Earth, Planets and Space
a SpringerOpen Journal

FULL PAPER   **Open Access**

Application of cluster analysis based on waveform cross-correlation coefficients to data recorded by ocean-bottom seismometers: results from off the Kii Peninsula

Takeshi Akuhara[*] and Kimihiro Mochizuki

Abstract

Waveform cross-correlation coefficients (CCs) have often been used to investigate clustering of earthquakes. Such techniques, however, have rarely been applied to waveform data recorded by ocean-bottom seismometers (OBSs). The data recorded by some OBSs are strongly influenced by site effects such as the existence of unconsolidated sediment layers. The resulting waveforms tend to have a monotone frequency so that the calculated CCs stay artificially high, and false delay times are obtained at certain multiples of the period. This effect also varies from place to place. To overcome such problems, CC measurements were first performed using multiple time windows with different lengths in order to reject unstable results. Station-specific CC thresholds were then objectively determined based on the CC distributions. This method was applied to seismic measurements obtained by OBSs deployed off the Kii Peninsula from 2003 to 2007 in order to identify characteristic hypocenter patterns. Two types of earthquake clusters were found. The first occurred within the oceanic mantle, and the component events were linearly aligned along the NNE direction. Such earthquakes are considered to occur on preexisting faults due to dehydration of the serpentinized mantle. The second type of cluster consisted of interplate earthquakes that occurred at the southern tip of the Kii Peninsula. These are thought to be the result of stress caused by local structural heterogeneity of the dense rock body.

Keywords: Cross-correlation coefficients; Ocean-bottom seismometers; Subduction zone; Earthquake cluster

Background

Cross-correlation coefficients (CCs) are used to determine the degree of similarity between seismograms. Because an observed seismogram is a convolution of the source term, path effects, site effects, and instrumental responses, earthquake pairs with high CC values are considered to have close hypocenters and similar focal mechanisms (Baisch et al. 2008). For a variety of reasons, researchers have attempted to link similar earthquakes and group them into clusters. One reason is to detect repeating earthquakes, which occur at the same asperities on the same fault planes (e.g., Geller and Mueller 1980; Igarashi et al. 2003). Their magnitudes are roughly comparable, and some of them are characterized by periodic

repetition. Another reason is to determine the spatial patterns of hypocenters and associate them with seismogenic structures such as faults (e.g., Cattaneo et al. 1999; Ferretti et al. 2005). In this case, it is not always necessary that similar earthquakes should have comparable magnitudes, and differences in location are also permitted to some extent. A third reason for grouping earthquakes is to more precisely identify their relative locations within clusters by measuring accurate differential travel times based on CC delay times (e.g., Lin and Shearer 2007; Shearer et al. 2005). Such studies require only that the waveforms associated with the first pulses be similar. Depending on the desired goal, different CC thresholds are introduced to judge the similarity between waveforms.

In spite of these applications, few researchers have applied such cluster analysis techniques to waveform data recorded by ocean-bottom seismometers (OBSs). The

* Correspondence: akuhara@eri.u-tokyo.ac.jp
Earthquake Research Institute, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

Springer

most critical difference between on-land seismometers and OBSs is the existence of thick unconsolidated sediment layers beneath the seafloor. As a result, monotone frequencies tend to be dominant in the waveforms recorded by some OBSs. Since these waveforms are intrinsically similar to each other, CC values are artificially high. There is also a possibility that false delay times are obtained at certain multiples of the period (the cycle skipping problem). Such site effects vary from place to place because of the different thickness of the sediment layer.

One approach to avoiding unstable CC results due to cycle skipping and other effects would be to directly measure the uncertainty in the time delays. However, there is no easy method of achieving this. As an alternative approach, Schaff and Waldhauser (2005) calculated two different CCs per event pair using different time window lengths and accepted only results that produced consistent delay times for both. Similar methods have been introduced for measuring S-wave splitting (e.g., Savage et al. 2010).

In most previous cluster analysis studies, a fixed CC threshold was used for all stations. It was implicitly assumed that there was no difference in site effects or instrumental responses among stations. This assumption might be appropriate for on-land stations, because they are usually installed on stiff outcrops or inside boreholes in order to minimize site effects.

In the present study, we first calculated CCs among seismograms recorded by OBSs, incorporating a method designed to prevent cycle skipping, and obtained the CC distribution for each OBS. An objective method is then proposed for determining thresholds (P- and S-waves) for each OBS based on the CC distribution. In this method, the threshold can be arbitrarily selected according to the study objectives. Finally, we applied the proposed cluster analysis method to earthquakes off the Kii Peninsula in order to clarify their hypocenter characteristics.

The Kii Peninsula is located at the central part of the Nankai subduction zone in southwest Japan, where the Philippine Sea Plate subducts along the direction N55°W (Miyazaki and Heki 2001). At the plate interface, megathrust earthquakes have occurred in a cycle of about 100 to 150 years (Ando 1975). The Nankai subduction zone is characterized by a highly deformed subducting slab (Shiomi et al. 2008) and has strong heterogeneity in its subsurface structure (e.g., Kodaira et al. 2000, 2006). This complexity can be expected to affect the development of future megathrust earthquakes in addition to rupture propagation during these events. It is therefore important to carry out a detailed investigation of the seismicity in order to obtain additional constraints on the complex tectonics in this region.

From November 2003 to November 2007, large-scale passive seismic monitoring was conducted around the Kii Peninsula using long-term OBSs (Figure 1). Mochizuki et al. (2010) used this data to determine the earthquake distribution and revealed the large-scale heterogeneous nature of the seismicity. In addition, Akuhara et al. (2013) carried out a tomographic analysis of local earthquakes using data obtained by both OBSs and on-land stations. They determined the three-dimensional (3D) P- and S-wave seismic velocity structures and revealed along-strike segmentation of the structure. Such segmentation may correspond to the large-scale heterogeneity of the seismicity. In the present study, we carried out a cluster analysis in order to focus on smaller-scale features of the seismicity in this region.
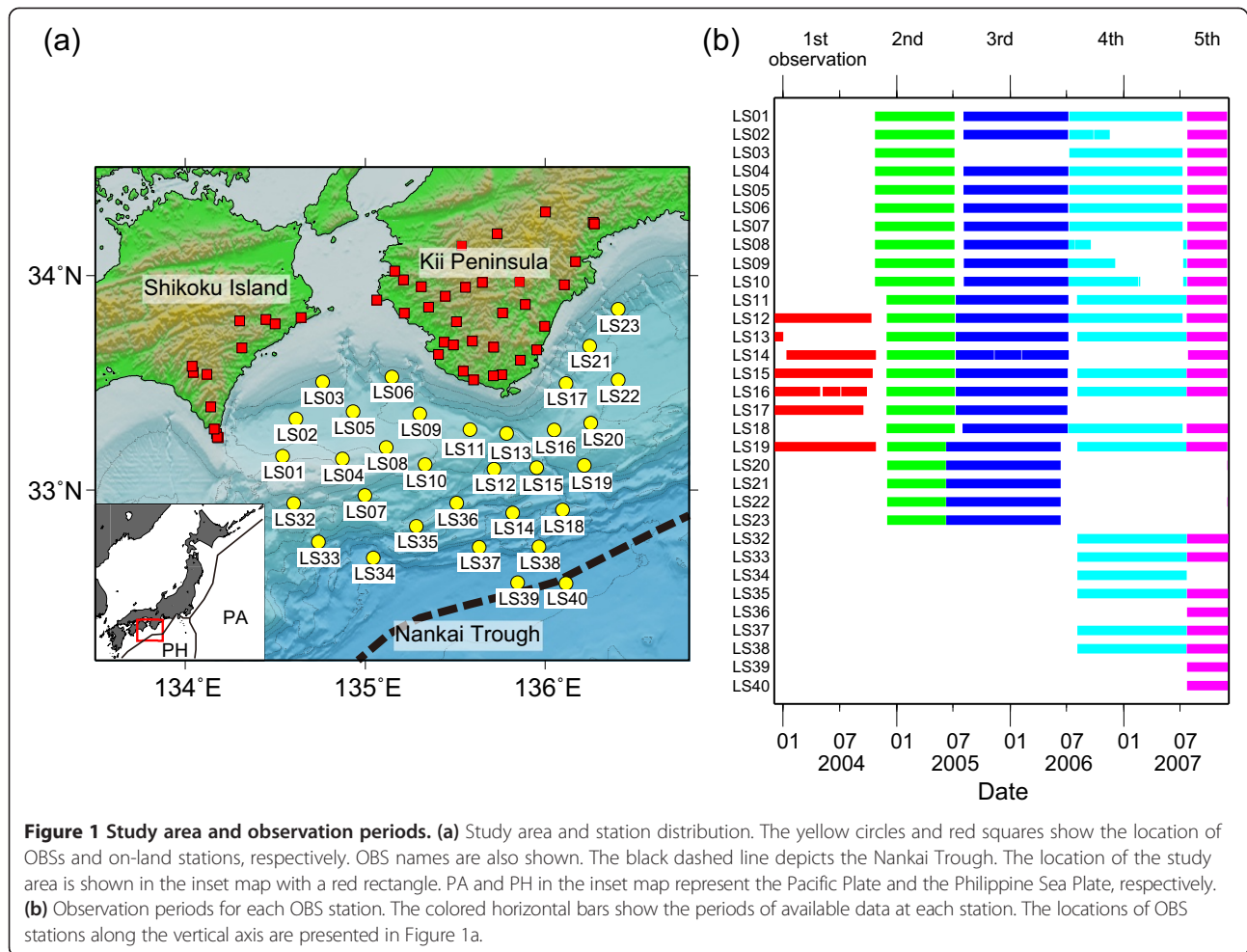
Methods
Data
The data used in the present study were obtained from long-term OBSs deployed off the Kii Peninsula (Figure 1a). The sampling rate was 200 Hz. The natural frequency of the velocity seismometers was 1 Hz. Each OBS was deployed for several months or about a year, after which it was replaced. In this way, continuous observations were carried out over 4 years from November 2003 to November 2007 (Figure 1b). Deployment involved releasing the OBS on the sea surface and allowing it to sink freely to the seafloor. For any given site, the OBS position was subject to a slight variation of several hundred meters during successive observation periods. In the present paper, to specify the station and the observation period, the convention used is the OBS name followed by the observation period number (defined in Figure 1b). For example, OBS LS04 deployed during the fourth observation period is referred to as LS04-4.

