stacking is not always perfect. Given the signal detection capability of DBS (Figure 1), the apparent (S/N) $n^{1/2}$ of 2, where n is the number of seismic traces, can be used as a threshold to identify significant signal with high recovery rate [*Korenaga*, 2014]. Figure 6 shows the depth distribution of such significant nodes measured from the DBS migration images for all events. The number of significant nodes varies among data sets, from ∼100 (event #3) to ∼9000 (event #14), and in general, an event with a greater magnitude yields a larger number of significant nodes (cf. Table 1).

Also shown in Figure 6 is the background noise level calculated as follows. For each data set, seismic traces are randomly time shifted with the maximum period of 20 s to generate an ensemble of 10 randomized data sets (i.e., $M = 10$ in Figure 4). Each of these randomized data sets is processed by migration in the same manner for the original data, and the depth distribution of significant nodes is measured. The noise level shown in Figure 6 is based on the statistics of such measurement from 10 randomized data sets. This estimate of background

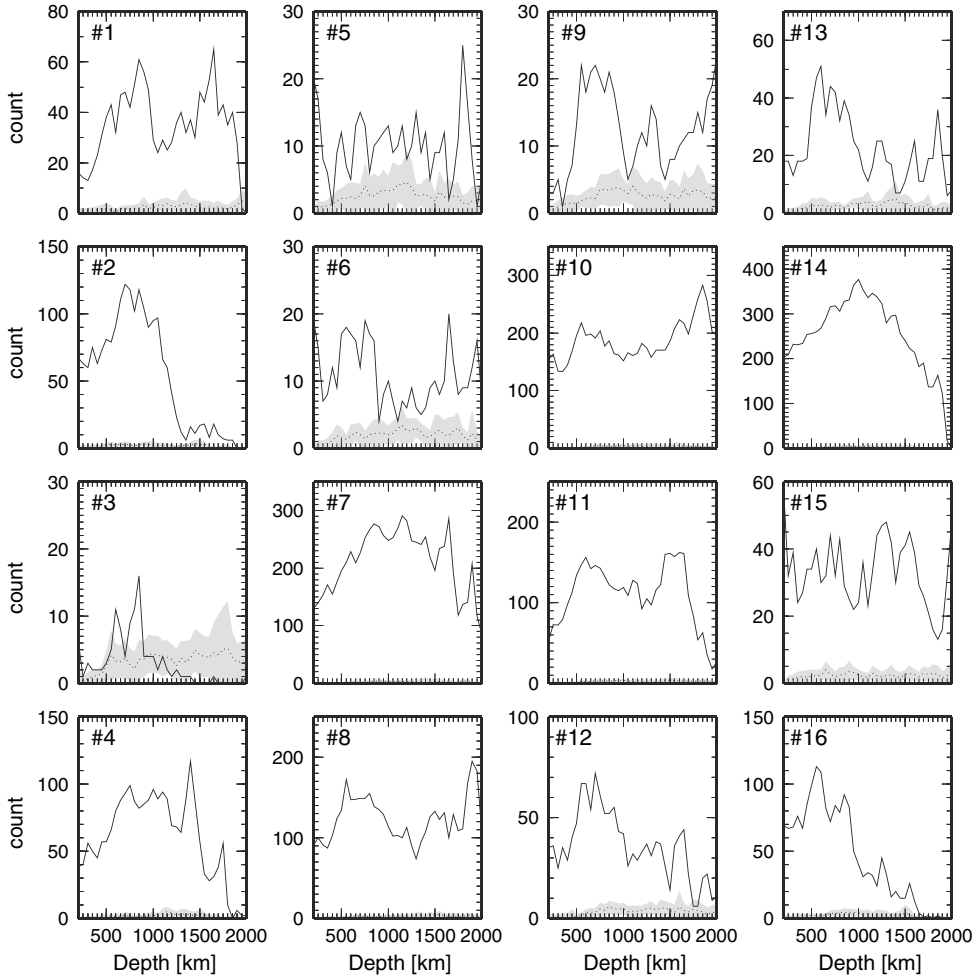

Figure 6. The depth distributions of significant nodes (defined as nodes with the apparent (S/N) $n^{1/2}$, where n is the number of traces, greater than 2), for DBS-based migration of all data sets. Solid line corresponds to results from the original data set, whereas dotted line and gray shading denote, respectively, the mean and one standard deviation of results from 10 randomized data sets.

noise level is important because the probability of having random noise happened to be aligned is rather high in migration imaging, in which the numerous combinations of trace shifting are tested. For the majority of the data sets, the background noise is negligible. The situation is contrasting to the previous study of DBS-based migration, where the relation between the significant nodes and the background noise level is similar to that for events #5 and #6 [see *Korenaga*, 2014, Figure 12]. This is not merely because the Mariana data sets exhibit some visible, high S/N scattered phases; a large number of significant nodes, with an insignificant noise level, are seen for events with no obvious scattered phase such as event #14 (Figure 3c).

3.2. Test for Reproducibility

One consistent feature among previous studies is that the appearance of scattered phases is highly variable from event to event, and this variability is also evident in Figure 6. Some events result in a large number of statistically significant nodes in migration, whereas others show much fewer significant nodes. This is not surprising because the efficiency of scattering depends on a source mechanism, the topology of a scatterer, and the relative location of a scatterer with respect to a source-receiver path. The detection of a scattered phase by stacking is further influenced by the number of seismic traces as well as the presence of noise in those data. Different seismic events therefore tend to illuminate the isochronal volume of a scatterer differently. Indeed, it is common to see high migration energy at certain nodes in one migration image, but much weaker or no energy at all at the same nodes in other images. This variable illumination of isochronal volume complicates the definition of reproducibility in teleseismic migration.

In *Korenaga* [2014], significant nodes in different migration images are considered to coincide if they are within the travel time variance of 1 s^2. The travel time variance is defined as

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

where j_1 and j_2 denote two different nodes in a migration volume, N_r is the number of seismic stations, and t_{ij}^t is the theoretical travel time from the hypocenter to the ith station through the jth node. The choice of the threshold variance was, however, ad hoc. It is desirable to avoid such a free parameter, especially when it causes a sharp cutoff in the definition of reproducibility.

In this paper, reproducibility among multiple migration images is assessed by calculating the following quantity called "duplicity" at each node j:

$$\mathcal{D}(j) = H\left(\sum_{i=1}^{N_s} \delta_{j,s_i(j)} - \frac{1}{2}\right)\left[\sum_{i=1}^{N_s} \mathrm{erfc}\left(\frac{\sigma_t^2(j, s_i(j))}{\sigma_{crit}^2}\right)\right], \tag{3}$$

where $H(\cdot)$ is the Heaviside step function, $\delta_{i,j}$ is the Kronecker delta, $\mathrm{erfc}(\cdot)$ is the complementary error function, $s_i(j)$ is the index for the nearest significant node in the ith image to the node j, and σ_{crit}^2 is the critical travel time variance. If a given node is not statistically significant in any of images, the argument for the step function becomes negative, so such a node has zero duplicity. The step function part thus ensures that a node with nonzero duplicity is registered as a significant node in at least one image. The summation in the square brackets measures how a given node is supported by a collection of images. If the node is statistically significant in the ith image (i.e., $s_i(j) = j$), $\sigma_t^2(j, s_i(j))$ is zero, contributing the value of unity to the summation. Otherwise, the travel time variance becomes nonzero, resulting in lower contribution. The travel time variance is normalized by the critical variance, and the contribution becomes lower than 0.1 when $\sigma_t^2(j, s_i(j)) > \sim 1.2\sigma_{\mathrm{crit}}^2$ and essentially zero when $\sigma_t^2(j, s_i(j)) > \sim 2\sigma_{\mathrm{crit}}^2$. The "nearest" significant node $s_i(j)$ in the ith image is the significant node with the smallest travel time variance to the node j; the proximity is not measured by the distance in a migration volume. The notion of travel time variance was first introduced by *Kito and Korenaga* [2010] to define the extent of an isochronal volume, and as a rule of thumb, the square of the half period of the stacking window can serve as a critical variance. The result of duplicity calculation with the critical travel time variance of 4 s^2 is shown in Figure 7. The use of the complementary error function in equation (3), along with the choice of critical travel time variance, introduces some level of subjectivity, but their influence can be examined by using different thresholds (see section 5.1).

