


FULL PAPER

Open Access



Noise classification for the unified earthquake catalog using ensemble learning: the enhanced image of seismic activity along the Japan Trench by the S-net seafloor network

Koji Tamaribuchi^{1*} , Fuyuki Hirose¹, Akemi Noda¹, Yuriko Iwasaki², Kazuhiro Iwakiri³ and Hiroshi Ueno³

Abstract

Homogeneous and accurate hypocenter distribution along the Japan Trench is important to better understand the process of plate subduction and the occurrence mode of large earthquakes. The seafloor seismic network (S-net) deployed recently along the Japan Trench has revealed new seismic activity including shallow slow earthquakes. However, conventional microseismic observations, such as those reported in the Japan Meteorological Agency (JMA) unified earthquake catalog, have been limited to land-based stations. Thus, steady seismic activity occurring at the shallow plate boundary, which is far from land is not always recorded. In the present study, we construct an automatic earthquake catalog using the nationwide observation network, including S-net. Because false detections caused by noise account for about 5% of automatically determined hypocenters, we attempted to reduce false detections using ensemble learning methods such as random forest and AdaBoost. First, we created a training dataset of earthquakes and noise by visual inspection based on the data recorded in the automatically determined catalog over a 2-month period, and we trained the dataset using the hypocenter and phase data as input. As a result of the training, AdaBoost was able to reduce the noise to about one-fifth of the total false detections, which is equivalent to 1%, while keeping the number of hypocenters above 99%. This method will contribute to significantly improving the efficiency of seismic activity monitoring and cataloging. In addition, the automatic denoised catalog data revealed that from January to August 2020, the completeness magnitude was M 1.6 along the Japan Trench. These microearthquakes were concentrated at depths of 20–50 km around the upper surface of the subducting Pacific Plate and are complementary to the slow earthquakes occurring at 10–20-km depths. Exceptionally, microearthquakes were observed off Iwate and Ibaraki prefectures, which correspond in location to areas of clustered foreshock activity. This spatial heterogeneity in microseismic activity is similar to the spatial complementary between the coseismic slip and afterslip of the 2011 Tohoku earthquake, which may be related to differences in interplate frictional properties and stress changes in the surrounding crust.

Keywords: Automatic hypocenter determination, Completeness magnitude, JMA unified seismic catalog, Machine learning, Japan Trench

Introduction

Homogeneous and accurate source distribution along the Japan Trench and other subduction zones is important to better understand the process of plate subduction

*Correspondence: ktamarib@mri-jma.go.jp

¹ Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan

Full list of author information is available at the end of the article

and the occurrence mode of large earthquakes. However, because the observation networks are located mainly on land, the detectability of earthquakes and the accuracy of focal depths decrease with distance from the land. Therefore, details of seismic activity occurring at the shallow plate boundary away from land are not well known. In particular, not all earthquakes were identified by visual inspection after the 2011 M_w 9.0 off the Pacific coast of Tohoku earthquake, referred to hereafter as the Tohoku earthquake, owing to the huge number of aftershocks.

To improve the accuracy of the hypocenter locations along the Japan Trench, various studies have been conducted. Gamage et al. (2009) used the sP depth phase to refine the focal depths and found that a double seismic zone had also formed in the vicinity of the Japan Trench. Seismic surveys (e.g., Tsuru et al. 2002, Miura et al. 2003) and seismic tomography (e.g., Zhao et al. 2011, Liu and Zhao 2018, Hua et al. 2020) have also revealed the depth of the plate boundary and the velocity structure of seismic waves. Frequent microseismic observations conducted regionally using pop-up ocean bottom seismometers (OBSs) have revealed seamount subduction and outer-rise seismic activity (e.g., Mochizuki et al. 2008, Obana et al. 2012). However, because the pop-up OBS observation period is between 3 and 12 months, the installation and recovery of these instruments must be repeated for regular seismic observation. In addition, the data can be analyzed only after collection; therefore, real-time analysis of the seismic waveforms is impossible.

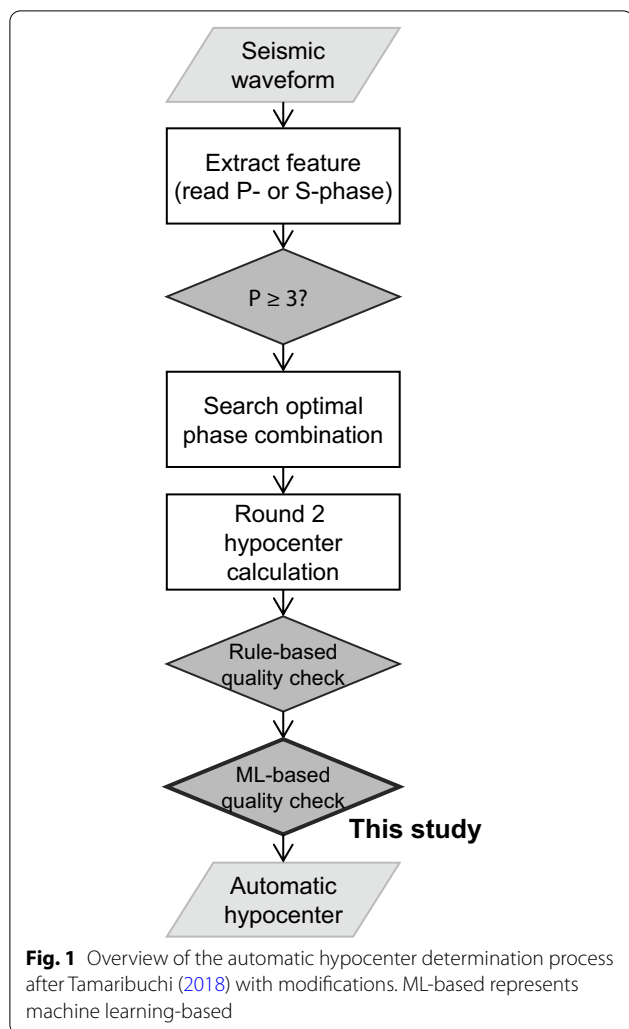
After the Tohoku earthquake, S-net, a closely spaced and highly sensitive ocean seismic network (Aoi et al. 2020) was deployed to capture earthquakes and tsunamis occurring along the Japan Trench (Fig. 7d, will be shown later). This project contributes to improving the speed and accuracy of early earthquake and tsunami warnings and provides new information on seismic phenomena, such as slow seismic activity occurring at the shallow part of the plate boundary (Nishikawa et al. 2019). However, the relationship between slow earthquakes and regular earthquakes, as well as their link to large earthquakes, has not been addressed thus far.

The Japan Meteorological Agency (JMA) collects real-time seismic waveform data from the National Research Institute for Earth Science and Disaster Resilience (NIED), universities, and other related institutions and releases a publically available earthquake catalog in cooperation with the Ministry of Education, Culture, Sports, Science and Technology (MEXT). We began to process the JMA unified earthquake catalog in October 1997. Through the introduction of Hi-net (Okada et al. 2004) in the early 2000s and the enhancement of automatic processing in April 2016 (Tamaribuchi 2018), we currently process data of about 200,000–300,000

hypocenters annually, excluding low-frequency earthquakes. The introduction of S-net is expected to provide a comprehensive and real-time view of seismic activity along the Japan Trench. However, regarding data continuity, the changes in completeness magnitude and hypocentral distribution of the JMA unified catalog before and after the use of S-net need to be clarified. As previously mentioned, it is difficult to identify all earthquakes occurring in this area by visual inspection; therefore, it is essential to process them automatically. To identify the fault geometry using aftershock distribution and to forecast the number of aftershocks, automatically determined hypocentral distribution is particularly important immediately following a major earthquake. This method of automated processing is used to quickly and accurately evaluate the number of earthquakes and is thus essential for disaster prevention.

The purpose of the present study is to evaluate the automatic hypocentral distribution using S-net and to elucidate the details of seismic activity occurring along the Japan Trench.

Here, we introduce the enhancements in automated processing introduced in April 2016 for the JMA unified catalog. To determine the simultaneous occurrences of earthquakes that were prominent after the Tohoku earthquake, we improved the processing by enacting two measures. Specifically, we continually detect discontinuities that can be considered as candidates for P- and S-phases without using short-time average over long-time average (STA/LTA) triggers, and we search for the phase combinations that best match the theoretical values based on the observed values of the arrival time of the phase candidates and the maximum amplitude after the arrival time. Importance sampling was used to search for the best combination of phases. As the first step, we randomly assume the location of the earthquake near a station in which the P-wave detection time is early, and we calculate the likelihoods based on residuals of travel time and magnitude for the assumed locations. On the basis of the results, the hypocenters are then assumed to be around the locations of high likelihoods, and the likelihoods are calculated again. By repeating this process, we can search for locations with high likelihoods, i.e., those with small residuals between the theoretical and observed values. This approach is referred to as the phase combination forward search (PF) method (Fig. 1). By employing this method, almost 100% of shallow inland earthquakes ($M \geq 1$) can be identified under normal conditions, which greatly contributes to speeding the hypocentral determination and improving the processing efficiency. The obtained automatic hypocenters are visually checked to ensure that they are not false data caused by noise, and the data are then listed in the JMA unified



catalog. In addition, a careful visual inspection of the phase is conducted at a particular scale of $M \geq 2$ inland, as detailed by Tamaribuchi (2018).

However, the unavoidable problem of earthquake misdetermination owing to noise in automatic processing is not limited to the PF method. A trade-off is present between the number of false detections and the determination rate of automatically identified earthquakes, the latter of which is the fraction of the number of earthquakes listed in the JMA unified catalog that were determined by the automatic process. Therefore, parameter tuning has been employed for many years to reduce false detection caused by noise while maintaining the earthquake determination rate. The following criteria are used for outputting the current PF method. (1) The sum of the number of P- and S-phases at 20 stations near the epicenter must be 5 or more. (2) The number of stations from which both the P- and S-phases are selected must be two or more; that for only the P-phase must be ten or

more. (3) The root mean square of the P-wave travel time residuals must be 0.6 s or less, and (4) that of the S-wave travel time residuals must be 1.2 s or less. In addition, (5) the latitude and longitude error remaining after the last hypocenter calculation should be less than 10 min, (6) the origin time error should be less than 2 s, and (7) the magnitude needs to be determined. Some may wonder why there are events whose magnitude is not determined. The magnitude is calculated when each station observes a several times larger signal than the noise level. Therefore, when observed amplitude changes are as small as the noise level, the events are marked as magnitude undetermined. If even one of these conditions (1)–(7) is not met, the hypocenter is judged to be a noise misdetermination, and no output is made. Although such rule-based discrimination between an earthquake and noise is easily understood by operators, it requires extensive trial and error owing to the arbitrary nature of the parameters.

In recent years, the success of deep learning has led to a worldwide boom in artificial intelligence and its resulting real-world application. The field of seismology has also successfully introduced machine learning, particularly deep learning methods, such as clustering of earthquake locations (Perol et al. 2018) and detection of seismic arrival time (Zhu and Beroza 2019; Ross et al. 2018). However, the black box nature of this process makes the estimation results difficult to interpret. Although explainable artificial intelligence (XAI) is not discussed in the present study, it should be noted that abundant research is being conducted on this topic. With the development of such machine learning methods and libraries, various attempts have been made on an operational basis in deep learning and other applications.