We used the hypocenter list compiled by Akuhara et al. (2013), which contains offshore microearthquakes that are not listed in the Japan Meteorological Agency (JMA) catalog. These hypocenters were determined by solving an inverse problem for both the hypocenters and the 3D seismic velocity structure, using a double-difference tomography method (Zhang and Thurber 2003). Arrival times recorded at permanent on-land stations (red squares in Figure 1a) were also included in the process. The list contained 4,533 hypocenters and 19,187 P-wave and 32,167 S-wave manual arrival time picks on the OBS records. Their magnitudes ranged from −0.7 to 5.1, as determined from the maximum amplitude and the distance to the hypocenter (Watanabe 1971), and the estimated completeness magnitude was about 1.5.

We extracted P- and S-wave seismograms based on the manual arrival time picks from continuous OBS data. A 3-s-long time window starting 1 s before the arrival time picks was extracted and bandpass-filtered in the

Figure 1 Study area and observation periods. (a) Study area and station distribution. The yellow circles and red squares show the location of OBSs and on-land stations, respectively. OBS names are also shown. The black dashed line depicts the Nankai Trough. The location of the study area is shown in the inset map with a red rectangle. PA and PH in the inset map represent the Pacific Plate and the Philippine Sea Plate, respectively. **(b)** Observation periods for each OBS station. The colored horizontal bars show the periods of available data at each station. The locations of OBS stations along the vertical axis are presented in Figure 1a.

frequency range 3.0 to 15.0 Hz. Such a wide frequency range was selected in order to avoid cycle skipping and because, on the ocean bottom, seismic noise tends to be low in this frequency band (e.g., Walker 1984; Webb 1998). In addition, the length of the time windows was chosen so as to exclude S-wave onsets from the time windows for P-waves. Hereafter, these 3-s-long time windows are referred to as 'parent' time windows. From these parent time windows, six shorter 'child' time windows were selected, as shown in Table 1. These child time windows had different lengths and were used to perform CC measurements. Although all the OBSs were equipped with three-component velocity sensors, we used only the vertical component in the present study. That is

because it was found that the S-waves of the horizontal components were more subject to cycle skipping than those of the vertical components.

CC calculation using different window lengths

To calculate the CCs, the algorithm reported by Schaff and Waldhauser (2005), which is based on the use of a correlation detector, was employed. A variant of the cross-correlation function, $CC(\tau)$, was calculated by sliding the seismogram of a child time window over a different parent time window, namely

$$CC(\tau) = \sum_{t=\tau-t_{pre}}^{\tau+t_{sig}} y_1(t)y_2(t-\tau) \Big/ \sqrt{\sum_{t=-t_{pre}}^{t_{sig}} y_1^2(t) \sum_{t=-t_{pre}}^{t_{sig}} y_2^2(t)},$$

(1)

where $y_1(t)$ and $y_2(t)$ represent the discrete time series for the parent and child time windows, respectively, and τ denotes the delay time. Both seismograms $y_1(t)$ and $y_2(t)$ are aligned using the manual arrival time pick at

Table 1 Time lengths of six types of child time windows

Window type	1	2	3	4	5	6
Time window length $t_{pre} + t_{sig}$ (s)	2	1.8	1.6	1.4	1.2	1.0
Sample time periods preceding phase onsets t_{pre} (s)	0.5	0.45	0.4	0.35	0.3	0.25

$t = 0$. The time range for the summations in Equation 1 is the same as that for the child time windows, and t_{pre} and t_{sig} represent the sample time periods preceding and succeeding the arrival time ($t = 0$), respectively. From the $CC(\tau)$ values calculated in this way, the maximum CC is given by

$$CC_{max} = \max[CC(\tau)], \tag{2}$$

and the corresponding delay time is written as τ_{max},

$$CC(\tau_{max}) = CC_{max} \tag{3}$$

To avoid cycle skipping, we repeated the above CC calculations with different child time window lengths (Table 1) and extracted only stable results using the following procedure:

- For two given seismograms A and B, CC measurements were performed six times using the parent time window for seismogram A and the six child time windows for seismogram B. The resulting CCs and delay times are referred to as $CC_{max}^{(i)}$ and $\tau_{max}^{(i)}$ respectively ($i = 1, 2, ..., 6$).

- These calculations were then repeated following switching of seismograms A and B, which yielded six additional sets of $CC_{max}^{(i)}$ and $\tau_{max}^{(i)}$ ($i = 7, 8, ..., 12$).
- If the conditions

$$\max\left[\left|s(i)\tau_{max}^{(i)} - s(j)\tau_{max}^{(j)}\right|\right]_{i\neq j} \leq 0.02 \,(\text{ sec}),$$
$$s(i) = \begin{cases} 1 & (i = 1, 2, ..., 6) \\ -1 & (i = 7, 8, ..., 12) \end{cases} \tag{4}$$

were satisfied, the measurements were regarded as stable, and the results obtained using a 2-s-long child time window were used.

Figure 2 shows examples of accepted and rejected measurements using this procedure.

We evaluated the effectiveness of the above method by applying it to closely located source pairs in the hypocenter list of Akuhara et al. (2013), whose hypocenter separation was less than 5 km. Figure 3 shows both accepted and rejected results for three OBSs (LS05-3, LS08-3, and LS09-3) for both P- and S-waves. The accepted results are centered on $\tau_{max} = 0$, which indicates the consistency of the manual arrival time picks. For S-waves, the accepted τ_{max} values are more scattered about $\tau_{max} = 0$ than those for P-waves. This implies that the
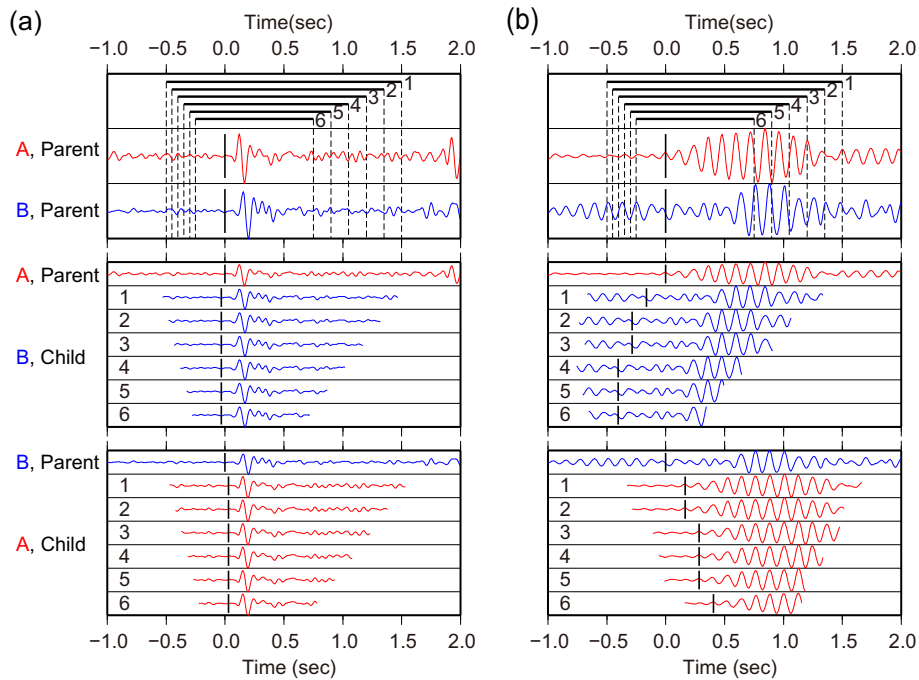

