

FULL PAPER

Open Access



Detection of hidden earthquakes after the 2011 Tohoku earthquake by automatic hypocenter determination combined with machine learning

Koji Tamaribuchi^{1*} , Shota Kudo², Kengo Shimojo¹ and Fuyuki Hirose¹

Abstract

After the 2011 M_w 9.0 Tohoku earthquake, seismicity became extremely active throughout Japan. Despite enormous efforts to detect the large number of earthquakes, microearthquakes ($M < 2$ inland, $M < 3$ offshore) were not always cataloged and many have remained undetected, making it difficult to understand the detailed seismicity after the 2011 Tohoku earthquake. We developed an automatic hypocenter determination method combined with machine learning to detect microearthquakes. Machine learning was used for phase classification with convolutional neural networks and ensemble learning to remove false detections. We detected > 920,000 earthquakes from March 2011 to February 2012, triple the number of the conventional earthquake catalog (~ 320,000). This represents a great improvement in earthquake detection, especially in and around the Tohoku region. Detailed analysis of our merged catalog more clearly revealed features such as (1) swarm migrations, (2) small foreshock activity, and (3) increased microseismicity preceding repeating earthquakes. This microseismic catalog provides a magnifying glass for understanding detailed seismicity.

Keywords Automatic hypocenter determination, Completeness magnitude, JMA unified earthquake catalog, Machine learning, The 2011 Tohoku earthquake

*Correspondence:

Koji Tamaribuchi

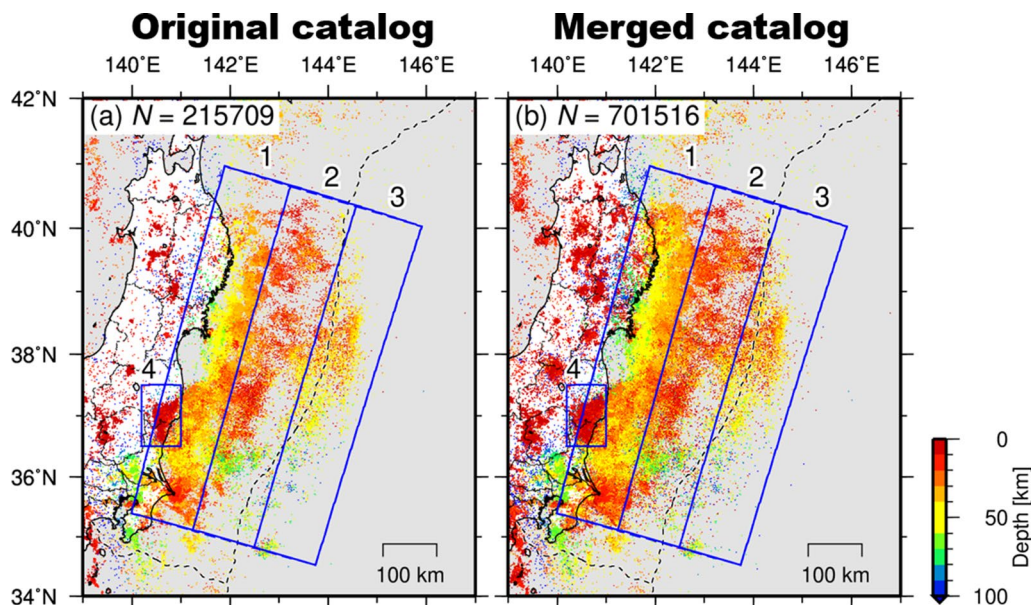
ktamarib@mri-jma.go.jp

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



© The Author(s) 2023, corrected publication 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Graphical Abstract



Introduction

Earthquake catalogs record hypocenter information such as the earthquake origin time, location, and magnitude (M). These extremely important datasets are used in many areas of seismology and disaster mitigation, including for understanding the physics of earthquakes from their fault geometry and nucleation processes, monitoring interplate slip, and statistically estimating the probability of foreshocks and aftershocks (Arrowsmith et al. 2022). Performing these analyses at high resolution requires many hypocenters.

The earthquake catalog in Japan (hereafter, the JMA unified catalog) has been compiled by the Japan Meteorological Agency (JMA) in cooperation with the Ministry of Education, Culture, Sports, Science and Technology through visual inspection of seismic waveforms by experts. Tamaribuchi et al. (2016) and Tamaribuchi (2018) developed the phase combination forward search (PF) method to automatically determine hypocenters for the JMA unified catalog. The PF method, which can efficiently determine hypocenters even during high seismicity, was put into operation in JMA's seismic monitoring system on 1 April 2016. The monitoring system achieved remarkable results during the 2016 Kumamoto earthquake sequence that started on 14 April 2016, automatically determining around 70,000 hypocenters over two months, and it has contributed to real-time seismic monitoring.

In retrospect, however, many small earthquakes remain undetected because of limited human resources. In particular, the 2011 M_w 9.0 off the Pacific coast of Tohoku Earthquake (hereafter, the 2011 Tohoku earthquake) caused many earthquakes over a wide area (Hirose et al. 2011). The magnitude of completeness (M_c , i.e., the minimum magnitude at which earthquakes are completely detected, which depends on the location of the stations and the seismicity) along the Tohoku coast is $M_c \approx 1$ under normal conditions, but increased to $M_c \geq 2-3$ after the 2011 Tohoku earthquake (Tamaribuchi 2018). Therefore, the details of the widespread microseismicity following the 2011 Tohoku earthquake remain unclear.