In the present study, we use machine learning to evaluate and discuss the reduction in the number of false detections recorded during the automatic determination process as well as the resulting seismic activity along the Japan Trench that benefits from S-net usage. “[Data and features](#)” section presents the data and the features used for the training. In “[Denoising method](#)” section, we discuss random forest and AdaBoost ensemble learning, which are types of supervised machine learning. The performance of automatic earthquake and noise classification is evaluated in “[Results of denoising method](#)” section, and the results obtained for March 2020 are presented in “[Application of denoising method](#)” section. In “[Seismic activity along the Japan Trench](#)” section, we discuss the seismic activity occurring in the shallow part of the plate boundary based on automatically processed catalog data recorded between January and August 2020, and a summary of the study is presented in “[Summary](#)” section.

Data and features

Training and test data

The automatically determined hypocenters obtained by employing the PF method from January 1 to February 29, 2020, were used as training and test data. In addition, S-net was used, as were several modifications prescribed for the S-net data, such as vector transformation, the addition of velocity structure by sea area, consideration of station height, and velocity magnitude correction for seafloor seismographs. Details of these modifications have been presented by Ueno et al. (2019). It should be noted that air gun surveys conducted in an ocean can be incorrectly identified as automatic hypocenters. Therefore, in this study, we used autocorrelation of the waveforms to avoid detecting events caused by such instruments (Additional file 1: S1).

During this period, the JMA unified catalog did not use S-net and did not include false determinations caused by noise because the period occurred after the scrutiny of automatically determined hypocenters. In contrast, the automatically determined hypocenters discussed in the present study employed S-net and contain false determinations owing to noise. Therefore, we first compared our results with the JMA unified catalog data. We considered automatic hypocenters generated within 5 s of the origin time and 50 km from the epicenter to be coincidental (category 0). All of the automatically determined hypocenters that disagreed with the JMA unified catalog were visually checked for phase readings and were classified into the following four categories: (1) earthquakes with generally correct readings, (2) earthquakes with one or two clear errors in the readings, (3) earthquakes with large errors in the readings, and (4) false detections caused by noise, hereinafter referred to simply as noise in this context. Of artificial sources in the automatically determined hypocenters, recordings of quarry blasting and air gun were classified as earthquakes and noise, respectively, because we could not distinguish the former from an earthquake by visual inspection. Of course, most of the air gun recordings were removed from data set beforehand (see Additional file 1: S1). Examples and the results of the classification are shown in Figs. 2 and 3, and the number of events in each category is shown in Table 1. It should be noted that many category 1 earthquakes were found in the Tohoku region, as shown in Fig. 3b, because many automatic hypocenters excluded from the JMA unified catalog were identified using S-net. Although earthquakes were distributed in the inland areas of western Japan, most were less than M 1, which is at the lower limit of detectability. Many category 4 noise events were also found around Ryukyu islands, southwestern Japan, as shown in Fig. 3e, because there are noisy stations and it is forced to determine the

hypocenter with fewer stations than inland. A cluster of false detection off the coast of Kochi Prefecture is due to noises caused by air gun surveys.

This problem was considered to be a classification problem with supervised learning. Therefore, categories 0 and 1 and category 4 were labeled as “Earthquake” and “Noise,” respectively. The labeled data were then split so that the training data and the test data were given in a ratio of 8:2. The training data were used to optimize the training parameters, and the test data were used to evaluate the generalization performance. These data must be independent of each other to avoid overlearning which means overfitting the data. To facilitate comparison between two types of ensemble learning methods, the datasets used for training and testing were the same between the methods.

The hypocenter and phase data include the origin time, source location, and magnitude as well as the arrival time and maximum amplitude of each station. In this study, 229 features used for training were extracted from the hypocenter and phase data.

Hypocenters and error

Hypocenters have five main features: origin time t_0 , latitude lat , longitude lon , depth dep , and magnitude M (Table 2). The origin time t_0 refers to the time in seconds elapsed from 00:00. This is because the difference in noise occurrences between daytime and nighttime and noises that observed at a particular time of day are considered in the learning. The date of the origin time was not used for learning in this study.

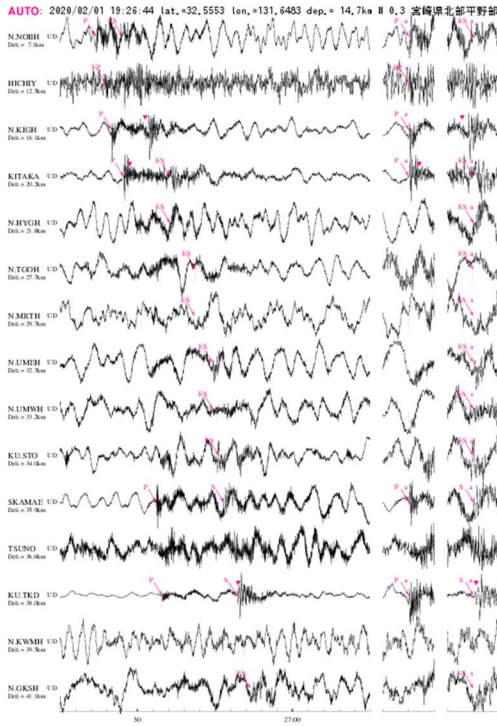
Hypocentral error includes four main features: origin time error σ_t , latitude error σ_{lat} , longitude error σ_{lon} , and depth error σ_{dep} .

Phase data

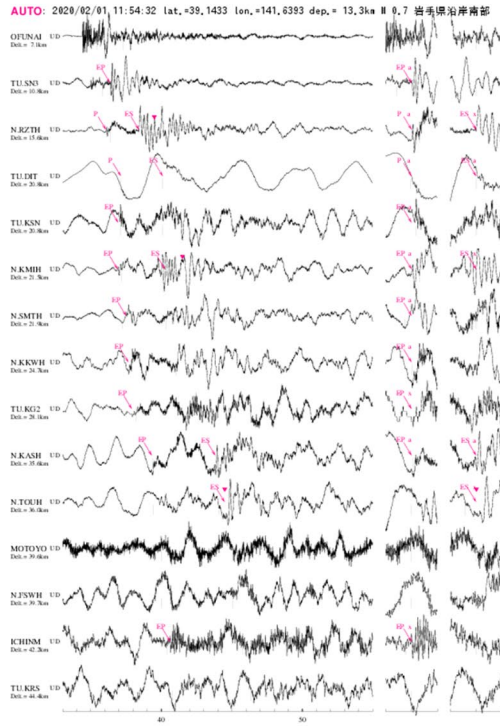
The phase data are listed in the order of location closest to the hypocenter ($i = 1, \dots, 20$), including stations in which no phase was read. These data include 11 main features at each station: cumulative number of P-phases N_p^i , cumulative number of S-phases N_s^i , cumulative number of stations N_{ps}^i in which both the P- and S-phases are read, epicentral distance D^i , back azimuth angle θ^i , travel time residuals of the P-phase r_p^i , travel time residuals of the S-phase r_s^i , magnitude residuals r_m^i , root mean square of the P-phase residuals up to the i th station σ_p^i , root mean square of the S-phase residuals σ_s^i , and root mean square of the magnitude residuals σ_m^i . Considering that 20 stations were included, a total of 220 features were obtained.

Here, the residuals in the absence of P- and S-phases r_p^i, r_s^i , and r_m^i were set to 0.0. Similarly, when the

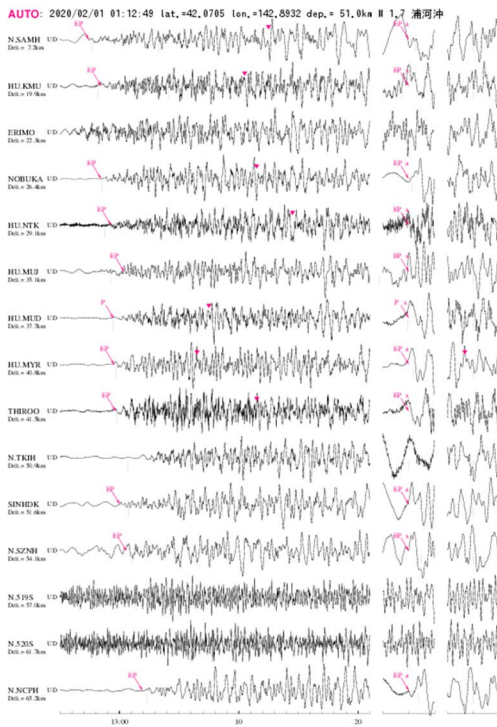
a Category 1



b Category 2



c Category 3



d Category 4

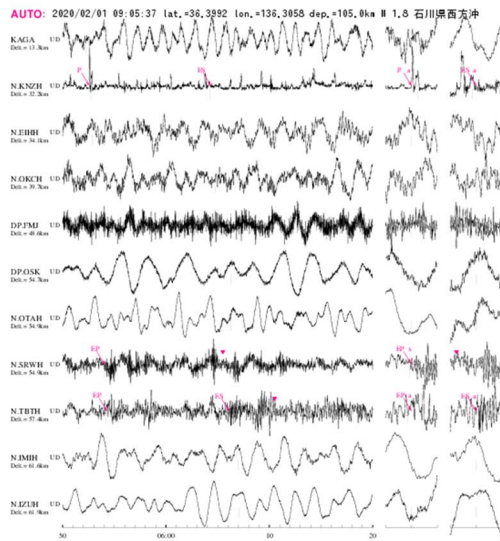


Fig. 2 Classification examples for each category. Each figure shows the vertical seismic waveforms in the order of epicentral distance. The right side shows enlarged views of 1 s before and after the P-wave and S-wave arrival times, for which we basically used reading times. We used theoretical values in the absence of a reading