Figure 2 Examples of (a) accepted and (b) rejected results by CC measurements with multiple time windows. The top panels show seismogram pairs for the 'parent' time window, where the black vertical bars on the seismograms represent manual arrival time picks. The thick horizontal bars above the seismograms represent ranges of 'child' time windows. In the lower two panels, the seismograms for six child time windows are slid over a different parent time window. Each seismogram is aligned at $\tau = \tau_{max}$, where $CC(\tau)$ is a maximum. All the seismograms are bandpass-filtered between 3.0 and 15.0 Hz.
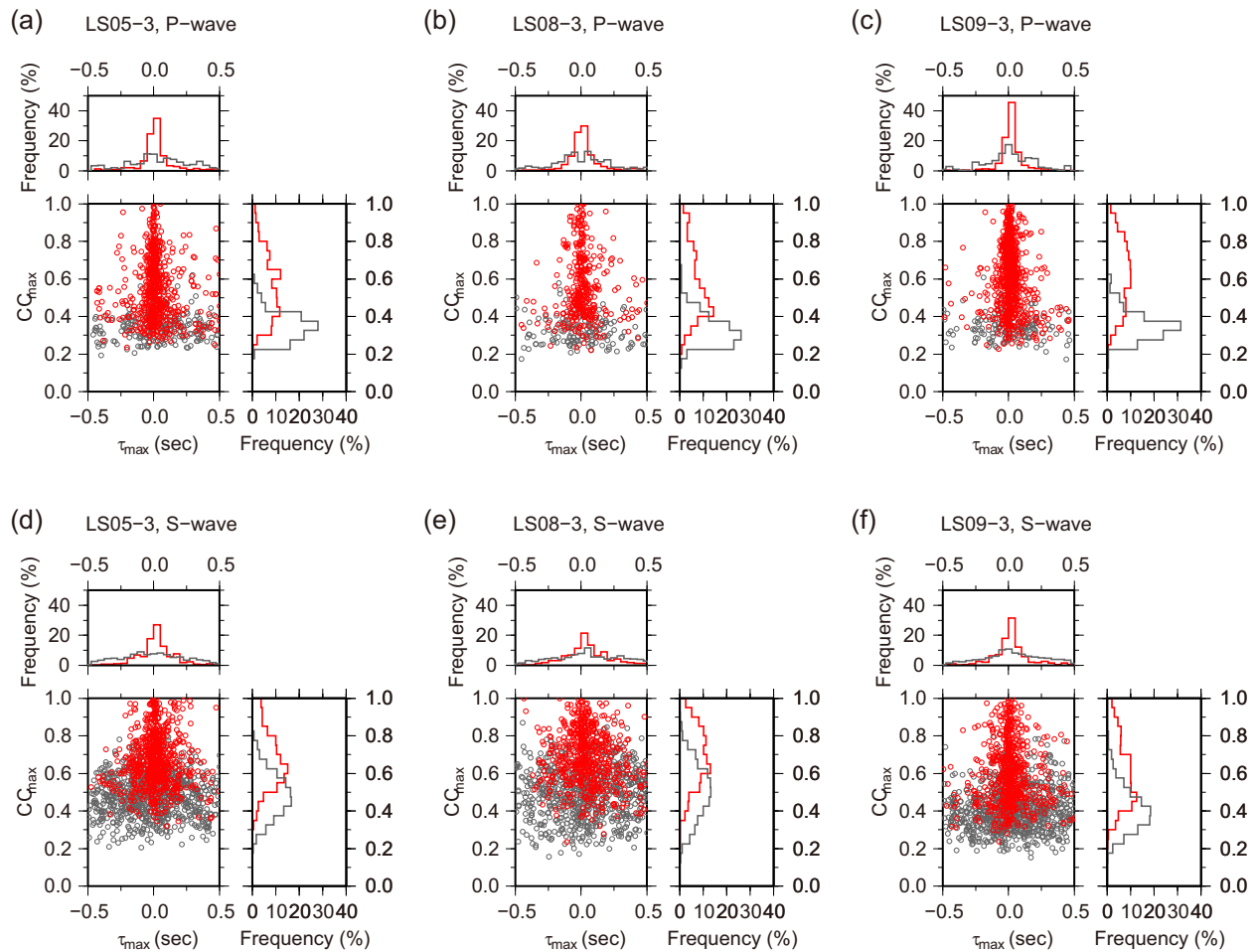
Figure 3 Results of CC measurements (a to f) with multiple time windows at three OBS stations. The red and gray dots indicate accepted and rejected results for pairs of closely located sources (within 5 km), respectively. The histograms for CC_{max} and τ_{max} are also shown at the right and the top of these plots using the same colors as the corresponding data plots.

manual pick values are less accurate for S-waves, probably because of contamination by the P-wave coda at the S-wave onset.

In Figure 3, it can also be seen that the rejected results tend to have low CC_{max} values and more scattered about $\tau_{max} = 0$ than the accepted ones. This is considered to be because these phase pairs are so dissimilar that the maximum correlation does not always occur at the phase onsets. In contrast, the rejected S-wave results for LS05-3 and LS08-3 tend to have high CC_{max} values (>0.6). This is believed to be due to the dominance of a specific range of frequencies in the spectra of S-waves due to the presence of unconsolidated sediments beneath these OBSs, which gives rise to cycle skipping. For LS09-3, the S-wave results appear to be free from such tendencies, which suggests that there are different site effects at this station than at LS05-3 and LS08-3. Note that the OBS positions were close to each other, and the observation

periods were identical, so almost the same set of earthquakes was used at each station in the above analysis.

For each OBS, we applied the algorithm to earthquake pairs with hypocenter separations of less than 75 km. The number of input phase pairs was about 8.5 million, from which about 2.4 million results were accepted using the above process. In addition, we also applied the method to about 17.0 million phase pairs recorded by OBS pairs at the same sites, but for different observation periods (LS01-1 and LS01-2, for example) to acquire about 4.4 million accepted results. In the following analysis, only the accepted results are used.