Tamaribuchi and Nakagawa (2020) attempted to apply the PF method to seismic waveforms during March 2011 to detect small earthquakes after the 2011 Tohoku earthquake. They detected twice as many earthquakes as recorded in the conventional JMA unified catalog, but up to 10% of the automatically determined hypocenters were false detections due to noise. Of course, the rate of false detections depends on the noise environment and the seismicity. Furthermore, correct phase picking of P- and S-waves is important because falsely picked phases generally require additional computational time for hypocenter estimation.

Traditionally, many statistical approaches have been used for seismic wave detection and phase picking

(Allen 1978; Baillard et al. 2013; Nakamura et al. 2007; Saragiotis et al. 2002; Yokota et al. 1981). However, statistical approaches can also detect non-stationary noise that needs to be eliminated after phase picking, which requires feature engineering. Later, Gibbons and Ringdal (2006) developed a template-matching technique that uses the seismic waveform itself as a template to search for similar waveforms. For example, Ross et al. (2019a) used template-matching to detect 1.81 million earthquakes in Southern California from 2008 to 2017, successfully lowering M_c from 1.7 to 0.3. Machine learning, including deep learning, has also been applied to seismology, and many deep learning approaches have been proposed for phase picking (Feng et al. 2022; Mousavi et al. 2020; Ross et al. 2018a, b; Yano et al. 2021; Zhu and Beroza 2018). Machine learning methods have also advanced fundamental techniques required for earthquake monitoring, such as phase association, which compiles the phase picks for each earthquake (Ross et al. 2019b; Zhang et al. 2019; Zhu et al. 2022). These methods are similar to the PF method in that they search for the best combination from a large number of phase picks, and each has been successful. For a summary of applications of machine learning to earthquake monitoring, see the review by Mousavi and Beroza (2023).

Deep learning approaches have also been applied retrospectively to earthquake catalogs. Tan et al. (2021) applied deep learning phase picking and phase association to 1-year waveforms following the 2016 Amatrice earthquake in Italy to determine ~900,000 earthquakes, which lowered M_c from 2.2 to 0.5 (Beroza et al. 2021). Yoon et al. (2023) also used an automated workflow based on deep learning to detect ~180,000 earthquakes in Southwestern Puerto Rico from late 2019 to 2023, successfully adding ~7 times as many events as the conventional catalog.

Using state-of-the-art technology to reanalyze earthquakes previously undetected by experts is extremely important for understanding microseismicity. High-resolution catalogs can provide insights into important physical processes such as dynamic and static earthquake triggering and nucleation processes. For more information on big-data-driven advances in seismology, see, for example, the review by Arrowsmith et al. (2022).

Understanding how seismicity changed after the 2011 Tohoku earthquake is an extremely important and urgent issue when considering large earthquake processes and the risks of aftershocks and induced earthquakes along the Nankai Trough and the Kuril Trench, which are feared to occur in the near future. Therefore, we here incorporated machine learning into the PF method to further improve the accuracy of automatic hypocenter

determinations. The objective of this study was to clarify the more detailed seismicity after the 2011 Tohoku earthquake by detecting hidden earthquakes using a fully automated process.

Data and methods

Dataset

For our analysis, we used seismic waveforms from around 1,400 stations throughout Japan (Fig. 1) and recorded during the year from 00:00 on 1 March 2011 to 24:00 on 29 February 2012 (all times in this article are reported in Japan Standard Time, JST = UTC + 9 h). Most of the seismometers were high-sensitivity, three-component velocity meters sampling at 100 Hz, which were also used for the JMA unified catalog. After the 11 March 2011 Tohoku earthquake, the widespread power outage in the Tohoku region rendered the observation network within an epicentral distance of about 400 km temporarily unavailable. The period of unavailability depends on the station. At 14:00 on 11 March, just before the Tohoku earthquake, the number of available stations was 1374, while at 17:00 on 11 March, it was 1129, with more than 200 stations temporarily unavailable. The number of available stations recovered to 1181 at 17:00 on 12 March, 1255 at 17:00 on