To determine how high duplicity should be to call a given significant node reproducible, the background noise level must be estimated for duplicity. To this end, the migration images previously generated for randomized data are used. Each event has 10 images from randomized data, and out of these 10 images, one image is randomly chosen to compose a random set of multiple images, for which duplicity is calculated. This procedure is repeated 10^3 times (i.e., $L = 10^3$ in Figure 4) to estimate reproducibility among randomized data sets. The result is summarized in Figure 8a. For both the original data sets and the randomized ones, the number of reproducible nodes decreases as required duplicity increases, but more rapidly so for the randomized data sets. At low duplicity (<5), even the randomized data sets yield high node counts ($\sim 10^4$), indicating that randomly located nodes in different images can easily be connected through isochronal volumes. At the required duplicity of 10, however, the probability of having a reproducible node becomes essentially zero for the randomized data sets, whereas the original data sets still have $\sim 10^3$ reproducible nodes, which are analyzed further in the next section.

Figure 8b shows how these reproducible nodes are supported by different data sets. Despite a large variation in the number of significant nodes among data sets, the reproducible nodes are supported more or less equally by all data sets. Event #3, for example, has only $\sim 10^2$ significant nodes, and only ~20 of them coincide with some of the reproducible nodes. Its total node contribution is, however, much higher (~500) because the rest of the significant nodes are connected to many of the reproducible nodes with small travel time variances. This is interesting because the depth distribution of the significant nodes for event #3 is nearly indistinguishable from the background noise (Figure 6). Without this test for reproducibility, therefore, one could easily disregard the event #3 data as statistically insignificant. It is also notable that node contribution from events

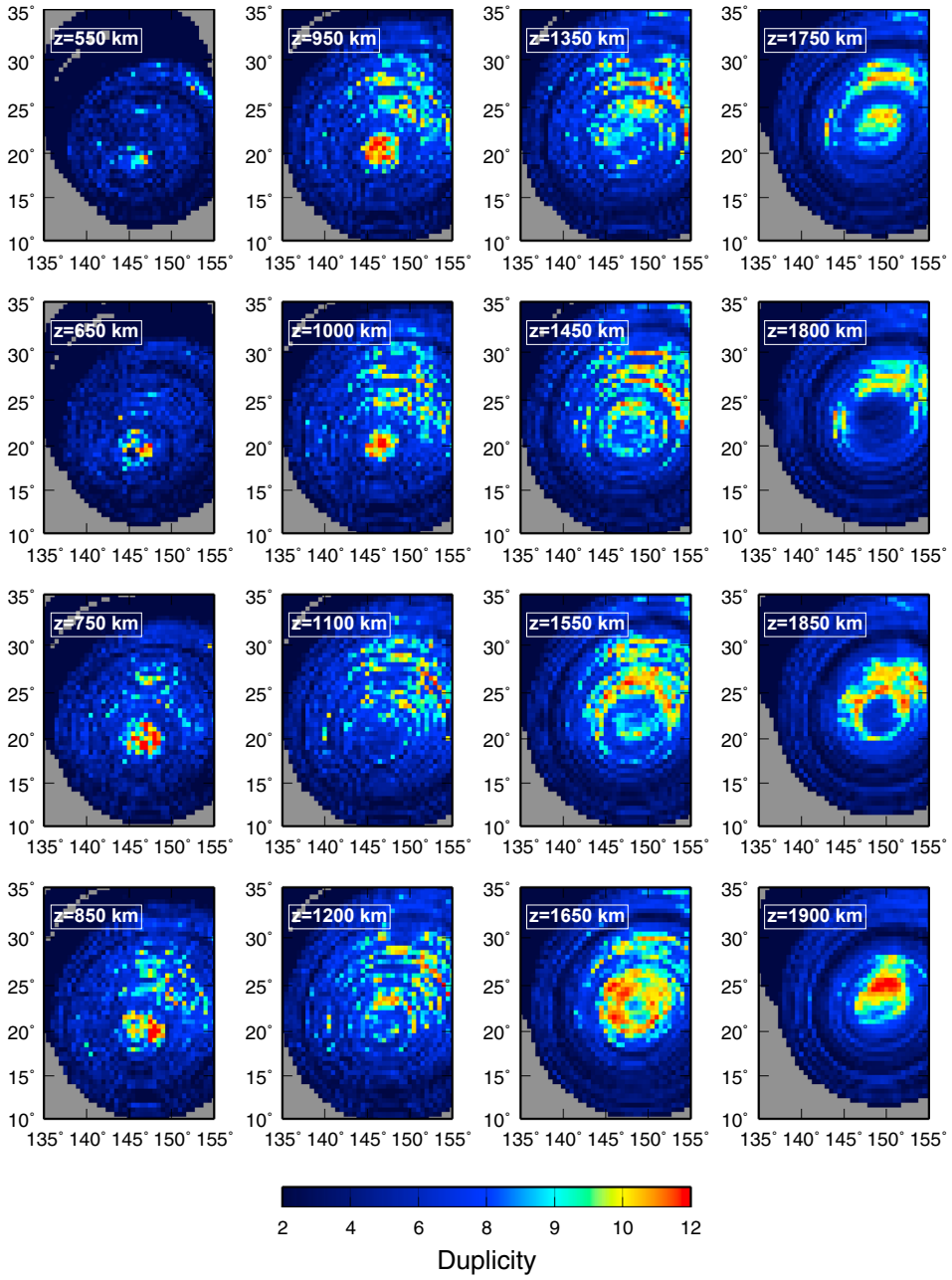

Figure 7. Duplicity at selected depths.

with a large number of significant nodes (#7, #8, #10, #11, and #14) is also at the level of $\sim 10^3$. The number of the significant nodes that coincide exactly with reproducible nodes is approximately 10 $\pm$ 5% of the total number of significant nodes for all events, so the majority of the significant nodes in those events are not sufficiently reproducible.

Given the spread of source locations (Table 1), this screening with duplicity establishes that these reproducible nodes represent potential point scatterers for the assumed *S*-to-*P* scattering, and not artifacts caused by erroneously migrating receiver-side reverberations. When source depths vary by ~500 km, a scattered phase from a near-source scatterer should have different delay times (with respect to the direct *P*) for different events. On the other hand, the delay times of receiver-side reverberations should not vary much among different events. Receiver-side reverberations as well as other types of scattering can still exist in the seismic

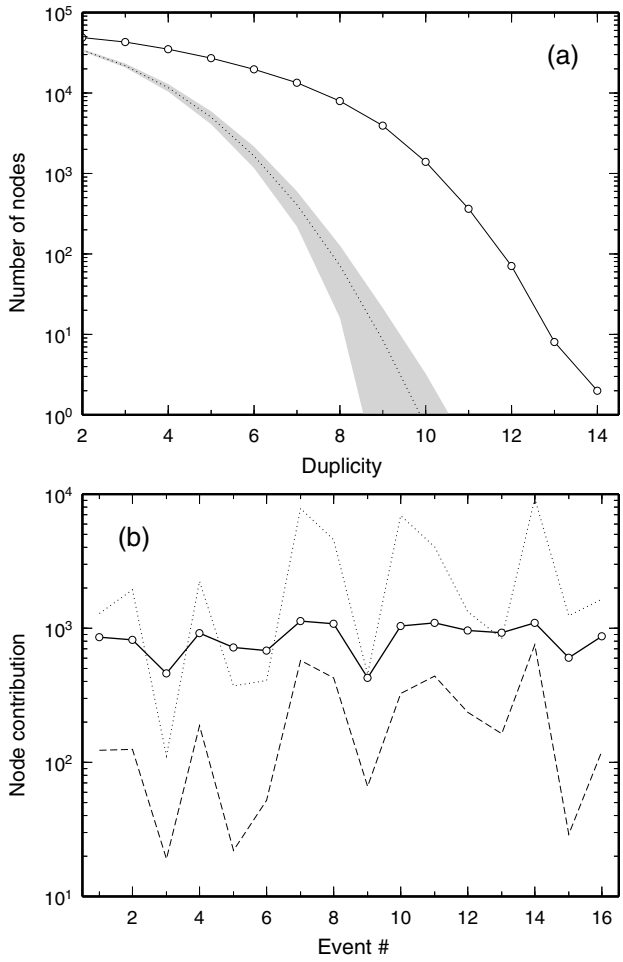

Figure 8. (a) The number of nodes that satisfy a given duplicity requirement (solid with circles). For comparison, dashed line and gray shading denote, respectively, the mean and one standard deviation of 10^3 reproducibility test results using randomized data sets. (b) Contribution from different events to duplicity (solid with circles), shown for the case of minimum duplicity of 10. Dotted line denotes the number of significant nodes per event, whereas dashed line shows how many of them contribute directly to duplicity.

data used in this study, but the reproducibility test allows us to extract scattered phases with the assumed type of scattering.