CC thresholds
The next task was to determine the threshold for CC_{max}, which should vary from one OBS to another due to site effects. CC_{max} represents the degree of similarity between two phases with respect to the source terms (focal
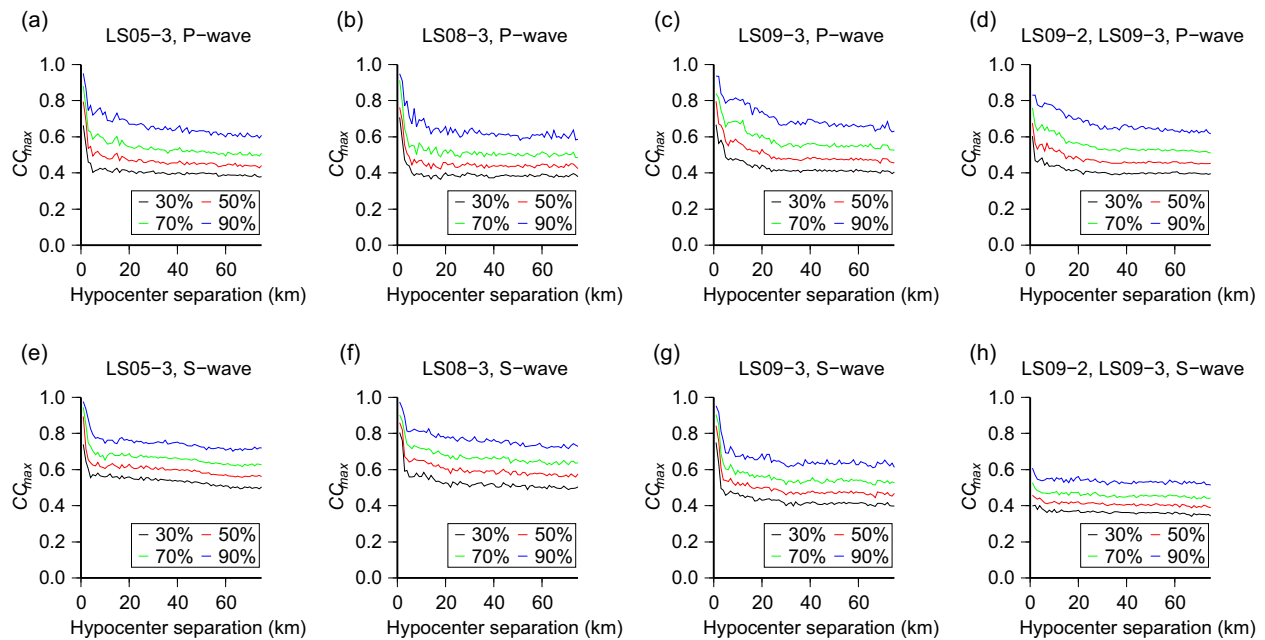
Figure 4 Relationship between CC$_{max}$ and hypocenter separation (a to h). The computed CC$_{max}$ values are divided into hypocenter separation bins with widths of 1 km. The cumulative frequency is then calculated for each bin. The cumulative frequency results for 30%, 50%, 70%, and 90% are plotted.
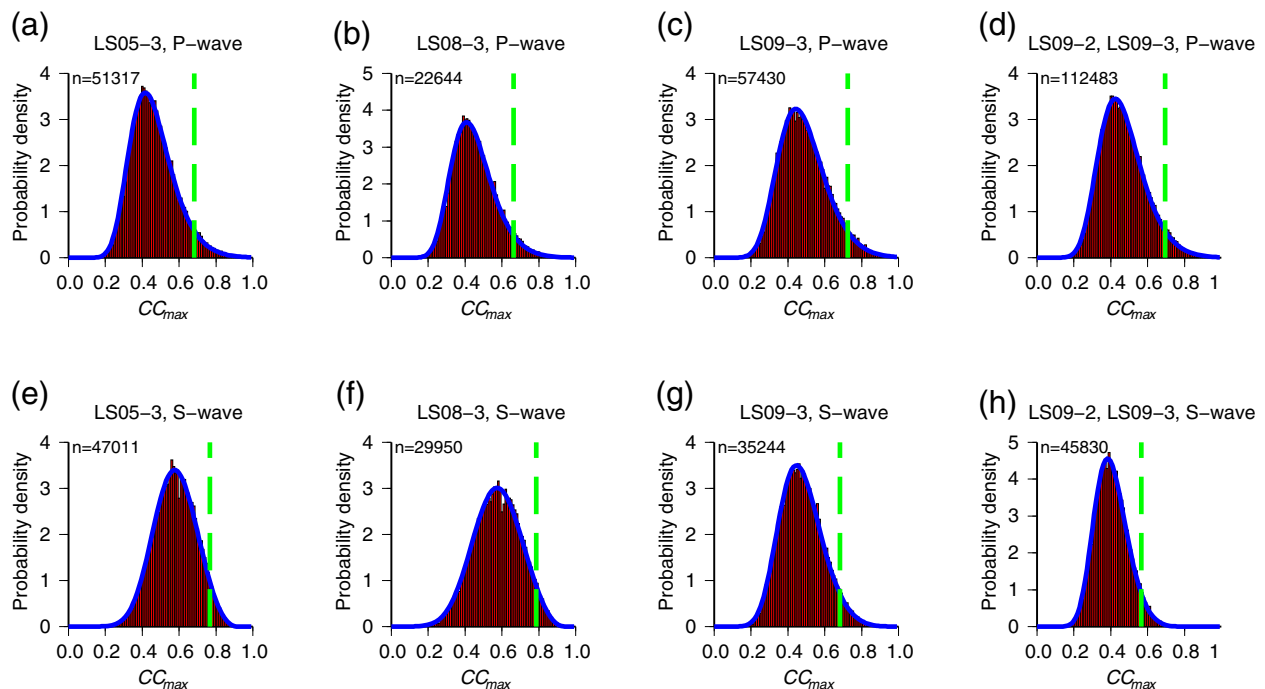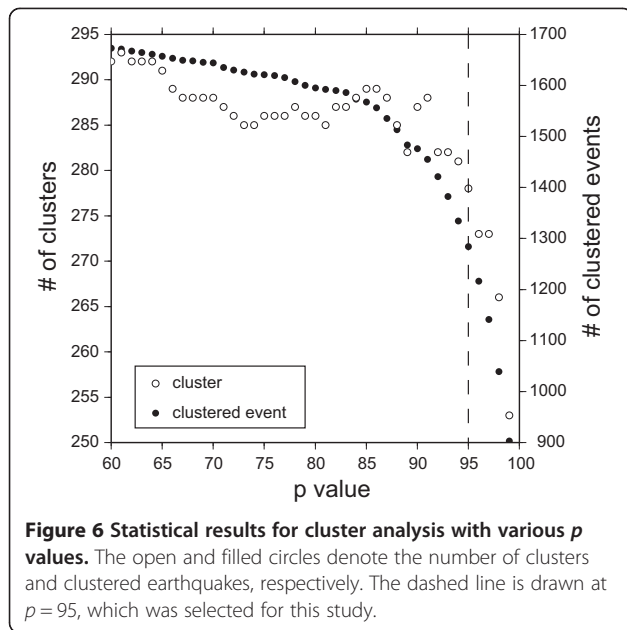

Figure 5 Distributions of CC$_{max}$ (a to h) for earthquake pairs with a large hypocenter separation. The red bars represent the probability density for CC$_{max}$ calculated for widely separated (>30 km) phase pairs. They are normalized so that their total area is unity. The blue curves are GEV distributions fitted to the probability density distributions. The green dashed lines show the 95th percentile of the distribution, which is used to determine the thresholds. The number of phase pairs used is shown at the upper-left corner of each panel.

Figure 6 Statistical results for cluster analysis with various *p* values. The open and filled circles denote the number of clusters and clustered earthquakes, respectively. The dashed line is drawn at $p = 95$, which was selected for this study.

mechanisms and radiation patterns) and the propagation paths. Site effects also influence CC_{max}, acting as filters in the frequency domain.

We investigated the effect of the propagation path by examining the relationship between CC_{max} and hypocenter separations between two events. All phase pairs were first divided into bins based on their hypocenter separations, as determined from the hypocenter list. The cumulative frequency of CC_{max} was then calculated for each bin, which had a width of 1 km. Here, the cumulative frequency is defined as the percentage of CC_{max} values that are lower than a given value. Figure 4 shows typical results for the cumulative frequency. It can be seen that it rapidly decreases as the hypocenter separation becomes more than a few kilometers and then remains almost constant. From these results, we assumed that the statistical distributions of CC_{max} at each OBS

become stable (independent of the hypocenter separation) if the sources are separated by more than 30 km. By using such stable CC_{max} distributions, the bias due to the inhomogeneous source distribution can be avoided. For example, for an earthquake swarm occurring beneath an OBS, the CC_{max} distribution will be biased towards higher values if closely located pairs are included.

The red histograms in Figure 5 show the CC_{max} distributions (observed probability density) for source pairs separated by more than 30 km. It is considered that the differences between the P- and S-wave distributions for the same OBS reflect differences in the radiated energy spectrum for the sources, whereas differences among OBSs are most likely due to site effects. Therefore, we used these distributions to determine the CC_{max} threshold for each OBS.

The distributions shown in Figure 5 are non-Gaussian in that they have asymmetric lobes. However, we found that they were well described using a generalized extreme value (GEV) distribution (Jenkinson 1955), which is a statistical distribution of the maximum values of random samples from an arbitrary probability density function. The blue curves in Figure 5 represent GEV fits to the CC_{max} distributions using the L-moment method (Hosking 1990). The CC_{max} threshold can then be determined as the p-th percentile of the fitted GEV distribution, which represents the value of CC_{max} for which the cumulative frequency becomes p%. The p value can be chosen in accordance with the study objectives.