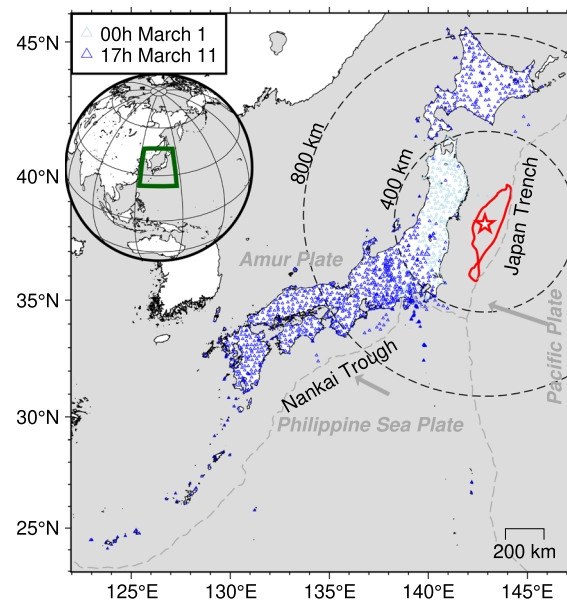


Fig. 1 Tectonic setting. Triangles indicate the seismic observation network; stations in light blue were available at 00:00 JST on 1 March 2011, and stations overlaid by dark blue are those that were still available at 17:00 on 11 March, when much of the seismic network was offline. The red star indicates the epicenter of the M_w 9.0 Tohoku earthquake, and the red contour indicates the area of >10 m of coseismic slip according to Yoshida et al. (2011). Bold gray arrows indicate the movement of the Pacific Plate and the Philippine Sea Plate relative to the Amur Plate (DeMets et al. 2010)

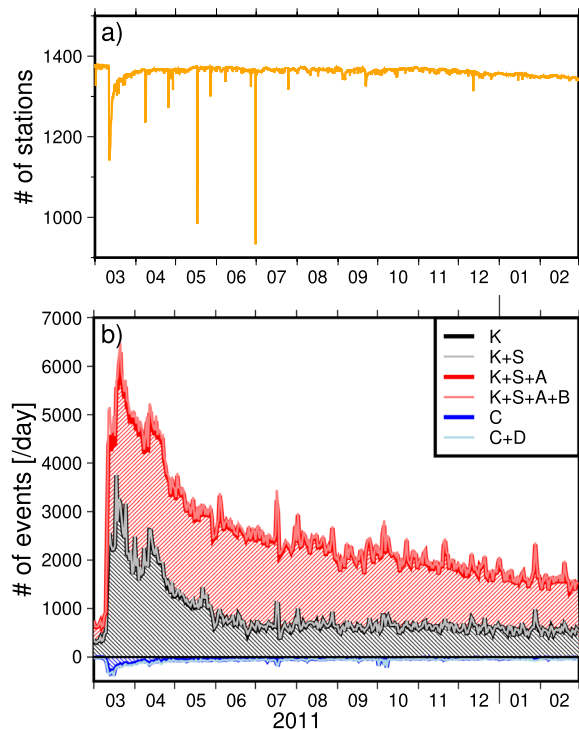


Fig. 2 Number of automatically determined hypocenters per day during the study period. **a** Number of stations available, and **b** number of events for each label. Labels K and S are quality labels from the JMA unified catalog and labels A–D are quality labels from this study (see “Quality control label classifier”); only labels K, S, A, and B are included in our merged catalog

13 March, and 1300 at 17:00 on 14 March. The number of available stations is shown in Fig. 2a.

Method overview

Our analytical procedure is illustrated in Fig. 3. The automatic hypocenter determination (Fig. 3a) consisted of (1) phase picking, (2) phase association, and (3) hypocenter calculation. (1) Phase picking was performed using conventional phase pickers such as the Autoregressive-Akaike’s Information Criterion (AR-AIC) method (Yokota et al. 1981). We then used a convolutional neural network (CNN) to discriminate picks based on the waveforms around the arrival times (hereafter phase classifier; see “Phase classifier” section). (2) Phase association was used to search for the optimal combination of phase picks using the importance sampling algorithm. For details on phase picking and phase association in the PF method, see Tamaribuchi (2018) and Tamaribuchi et al. (2021). (3) Hypocenters were then calculated using the same method used for the JMA unified catalog (Ueno et al. 2002; Hamada et al. 1983).

The automatically determined hypocenters were further processed for (4) blast removal, (5) quality control,

and (6) comparison with the JMA unified catalog, and then they were merged with the JMA unified catalog (Fig. 3b). (4) Because the automatically determined hypocenters included blasts, the blasts were identified by waveform correlation and removed (“Removal of blasts” section). (5) For quality control, each hypocenter was assigned a quality label (A–D, in order of decreasing quality) based on various hypocentral and phase picking features (e.g., hypocentral error, the number of phase picks) using a light gradient-boosting machine (LightGBM) algorithm (“Quality control label classifier” section). (6) Finally, we compared our catalog of automatically determined hypocenters with the JMA unified catalog, and added the unmatched automatically determined hypocenters to the JMA unified catalog to create a merged catalog (“Creation of the merged catalog” section).

Phase classifier

After phase picking, we fed 4-s-duration (i.e., 2 s before and after the P and S picks) three-component seismic waveforms to a convolutional neural network (CNN) to obtain the probabilities that each pick was a P-wave, S-wave, or noise using the generalized phase detection model architecture of Ross et al. (2018b). The specific model used in this study was trained on approximately 900,000 waveforms by Kudo et al. (2023). The waveforms were high-pass-filtered at 2 Hz; if any of the P-wave, S-wave, or noise probabilities were above the threshold probability ($p_{th}=0.9$, see following paragraphs), the other phase types detected within ± 2 s were removed. That is, if the P-wave probability was greater than 0.9, the S-wave phase picks within ± 2 s were discarded. Similarly, if the S-wave probability was greater than 0.9, the P-wave phase picks within ± 2 s were discarded, and if the noise probability was greater than 0.9, the P- and S-wave phase picks within ± 2 s were discarded. This process is called the phase classifier. Although the CNN can be used for phase picking, as originally proposed by Ross et al. (2018b), a preliminary study by Kudo et al. (2023) on Japanese continuous waveforms showed an increase in both false negatives and false positives compared to the conventional picker. Therefore, we used the CNN as the phase classifier in this study.