4. Cluster Analysis

The semicircular isochronal artifact is still visible even in the highly reproducible nodes with the minimum duplicity of 10 (Figure 7). The extent of the artifact generally increases with greater depths, because isochronal volumes tend to become larger for deeper scatterers. Visual impressions can be misleading; a tiny clump of reproducible nodes at the depth of 650 km, for example, is no less significant than the pronounced red spot at the depth of 1900 km. Also, an isochronal volume is 3-D, so the extent of isochronal artifact is not confined in a given depth slice. Seeing a bright spot at a certain depth does not necessarily mean the presence of a scatterer at that depth.

One way to address the isochronal volume issue is to estimate the minimum set of scatterers that can explain all of these reproducible nodes. This is not an easy problem. One may compute an isochronal volume for each of the reproducible nodes to see how many other nodes can be contained in it, but how to proceed from there is not clear. It would be easy to find the node that can explain the largest number of reproducible nodes, but unless it can explain all of the reproducible nodes, it may not be the best to accept it as a scatterer location. To explain the other nodes that are left over, one may need many scatterers, whereas all of the reproducible nodes could be explained by much fewer scatterers if their locations are optimally

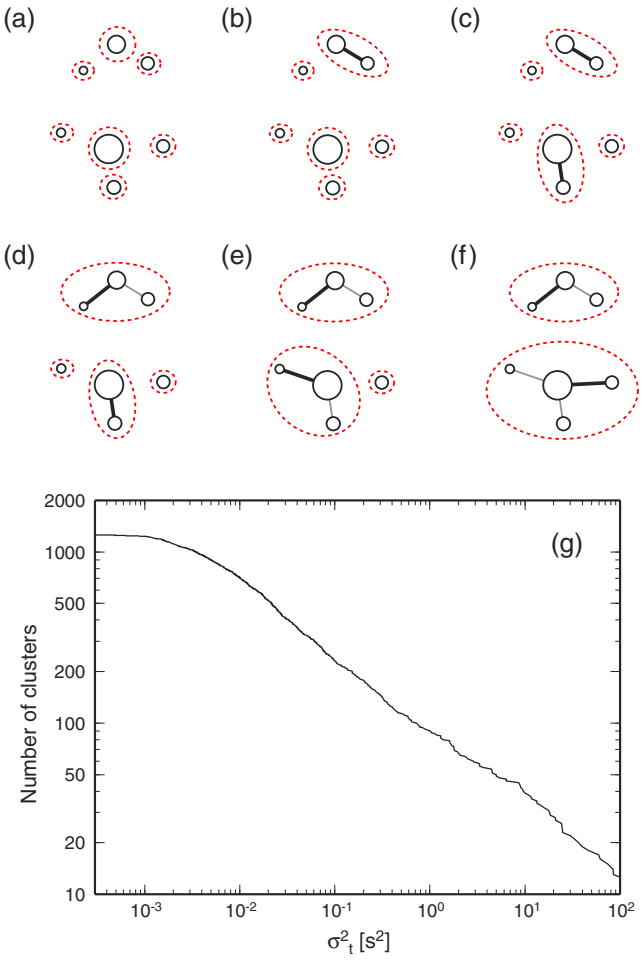

Figure 9. (a–f) Cartoon illustrating the progression of hierarchical clustering with the centroid-based linkage adopted in this study. Clusters are indicated by dashed ovals. The size of a node indicates its amplitude, and a distance between nodes reflects a travel time variance between them. A centroid in a cluster is the node with the largest amplitude. At the beginning, all nodes are clusters of their own, and at every subsequent step, two clusters are merged when the greatest distance from the centroid in a new cluster (shown by thick solid line) is the minimum among all possible (hypothetical) cluster pairs. (g) The number of clusters as a function of travel time variance, starting with ~1300 significant nodes with the minimum duplicity of 10 (i.e., reproducible nodes).

chosen. To solve this large-scale combinatorics problem in a reasonable time frame, I choose to use hierarchical cluster analysis [e.g., *Defays*, 1977]. *Korenaga* [2014] also used cluster analysis for the same purpose, but its algorithm adopted complete linkage so that the distance between clusters is defined as the maximum travel time variance between all possible pairs for all nodes contained in the two clusters. Such a linkage does not yield the minimum set of scatterers. Also, amplitude information was ignored in clustering. The results of cluster analysis were not readily interpretable for the geometries of scatterers, either. An improved procedure, which involves a two-step application of cluster analysis, is therefore presented in this section.

4.1. Amplitude-Based Hierarchical Clustering

Hierarchical clustering defines clusters in a top-down or bottom-up manner based on a certain metric, which is a measure of distance between pairs of points. In the agglomerative hierarchical clustering, which is adopted here, each datum 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 metric among all possible cluster pairs are combined. The definition of a metric is thus central to hierarchical clustering, and I choose to define it as the travel time variance with respect to a node with the largest amplitude among all the nodes in the cluster pair under consideration (Figure 9). At any step in clustering, a node with the largest amplitude in a cluster serves as the centroid of the cluster, so a distance between two clusters is measured with respect to a centroid with the larger amplitude. As clustering proceeds, the number of clusters decreases, and the minimum metric increases; clustering is

Table 2. Cluster Centroids and Scatterers

Scatterer #	Centroid #	Latitude (°N)	Longitude (°E)	Depth (km)	Amplitude (dB)	N^a
1	1	23.0	147.0	200	−19.7	3
2	2	19.5	145.0	500	−9.3	11
3	3	19.5	146.0	650	−12.7	3
4	4	21.5	147.0	700	−12.7	9
5	5	20.5	147.5	700	−12.2	12
6	6	19.0	145.0	750	−13.0	16
7	7	20.0	145.5	750	−3.4	15
	8	19.5	146.0	750	−10.1	4
8	9	26.5	147.5	750	−29.0	5
9	10	20.5	147.0	800	−12.1	13
10	11	25.5	149.0	800	−24.3	14
11	12	19.5	149.0	850	−24.8	6
12	13	19.0	146.5	950	−17.6	10
	14	19.5	146.5	950	−17.0	7
	15	20.0	147.0	900	−15.4	7
	16	20.5	147.5	900	−11.5	19
	17	21.0	147.5	900	−15.9	6
	18	19.5	147.5	850	−15.9	8
	19	19.5	148.0	850	−11.4	20
	20	20.0	148.0	800	−10.9	34
	21	19.5	147.5	950	−14.7	23
13	22	24.5	151.5	900	−25.6	9
14	23	23.5	150.5	950	−21.3	9
	24	24.0	150.5	900	−21.8	28
15	25	20.5	145.5	950	−14.4	10
	26	21.0	146.0	1000	−19.0	13
16	27	23.5	152.5	1000	−23.2	9
17	28	23.5	147.5	1200	−18.8	44
18	29	21.5	152.5	1250	−26.8	5
19	30	30.0	147.5	1300	−20.1	50
20	31	25.0	150.0	1400	−23.8	28
21	32	25.0	148.5	1400	−23.7	23
	33	24.5	149.0	1450	−22.4	6
22	34	20.0	142.5	1450	−24.0	21
23	35	26.5	151.0	1450	−21.0	68
24	36	27.5	147.5	1500	−21.3	27
25	37	27.5	149.5	1500	−20.7	18
26	38	25.5	151.5	1500	−19.1	47
27	39	26.5	148.5	1500	−20.6	45
	40	26.5	148.5	1550	−19.4	25
	41	26.0	149.0	1550	−15.0	64
	42	26.0	149.5	1600	−10.5	74
28	43	27.5	151.0	1550	−20.4	35
29	44	25.5	145.0	1600	−24.2	24
30	45	23.0	151.0	1600	−23.1	14
31	46	25.0	146.0	1650	−24.8	15
32	47	25.0	148.5	1650	−16.3	64
33	48	29.5	148.0	1700	−25.1	38

Table 2. (continued)

Scatterer #	Centroid #	Latitude (°N)	Longitude (°E)	Depth (km)	Amplitude (dB)	N^a
34	49	23.0	149.5	1700	−23.3	10
35	50	24.5	148.5	1750	−23.5	38
36	51	27.5	150.0	1800	−26.5	11
37	52	26.5	151.5	1800	−26.4	10
38	53	26.5	150.5	1850	−20.6	56
39	54	25.5	150.5	1900	−21.1	93

[a]Number of reproducible nodes in a cluster.

terminated when the minimum metric exceeds a given threshold, which is set to the critical travel time variance used for duplicity. All of reproducible nodes are contained in resulting clusters, and all nodes in a given cluster are within the isochronal volume of a cluster centroid.