The different observation periods (Figure 1b) must also be taken into consideration. Site effects are thought to be different among observation periods, even if the locations are almost the same, because different conditions exist for the mechanical coupling between the OBS and the seafloor each time an OBS is deployed. To account for this, we determined the thresholds for all combinations of observation periods in the manner described above (Figures 4d,h and 5d,h).
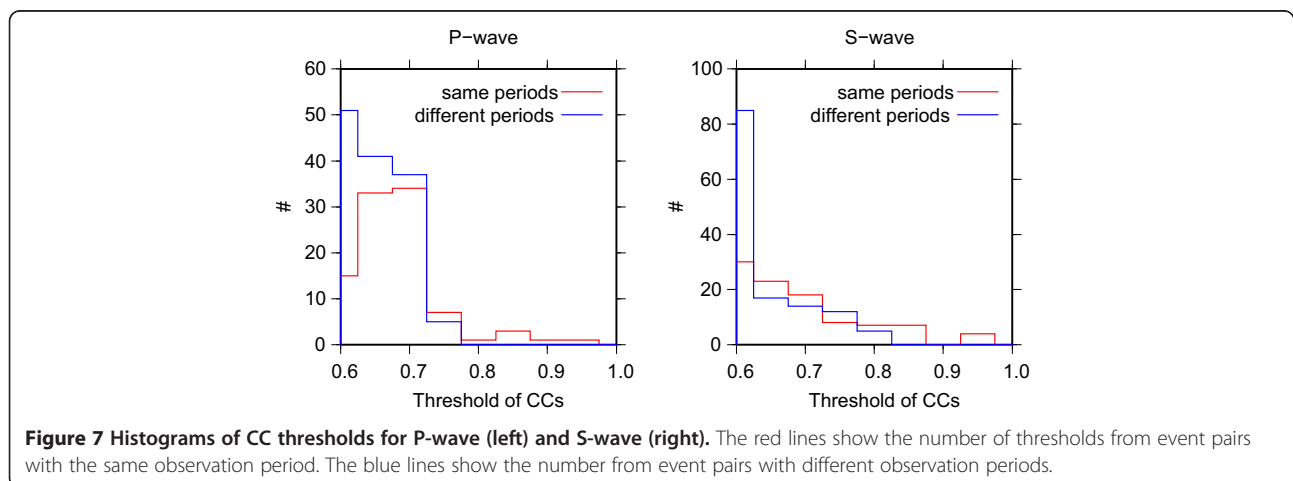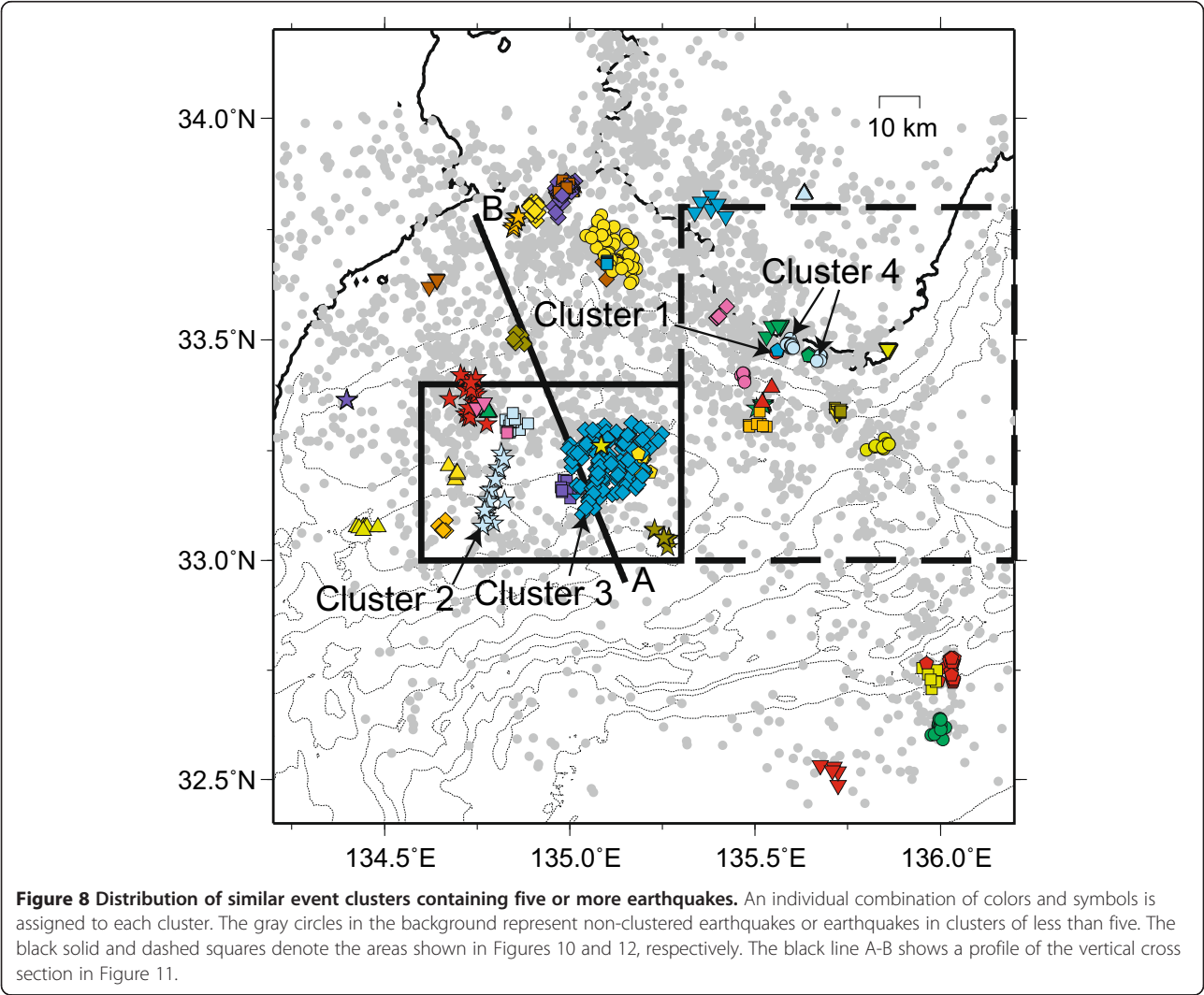

Figure 7 Histograms of CC thresholds for P-wave (left) and S-wave (right). The red lines show the number of thresholds from event pairs with the same observation period. The blue lines show the number from event pairs with different observation periods.

Table 2 Conditions for double-difference hypocenter relocation

	Phase	Station type	Number	Root mean square of travel time residuals after relocation (s)
Events	-	-	1,284	-
Absolute travel time	P	On-land stations and OBSs	75,769	0.078
	S		109,616	0.103
Differential travel time from hypocenter list	P	On-land stations and OBSs	75,536	0.052
	S		123,236	0.080
Differential travel time from CC measurements	P	OBSs	5,036	0.017
	S		12,509	0.020

Cluster analysis of earthquakes

We next performed a cluster analysis based on the CCs in order to identify characteristic patterns in the hypocenters around the Kii Peninsula. We defined clusters of similar earthquakes based on hypocenter separations and waveform similarities. First, any two events satisfying the following conditions were linked together: (1) the hypocenter separation was less than 5 km, and (2) CC_{max} was greater than or equal to the threshold value for three or more phase pairs, including at least one S-wave pair. The linked earthquakes were then grouped into clusters using the following simple rule: if earthquakes A and B are linked and if B and C are also linked, then A, B, and C are included in the same cluster irrespective of

Figure 8 Distribution of similar event clusters containing five or more earthquakes. An individual combination of colors and symbols is assigned to each cluster. The gray circles in the background represent non-clustered earthquakes or earthquakes in clusters of less than five. The black solid and dashed squares denote the areas shown in Figures 10 and 12, respectively. The black line A-B shows a profile of the vertical cross section in Figure 11.

the waveform similarity between A and C. The CC_{max} thresholds were determined for a given p value in the manner described above. However, if the calculated threshold was lower than 0.6, it was changed to 0.6 because we noticed that waveforms began to appear dissimilar below this value.

Figure 6 shows the p value dependence of the number of clusters and component earthquakes. Both are seen to decrease with increasing p. Note that the lower limit for the number of clusters is determined by the maximum hypocenter separation (5 km) and the minimum CC_{max} threshold value (0.6). Otherwise, one large cluster containing all events would be produced at an extremely low p value. When selecting the appropriate value of p, there is a trade-off between the number of clusters and clustered events, and the CC_{max} threshold. In the present study, a p value of 95 was chosen, since this is where the two numbers begin to drop rapidly as shown in Figure 6. Figure 7 shows the histograms of the obtained thresholds.

We then relocated the hypocenters of the linked earthquakes using the cross-correlation results in order to improve the accuracy of their relative positions. This was performed by means of the double-difference algorithm (Waldhauser and Ellsworth 2000) using the tomoDD program (Zhang and Thurber 2003), which supports 3D velocity structures. The 3D P- and S-wave velocity model used was obtained from a previous tomographic analysis (Akuhara et al. 2013). As input data for tomoDD, we used the differential travel time calculated from τ_{max}, in addition to the absolute and differential travel times from the hypocenter list. Additional information concerning the relocation process is listed in Table 2. In spite of the use of CC delay times, there was no remarkable change in appearance following relocation. This may be due to the relatively small number of differential travel times from the CC measurements. The root mean square values of the differential travel time residuals from the CC measurements were 0.017 and 0.020 s for P- and S-waves, respectively. If it is assumed that the P- and S-wave velocities are 7.0 and 3.8 km/s, respectively, which are actually somewhat larger than typical values for the oceanic crust, the relative location error can be estimated to be about 100 m.

Results and discussion

As a consequence of the cluster analysis, we obtained 278 clusters and 1,284 clustered earthquakes from a total of 4,523 earthquakes. Figure 8 shows the 49 clusters that contain five or more earthquakes. Figure 9a shows seismograms for the earthquakes in cluster 1 in Figure 8; it can be seen that they are extremely similar.