To evaluate the performance of the phase classifier under normal and swarm conditions, we applied different p_{th} values to 12-h waveforms under normal (00:00–12:00 on 1 March 2011) and swarm conditions (12:00–24:00 on 11 March 2011), respectively. Figure 4a–d shows an example over a 3-min period under swarm conditions. The waveforms at station ODAWA2 near the epicenter (Fig. 4b) show that when $p_{th}=1.0$ (i.e., without the phase classifier), the P- and S- waves were incorrectly read

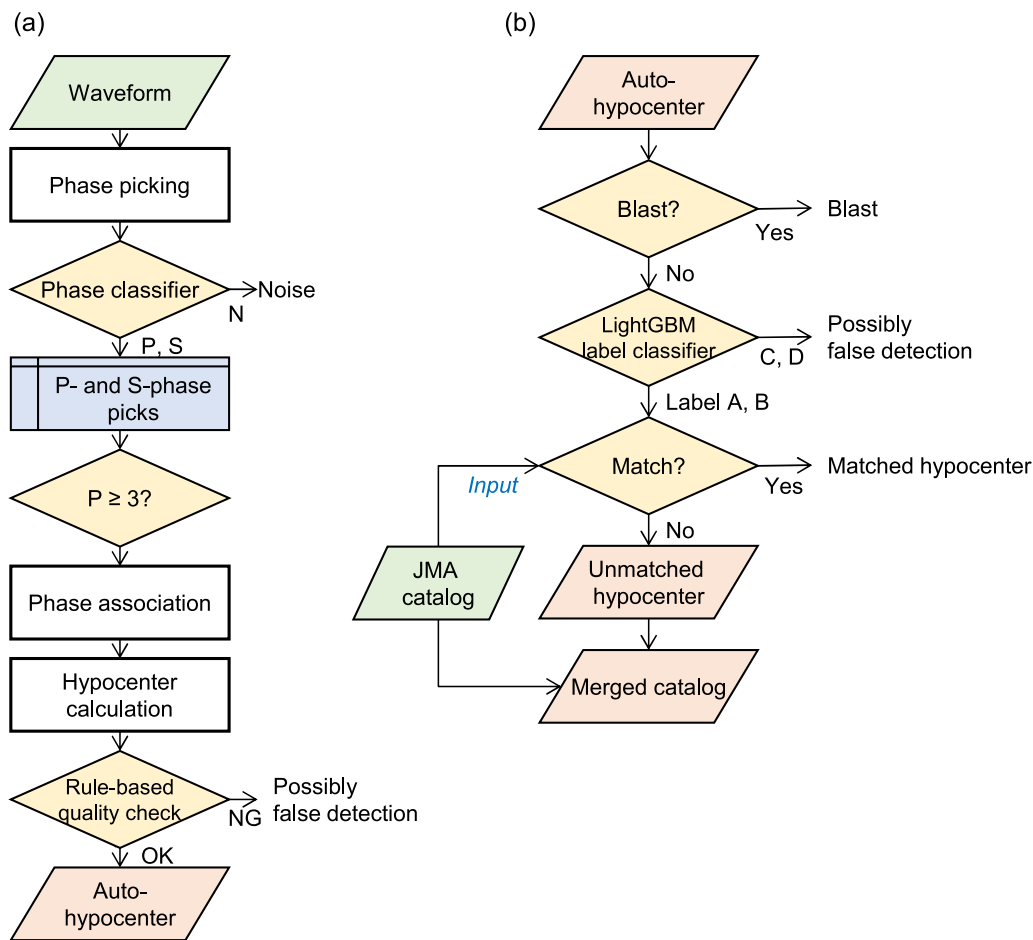


Fig. 3 Flowcharts of our automatic hypocenter determination and catalog merging procedures. **a** Hypocenter determination. **b** Procedure for creating the merged catalog

simultaneously upon arrival of the P-wave, whereas when $p_{th}=0.9$, the P- and S-waves were appropriately read separately. In the case of the conventional picker without the phase classifier, the AR-AIC method simply detected discontinuities in the seismic waveform, which was insufficient to classify P- and S-waves (Tamaribuchi 2018). Figure 4c, d show the same comparison for the other stations as shown in Fig. 4a.

We also compared our automatically determined hypocenters with those in the JMA unified catalog. Figure 4e, f show the number of phase picks and the precision under normal and swarm conditions, respectively, where the precision is the ratio of the number of automatically determined hypocenters that matched the JMA unified catalog to the total number of automatically determined hypocenters. With decreasing p_{th} under both conditions, the number of detections decreased significantly, but the precision increased until $p_{th}=0.90$. This result indicates that the phase classifier removed noise and discriminated P- and S-waves adequately. Because the highest precision

was obtained at $p_{th}=0.90$ under both normal and swarm conditions, we set to $p_{th}=0.90$. Compared to $p_{th}=1.0$ under both conditions, the number of automatically determined hypocenters that matched the JMA unified catalog remained almost the same, but the total number of P- and S-wave picks was reduced by about 30%. The reason for the low precision (Fig. 4e, f, up to ~40%) is that the automatically determined hypocenters contained many true microearthquakes that are not included in the JMA unified catalog (discussed in “Results” section).