As duplicity is based on multiple migration images, there can be different ways of defining the representative amplitude of a given node, and I adopt the maximum amplitude among different images because it likely represents the case of the highest scattering efficiency as well as the highest S/N among different data sets. The data sets correspond to different source-receiver combinations, so a travel time variance between nodes is also nonunique. Here I choose the minimum travel time variance, which should lead to the smallest number of clusters.

With the critical travel time variance of 4 s^2, a total of 54 clusters are formed (Table 2), and some characteristics of resulting clusters are shown in Figure 10. The amplitudes of cluster centroids are mostly in the range

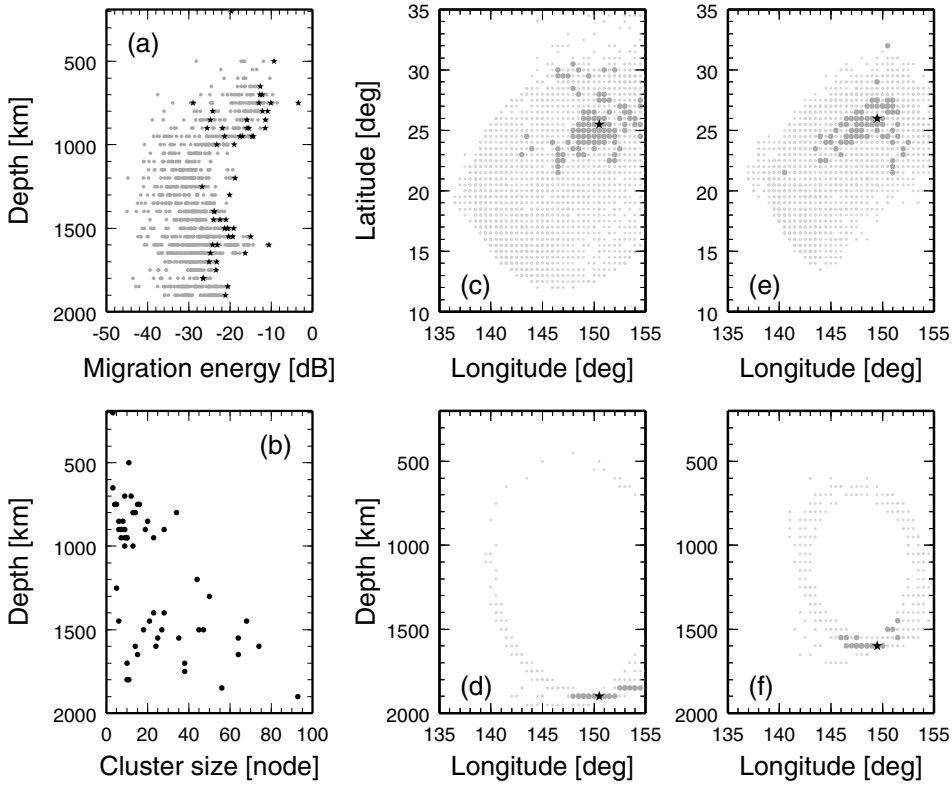

Figure 10. (a) Covariation of the depth and migration energy of reproducible nodes (gray circles) and the centroid of clusters (stars). (b) Covariation of the size of clusters and their centroid depth. (c and d) The distribution of the largest cluster (with 93 reproducible nodes) is shown. Its centroid is denoted by a star and all other nodes by gray circles. Light gray dots denote nodes in the isochronal volume of the centroid with the travel time variance of <0.5 s^2. All nodes are projected onto the surface in Figure 10c, whereas only nodes within the latitudes of 25°N and 26°N are shown in Figure 10d. (e and f) The second largest cluster (with 74 nodes) is shown similarly.

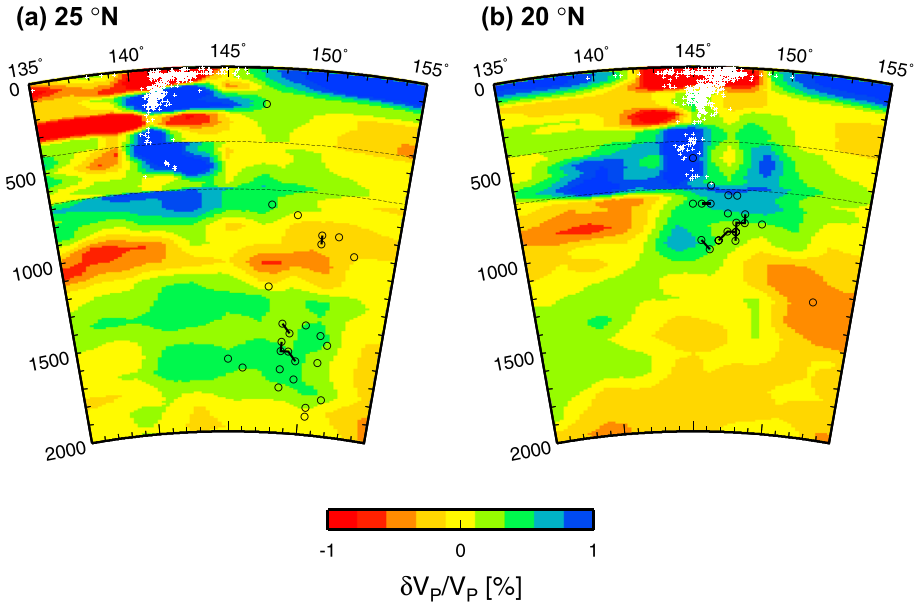

Figure 11. Spatial distribution of cluster centroids (circles), shown on the *P* wave tomography model of *Simmons et al.* [2012]. Neighboring centroids are connected by solid lines. In two vertical cross sections at the latitude of (a) 25°N and (b) 20°N, centroids with nearby latitudes (±2°) are included. Dotted lines are drawn at the depths of 410 km and 670 km, and white crosses denote hypocenters of earthquakes since 1980 (from the NEIC PDE catalog).

between −10 dB and −25 dB; i.e., the amplitudes of corresponding scattered phases are in the range between ∼30 % and ∼6 % of that of the direct *P*. The number of nodes contained in a cluster generally increases as the depth increases, as expected from how an isochronal volume varies with the location of a scatterer. Examples of a large cluster are also shown in Figure 10; it can be seen that only a small fraction of an isochronal volume is illuminated by reproducible nodes. These isochronal volumes exhibit somewhat complicated shapes because they are based on multiple source-receiver configurations.

4.2. Connecting Cluster Centroids

In the 3-D migration grid used in this study, every internal node is surrounded by 26 neighbor nodes. When two cluster centroids are neighbors, it is reasonable to postulate that they collectively represent a continuous scatterer spanning them. These centroids carry their own reproducible nodes, and the former cluster analysis guarantees that the combination of these reproducible nodes cannot be explained by a single scatterer. Cluster centroids are thus connected whenever two are neighboring nodes; this connection is done by another cluster analysis, in which the nodal distance is used as a metric. This is the final step in determining the geometries of scatterers (Figure 4). This second cluster analysis yields 39 scatterers (Table 2). Though most of the 54 centroids are stand-alone scatterers, there are two large scatterers, one with nine centroids (at the depth of ∼900 km) and the other with four centroids (at ∼1550 km). The Mariana region has been studied repeatedly in the past [e.g., *Kaneshima and Helffrich*, 1998; *Castle and Creager*, 1999; *Krüger et al.*, 2001; *Niu et al.*, 2003; *Rost et al.*, 2008; *Bentham and Rost*, 2014], but this is the first time to identify this many scatterers with their locations and sizes (albeit crudely) estimated.

The spatial distribution of these scatterers is compared with the *P* wave tomography model of *Simmons et al.* [2012] (Figure 11), and most scatterers appear to be found within high-velocity anomalies. These velocity anomalies are commonly interpreted as the remnant of subducted slabs, so the detected scatterers may well represent the fragments of oceanic crust from past subduction history [e.g., *Kaneshima and Helffrich*, 1999, 2003; *Rost et al.*, 2008]. Stand-alone scatters (i.e., with just one centroid each) may look insignificant in Figure 11, but each of them is supported by a collection of reproducible nodes with the minimum duplicity of 10. A few scatterers in low-velocity anomalies, therefore, also deserves some attention, and they may represent long-lived chemical heterogeneities originating in ancient subduction or continental delamination.