To investigate the characteristics of each cluster, we determined the focal mechanisms from the P-wave polarities, which were observed at OBSs, on-land stations,
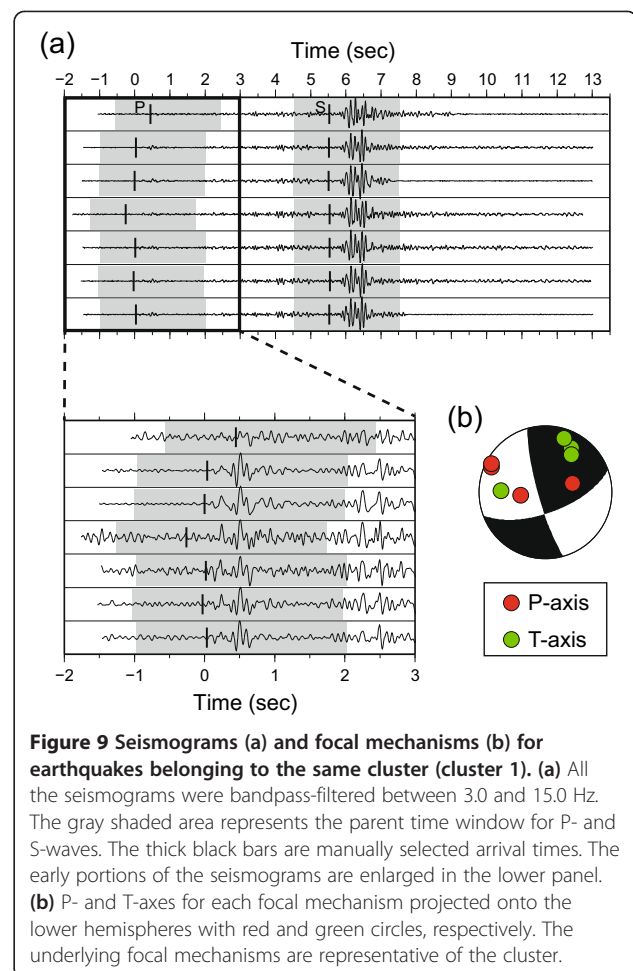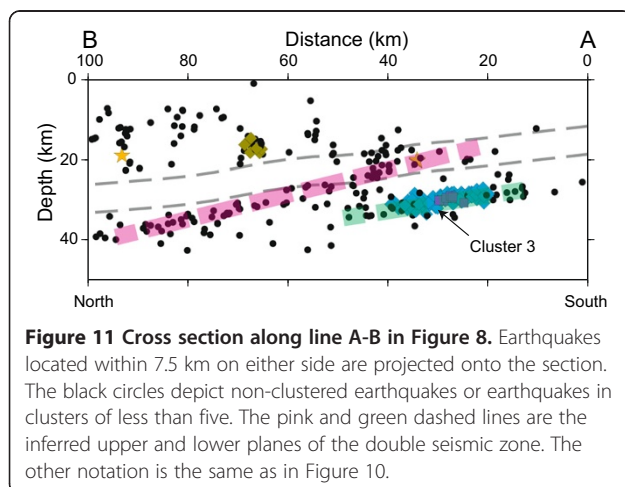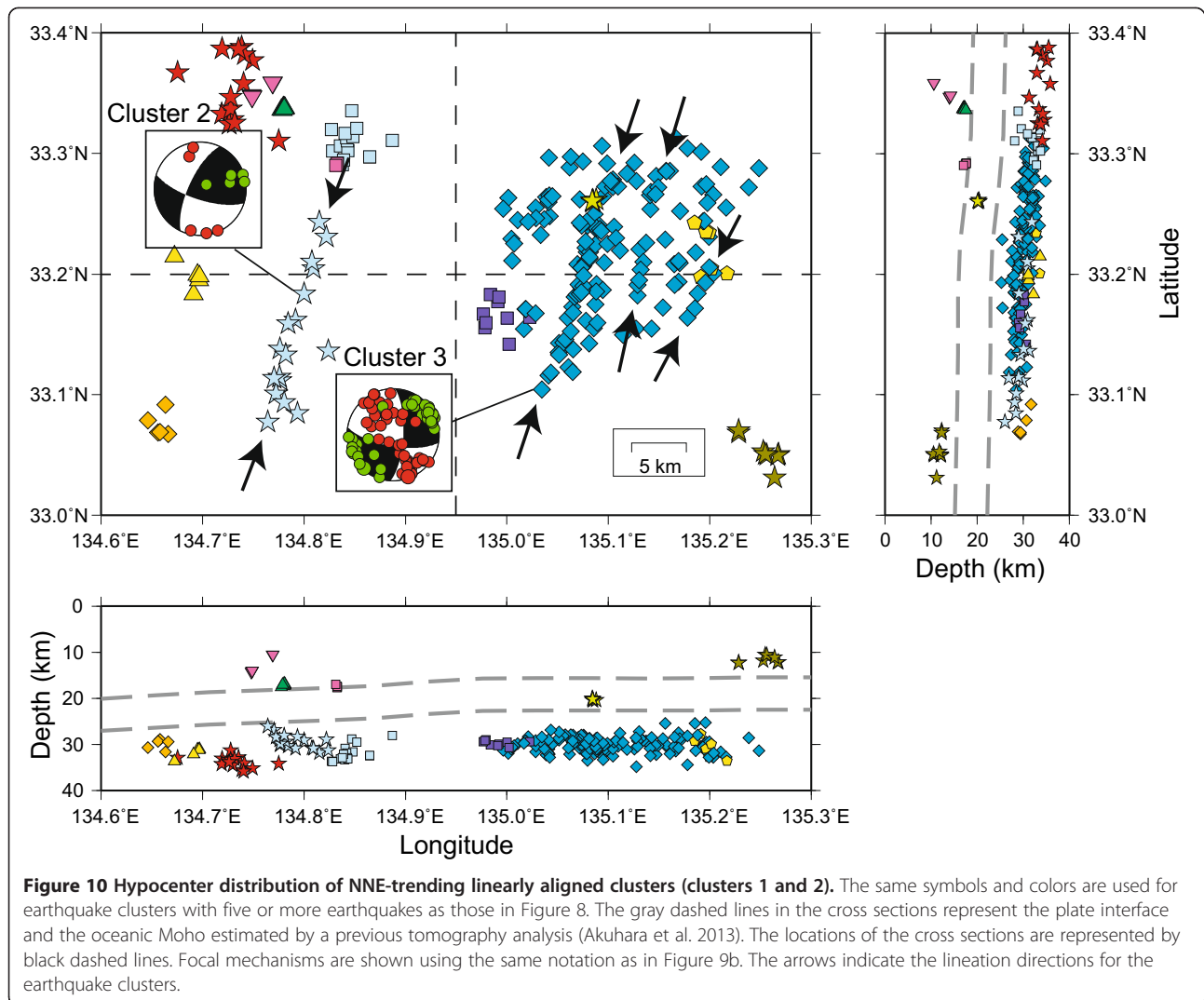

Figure 9 Seismograms (a) and focal mechanisms (b) for earthquakes belonging to the same cluster (cluster 1). (a) All the seismograms were bandpass-filtered between 3.0 and 15.0 Hz. The gray shaded area represents the parent time window for P- and S-waves. The thick black bars are manually selected arrival times. The early portions of the seismograms are enlarged in the lower panel. **(b)** P- and T-axes for each focal mechanism projected onto the lower hemispheres with red and green circles, respectively. The underlying focal mechanisms are representative of the cluster.