Removal of blasts

In general, automatically determined hypocenters include blast events. Therefore, we performed a waveform correlation analysis using 1655 blast waveforms collected by JMA during 2017–2021 as templates to remove events with similar waveforms. First, we extracted automatically determined hypocenters within 5 km of the known blast sites (Fig. 5a). We calculated cross-correlation coefficients between the blast templates and the

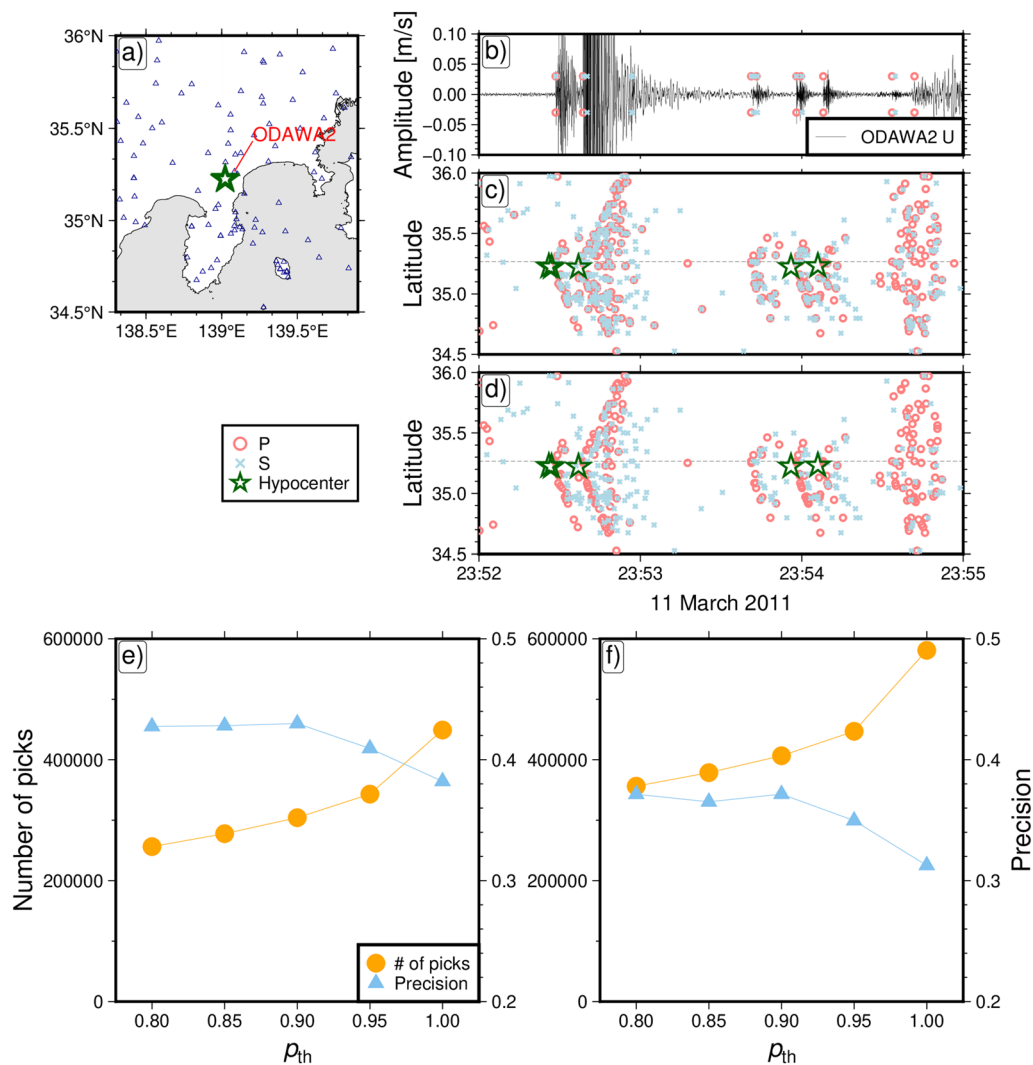


Fig. 4 Example of the phase classifier. **a** Station distribution (blue triangles). Green stars are epicenters from the JMA unified catalog during 23:52–23:55 on 11 March 2011. **b** The waveform of the unfiltered vertical component at station ODAWA2. Symbols plotted above the waveform are pick times when $\rho_{th} = 1.0$, and those below are pick times when $\rho_{th} = 0.9$. Circles and crosses are P- and S-wave picks, respectively. **c, d** P- and S-phase picks when $\rho_{th} = 1.0$ and 0.9, respectively. The dashed lines indicate the latitude of station OADAWA2. **e, f** Total number of picks and precision (i.e., the percentage of automatically determined hypocenters that coincide with the JMA unified catalog) during 00:00–12:00 on 1 March 2011 and 12:00–24:00 on 11 March 2011, respectively

automatically determined hypocenters, and hypocenters with correlation coefficients ≥ 0.5 at three or more stations were considered blast candidates. A 2–8 Hz band-pass filter was applied to the waveforms, the waveform window for calculating the correlation coefficient was 3 s, and the search range for the maximum correlation coefficient was within ± 2 s of the P- and S-wave arrival times.