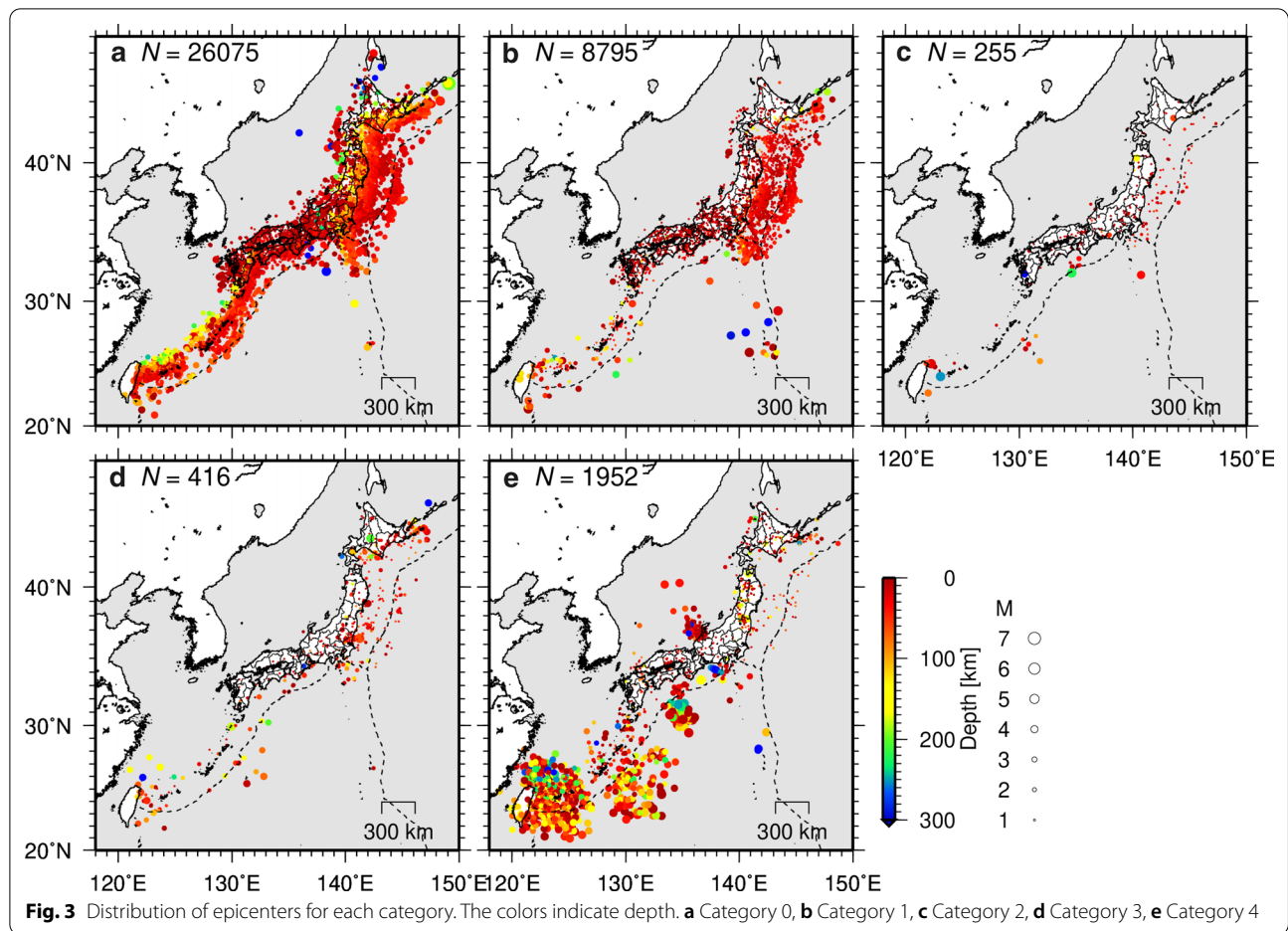


Table 1 Total number of events in each category

Category	<i>N</i>	Rate (%)
0: Matched	26,075	69.5
1: Earthquake and phases are correct	8795	23.5
2: Earthquake and a few phases are incorrect	255	0.7
3: Earthquake and phases are incorrect	416	1.1
4: Noise	1952	5.2
Total	37,493	100.0

N and rate indicate the number of events and percentage to total events, respectively

cumulative number of N_p^i , N_s^i , and N_m^i was zero, the root mean square σ_p^i , σ_s^i , and σ_m^i was set to 0.0, respectively.

Denosing method

Ensemble learning is a method used to create high-performance classifiers by combining low-performance classifiers (e.g., decision trees). Two main methods are employed in this type of learning: bagging and boosting. Random forest, which is a type of bagging, creates

multiple sub-datasets of training data by random sampling with replacements and trains a decision tree for each sub-dataset (Breiman 2001). Although the classification performance of individual decision trees, referred to as weak learners, might not be high, the total classification performance can be improved using their majority votes or the average of their probability values. AdaBoost, a type of boosting, differs from bagging because it does not create multiple sub-datasets of training data. This method is used for boosting weights of data that misclassified by a weak learner to adapt to more difficult data (Freund and Schapire 1997). No one method is better than the other and thus the method should be chosen depending on the data for appropriate application. Therefore, in this study, we tested both methods and selected that with the highest classification performance.

Random forest

In this study, the number of decision trees was set to 100, and the classification and regression trees (CART) decision tree algorithm was used to find the bifurcation point at which the impurity showed the greatest

Table 2 Types of features

Features	
t_0	Elapsed time of origin time from 00:00 (s)
lat	Latitude (°)
lon	Longitude (°)
dep	Depth (km)
M	Magnitude (no unit)
σ_t	Origin time error (s)
σ_{lat}	Latitude error (min)
σ_{lon}	Longitude error (min)
σ_{dep}	Depth error (km)
Features	
N_p^i	Cumulative number of P-phase
N_s^i	Cumulative number of S-phase
N_{ps}^i	Cumulative number of P-phase and S-phase
D^j	Epicentral distance (km)
θ^j	Back azimuth (°)
r_p^i	Residual of P-phase (s)
r_s^i	Residual of S-phase (s)
r_m^i	Residual of magnitude (no unit)
σ_p^i	Root mean square of P-phase residual (s)
σ_s^i	Root mean square of S-phase residual (s)
σ_m^i	Root mean square of magnitude residual (no unit)

i indicates the station index in the order of location closest to the hypocenter ($i = 1, \dots, 20$), including stations in which no phase was read

reduction. For example, one of the decision trees for the case in which the branching depth was up to 3, which is easy to visualize, is shown in Fig. 4.

Here, we give a brief overview of CART. The feature vector used for label estimation is $X = \{x_1, \dots, x_k\}$, where k denotes the total number of features, and the label is y . The labels were classified into two categories in this study: earthquake and noise. For the sake of convenience, we defined an earthquake as -1 and noise as 1; then, $y \in \{-1, 1\}$. For each training dataset, the number of datasets with earthquake and noise labels are N_1 and N_2 , respectively. Then, the Gini impurity is defined by the following equation:

$$H(X, y) = \sum_{l=1}^2 \frac{N_l}{N} \left(1 - \frac{N_l}{N}\right). \quad (1)$$

Here $N = N_1 + N_2$. The Gini impurity has the maximum value, $H(X, y) = 1/2$, when the dataset has same number of earthquakes and noise events, $N_1 = N_2 = N/2$. On the contrary, if the dataset has only earthquakes or noise events, $N_1 = N, N_2 = 0$ or $N_1 = 0, N_2 = N$, and the Gini impurity has the minimum value, $H(X, y) = 0$.

Then, we split this training dataset (X, y) with the threshold q_t of the j th feature to the left and right. It is represented by an equation with the splitting parameters $\theta = (j, q_t)$ as

$$\begin{cases} Q_{\text{left}}(\theta) = (X, y) | x_j \leq q_t \\ Q_{\text{right}}(\theta) = (X, y) | x_j > q_t \end{cases}. \quad (2)$$

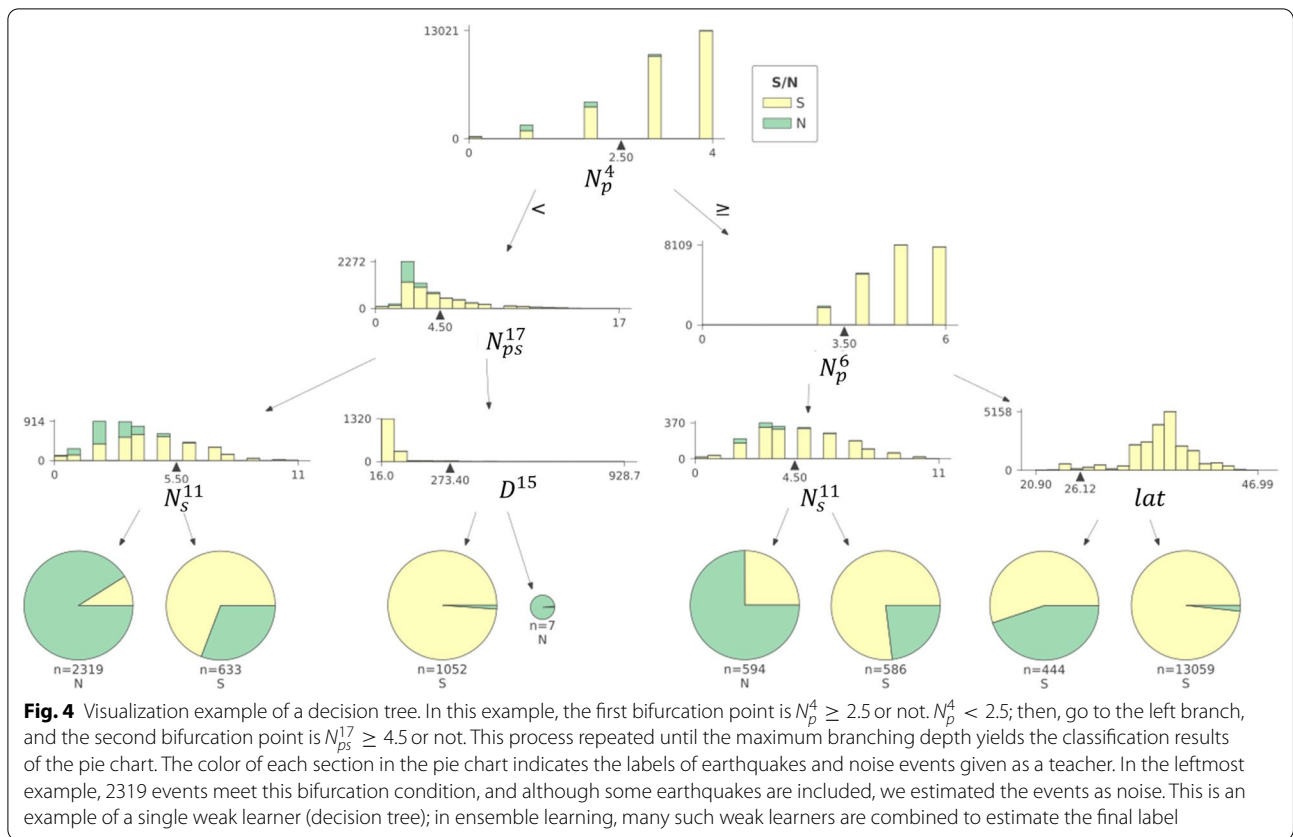
The gain by the split is

$$G(Q, \theta) = H(X, y) - \frac{n_{\text{left}}}{N} H(Q_{\text{left}}) - \frac{n_{\text{right}}}{N} H(Q_{\text{right}}). \quad (3)$$

That is, when we split a highly impure dataset with a mixture of earthquake and noise labels with the parameter θ , the weighted sum of impurities remaining after the split is minimized. The parameter θ^* with the maximum gain is

$$\theta^* = \operatorname{argmax}_{\theta} G(Q, \theta). \quad (4)$$

By repeating these partitions up to the maximum depth of branching in the decision tree, we split the dataset and work to reduce the impurity. Maximum branching depths of 3, 5, and 7 were attempted for each decision



tree. To resolve the imbalance in the number of data for each label of earthquake and noise, we set weights normalized by the number of earthquakes and noise during training. After training, we can estimate the label, and the output is $\hat{y} \in \{-1, 1\}$. The above configuration can be written in a few lines using the scikit-learn library in Python (Pedregosa et al. 2011).

AdaBoost

The feature vector for each dataset used for label estimation is $X_i = \{x_{i1}, \dots, x_{ik}\}$, where $i = 1, \dots, N$ is the number of training datasets, and k indicates the total number of features. The earthquake and noise labels corresponding to the i th data are $y_i \in \{-1, 1\}$.

1. In the first step ($t = 1$), we trained with a decision tree (weak learner), with all data weights uniform. That is, when we defined the weight of dataset at step t as $w_i^{(t)}$ ($i = 1, \dots, N$), $w_i^{(1)} = 1/N$.
2. For weighted data, the decision tree was trained on the weighted data, and the weak learner $f_t(X_i)$ was obtained.

3. Using the obtained weak learner, the misclassification rate E_t of the labels was calculated; E_t is a weighted average of the number of misclassifications.
4. If $E_t > 0.5$, then this loop was broken; $E_t > 0.5$ indicates that more than half of the data estimated the wrong label.
5. We calculated $\alpha_t = \frac{1}{2} \log((1 - E_t)/E_t)$, which is the coefficient of $f_t(X_i)$. When the misclassification rate was high, the coefficient was low.
6. On the basis of the classification results, we updated the weights according to Eq. (5).

$$w_i^{(t+1)} \propto w_i^{(t)} \exp(-\alpha_t y_i f_t(X_i)). \tag{5}$$

Here, the reason of proportionality is to multiply the normalization factor so that $\sum_i w_i^{(t+1)} = 1$. The $y_i f_t(X_i)$ term is 1 (-1) if the predictions are consistent (inconsistent) with the labels of the observation. That is, Eq. (5) increases (decreases) the weight of the incorrect (correct) answer data.