or both, using the FPFIT program (Reasenberg and Oppenheimer 1985). We accepted focal mechanisms only if they were determined from eight or more P-wave polarities, observed at stations whose maximum azimuthal gap from the hypocenter was less than 180°. A total of 541 focal mechanisms were acquired in this manner, for which the average half width of the 90% confidence region for the strike, dip, and rake was 6°, 7°, and 9°, respectively. Figure 9b shows the P- and T-axes of the focal mechanisms for cluster 1. Unlike the case for the waveforms themselves, large differences are observed, even if the uncertainties are taken into account. This may be caused by misidentification of the P-wave onsets or polarities due to a low signal-to-noise ratio in the OBS records. Although individual focal mechanisms are not sufficiently reliable, useful information concerning the tectonics can be extracted by gathering them within each cluster.
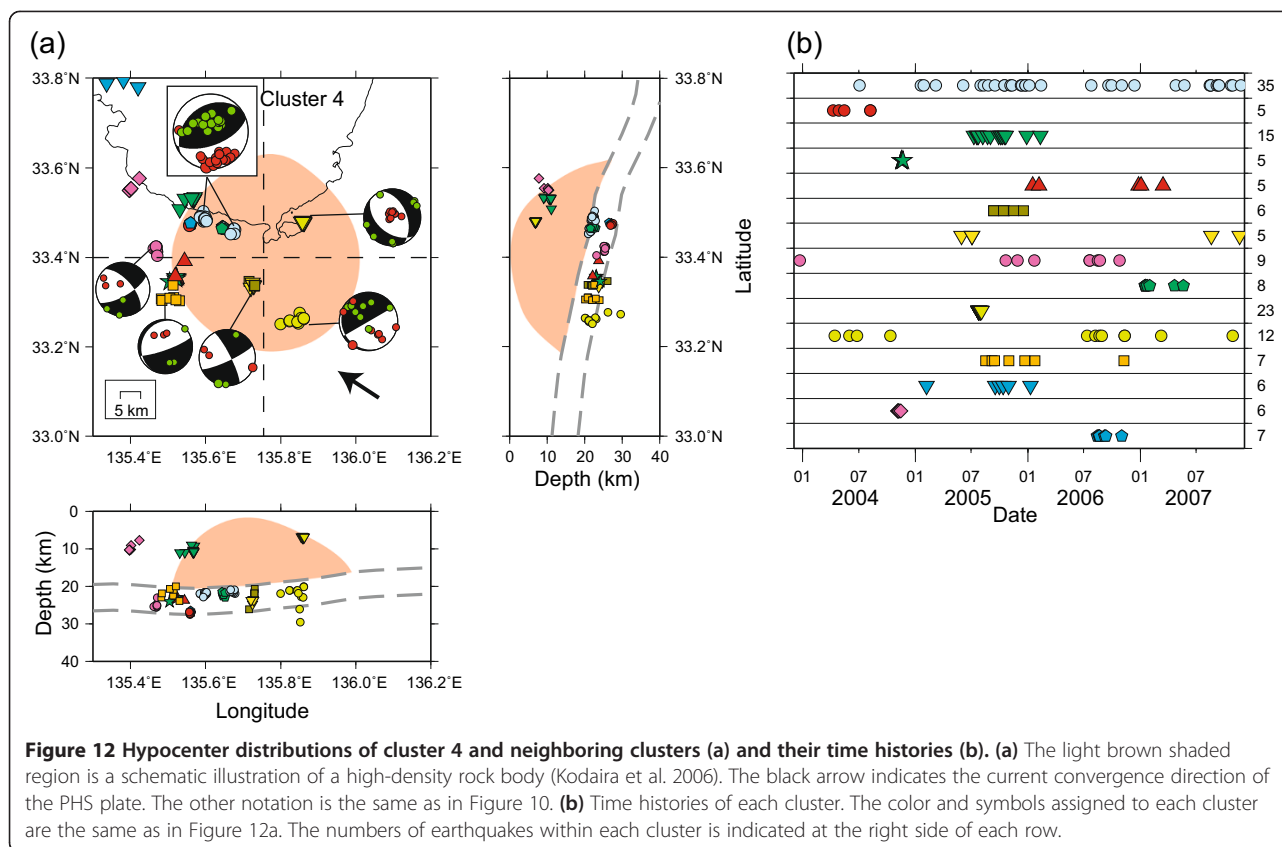
NNE-trending faults within the oceanic mantle
Figure 10 shows an enlarged view of clusters 2 and 3 in Figure 8, located offshore to the southwest of the Kii Peninsula, which are seen to be aligned along the NNE

Figure 10 Hypocenter distribution of NNE-trending linearly aligned clusters (clusters 1 and 2). The same symbols and colors are used for earthquake clusters with five or more earthquakes as those in Figure 8. The gray dashed lines in the cross sections represent the plate interface and the oceanic Moho estimated by a previous tomography analysis (Akuhara et al. 2013). The locations of the cross sections are represented by black dashed lines. Focal mechanisms are shown using the same notation as in Figure 9b. The arrows indicate the lineation directions for the earthquake clusters.

Figure 11 Cross section along line A-B in Figure 8. Earthquakes located within 7.5 km on either side are projected onto the section. The black circles depict non-clustered earthquakes or earthquakes in clusters of less than five. The pink and green dashed lines are the inferred upper and lower planes of the double seismic zone. The other notation is the same as in Figure 10.

direction. A comparison with the velocity structure model obtained in a previous tomographic analysis (Akuhara et al. 2013) reveals that these earthquakes occurred within the oceanic mantle. The P- and T-axes are oriented along the directions normal and parallel to the through, respectively (Figure 10). These axes are consistent with the regional stress field, which is considered to be due to slab pulling beneath Kyushu Island and the resistance of the mantle to the subduction motion (Wang et al. 2004). The strike for each focal mechanism almost agrees with the cluster alignment direction. As shown in Figure 11, it was also found that the clusters form the lower plane of a double seismic zone, separated from the upper plane by a distance of about 15 km.

Clear lineation of earthquakes has often been interpreted as being associated with faults, on the basis of additional geological evidence (e.g., Waldhauser et al. 1999). Unfortunately, in the present case, no geological evidence is available that directly supports the existence of faults,

Figure 12 Hypocenter distributions of cluster 4 and neighboring clusters (a) and their time histories (b). (a) The light brown shaded region is a schematic illustration of a high-density rock body (Kodaira et al. 2006). The black arrow indicates the current convergence direction of the PHS plate. The other notation is the same as in Figure 10. **(b)** Time histories of each cluster. The color and symbols assigned to each cluster are the same as in Figure 12a. The numbers of earthquakes within each cluster is indicated at the right side of each row.

because of the difficulty of offshore exploration at the depths involved. However, the geological features of the shallow part of the Philippine Sea slab can be helpful in placing some constraints on the cause of cluster lineation. Tsuji et al. (2013) conducted reflection seismic surveys seaward of the Nanakai Trough, off the Kii Peninsula, to identify a fault network within the oceanic crust, where some of the faults cross the Moho to the upper mantle. The fault network is similar to the clusters in the present study in that both comprise a sequence of parallel lineaments. In addition, the NNE trend of the lineaments is different from that of the fossil spreading axis of the Shikoku Basin with a NW-SE strike (Okino et al. 1994; Kido and Fujiwara 2004), along which faulting is expected to occur easily. This falsifies the hypothesis that the faults were formed long after subduction began, because once the slab has been subducted to a certain depth, the only possible mechanism for creating such a large fault is by trough-parallel regional tension. However, such tension is more likely to produce NW-SE faults along the direction of the fossil spreading axis. Consequently, it is considered that the earthquake clusters represent preexisting faults that were formed before subduction occurred.

It is well established that faults act as conduits in the outer trench regions, allowing fluids to seep into the mantle, and promote its serpentinization (e.g., Ranero et al.

2003; Faccenda et al. 2009). The serpentinized mantle dehydrates under certain conditions of temperature and pressure to cause intraslab earthquakes by overcoming the high confining pressure, which is a plausible explanation for the double seismic zones (Hacker 2003). Yamasaki and Seno (2003) calculated the temperature structure beneath the Kii Peninsula, and demonstrated that dehydration of the metamorphosed crust and serpentinized mantle could lead to upper and lower planes, respectively. The expected separation of the two planes is approximately 10 to 20 km, which is consistent with the results obtained in the present study. Therefore, it can be concluded that these clustered earthquakes occur due to dehydration of the serpentinized mantle along preexisting faults. Miyoshi and Obara (2010) also reported the existence of a lower plane at the northeastern part of the Kii Peninsula and concluded that it is the result of slab bending. However, on the basis of the number of earthquakes and the separation between the upper and lower planes, the lower plane identified by Miyoshi and Obara (2010) is considered to have a different origin to that in the present study.

Interplate earthquake cluster
As shown in Figures 8 and 12, cluster 4 at the southern tip of the Kii Peninsula contains more than 30 earthquakes with a low-angle thrust-fault focal mechanism.

The magnitude range is from 1.0 to 3.4. By comparing the velocity structure model and the convergence direction of the PHS plate, it can be concluded that these earthquakes occur at the plate interface. Focal mechanisms for neighboring clusters are also shown in Figure 12, revealing that interplate earthquakes are specific characteristic of cluster 4. Note that the cluster indicated by lime circles also has a low-angle thrust focal mechanism, but it is located within the oceanic crust. In addition, compared to the other clusters, cluster 4 is characterized by a large number of earthquakes and long periods of seismic activity (Figure 12b).

The repeated occurrence of similar earthquakes on the plate interface has been interpreted as slip at localized asperities (Igarashi et al. 2003). Kodaira et al. (2006) discovered a high-density rock body embedded in the overriding plate at almost the same location as the interplate clusters, although its center is shifted slightly southeastward (Figure 12). They estimated the excess normal stress to be more than 30 MPa, taking account of the results obtained from gravity observations (Honda and Kono 2005), which suggests that the embedded rock is a strong asperity at the plate interface. Moreover, the rate of interplate coupling drastically changes at the northeastern edge of the rock body: the coupling rate is higher seaward and lower landward (Miyazaki and Heki 2001; Saiga et al. 2011). Hence, it is considered that both sides of the plate interface are strongly coupled due to the weight of the rock body, with the strain being concentrated at the northwestern edge, where the interplate cluster is located. Long-term observations of these interplate earthquakes might provide insights into the process of strain accumulation during interseismic periods, which leads to megathrust earthquakes.