Although the extracted blast candidates were highly correlated with the blast templates, they sometimes contained true earthquake clusters. Therefore, we plotted a time–frequency histogram of blast candidates for each template; blast candidates observed throughout the day

and night were considered to be true seismicity, whereas those observed only during the day were considered to be artificial blasts and were removed from the automatically determined hypocenters. Examples of temporally biased and unbiased candidates are shown in Fig. 5b, c, respectively. For example, template index #107 was a blast in central Oita Prefecture and was removed because the activity was concentrated during morning working hours. In contrast, template index #1598 was not actually a blast in Aizu, Fukushima Prefecture, and was kept because it detected a true earthquake cluster occurring in the vicinity of a blast template site.

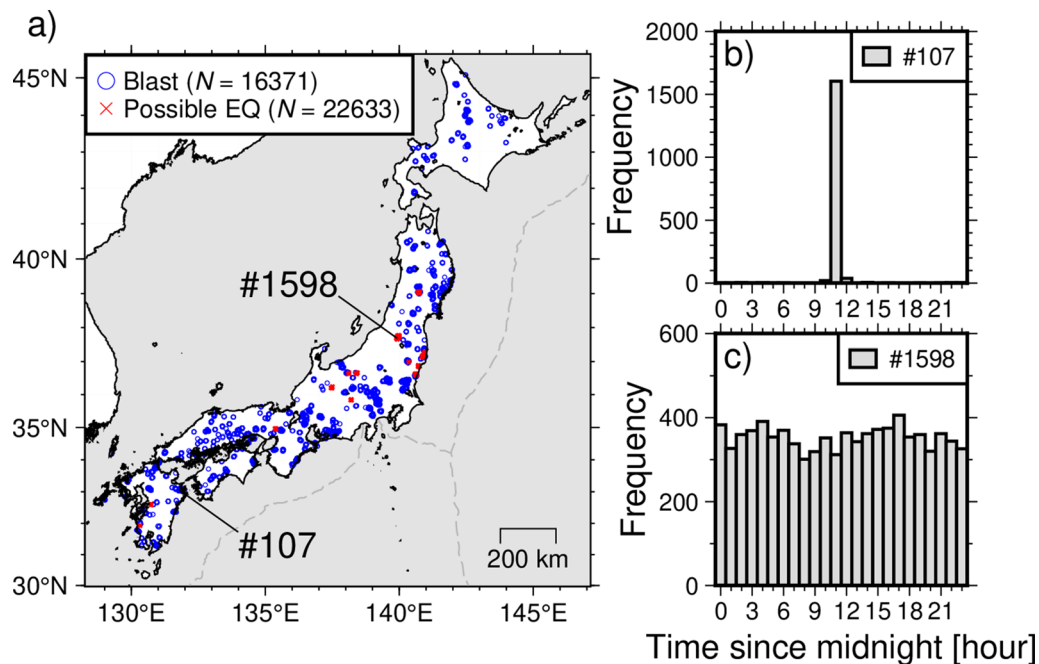


Fig. 5 Distribution of blast templates. **a** Blue circles indicate blasts, and red crosses indicate possible earthquakes that were detected by cross-correlation but are not considered to be blasts. # is the template number. **b** Histogram of events with high cross-correlation with template #107; these were considered blast events in central Oita Prefecture and removed because the activity was concentrated during morning working hours. **c** Histogram of events with high cross-correlation with template #1598 in Aizu, Fukushima Prefecture; these events were not removed from the catalog because they did not vary with time of day and because a nearby earthquake swarm was detected

Overall, 16,371 of 927,899 automatically determined hypocenters from March 2011 to February 2012 were determined to be blasts and removed, leaving 911,528 hypocenters for further processing. Based on our time–frequency analysis, 22,633 events were highly correlated with the blast templates, but were determined to be earthquake clusters.

Quality control label classifier

After removing blasts, we further eliminated characteristic false detections and mislocations within the automatically determined hypocenters. Tamaribuchi et al. (2021) proposed a supervised ensemble learning method to discriminate between earthquakes and noises based on the hypocenter and its phase data determined at 20 stations near the epicenter. In addition to their recommended input phase features, we added maximum amplitudes, the period of the maximum amplitude, and the time of the maximum amplitude after the arrival time at each station (Additional file 1: Table S1). We also checked the status of available stations every hour and selected 20 stations in the vicinity of the epicenter, excluding any stations that were unavailable.