7. $t \leftarrow t + 1$, the process is repeated from step 2.

Here, we used the same decision tree as that used in “Random forest” section. The maximum number of the decision trees was set to 100, which is the same as

that used for the random forest, and for the maximum depth of branching of each decision tree 3, 5, and 7 was attempted. Label estimation was based on the weighted sum of each decision tree:

$$f(X) = \text{sign} \left\{ \sum_t \alpha_t f_t(X) \right\}, \quad (6)$$

where sign is a sign function in which 1 (−1) is used for a positive (negative) value.

Feature selection

As mentioned in “[Data and features](#)” section, 229 features were used in this study. However, not all contributed to discriminate noise and earthquakes. Sorting the features that truly contribute to the discrimination from the 229 features can improve the readability of the discrimination process in machine learning which is often referred to as black box. Furthermore, this process is expected to be effective in reducing overlearning and improving the classification performance. Therefore, we attempted to use Boruta’s method of feature selection (Kursa and Rudnicki 2010) in the present study.

The outline of this method is described as follows. First, we randomly shuffled each dataset for each feature from the training dataset. This created a shuffled dataset with a spurious feature that should not be useful for label estimation. We trained this together with the original training dataset in a random forest. If the feature importance was less than the importance of the spurious feature, the feature had no significance and was considered to be random. Therefore, we repeated this test many times up to 500 times to determine whether we can reject the null hypothesis that the importance of a feature is the same as the importance of a spurious feature. That is, this method is used for selecting only significant features.

To maintain stability, we used the 80% quantile of the spurious features when comparing importance. A significance level of $p=5\%$ was used to determine the significance of the features. However, we performed a Bonferroni correction to test the same data repeatedly because it eliminates the effect of multiple testing by dividing the p value threshold by the number of tests.

In this study, we attempted to classify automatically determined hypocenters by the random forest and AdaBoost after variable selection.

Results of denoising method

Variable selection

As mentioned in “[Training and test data](#)” section, 29,457 earthquakes and noise events were used as training data for variable selection. The maximum branching depth of the decision tree in the random forest used for variable

selection was set to 7. The variable selection reduced the number of features from 229 to 183 (Table 3). The excluded features were travel time residuals and magnitude residuals far from the epicenter. The S-phase at the station closest to the epicenter was excluded because they are difficult to detect owing to their short S–P times, and they have not been identified in a significant number of earthquakes.

Random forest

For the random forest method, 29,457 (7365) earthquakes and noise events were used for training (testing). Table 4a shows the results for a case in which the maximum branching depth of the decision tree was set to 7. Of the total number of test data, the percentage of data correctly classified as noise events and earthquakes (accuracy) was 96.39%. In addition, the percentages of test data for earthquakes and noise events that could be correctly classified as such were 96.52% and 94.23%, respectively. As shown in Table 1, false detections caused by noise were originally about 5% of all automatically determined catalog. Therefore, machine learning can be used for about 94% suppression. Simple calculation revealed that the ratio of the noise events to the total was 0.3%. On the contrary, 3.5% of the earthquakes were misclassified as noise.

In this study, we showed that the number of decision trees was 100, and the maximum depth of branching was 7. At maximum branching depths of 5 and 3, the accuracies were 94.58% and 91.19%, respectively.

Table 5a shows the top ten factors of permutation importance, which indicates which important features are considered for discrimination. Permutation importance is a measure of the decrease in accuracy that occurs when the order of the data for a feature is shuffled. A lower accuracy reflects higher feature importance. In this study, we shuffled ten times for each feature, and the difference between the mean of the accuracy in the shuffled data and the original training data was taken as the importance. The results showed a high importance of magnitude. In fact, many of the noise events around Ryukyu islands, southwestern Japan, are misclassified as large-magnitude events even though the number of phases is low. Therefore, this condition is considered to be a learning experience.

AdaBoost

For the AdaBoost method, 29,457 (7365) earthquakes and noise events were used for training (testing). For the sake of fairness, the dataset was kept the same as that discussed in the previous section. The results for a maximum branching depth of 7 are shown in Table 4b. Of the total number of test data, the accuracy for noise

Table 3 Variable selection

(a) Hypocenter											
t_0	lat	lon	dep	M	σ_t	σ_{lat}	σ_{lon}	σ_{dep}			
TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE			
(b) Phase											
i	N_p^i	N_s^i	N_{ps}^i	D^i	θ^i	r_p^i	r_s^i	r_m^i	σ_p^i	σ_s^i	σ_m^i
1	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
2	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
3	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
4	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
5	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE
6	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
7	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE
8	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE
9	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
10	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
11	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
12	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE
13	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
14	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
15	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
16	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
17	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
18	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
19	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE
20	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE

True indicates the selected feature. The meaning of each feature is described in Table 2
 FALSE indicates a non-significant feature

events and earthquakes was 98.51%. The percentages of earthquakes and noise events that could be correctly classified as such were 99.53% and 81.49%, respectively.

Table 4 Classification results (confusion matrix)

	Prediction		
	Noise	Earthquake	Total
(a) Random forest, depth = 7			
Observation			
Noise	392	24	416
Earthquake	242	6707	6949
Total	634	6731	7365
(b) AdaBoost, depth = 7			
Observation			
Noise	339	77	416
Earthquake	33	6916	6949
Total	372	6993	7365

The number indicates events classified in each category. All correct predictions are located in the diagonal of the table

When the maximum branching depths of the decision tree were 5 and 3, the accuracies were 98.49% and 98.48%, respectively. These results are better than those obtained by the random forest method even when using a low maximum branching depth.

The top ten permutation importance factors are shown in Table 5b. Similar to that shown in the random forest results, magnitude was the most important feature.

Application of denoising method

We applied the denoising method tuned in the previous section to 1 month of data recorded in March 2020. The number of automatically determined events during this period was 18,982. For both random forest and AdaBoost methods, the time required for noise and earthquake classification was about 6 min (CPU: Intel Xeon Silver 4214 2.20 GHz with one core). That is, the time for one event was about 0.02 s, which is sufficient for practical purposes. Using the random forest method, 17,175 (1807) data were classified as earthquakes (noise events),

Table 5 Top 10 permutation importance features

Feature	Importance
(a) Random forest	
M	0.002471
σ_p^4	0.000299
σ_s^{11}	0.000244
σ_p^1	0.000190
σ_s^{10}	0.000177
σ_s^{14}	0.000177
σ_m^{19}	0.000177
σ_p^5	0.000163
σ_p^2	0.000163
θ^1	0.000136
(b) AdaBoost	
M	0.003585
lon	0.001439
t_0	0.001371
N_p^{14}	0.001181
N_p^5	0.001154
N_p^{20}	0.000950
N_p^{19}	0.000815
N_p^{10}	0.000815
N_p^{11}	0.000692
N_p^4	0.000597

or 90.5% (9.5%). When the AdaBoost method was used, however, 17,991 (991) were classified as earthquakes (noise events), or 94.8% (5.2%).

Figure 5 shows the epicentral distribution after classification by AdaBoost. The frequency magnitude distribution showed that the rate of noise events of $M \sim 3-4$ was relatively higher than that of earthquakes. The automatic denoising by this study enabled us to suppress the noise and to obtain reliable epicentral distribution. The number of automatically determined epicenters identified within 5 s of the origin time and 50 km from the epicenter was reduced by 49 (0.4%), from 13,082 to 13,033, compared with that given in the JMA unified catalog. This value is sufficiently practical in terms of early detection of seismic activity. It should be noted that there were 4958 (= 17,991-13,033) earthquakes which did not coincide with the JMA unified catalog. More than half of them occurred off the Tohoku region that received the benefits of S-net. This will be discussed in the next section.

A comparison of Figs. 3e and 5a revealed that the distributions of noise classified by visual inspection and those by machine learning showed a similar trend. However, some areas, such as off the coast of Kochi Prefecture (31 °N, 135 °E) and near Ishikawa Prefecture (37 °N, 136 °E), showed slightly different noise distributions. There is no false detection off Kochi Prefecture in Fig. 5a because

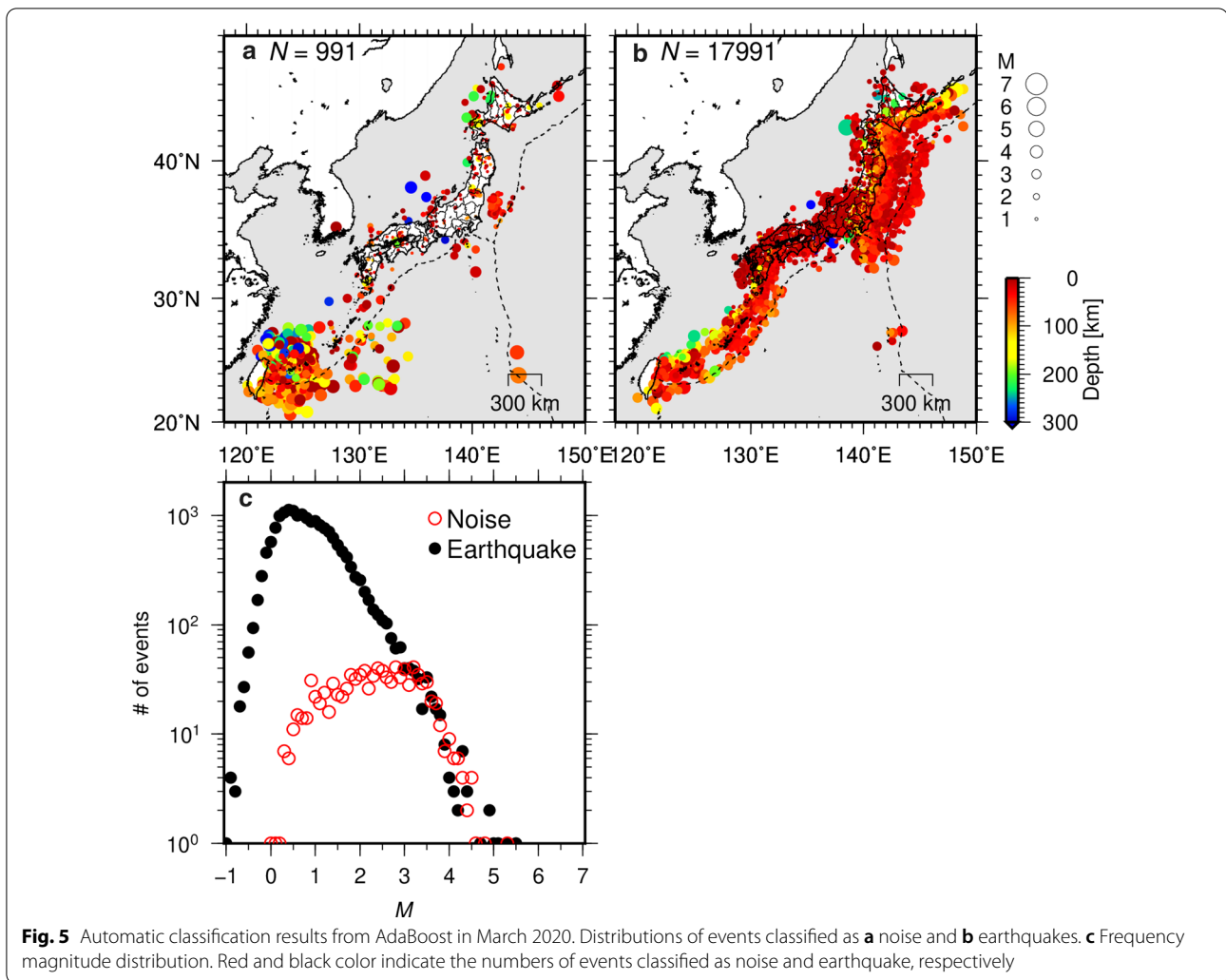
airgun surveys were not conducted during this period. It also seems that few false detections near Ishikawa Prefecture due to changes in the noise environment caused by construction work near the station. This shows that machine learning does not learn only the spatial distribution of noise. In addition, off the coast of Fukushima and Ibaraki (37 °N, 142 °E), the waveform data were not available between March 2 and 26 owing to S-net maintenance. Therefore, some earthquakes which occurred near the region were classified as noise events. In this study, our discrimination between noise and earthquakes was based on hypocenter and phase data only. Therefore, it is necessary to develop a method that considers the missing data in the future research. It is also desirable to continuously add (update) training data since the frequently observed noise (construction and air gun) at a particular station may change with the period or season.