With the grid spacing used in this study, each node can represent up to a volume of $\sim 1.5 \times 10^5$ km^3 (at the depth of 200 km) to $\sim 8 \times 10^4$ km^3 (at the depth of 2000 km). Using this nodal volume, the volume of

scatterers may be roughly estimated to vary from $< 8 \times 10^4$ km^3 to $\sim 1 \times 10^6$ km^3. This is obviously an uncertain estimate, and it could be an upper bound if scatterers are more two-dimensional (e.g., a thin plate). The seismic illumination of a scatterer is likely to be incomplete, however, and given the adopted strategy of estimating the minimum number of scatterers by suppressing isochronal volumes, such a volume estimate may not be considerably incorrect. An independent check may be done with the amplitude of scattered phases. For *S*-to-*P* scattering, the maximum amplitude of a scattering wave due to a spherical heterogeneity may be expressed as [*Wu and Aki*, 1985]

$$\max\left(\frac{A'}{A_P}\right) \sim \frac{A_S}{A_P}\frac{V}{4\pi r}\left(\frac{2\pi f}{\alpha_0}\right)^2\left|\frac{\delta\rho}{\rho_0}\right|, \tag{4}$$

where A', A_P, and A_S are the amplitudes of a scattered wave, the direct *P*, and the direct *S*, respectively, V is the volume of a spherical heterogeneity, r is the distance between the heterogeneity and a receiver, f is the frequency of a seismic wave, α_0 and ρ_0 are, respectively, the *P* wave velocity and density of the ambient medium, and $\delta\rho$ denotes density perturbation due to the heterogeneity. With $A_S/A_P \sim 2$, $r \sim 9000$ km, $f \sim 1$ Hz, $\alpha_0 \sim 12$ km s^{-1}, and $\delta\rho/\rho_0 \sim 0.05$, a spherical scatterer with the volume of 10^6 km^3 would result in A'/A_P of ~ 0.24, or in terms of migration energy, around -12 dB. The volume estimate is thus in a broad agreement with the amplitude information (Table 2).

5. Discussion

5.1. Sensitivity to Critical Travel Time Variance

The critical travel time variance σ^2_{crit}, used in the duplicity calculation as well as in the first cluster analysis, is the only free parameter in this study. The adopted value of 4 s^2 may be justified as follows: It means that the two sets of synthetic travel times (at an array of seismic stations) for different scatterer locations are different by 2 s on average, which seems long enough to call them dissimilar when the dominant frequency of filtered seismic data is around 1 Hz (section 2). Nevertheless, it is important to investigate how results depend on the definition of isochronal volume, which plays an essential role in handling isochronal artifacts. The effect of using a higher threshold is obvious; it means assuming larger isochronal volumes, which lead to higher duplicity not only for the original data sets but also for the randomized ones, and the background noise level increases in a rather artificial way. The outcome of lowering the threshold is less clear and thus of primary concern.

Results with lower critical travel time variances are summarized in Figure 12. A lower threshold leads to a smaller number of nodes for a given duplicity, but this effect is more severe for the randomized data sets. The critical duplicity, at which the randomized data sets yield essentially no reproducible node, decreases from 10 (with σ^2_{crit} of 4 s^2) to 8 (with 2 s^2) and 6 (with 1 s^2) (Figure 12a). So the net effect of lowering the threshold is an increase in the number of reproducible nodes: $\sim$1500 and $\sim$2300 for σ^2_{crit} of 2 s^2 and 1 s^2, respectively. The use of a lower critical travel time variance is expected to result in a larger number of clusters, and indeed, the subsequent cluster analyses yield 89 centroids and 53 scatterers for σ^2_{crit} of 2 s^2 and 151 centroids and 85 scatterers for 1 s^2. The overall depth distribution of these scatterers is, however, maintained with different thresholds (Figure 12b). There are at least two large scatterers at the depths of $\sim$900 km and $\sim$1600 km, with a relatively uniform distribution of smaller scatterers from $\sim$500 km to $\sim$1900 km. The scatterers listed in Table 2 may thus be regarded as a conservative estimate.

5.2. Comparison With Previous Studies

Using the same data, *Kaneshima and Helffrich* [1998, 1999, 2003] identified three midmantle scatterers: (1) 25 $\pm$ 4°N, 146 $\pm$ 4°E, 1300 $\pm$ 100 km, (2) 27 $\pm$ 4°N, 149 $\pm$ 4°E, 1550 $\pm$ 100 km, and (3) 28 $\pm$ 4°N, 145 $\pm$ 4°E, 1750 $\pm$ 100 km (as summarized in *Kaneshima and Helffrich* [2010]), and *Kaneshima* [2003] found three more at shallower depths: (4) 19.8 $\pm$ 0.3°N, 145.7 $\pm$ 0.3°E, 710 $\pm$ 30 km, (5) 20.4 $\pm$ 0.4°N, 145.4 $\pm$ 1.0°E, 900 $\pm$ 80 km, and (6) 20.6 $\pm$ 0.4°N, 147.4 $\pm$ 0.5°E, 860 $\pm$ 40 km. These scatterers are labeled as K1 to K6 in the following. Comparing with the scatterers listed in Table 2, K1–K4 probably correspond to the scatterers #21, #27, #32, and #7, respectively, and K5 and K6 collectively to #12 (the largest scatterer in the list). It is unsurprising that except for K3, these previously found scatterers correspond to the scatterers with multiple centroids.

Castle and Creager [1999] also used Mariana earthquakes (during 1982 and 1993) and seismic networks in the U.S. West Coast and suggested the presence of a north-south trending, steeply dipping discontinuity at 30°N, 145°E, and 1000 km, which does not correspond to any of the scatterers detected in this study. Their migration

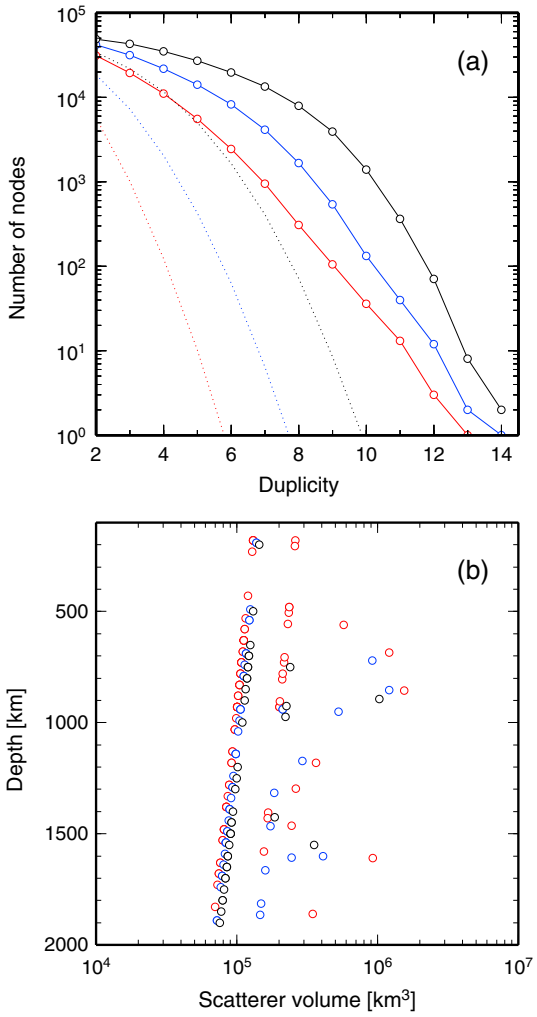

Figure 12. Sensitivity to the critical travel time variance, shown in terms of (a) duplicity and (b) the depth-volume covariation of individual scatterers. The cases of 1, 2, and 4 s^2 are shown in red, blue, and black, respectively. As in Figure 8a, the number of nodes that satisfy a given duplicity requirement is shown for both original (solid) and randomized (dotted) data sets in Figure 12a. In Figure 12b, depths and volumes for the case of 1 s^2 are reduced, respectively, by 20 km and 10%, and those for 2 s^2 by 10 km and 5%, for display purposes.

images (shown in their Figures 5 and 6) suggest that their inference is affected by isochronal artifacts. Using a cluster of deep Mariana earthquakes in 1995 and the Warramunga array in Australia, *Krüger et al.* [2001] detected two scatterers around (18°N, 145.5°E, 790 km) and (17.5°N, 143°E, 720 km). Using a similar set of events but with a Japanese seismic network, *Niu et al.* [2003] detected a dipping seismic reflector at 24.25°N, 145.75°E, and 1115 km. These scatterers are not detected in this study, probably owing to the differences in the source-receiver path.