We used the automatically determined earthquake catalog (36,806 events from 1–16 March 2011) developed by Tamaribuchi and Nakagawa (2020) as training

data. We considered automatically determined hypocenters within 5 s of the origin time and 50 km of a hypocenter in the JMA unified catalog to be matched, resulting in 12,982 matched hypocenters. The remaining 23,824 hypocenters were visually checked and given one of four labels: A, earthquakes with generally correct picks (12,633); B, earthquakes with one or two clear errors in the picks (4785); C, earthquakes with three or more errors in the picks or one earthquake erroneously separated into several events (3843); and D, false detections caused by noise (2563). Refer to Fig. 2 in Tamaribuchi et al. (2021) for the corresponding waveform examples mentioned in each label. That is, 7% of the hypocenters automatically determined by Tamaribuchi and Nakagawa (2020) contained noise during this period. The 12,982 matched earthquakes were also assigned label A. Figure 6a–d shows the distribution of the training data. This dataset was then randomly separated in a 6:2:2 ratio for training, validation, and testing, respectively. We used LightGBM (Ke et al. 2017) because it had the highest accuracy after a preliminary comparison of several models in the Pycaret library (Ali 2020). LightGBM is a robust and efficient gradient boosting framework. It employs tree-based models to handle small and large datasets promptly, making it popular for tasks such as classification and regression.

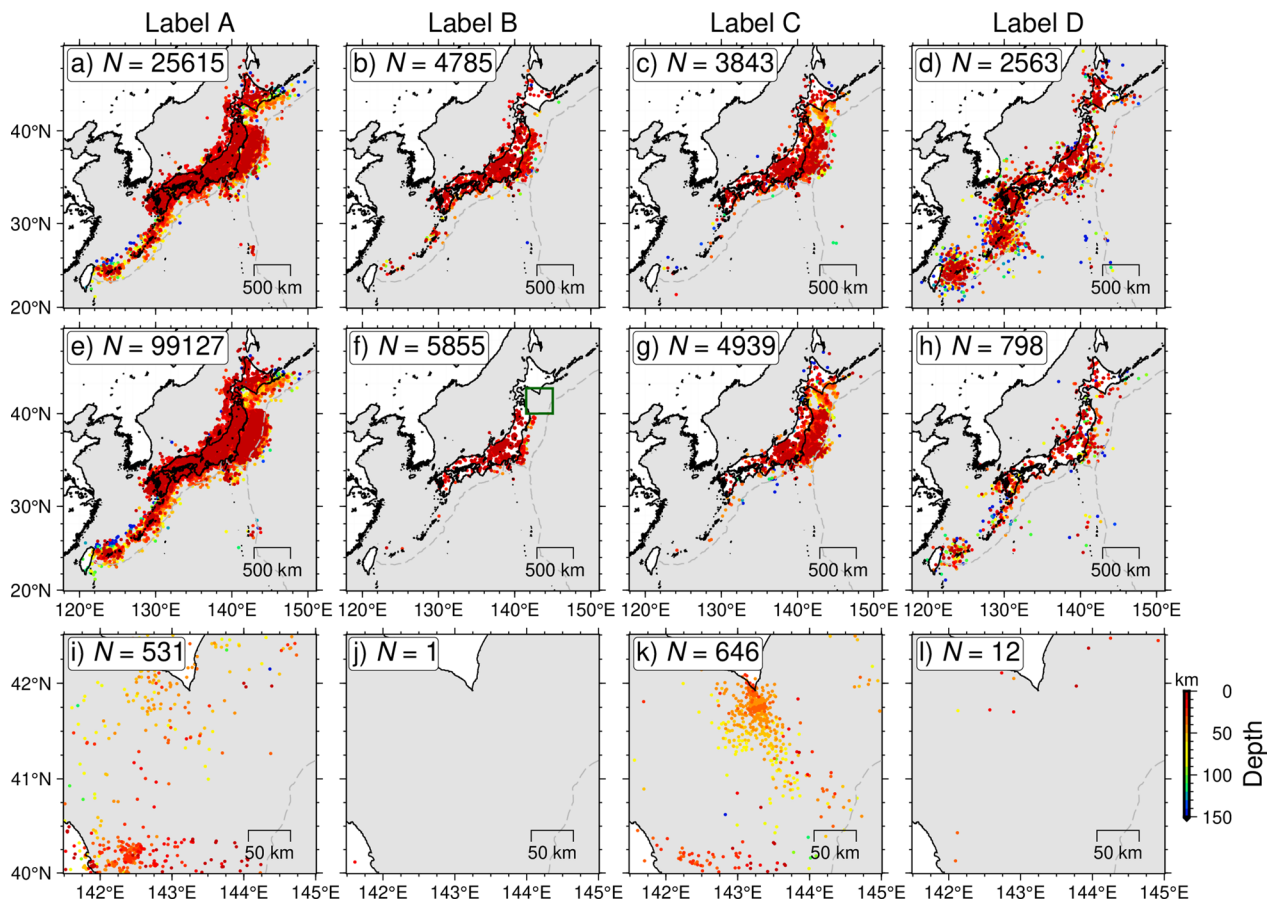


Fig. 6 Epicentral distributions for quality labels (a, e, i) A, (b, f, j) B, (c, g, k) C, and (d, h, l) D. Symbol colors indicate depth. a–d Training dataset, i.e., automatically determined hypocenters from Tamaribuchi and Nakagawa (2020) that we visually classified into each label. e–l Automatically classified hypocenters using LightGBM label classifier in this study during the period 1–31 March 2011. i–l An enlarged view of the area shown by the green box in (f)

Table 1 Confusion matrix of the classification results using our LightGBM quality control classifier

Observation	Prediction				Precision	Recall	F1-score	Accuracy
	A	B	C	D				
Label A	7280	231	120	68	0.859	0.946	0.900	0.819
Label B	724	570	109	6	0.557	0.405	0.469	
Label C	357	203	595	23	0.703	0.505	0.588	
Label D	112	19	22	603	0.861	0.798	0.828	

Its primary advantages include velocity, low memory consumption and high accuracy.