In the present study, we used the random forest and Adaboost methods, in which libraries are readily available for performing the calculations. Ultimately, we adopted AdaBoost because it had a lower rate of earthquake misclassification as noise. The major advantage of this study is that this method was easily applied because the data size was smaller than that of the seismic waveform itself. The maximum number of features in this study was about 200, and we were able to improve the classification performance sufficiently without the use of complex deep learning. By applying and validating other machine learning methods, we can further improve the classification accuracy within the same framework. However, to more effectively improve the noise classification performance, machine learning detection or denoising methods using seismic waveforms would also need to be integrated, as discussed by Zhu and Beroza (2019) and Ross et al. (2018). This would contribute to the early detection of seismic activity using automatically determined hypocenters and would improve the cataloging efficiency by providing the data as guidance for the operator.

The application of the classifier developed in the present study to other observation networks needs to be validated on a case-by-case basis. In the future, we would like to consider, for example, training on observation networks in eastern Japan and testing in western Japan to improve the generalization performance (Hara et al. 2019).

Seismic activity along the Japan Trench

In this section, we show the changes in hypocenter distribution based on S-net usage between January 1 and August 31, 2020; denoising by visual inspection for the hypocenter between January 1 to February 29, 2020: that is, categories 0 and 1 made in “Data and features” section;

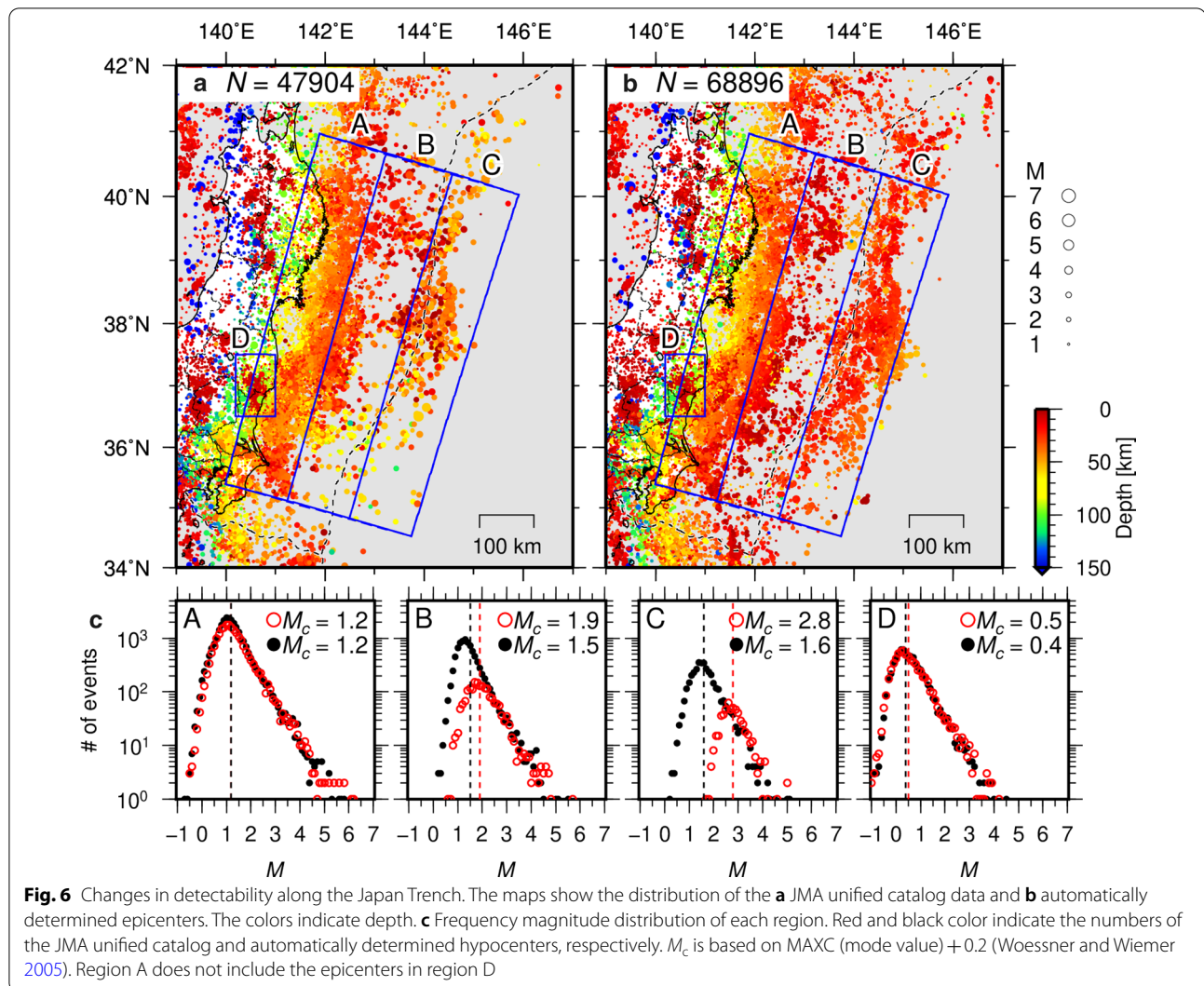


and ensemble learning with AdaBoost after March 1, 2020.

Figure 6 shows the detectability of the JMA unified catalog and automatically determined earthquakes. Additional file 1: Table S1 shows the residuals of hypocenter locations and magnitudes between coincidental events. Compared with the JMA unified catalog data, the completeness magnitude (M_c) near the trench axis (region C) decreased significantly from 2.8 to 1.6, so that events of $M \geq 2$ could be detected throughout the entire Tohoku region. Of course, the epicenter near the trench axis (region C) becomes shallower when using OBS, which makes M become systematically smaller. A comparison of the JMA unified catalog data with the automatically determined hypocenters that coincided with each other within 5 s of the origin time and 50 km from the epicenter revealed that the average focal depth with S-net was 8 km shallower than that without, and its use contributed to a 0.3 reduction in magnitude (Additional

file 1: Table S1). Therefore, the reduction of M_c by 1.2 ($=2.8-1.6$) includes not only the effect of the improved detectability with S-net but also that of the change in depth. Also, the earthquake decision rate is more than 94% for $M \geq 1$ throughout the entire Tohoku region (Additional file 1: Table S2). In regions A and D, which are close to the coast, the frequency magnitude distributions of the JMA unified catalog data and automatically determined earthquakes were almost identical and were almost unaffected by S-net. Therefore, it is possible to obtain statistics of seismic activity such as the b -value of the Gutenberg–Richter law (Gutenberg and Richter 1944) continuously for past earthquakes.

The seismic activity and tectonics along the Japan Trench are shown in Fig. 7. In terms of the along-dip direction, the microseismic activity was concentrated within 20–50 km depths along the upper surface of the subducting Pacific Plate (Fig. 7a). On the contrary, many slow earthquakes have been detected at 10–20 km depths



in the upper surface of the subducting Pacific Plate (Nishikawa et al. 2019) with distribution complementary to that of the microearthquakes (Fig. 7b). According to the Japan integrated velocity structure model (JIVSM) developed by the Headquarters for Earthquake Research Promotion and Koketsu et al. (2012), the isodepth contours of 20–30 km in the upper surface of the Pacific Plate roughly correspond to the intersection of the Pacific Plate and the continental Moho.

On the contrary, in the along-strike direction, almost no slow earthquakes or microearthquakes occurred in the large coseismic slip area of the 2011 Tohoku earthquake off the coast of Miyagi Prefecture. We emphasize that few seismic activities are visible in the coseismic slip area, even though the use of S-net enabled us to detect smaller earthquakes in this study (Fig. 6). The surface projection of seismic activities is similar to that of afterslip and aftershocks at the deep extension of the coseismic

slip area revealed from pre- and post-seismic geodetic data (Yamagiwa et al. 2015; Iinuma et al. 2016; Noda et al. 2018; Tomita et al. 2020). This spatial complementarity between the coseismic slip and afterslip may be caused by differences in frictional properties. To understand the physical mechanism of seismic activity complementary to the coseismic slip, additional studies are needed, such as distinguishing between interplate and intraplate events based on focal mechanisms and hypocenter locations (e.g. Hasegawa et al. 2012; Nakamura et al. 2016). However, it would reflect the diversity of interplate slip motion due to spatial heterogeneity of friction and the associated stress changes in the surrounding crust.

Seismic activity also occurred occasionally at depths of 10–20 km off the coast of Iwate and Ibaraki prefectures adjacent to the coseismic slip area (Clusters N, S in Fig. 7a). These clusters are restricted to the fringes of low-velocity anomalies (Fig. 7c; Hua et al. 2020)