Recently, *Weber et al.* [2015] challenged the interpretation of *Kaneshima and Helffrich* [1998]. Using events #7, 10, 12, and two more recent events, they suggested the presence of a reflector near 400 km depth, instead of the scatterer K2 located around 1600 km depth (which probably corresponds to the scatterer #27 in Table 2). Their argument, however, relies on the travel time information of a scattered *SH* phase identified only in event #12. It is unclear whether this *SH* phase corresponds to scattered *P* phases seen in other events, so these contrasting interpretations remain to be resolved. Nonetheless, based on the reproducibility test conducted in this study, it is unlikely that the scatterer #27 is somehow mislocated from the upper mantle.

It has been common to use beamforming when searching for scattered phases [e.g., *Kaneshima and Helffrich*, 1998, 2003; *Rost et al.*, 2008], whereas this study relies exclusively on teleseismic migration. Figure 13 shows some examples of beamforming using the event #8 data, together with cluster centroids determined in this study. As in Figure 5, three stacking schemes are compared, and the superior performance of DBS may be recognized. Even with DBS, however, the interpretation of beamforming is not simple. A few centroids seen in the time slice of 30–34 s belong to the largest scatterer (#12), indicating that a broad energy peak in beamforming may correspond to a large scatterer, but such a plain correspondence is not common. The centroid in the time slice of 40–44 s, for example, is not associated with detectable energy. This particular centroid is supported directly by other data sets but also indirectly by this data set as well through the isochronal volume, as seen in Figure 14. Each of 54 centroids listed in Table 2 can be identified with an energy peak in one or more beamforming results, but these examples also indicate that how to interpret energy peaks in beamforming is not straightforward. The effect of isochronal volume may still be considered in beamforming, but the same scatterer has different combinations of time, slowness, and back azimuth for different data sets, making it cumbersome to assess reproducibility. This difficulty is detrimental because as mentioned in section 3.2, a test for reproducibility among different events is a key to distinguish between different scattering mechanisms.

Teleseismic migration with composite semblance used by *Kaneshima and Helffrich* [2003] may look similar to the approach adopted in this study, though they differ in some details. First, semblance is a measure of coherency based simply on the instantaneous polarity of data as a whole, so as for NRS and PWS, it is not

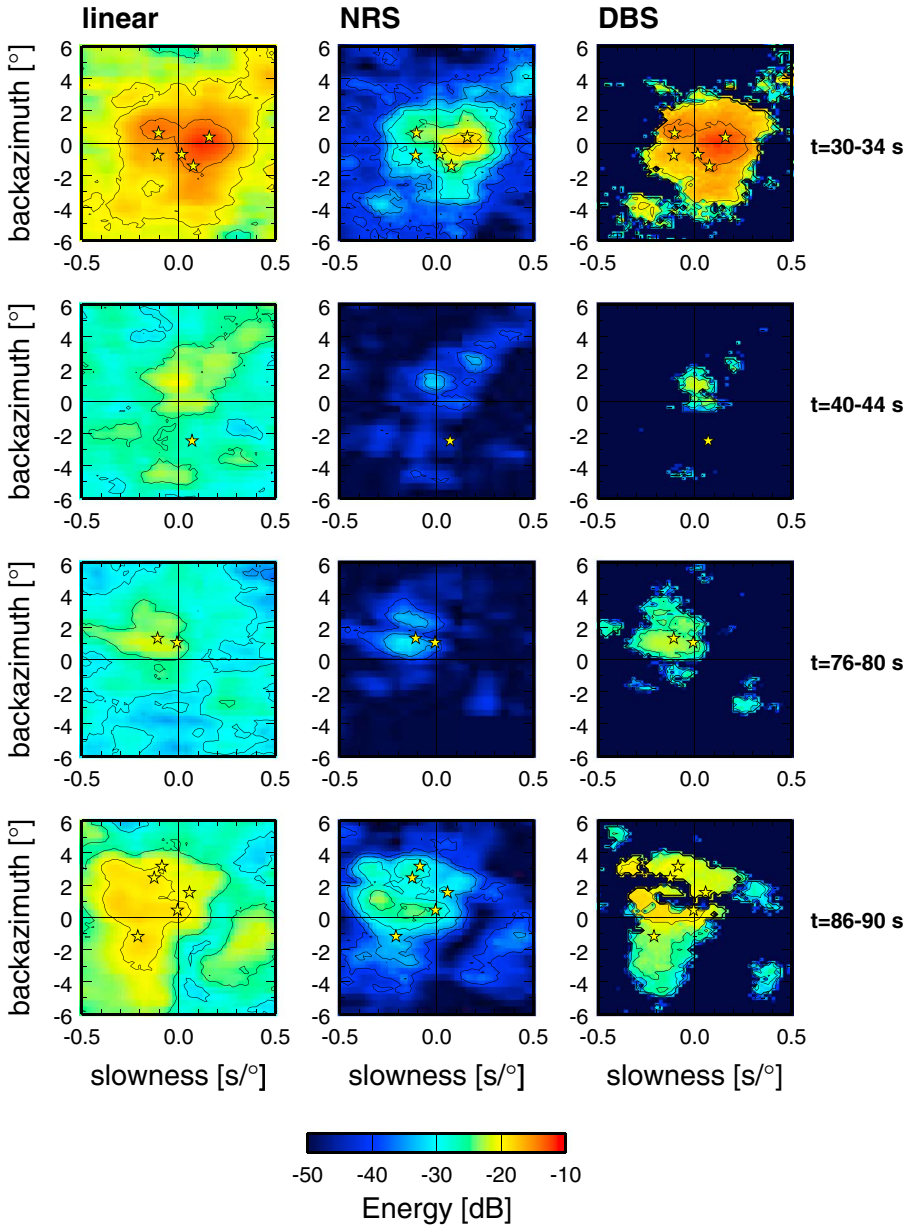

Figure 13. Selected beamforming results with event #8 data. Stacking is done using data within the time window given at the right. As in Figure 5, three stacking schemes are compared. Stars denote cluster centroids determined in this study. Only the centroids whose times fall within the middle 2 s of the window (e.g., 31–33 s for the first row) are shown.

suitable for detecting low S/N signals. When S/N is lower than unity, noise starts to govern the polarity of data, making the detection of any buried signal more difficult (Figure 1) [*Korenaga*, 2013]. Second, an isochronal volume is different for different source-receiver geometries, so simply merging semblance information from multiple data sets would work only when a true scattering point is successfully imaged in the majority of individual migration images. In other words, indirect support through isochronal volume is not possible with composite semblance, but such support is important for low S/N signals. Also, without explicitly quantifying the effect of isochronal volume, it is difficult to constrain the geometries of detected scatterers.

5.3. Remaining Issues

The volume of the largest scatterer detected in this study can be as large as $\sim 10^6$ km^3, which is comparable to the crustal volume of a typical large igneous province [e.g., *Coffin and Eldholm*, 1994]. If such a scatterer represents a blob of subducted oceanic crust, for example, its upwelling could result in nearly complete melting

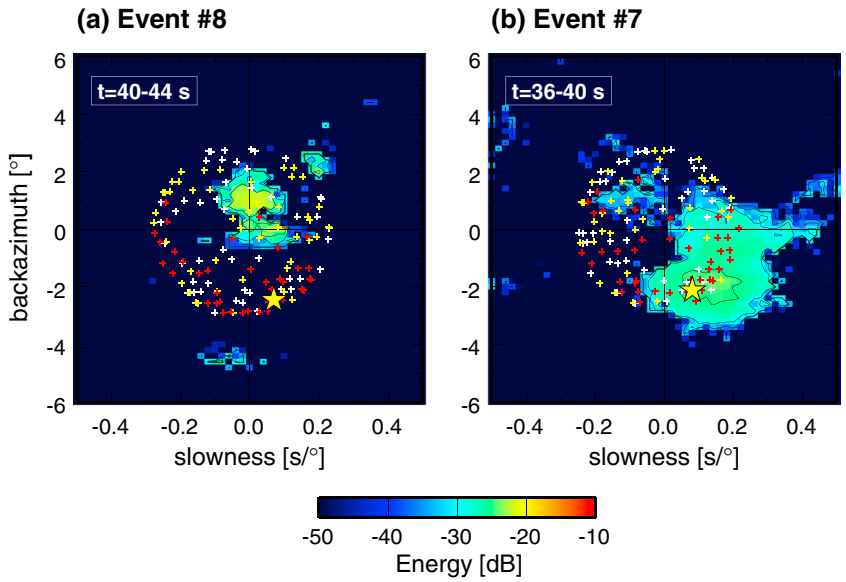

Figure 14. Comparison of DBS-based beamforming for (a) event #8 data and (b) event #7 data. Figure 14a is same as the one shown in Figure 13. The star in Figure 14a and that in Figure 14b correspond to the same centroid (centroid #12 and scatterer #11 in Table 2). Because of the difference in the source parameters, the same centroid shows up in a different time window for event #7. Red, yellow, and white crosses denote the locations of the nodes (in the migration grid) those travel time variance with respect to the centroid is less than 0.25, 0.5, and 1 s^2, respectively. As for centroids, only the nodes whose times fall within the middle 2 s of the window are shown.