Conclusions

In the present study, CC results for records obtained by OBSs deployed off the Kii Peninsula were used to carry out a cluster analysis of earthquakes. We proposed methods to deal with problems intrinsic to OBS data, namely the predominance of monotone frequencies and varying site effects among stations. These methods involved multiple CC calculations, using time windows with different lengths, and appropriately determining the CC threshold, which varied among OBSs. By applying the appropriate threshold, we identified clusters of similar earthquakes.

There was found to be two types of characteristic clusters. The first existed within the oceanic mantle and was aligned along the NNE direction, forming the lower plane of a double seismic zone. This is considered to represent a preexisting fault formed before subduction began. It is proposed that these earthquakes are the result of dehydration of the serpentinized mantle along this fault. The second finding is a cluster of interplate earthquakes at the southern tip of the Kii Peninsula. We proposed that the primary generation mechanism for earthquakes in this cluster is strain accumulation beneath a high-density rock body.

The tectonics of most ocean areas remains unclear. We believe that the proposed method can be easily applied to OBS data from other areas, allowing new insights into the surrounding tectonics to be obtained.

References
Akuhara T, Mochizuki K, Nakahigashi K, Yamada T, Shinohara M, Sakai S, Kanazawa T, Uehira K, Shimizu H (2013) Segmentation of the Vp/Vs ratio and low-frequency earthquake distribution around the fault boundary of the Tonankai and Nankai earthquakes. Geophys Res Lett 40:1306–1310, doi:10.1002/grl.50223
Ando M (1975) Source mechanisms and tectonic significance of historical earthquakes along the Nankai Trough, Japan. Tectonophysics 27:119–140, doi:10.1016/0040-1951(75)90102-X
Baisch S, Ceranna L, Harjes HP (2008) Earthquake cluster: what can we learn from waveform similarity? Bull Seismol Soc Am 98:2806–2814, doi:10.1785/0120080018
Cattaneo M, Augliera P, Spallarossa D, Lanza V (1999) A waveform similarity approach to investigate seismicity patterns. Nat Hazards 19:123–138, doi:10.1023/A:1008099705858
Faccenda M, Gerya TV, Burlini L (2009) Deep slab hydration induced by bending-related variations in tectonic pressure. Nat Geosci 2:790–793, doi:10.1038/ngeo656
Ferretti G, Massa M, Solarino S (2005) An improved method for the recognition of seismic families: application to the Garfagnana-Lunigiana Area, Italy. Bull Seismol Soc Am 95:1903–1915, doi:10.1785/0120040078
Geller RJ, Mueller CS (1980) Four similar earthquakes in central California. Geophys Res Lett 7:821–824, doi:10.1029/GL007i010p00821
Hacker BR (2003) Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? J Geophys Res 108:2030
Honda R, Kono Y (2005) Buried large block revealed by gravity anomalies in the Tonankai and Nankai earthquakes regions, southwestern Japan. Earth Planets Space 57:e1–e4
Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc Ser B 52:105–124
Igarashi T, Toru M, Hasegawa A (2003) Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone. J Geophys Res 108:2249, doi:10.1029/2002JB001920
Jenkinson A (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q J R Meteorol Soc 81:158–171, doi:10.1002/qj.49708134804

Kido Y, Fujiwara T (2004) Regional variation of magnetization of oceanic crust subducting beneath the Nankai Trough. Geochem Geophys Geosystems 5:Q03002, doi:10.1029/2003GC000649

Kodaira S, Hori T, Ito A, Miura S, Fujie G, Park J-O, Baba T, Sakaguchi H, Kaneda Y (2006) A cause of rupture segmentation and synchronization in the Nankai trough revealed by seismic imaging and numerical simulation. J Geophys Res 111:B09301, doi:10.1029/2005JB004030

Kodaira S, Takahashi N, Nakanishi A, Miura S, Kaneda Y (2000) Subducted seamount imaged in the rupture zone of the 1946 Nankaido earthquake. Science 289:104–106, doi:10.1126/science.289.5476.104

Lin G, Shearer P (2007) Estimating local Vp/Vs ratios within similar earthquake clusters. Bull Seismol Soc Am 97:379–388, doi:10.1785/0120060115

Miyazaki S, Heki K (2001) Crustal velocity field of southwest Japan: subduction and arc-arc collision. J Geophys Res 106:4305–4326, doi:10.1029/2000JB900312

Miyoshi T, Obara K (2010) Double seismic zone within the ridge-shaped slab beneath southwest Japan. Earth Planets Space 62:949–954, doi:10.5047/eps.2010.11.001

Mochizuki K, Nakahigashi K, Kuwano A, Yamada T, Shinohara M, Sakai S, Kanazawa T, Uehira K, Shimizu H (2010) Seismic characteristics around the fault segment boundary of historical great earthquakes along the Nankai Trough revealed by repeated long-term OBS observations. Geophys Res Lett 37:L09304, doi:10.1029/2010GL042935

Okino K, Shimakawa Y, Nagaoka S (1994) Evolution of the Shikoku Basin. J Geomagn Geoelectr 46:463–479, doi:10.5636/jgg.46.463

Ranero CR, Morgan JP, McIntosh K, Reichert C (2003) Bending-related faulting and mantle serpentinization at the Middle America trench. Nature 425:367–373, doi:10.1038/nature01961

Reasenberg PA, Oppenheimer D (1985) FPFIT, FPPLOT and FPPAGE: Fortran computer programs for calculating and displaying earthquake fault-plane solutions. US Geol Surv Open File Rep 85–739:1–109

Saiga A, Kato A, Sakai S, Iwasaki T, Hirata N (2011) Crustal anisotropy structure related to lateral and down-dip variations in interplate coupling beneath the Kii Peninsula, SW Japan. Geophys Res Lett 38:L09307, doi:10.1029/2011GL047405

Savage MK, Wessel A, Teanby NA, Hurst AW (2010) Automatic measurement of shear wave splitting and applications to time varying anisotropy at Mount Ruapehu volcano, New Zealand. J Geophys Res 115:B12321, doi:10.1029/2010JB007722

Schaff DP, Waldhauser F (2005) Waveform cross-correlation-based differential travel-time measurements at the Northern California seismic network. Bull Seismol Soc Am 95:2446–2461, doi:10.1785/0120040221

Shearer P, Hauksson E, Lin G (2005) Southern California hypocenter relocation with waveform cross-correlation, part 2: results using source-specific station terms and cluster analysis. Bull Seismol Soc Am 95:904–915, doi:10.1785/0120040168

Shiomi K, Matsubara M, Ito Y, Obara K (2008) Simple relationship between seismic activity along Philippine Sea slab and geometry of oceanic Moho beneath southwest Japan. Geophys J Int 173:1018–1029, doi:10.1111/j.1365-246X.2008.03786.x

Tsuji T, Kodaira S, Ashi J, Park JO (2013) Widely distributed thrust and strike-slip faults within subducting oceanic crust in the Nankai Trough off the Kii Peninsula, Japan. Tectonophysics 600:52–62, doi:10.1016/j.tecto.2013.03.014

Waldhauser F, Ellsworth WL (2000) A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California. Bull Seismol Soc Am 90:1353–1368, doi:10.1785/0120000006

Waldhauser F, Ellsworth WL, Cole A (1999) Slip-parallel seismic lineations on the northern Hayward fault, California. Geophys Res Lett 26:3525–3528, doi:10.1029/1999GL010462

Walker DA (1984) Deep ocean seismology. Eos, Trans Am Geophys Union 65:2–3, doi:10.1029/EO065i001p00002

Wang K, Wada I, Ishikawa Y (2004) Stresses in the subducting slab beneath southwest Japan and relation with plate geometry, tectonic forces, slab dehydration, and damaging earthquakes. J Geophys Res 109:1–15, doi:10.1029/2003JB002888

Watanabe H (1971) Determination of earthquake magnitude at regional distance in and near Japan (in Japanese). J Seism Soc Jpn 24:189–200

Webb SC (1998) Broadband seismology and noise under the ocean. Rev Geophys 36:105–142, doi:10.1029/97RG02287

Wessel P, Smith WHF (1991) Free software helps map and display data. Eos Trans Am Geophys Union 72:441–441, doi:10.1029/90EO00319

Yamasaki T, Seno T (2003) Double seismic zone and dehydration embrittlement of the subducting slab. J Geophys Res 108:2212, doi:10.1029/2002JB001918

Zhang H, Thurber CCH (2003) Double-difference tomography: the method and its application to the Hayward fault, California. Bull Seismol Soc Am 93:1875–1889, doi:10.1785/0120020190