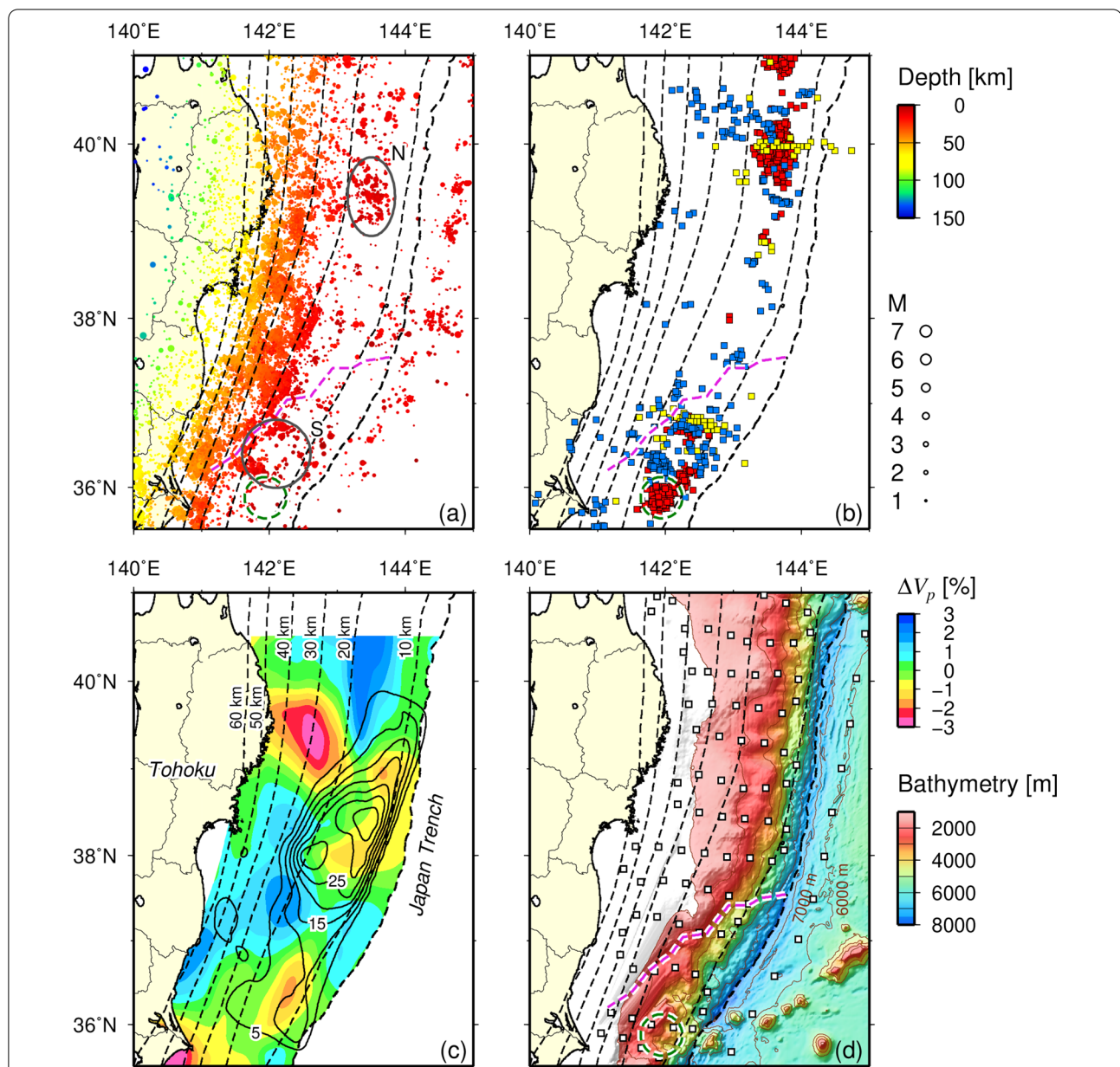


Fig. 7 Seismic activity and tectonics along the Japan Trench. **a** Epicentral distribution of automatically determined epicenters (M all, within 10 km distance from the upper surface of the Pacific Plate). The colors indicate depth. Ellipses N and S indicate clusters of shallow microearthquakes. **b** Slow earthquake distribution determined by Nishikawa et al. (2019). Red color indicates low-frequency tremors of 80 s or longer duration; yellow indicates VLFs; and blue indicates background swarms. **c** P-wave velocity structure perturbation of the top surface of the subducting Pacific Plate determined by Hua et al. (2020). The contour lines show the distribution of coseismic slip determined by Yoshida et al. (2011) for the 2011 Tohoku earthquake. The unit of the contour line is meters. **d** Distribution of S-net stations. Black dashed lines indicate the plate isodepth lines from the JIVSM (Koketsu et al. 2012). The purple dashed line shows the forearc segment boundary according to Bassett et al. (2016). The green dotted circle indicates the location of the subducted seamount according to Mochizuki et al. (2008)

among the background swarms (blue square in Fig. 7b; Nishikawa et al., 2019) that are suggested to be related to slow slip. Maeda (1996) reported that swarm-like seismic activity in these areas tends to precede major earthquakes. Hirose et al. (2020) pointed out that the

seismic activity in these regions tends to trigger a main shock more easily than that expected by the epidemic-type aftershock sequence (ETAS) model, which reflects some physical phenomena (for example, a nucleation process like a precursory slip) that cannot be explained

by mere triggering effects. Therefore, microearthquakes occurring in these regions are considered to reflect changes in interplate coupling caused by slow slips or precursory slips.

Bassett et al. (2016) defined the abrupt change in geological structure of the upper plate off Fukushima and Ibaraki prefectures as the forearc segment boundary (FSB) (the purple dashed line in Fig. 7) on the basis of the residual topography and gravity anomaly. Nishikawa et al. (2019) discussed that the FSB corresponded to the boundary between the coseismic slip area of the Tohoku earthquake and the slow earthquake activity area. On the other hand, the present study did not show a clear contrast in microseismic activity across the FSB.

In the next section, we detail the seismic activity occurring between January and August 2020 for each region. In Additional file 1: S2, we also discussed the

characteristics of $M \geq 5$ earthquakes marked as stars in Figs. 8, 9, 10.

Seismic activity off Aomori and Fukushima prefectures

In areas off Aomori to Fukushima prefectures (Figs. 8 and 10), seismic activity was observed along the plate boundary. It can be clearly seen that the depth of the offshore earthquakes has improved significantly comparing Figs. 10 and Additional file 1: Figure S3. As previously mentioned, seismic activity occurring near the plate boundary between the Pacific Plate and the continental crust was almost completely absent in the shallower areas. However, in cross-section C, seismic activity was also observed at depths of 10–20 km along the plate boundary. Moreover, seismic activity was observed in the toe of the mantle wedge between the plate boundary and the continental Moho in cross-sections C–E. This is referred to as a supraslab earthquake, which Uchida et al.

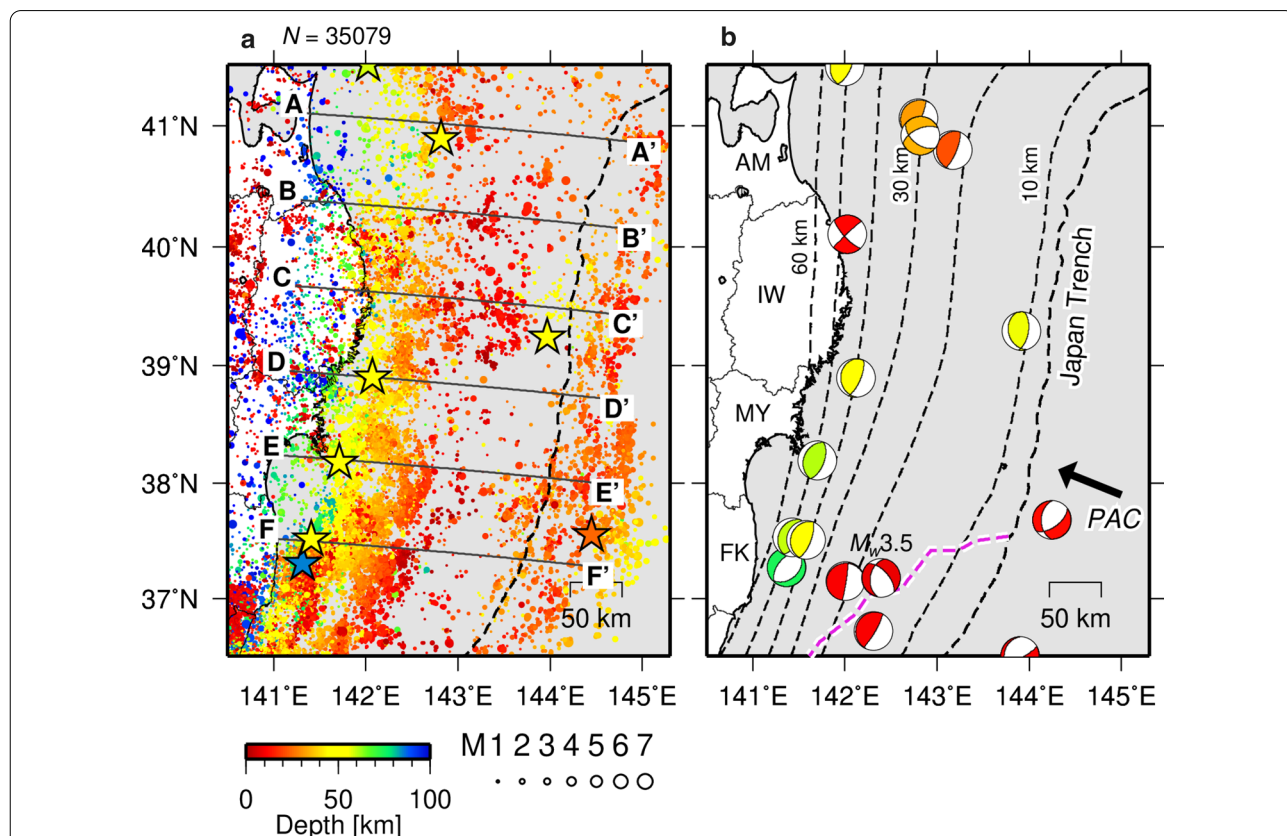
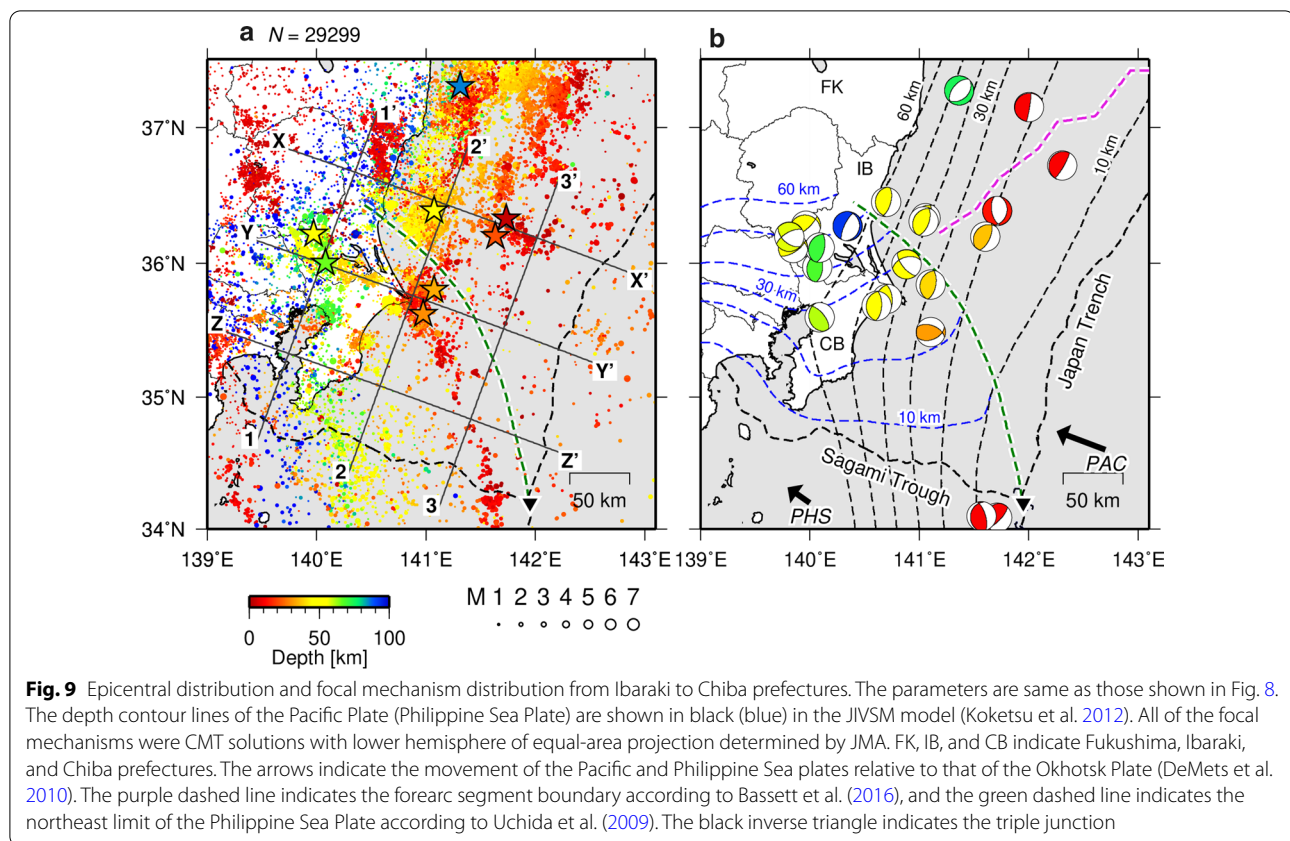


Fig. 8 Epicentral distribution and focal mechanism distribution from Aomori to Fukushima prefectures. The colors indicate depth. **a** Epicentral distribution. We show only the epicenter of events determined using ≥ 10 stations. The automatically determined hypocenters of $M \geq 5$ are indicated by stars. **b** Focal mechanism distribution. The focal mechanisms were CMT solutions with lower hemisphere of equal-area projection determined by JMA, except for the M_w 3.5 event which is a moment tensor solution determined by NIED. AM, IW, MY, and FK indicate Aomori, Iwate, Miyagi, and Fukushima prefectures. The depth contour lines of the Pacific Plate are shown in black in the JIVSM model (Koketsu et al. 2012). The arrows indicate the movement of the Pacific Plate relative to that of the Okhotsk Plate (DeMets et al. 2010). The purple dashed line in the maps indicates the forearc segment boundary according to Bassett et al. (2016)



(2010) interpreted to occur within seamounts detached from the Pacific Plate.