(because crust itself is the product of melting). In other words, just one scatterer, if large enough, can be an important source for anomalous magmatism such as continental flood basalts and oceanic plateaus. There are a few major uncertainties, however, that prevent us to speculate further on the significance of small-scale heterogeneities.

First of all, the chemical nature of such heterogeneities needs to be better constrained. *Kaneshima and Helffrich* [1999] and *Niu et al.* [2003], for example, analyzed the waveform of scattered phases to estimate the thickness of a plane-like heterogeneity and its velocity contrast. A similar attempt would be challenging for low S/N signals. Figure 15 shows some examples of stacked waveforms at chosen scatterers, using the data set for event #8, and the comparison of linear stacking, NRS, and DBS is disconcerting. Even for strong scatterers (i.e., with migration energy around −10 dB; Figure 15b), the waveforms by linear stacking and DBS are very different. DBS can retain the amplitude of the most negative peak, but other details of the linear stack are regarded as indistinguishable from residual noise according to the statistical tests implemented in DBS. When analyzing the waveform of scattered phases, therefore, we have to be careful to avoid overinterpretation. We may limit ourselves to more robust characteristics such as the maximum amplitude, and Figure 15 provides yet another example for excellent signal recovery by DBS, as expected by the scaling shown in Figure 1.

In this regard, the following observation may be of some use: Isochronal volumes are usually illuminated only partially by any single event and also differently by different events. Events #7 and #8 compared in Figure 14, for example, both belong to the cluster of Mariana earthquakes on 23 August 1995. Their source mechanisms are thus probably similar, but slightly different source depths result in such a difference. By investigating carefully the relation between partial illumination and source characteristics such as location and mechanism, therefore, we may be able to extract more constraints on the nature of heterogeneities. *S*-to-*P* scattering due to a spherical heterogeneity, for example, is expected to exhibit a $\cos\theta$ variation due to density perturbation and a $\sin 2\theta \sin\phi$ variation due to shear modulus perturbation, where θ and ϕ denote the angles between the direction of particle motion of the incident *S* wave and that of the scattered *P* wave [*Wu and Aki*, 1985]. Synthetic tests using the numerical simulation of wave propagation through heterogeneities would also be valuable.

Second, there is still some uncertainty in the number of scatterers and their sizes, even if we limit ourselves to the approach taken in this study (Figure 12). Table 2 is merely a conservative estimate. There are at least two directions to pursue to address this issue. We can continue to look at *S*-to-*P* scattering in other relevant

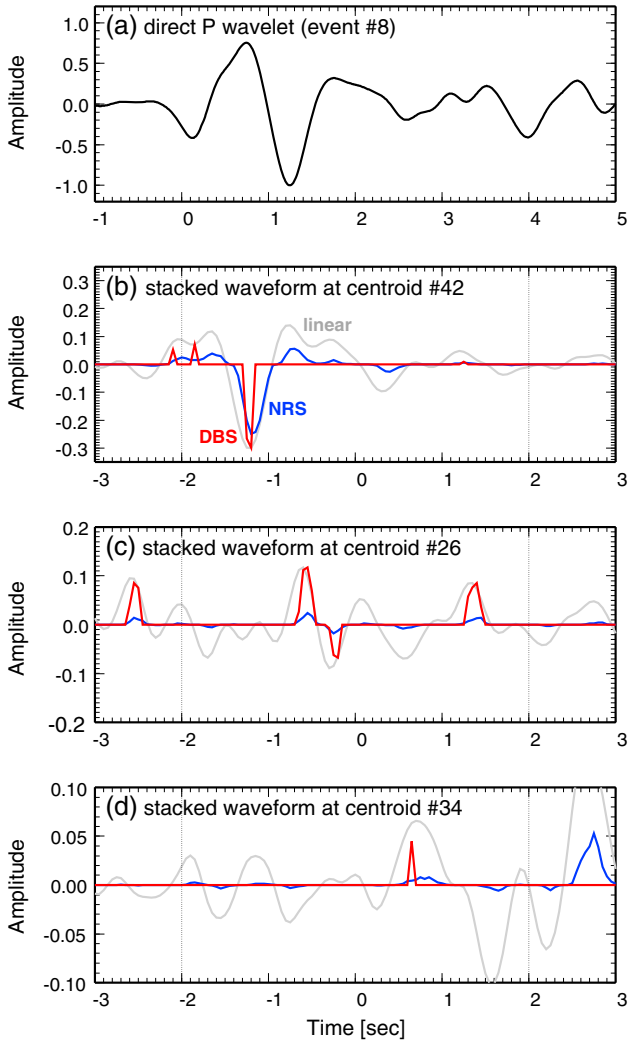

Figure 15. Examples of stacked waveform for event #8 data. (a) Direct *P* wavelet, average of all 79 stations. Stacked waveforms at centroids (b) #42, (c) #26, and (d) #34, which are representative of, respectively, strong, intermediate, and weak scatterers. Amplitudes are normalized to the maximum amplitude of the direct *P* arrival. Gray, blue, and red curves denote linear stacking, NRS, and DBS, respectively. Vertical dotted lines denote the range of the stacking window.

data sets by the new method. For the same old events used here, data at seismic networks in California are yet to be utilized, and there are also quite a few more recent earthquakes in the Mariana region, for which the USArray data exist. We can also study other types of scattering. *Bentham and Rost* [2014], for example, detected several scatterers in the Mariana region by analyzing *PP* precursors in the data recorded at a seismic array in Alaska. Though most of their scatterers are not supported by multiple data sets, it would be interesting to see how such low reproducibility may be improved by signal detection with DBS and duplicity calculation with isochronal volume.

At the same time, regional comparison can help to understand how unique the study region might be, and to this end, we need to study other regions with the new method. Even with the same Mariana earthquakes, other parts of the Mariana region can be studied by using different seismic arrays [*Krüger et al.*, 2001; *Niu et al.*, 2003], and there are a number of source-receiver combinations around the globe that allow us to probe contrasting tectonic regions.

6. Conclusion

The new array-based method, which is built on teleseismic migration with DBS, is applied to the classic data sets of the Mariana-UW path, and the results indicate the existence of at least 39 scatterers in the mantle near

the Mariana subduction zone, with the estimated volume of a scatterer ranging up to $\sim 10^6$ km^3. The reproducibility test among multiple events with different source depths ensures that these scatterers are indeed illuminated by the assumed *S*-to-*P* scattering.

The notable aspects of the new method include (1) the detection of faint scattered signals by DBS, (2) the assurance of high signal reproducibility by calculating duplicity for both original and randomized data sets, and (3) the minimization of isochronal artifacts by cluster analysis. The use of DBS is essential not only for the first point but also for the remaining two. It is the stringent statistical test built in DBS that provides us a set of sharply defined significant nodes. The subsequent duplicity calculation and cluster analysis would not be worthwhile without such statistical screening. Also, agglomerative hierarchical clustering is a numerically intensive algorithm with $O(n^3)$, where n is the number of nodes, so it would be too time consuming unless n is minimized by DBS and reproducibility.

Acknowledgments

This work was sponsored by the U.S. National Science Foundation under grant EAR-1416842. Figures were generated with GMT [*Wessel and Smith*, 1995]. Data used in this study were obtained from the Pacific Northwest Regional Seismic Network through the IRIS Data Management Center. Data of key migration results are presented in Table 2. 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 and by a visiting professorship at the Earthquake Research Institute of University of Tokyo. The author thanks Satoshi Kaneshima and Hitoshi Kawakatsu for a number of valuable suggestions made during the course of this project. Reviews by two anonymous reviewers were very helpful to improve the accuracy and clarity of the manuscript.

References

Allègre, C. J., and D. L. Turcotte (1986), Implications of a 2-component marble-cake mantle, *Nature*, *323*, 123–127.