In cross-section F, unlike that in other regions, the hypocenter became shallow upward from about 20 km in depth at the plate boundary (horizontal axis is from -50 km to -20 km in Fig. 10). Some moment tensor solutions by NIED F-net have normal faults, and these could have occurred in the continental crust. Although branching faults from the plate boundary are possible, detailed determination of the hypocenter must be based on three-dimensional velocity structures. Aftershocks of the M 7.4, 25 km depth intra-continental crust earthquake that occurred on November 22, 2016, are also evident in cross-section F, in which the horizontal axis of distance is about -120 km.

Seismic activity from Ibaraki to Chiba prefectures

Off the coast of Chiba Prefecture (Figs. 9 and 10), the Philippine Sea Plate is subducting from the south, and the Pacific Plate is subducting beneath the Philippine Sea Plate, creating complex tectonics. In this study, distinct seismic activity was apparent for each plate when using S-net. Along the northeastern edge of the Philippine Sea Plate shown by Uchida et al. (2009), the microseismic activity is linearly distributed in shallow areas from

the eastern coast of Chiba Prefecture to the triple junction. Ito et al. (2019) interpreted that low-angle thrust fault earthquakes occurring at depths of 20–30 km in this vicinity are associated with the subduction of the Philippine Sea Plate. However, because the depth of the determined microseismic activity is as shallow as 10 km in this study, it may be considered to have occurred in the intra-continental crust. Seismic activity was continually observed in the southern region of the subduction of the Philippine Sea Plate (Sagami Trough). Serpentinized peridotite present in the Philippine Sea Plate has been suggested to have formed prior to the subduction (Kamimura et al. 2002; Nakajima et al. 2009). This activity could be related to the eastern margin of the serpentinized peridotite region. However, the plate depth around the triple junction is unclear, and further research is needed to understand the relationship between the serpentinized peridotite region and the microseismic activity.

Summary

In this study, to understand the microseismic activity occurring along the Japan Trench, we used approximately 27,000 automatically determined hypocenters and their phase data as training data to automatically

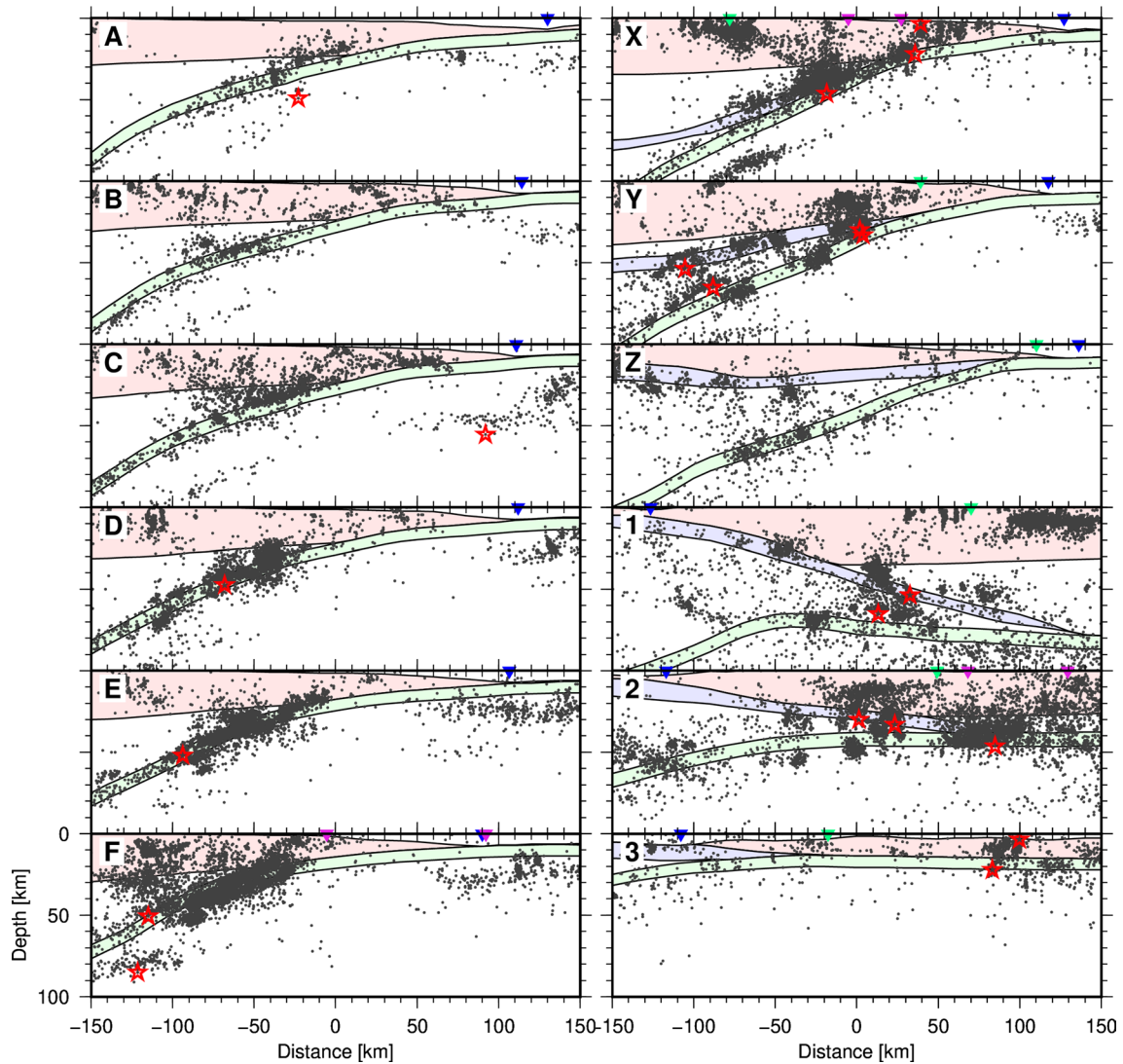


Fig. 10 Cross sections of automatically determined hypocenters. The cross-sectional view shows earthquakes occurring within 40 km of each line in Figs. 8 and 9. The automatically determined hypocenters of $M \geq 5$ events are indicated by stars. The pink area represents the continental crust, green area represents the Pacific Plate crust, and the blue area represents the Philippine Sea Plate crust. The purple inverse triangles indicate the forearc segment boundary according to Bassett et al. (2016), and the green inverse triangles indicate the northeast limit of the Philippine Sea Plate according to Uchida et al. (2009). The blue inverse triangle indicates the trench

discrimination between earthquakes and noise events using random forest and AdaBoost ensemble learning. The results showed that all the methods were able to correctly discriminate between earthquakes and noise with about 96–99% accuracy. In particular, AdaBoost had less than 1% probability of misclassifying earthquakes as noise. This will contribute to the early detection of seismic activity using automatically determined hypocenters and will increase the efficiency of cataloging. In addition, the seismic activity along the Japan Trench was scrutinized on the basis of the obtained hypocenters.

At the shallow plate boundary at depths of 10–20 km, microseismic activity was low, and the distribution was complementary to that of the slow earthquake and large slip areas of the 2011 Tohoku earthquake. The difference in the interplate frictional properties could define the microseismogenic zone, the mega seismogenic zone, or the slow seismogenic zone. Because S-net usage for the JMA unified catalog has been started from September 2020, further accumulation of data is expected to reveal detailed activities such as temporal variation at the shallow part of the plate boundary.

Abbreviations

CART: Classification and regression trees; CC: Cross-correlation coefficient; CMT: Centroid moment tensor; ETAS: Epidemic-type aftershock sequence; FSB: Forearc segment boundary; JIVSM: Japan integrated velocity structure model; JMA: Japan Meteorological Agency; M : Magnitude; M_c : Completeness magnitude; Moho: Mohorovičić discontinuity; M_w : Moment magnitude; MEXT: Ministry of Education, Culture, Sports, Science and Technology; NIED: National Research Institute for Earth Science and Disaster Resilience; OBS: Ocean bottom seismometer; PF: Phase combination forward search; STA/LTA: Short-term average/long-term average; VLFE: Very low-frequency earthquake.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-021-01411-6>.

Additional file 1: Text S1. Air gun waveform removal. **Figure S1.** Frequency distribution of scaled CC. **Figure S2.** Comparison of processing results with and without the CC denoiser. **Table S1.** Residuals between the JMA unified catalog and the corresponding automatically determined hypocenters. **Table S2.** Comparison of the JMA unified catalog and the automatically determined hypocenters. **Text S2.** Seismic activity above $M 5$ along the Japan Trench. **Figure S3.** Cross-sections of the JMA unified catalog obtained without the usage of S-net.

Acknowledgements

We thank Akio Kobayashi for helpful comments. To create the automatic determined hypocenters used in this study, we used the following data to create the seismographs: National Research Institute for Earth Science and Disaster Prevention (NIED), Hokkaido University, Hirosaki University, Tohoku University, the University of Tokyo, Nagoya University, Kyoto University, Kochi University, Kyushu University, Kagoshima University, National Institute of Advanced Industrial Science and Technology (AIST), the Geospatial Information Authority of Japan, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), the Association for the Development of Earthquake Prediction, Aomori Prefecture, the Tokyo Metropolitan Government, Shizuoka Prefecture, the Hot Springs Research Institute of Kanagawa Prefecture, and Incorporated Research Institutions for Seismology (IRIS). We used the JMA unified catalog [<https://www.data.jma.go.jp/svd/eqev/data/bulletin/index.html>]. We used the Python libraries: scikit-learn [<https://scikit-learn.org/stable/>], Boruta [https://github.com/scikit-learn-contrib/boruta_py], and dtreeviz [<https://github.com/parrr/dtreeviz>]. A catalog of low-frequency tremor, very-low-frequency earthquakes, and background seismic swarming along the Japan Trench is available from Dr. Tomoaki Nishikawa's website [<https://sites.google.com/site/tomoakinishikawahomepage/>]. The Japan integrated velocity structure model (JIVSM) is available from the Headquarters for Earthquake Research Promotion website [https://www.jishin.go.jp/evaluation/seismic_hazard_map/lpshm/12_choshuki_dat/]. The velocity structure model of the upper surface of the Pacific slab was provided by Dr. Dapeng Zhao and Yu Katayama. The bathymetric data is from the Japan Coast Guard's Information Utilization Promotion Division / Japan Marine Data Center, Marine Information Department [https://jdoss1.jodc.go.jp/vpage/depth500_file_j.html]. Plate convergence rate data were calculated by Plate Motion Calculator on the web site of UNAVCO [<https://www.unavco.org/>]. We used GMT (Wessel et al. 2013) to create the figure. This work was supported by JSPS KAKENHI Grant Number JP20K14565. We are also grateful to two anonymous reviewers, and Editor Ryosuke Ando for their constructive comments.

Authors' contributions

KT analyzed and wrote most of the manuscript. FH and AN helped with the preparation of the manuscript. YI, KI, and HU contributed to improvement of automatic processing and created the velocity structures. All authors read and approved the final manuscript.

Funding

This work was supported by JSPS KAKENHI Grant Number JP20K14565.

Availability of data and materials

The data used in this article were provided by the Data Management Center of NIED (<http://www.hinet.bosai.go.jp/?LANG=en>) and JMA (http://www.data.jma.go.jp/svd/eqev/data/bulletin/index_e.html).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Author details

¹ Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan. ² Osaka Regional Headquarters, Japan Meteorological Agency, 4-1-76, Otemae, Chuo-ku, Osaka 540-0008, Japan. ³ Seismology and Volcanology Department, Japan Meteorological Agency, 3-6-9 Toranomon, Minato-ku, Tokyo 105-8431, Japan.

Received: 23 December 2020 Accepted: 29 March 2021

Published online: 16 April 2021

References

- Aoi S, Asano Y, Kunugi T, Kimura T, Uehira K, Takahashi N, Ueda H, Shiomi K, Matsumoto T, Fujiwara H (2020) MOWLAS: NIED observation network for earthquake, tsunami and volcano. *Earth Planets Space* 72:126. <https://doi.org/10.1186/s40623-020-01250-x>
- Bassett D, Sandwell DT, Fialko Y, Watts AB (2016) Upper-plate controls on co-seismic slip in the 2011 magnitude 9.0 Tohoku-oki Earthquake. *Nature* 531:92–96. <https://doi.org/10.1038/nature16945>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- DeMets C, Gordon RG, Argus DF (2010) Geologically current plate motions. *Geophys J Int* 181:1–80. <https://doi.org/10.1111/j.1365-246X.2009.04491.x>
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Gamage SSN, Umino N, Hasegawa A, Kirby SH (2009) Offshore double-planned shallow seismic zone in the NE Japan forearc region revealed by sP depth phases recorded by regional networks. *Geophys J Int* 178(1):195–214. <https://doi.org/10.1111/j.1365-246X.2009.04048.x>
- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- Hara S, Fukahata Y, Iio Y (2019) P-wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets Space* 71:127. <https://doi.org/10.1186/s40623-019-1111-x>
- Hasegawa A, Yoshida K, Asano Y, Okada T, Iinuma T, Ito Y (2012) Change in stress field after the 2011 great Tohoku-Oki earthquake. *Earth Planet Sci Lett* 355–356:231–243. <https://doi.org/10.1016/j.epsl.2012.08.042>
- Hirose F, Tamaribuchi K, Maeda K (2020) Does an ETAS model dream of foreshocks?: Through the application of an earthquake forecasting method based on swarm-like activity. In: Abstracts of the 2020 Seismol Soc Jpn Fall meeting, Online, 29–31 Oct 2020.
- Hua Y, Zhao D, Toyokuni G, Xu Y (2020) Tomography of the source zone of the great 2011 Tohoku earthquake. *Nat Commun* 11:1163. <https://doi.org/10.1038/s41467-020-14745-8>
- Iinuma T, Hino R, Uchida N, Nakamura W, Kido M, Osada Y, Miura S (2016) Seafloor observations indicate spatial separation of coseismic and postseismic slips in the 2011 Tohoku earthquake. *Nat Commun* 7:13506. <https://doi.org/10.1038/ncomms13506>
- Ito A, Tonegawa T, Uchida N, Yamamoto Y, Suetsugu D, Hino R, Sugioka H, Obana K, Nakahigashi K, Shinohara M (2019) Configuration and structure of the Philippine Sea Plate off Boso, Japan: constraints on the shallow

- subduction kinematics, seismicity, and slow slip events. *Earth Planets Space* 71:111. <https://doi.org/10.1186/s40623-019-1090-y>
- Kamimura A, Kasahara J, Shinohara M, Hino R, Shiobara H, Fujie G, Kanazawa T (2002) Crustal structure study at the Izu-Bonin subduction zone around 31°N: Implications of serpentinized materials along the subduction plate boundary. *Phys Earth Planet Int* 132:105–129. [https://doi.org/10.1016/S0031-9201\(02\)00047-X](https://doi.org/10.1016/S0031-9201(02)00047-X)
- Koketsu K, Miyake H, Suzuki H (2012) Japan integrated velocity structure model version 1. In: Proceedings of the 15th world conference on earthquake engineering, Lisbon, 24–28 September 2012.
- Kursa MB, Rudnicki WR (2010) Feature selection with the Boruta package. *J Stat Softw* 36(11):1–13. <https://doi.org/10.18637/jss.v036.i11>
- Liu X, Zhao D (2018) Upper and lower plate controls on the great 2011 Tohoku-oki earthquake. *Sci Adv* 4(6):eaat4396. <https://doi.org/10.1126/sciadv.aat4396>
- Maeda K (1996) The use of foreshocks in probabilistic prediction along the Japan and Kuril Trenches. *Bull Seismol Soc Am* 86(1A):242–254
- Miura S, Kodaira S, Nakanishi A, Tsuru T, Takahashi N, Hirata N, Kaneda Y (2003) Structural characteristics controlling the seismicity crustal structure of southern Japan Trench fore-arc region, revealed by ocean bottom seismographic data. *Tectonophysics* 363(1–2):79–102. [https://doi.org/10.1016/S0040-1951\(02\)00655-8](https://doi.org/10.1016/S0040-1951(02)00655-8)
- Mochizuki K, Yamada T, Shinohara M, Yamanaka Y, Kanazawa T (2008) Weak interplate coupling by seamounts and repeating $M \sim 7$ earthquakes. *Science* 321(5893):1194–1197. <https://doi.org/10.1126/science.1160250>
- Nakajima J, Hirose F, Hasegawa A (2009) Seismotectonics beneath the Tokyo metropolitan area, Japan: effect of slab-slab contact and overlap on seismicity. *J Geophys Res Solid Earth* 114:B08309. <https://doi.org/10.1029/2008JB006101>
- Nakamura W, Uchida N, Matsuzawa T (2016) Spatial distribution of the faulting types of small earthquakes around the 2011 Tohoku-oki earthquake: a comprehensive search using template events. *J Geophys Res Solid Earth* 121:2591–2607. <https://doi.org/10.1002/2015JB012584>
- Nishikawa T, Matsuzawa T, Ohta K, Uchida N, Nishimura T, Ide S (2019) The slow earthquake spectrum in the Japan Trench illuminated by the S-net seafloor observatories. *Science* 365(6455):808–813. <https://doi.org/10.1126/science.aax5618>
- Noda A, Takahama T, Kawasato T, Matsu'ura M, (2018) Interpretation of offshore crustal movements following the 2011 Tohoku-oki earthquake by the combined effect of afterslip and viscoelastic stress relaxation. *Pure Appl Geophys* 175:559–572. <https://doi.org/10.1007/s00024-017-1682-z>
- Obana K, Fujie G, Takahashi T, Yamamoto Y, Nakamura Y, Kodaira S, Takahashi N, Kaneda Y, Shinohara M (2012) Normal-faulting earthquakes beneath the outer slope of the Japan Trench after the 2011 Tohoku earthquake: implications for the stress regime in the incoming Pacific plate. *Geophys Res Lett* 39:L00G24. <https://doi.org/10.1029/2011GL050399>
- Okada Y, Kasahara K, Hori S, Obara K, Sekiguchi S, Fujiwara H, Yamamoto A (2004) Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiK-net. *Earth Planets Space* 56(8):xv–xxviii. <https://doi.org/10.1186/BF03353076>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Perol T, Gharbi M, Denolle M (2018) Convolutional neural network for earthquake detection and location. *Sci Adv* 4(2):e1700578. <https://doi.org/10.1126/sciadv.1700578>
- Ross ZE, Meier MA, Hauksson E, Heaton TH (2018) Generalized seismic phase detection with deep learning. *Bull Seismol Soc Am* 108(5A):2894–2901. <https://doi.org/10.1785/0120180080>
- Tamaribuchi K (2018) Evaluation of automatic hypocenter determination in the JMA unified catalog. *Earth Planets Space* 70:141. <https://doi.org/10.1186/s40623-018-0915-4>
- Tomita F, Iinuma T, Ohta Y, Hino R, Kido M, Uchida N (2020) Improvement on spatial resolution of a coseismic slip distribution using postseismic geodetic data through a viscoelastic inversion. *Earth Planets Space* 72:84. <https://doi.org/10.1186/s40623-020-01207-0>
- Tsuru T, Park JO, Miura S, Kodaira S, Kido Y, Hayashi T (2002) Along-arc structural variation of the plate boundary at the Japan Trench margin: implication of interplate coupling. *J Geophys Res Solid Earth* 107(B12):2357. <https://doi.org/10.1029/2001jb001664>
- Uchida N, Nakajima J, Hasegawa A, Matsuzawa T (2009) What controls interplate coupling?: evidence for abrupt change in coupling across a border between two overlying plates in the NE Japan subduction zone. *Earth Planet Sci Lett* 283(1–4):111–121. <https://doi.org/10.1016/j.epsl.2009.04.003>
- Uchida N, Kirby SH, Okada T, Hino R, Hasegawa A (2010) Supraslab earthquake clusters above the subduction plate boundary offshore Sanriku, north-eastern Japan: Seismogenesis in a graveyard of detached seamounts? *J Geophys Res Solid Earth* 115:B09308. <https://doi.org/10.1029/2009JB006797>
- Ueno H, Ueda M, Misu H, Yokoyana H, Iwasaki Y, Iwakiri K (2019) Improvement of the accuracy of the unified Japan earthquake catalog using ocean bottom seismographs. In: Abstracts of JpGU2019, Makuhari Messe, Chiba, 26–30 May 2019.
- Wessel P, Smith WHF, Scharroo R, Luis J, Wobbe F (2013) Generic Mapping Tools: improved version released. *Eos Trans AGU* 94(45):409–410. <https://doi.org/10.1002/2013EO450001>
- Woessner J, Wiemer S (2005) Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bull Seismol Soc Am* 95(2):684–698. <https://doi.org/10.1785/0120040007>
- Yamagiwa S, Miyazaki S, Hirahara K, Fukahata Y (2015) Afterslip and viscoelastic relaxation following the 2011 Tohoku-oki earthquake (M_w 9.0) inferred from inland GPS and seafloor GPS/Acoustic data. *Geophys Res Lett* 42:66–73. <https://doi.org/10.1002/2014GL061735>
- Yoshida Y, Ueno H, Muto D, Aoki S (2011) Source process of the 2011 off the Pacific coast of Tohoku Earthquake with the combination of teleseismic and strong motion data. *Earth Planets Space* 63:10. <https://doi.org/10.5047/eps.2011.05.011>
- Zhao D, Huang Z, Umino N, Hasegawa A, Kanamori H (2011) Structural heterogeneity in the megathrust zone and mechanism of the 2011 Tohoku-oki earthquake (M_w 9.0). *Geophys Res Lett* 38:L17308. <https://doi.org/10.1029/2011GL048408>
- Zhu W, Beroza GC (2019) PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophys J Int* 216(1):261–273. <https://doi.org/10.1093/gji/ggy423>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.