Bentham, H. L. M., and S. Rost (2014), Scattering beneath Western Pacific subduction zones: Evidence for oceanic crust in the mid-mantle, *Geophys. J. Int.*, *197*, 1627–1641.

Castle, J. C., and K. C. Creager (1999), A steeply dipping discontinuity in the lower mantle beneath Izu-Bonin, *J. Geophys. Res.*, *104*, 7279–7292.

Coffin, M. F., and O. Eldholm (1994), Large igneous provinces: Crustal structure, dimensions, and external consequences, *Rev. Geophys.*, *32*, 1–36.

Cordery, M. J., G. F. Davies, and I. H. Campbell (1997), Genesis of flood basalts from eclogite-bearing mantle plumes, *J. Geophys. Res.*, *102*, 20,179–20,197.

Davies, G. F. (2002), Stirring geochemistry in mantle convection models with stiff plates and slabs, *Geochim. Cosmochim. Acta*, *66*, 3125–3142.

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

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.*, 2825–2852.

Ferrachat, S., and Y. Ricard (1998), Regular vs. chaotic mantle mixing, *Earth Planet. Sci. Lett.*, *155*, 75–86.

Gudmundsson, O., J. H. Davies, and R. W. Clayton (1990), Stochastic analysis of global traveltime data: Mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, *102*, 25–43.

Hauri, E. H. (1996), Major-element variability in the Hawaiian mantle plume, *Nature*, *382*, 415–419.

Hedlin, M. A. H., and P. M. Shearer (2000), An analysis of large-scale variations in small-scale mantle heterogeneity using Global Seismographic Network recordings of precursors to *PKP*, *J. Geophys. Res.*, *105*, 13,655–13,673.

Hedlin, M. A. H., P. M. Shearer, and P. S. Earle (1997), Seismic evidence for small-scale heterogeneity throughout the Earth's mantle, *Nature*, *387*, 145–150.

Hoffman, N. R. A., and D. P. McKenzie (1985), The destruction of geochemical heterogeneities by differential fluid motions during mantle convection, *Geophys. J. R. Astron. Soc.*, *82*, 163–206.

Hofmann, A. W. (1997), Mantle geochemistry: The message from oceanic volcanism, *Nature*, *385*, 219–229.

Ito, G., and J. J. Mahoney (2005), Flow and melting of a heterogeneous mantle: 2. Implications for a chemically nonlayered mantle, *Earth Planet. Sci. Lett.*, *230*, 47–63.

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

Kaneshima, S. (2003), Small-scale heterogeneity at the top of the lower mantle around the Mariana slab, *Earth Planet. Sci. Lett.*, *209*, 85–101.

Kaneshima, S., and G. Helffrich (1998), Detection of lower mantle scatterers northeast of the Mariana subduction zone using short-period array data, *J. Geophys. Res.*, *103*, 4825–4838.

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

Kaneshima, S., and G. Helffrich (2003), Subparallel dipping heterogeneities in the mid-lower mantle, *J. Geophys. Res.*, *108*(B5), doi:10.1029/2001JB001596.

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

Karato, S. (1997), On the separation of crustal component from subducted oceanic lithosphere near the 660 km discontinuity, *Phys. Earth Planet. Inter.*, *99*, 103–111.

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

Kennett, B. L. N., S. Widiyantoro, and R. D. van der Hilst (1998), Joint seismic tomography for bulk-sound and shear wavespeed in the Earth's mantle, *J. Geophys. Res.*, *103*, 12,469–12,593.

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

Korenaga, J. (2004), Mantle mixing and continental breakup magmatism, *Earth Planet. Sci. Lett.*, *218*, 463–473.

Korenaga, J. (2005), Why did not the Ontong Java Plateau form subaerially?, *Earth Planet. Sci. Lett.*, *234*, 385–399.

Korenaga, J. (2006), Archean geodynamics and the thermal evolution of Earth, in *Archean Geodynamics and Environments*, edited by K. Benn, J.-C. Mareschal, and K. Condie, pp. 7–32, AGU, Washington, D. C.

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

Korenaga, J. (2014), Teleseismic migration with dual bootstrap stack, *Geophys. J. Int.*, *196*, 1706–1723.

Korenaga, J., and S. Karato (2008), A new analysis of experimental data on olivine rheology, *J. Geophys. Res.*, *113*, B02403, doi:10.1029/2007JB005100.

Korenaga, J., and P. B. Kelemen (2000), Major element heterogeneity of the mantle source in the North Atlantic igneous province, *Earth Planet. Sci. Lett.*, *184*, 251–268.

Krüger, F., M. Baumann, F. Scherbaum, and M. Weber (2001), Mid mantle scatterers near the Mariana slab detected with a double array method, *Geophys. Res. Lett.*, *28*, 667–670.
Manga, M. (1996), Mixing of heterogeneities in the mantle: Effect of viscosity differences, *Geophys. Res. Lett.*, *23*, 403–406.
Margerin, L., and G. Nolet (2003), Multiple scattering of high-frequency seismic waves in the deep Earth: *PKP* precursor analysis and inversion for mantle granularity, *J. Geophys. Res.*, *108*(B11), 2514, doi:10.1029/2003JB002455.
Muirhead, K. J. (1968), Eliminating false alarms when detecting seismic events automatically, *Nature*, *217*, 533–534.
Niu, F., H. Kawakatsu, and Y. Fukao (2003), Seismic evidence for a chemical heterogeneity in the midmantle: A strong and slightly dipping seismic reflector beneath the Mariana subduction zone, *J. Geophys. Res.*, *108*(B9), 2419, doi:10.1029/2002JB002384.
Nolet, G., and T. J. Moser (1993), Teleseismic delay times in a 3-D Earth and a new look at the *S* direcrepancy, *Geophys. J. Int.*, *114*, 185–195.
Olson, P., D. A. Yuen, and D. Balsigar (1984), Convective mixing and the fine structure of mantle heterogeneity, *Earth Planet. Sci. Lett.*, *36*, 291–304.
Rost, S., and C. Thomas (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.
Rost, S., E. J. Garnero, and Q. Williams (2008), Seismic array detection of subducted oceanic crust in the lower mantle, *J. Geophys. Res.*, *113*, B06303, doi:10.1029/2007JB005263.
Schimmel, M., and H. Paulssen (1997), Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.*, *130*, 487–505.
Shearer, P. M. (2015), Deep Earth structure: Seismic scattering in the deep Earth, in *Treatise on Geophysics*, vol. 1, 2nd ed., edited by G. Schubert, pp. 759–787, Elsevier, Oxford, U. K.
Simmons, N. A., S. C. Myers, G. Johannesson, and E. Matzel (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.
Sleep, N. H. (1984), Tapping of magmas from ubiquitous mantle heterogeneities: An alternative to mantle plumes?, *J. Geophys. Res.*, *89*, 10,029–10,041.
Smith, W. H. F., and D. T. Sandwell (1997), Global sea floor topography from satellite altimetry and ship depth soundings, *Science*, *277*, 1956–1962.
Sobolev, A. V., et al. (2007), The amount of recycled crust in sources of mantle-derived melts, *Science*, *316*, 412–417.
Takahashi, E., K. Nakajima, and T. L. Wright (1998), Origin of the Columbia River basalts: Melting model of a heterogeneous plume head, *Earth Planet. Sci. Lett.*, *162*, 63–80.
Trampert, J., F. Deschamps, J. Resovsky, and D. Yuen (2004), Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science*, *306*, 853–856.
van Keken, P. E., and C. J. Ballentine (1998), Whole-mantle versus layered mantle convection and the role of a high-viscosity lower mantle in terrestrial volatile evolution, *Earth Planet. Sci. Lett.*, *156*, 19–32.
Weber, M., C. Wicks, Y. Le Stunff, B. Romanowicz, and F. Krüger (2015), Seismic evidence for a steeply dipping reflector—Stagnant slab in the mantle transition zone, *Geophys. J. Int.*, *200*, 1235–1251.
Wessel, P., and W. H. F. Smith (1995), New version of the generic mapping tools released, *Eos Trans. AGU*, *76*, 329.
Wu, R., and K. Aki (1985), Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, *50*, 582–595.
Xie, S., and P. J. Tackley (2004), Evolution of U-Pb and Sm-Nd systems in numerical models of mantle convection and plate tectonics, *J. Geophys. Res.*, *109*, B11204, doi:10.1029/2004JB003176.
Zindler, A., and S. Hart (1986), Chemical geodynamics, *Annu. Rev. Earth Planet. Sci.*, *14*, 493–571.