Table 1 shows the confusion matrix and performance evaluation metrics for the test data subset. We evaluated model performance based on accuracy, precision, recall, and the F1-score. Accuracy is the proportion of correctly predicted labels among all predictions, calculated for all labels as $(TP + TN) / (TP + FP + TN + FN)$, where TP represents true positives (i.e., the model correctly predicted

the label), TN true negatives (correctly predicted other labels), FP false positives (incorrectly predicted the label), and FN false negatives (incorrectly predicted other labels). Precision represents the proportion of correctly predicted labels among all predicted labels, calculated for a given label (or combination of labels) as $TP / (TP + FP)$. Recall represents the proportion of correctly predicted labels among all true labels, calculated as $TP / (TP + FN)$. The F1-score represents the harmonic mean of precision

and recall, calculated as $2 \text{ (Precision} \times \text{Recall)} / (\text{Precision} + \text{Recall})$. Overall accuracy was 81.9%. Although labels A and D had F1-scores of around 0.8–0.9, labels B and C had lower values of around 0.5, indicating that, even if there were some errors in the phase picks, they did not have a significant impact on the hypocenter calculation. Because we focused on noise elimination, accuracy was 97.7% when the binary label classification of A–C vs. D was considered, which is almost equivalent to the conventional noise classifier proposed by Tamaribuchi et al. (2021).

Figure 6e–h shows the classification results when this classifier was applied to our refined automatically determined hypocenter dataset during March 2011. Most of the automatically determined hypocenters were classified as label A, whereas those in island areas were mostly classified as label D. Figure 6i–l shows an enlarged image of the off-Tokachi region, where hypocenters were most often assigned label C. This result is due to incorrect phase association. The phase picks of an earthquake that occurred off the coast of the Tohoku region were associated with multiple earthquakes, and one of the hypocenters was incorrectly determined to be off Tokachi. Therefore, labels C and D were excluded from further consideration to eliminate such characteristic false determinations.

Creation of the merged catalog

We created our merged catalog by merging the JMA unified catalog with our refined automatically determined hypocenters. Upon comparison of the two catalogs, hypocenters with origin times and epicentral locations that differed by ≤ 5 s and ≤ 50 km, respectively, were considered to be coincident. If two hypocenters were coincident, we adopted the one in the JMA unified catalog. If there were two or more automatically determined hypocenters within 5 s and 50 km of the hypocenter in the JMA unified catalog, the one with the shortest origin time difference was considered to be coincident. This means that if there are several automatically determined hypocenters within 5 s, one is coincident and the rest are not. Finally, automatically determined hypocenters with labeled A or B that did not match any hypocenter in the JMA unified catalog were considered to be earthquakes and merged with the JMA unified catalog.

Results

Number of earthquakes

When we applied our analysis to seismic waveforms recorded from 1 March 2011 to 29 February 2012, 927,899 hypocenters were automatically determined. We eliminated 16,371 events identified as blasts by waveform correlation, and 21,556 and 7,323 events classified

as having quality control labels C and D, respectively. Among the remaining earthquakes with quality control labels A and B, 583,561 and 18,892 hypocenters, respectively, did not match the JMA unified catalog, i.e., they were not already listed in the JMA unified catalog. These were merged with 320,427 events in the JMA unified catalog (labels K, S, excluding low-frequency earthquakes) to obtain a total of 922,880 events in the merged catalog. Here, quality labels K or S are assigned for high or low accuracy, respectively, in the JMA unified catalog. The obtained 922,880 events in our merged catalog (labels K, S, A, and B) were roughly triple the number in the JMA unified catalog alone (320,427 events, labels K and S, excluding low-frequency earthquakes). Figure 2 presents the daily numbers of earthquakes and available stations. Seismic network outages in much of the Tohoku region immediately after 11 March 2011 reduced seismic detectability for a few days. The number of available stations also temporarily dropped to around 1000 twice after April 2011, but that effect was only one to several hours and thus had only minor impacts on the overall catalog; therefore, we will not go into further detail here.

The daily number of earthquakes in the merged catalog (labels K, S, A and B in Fig. 2b) gradually decreased, whereas that in the JMA unified catalog (labels K and S) remained almost constant after June 2011. The constant number seen in the JMA unified catalog is an artifact resulting from the gradual increase in the number of $M \sim 1$ earthquakes, and there is no difference in the number of earthquakes above $M 2 \sim 3$ between the two catalogs. This indicates the upper limit of manually determinable earthquakes per day.

Improvements achieved using the phase classifier

Compared to the seismicity reported during March 2011 by Tamaribuchi and Nakagawa (2020), the hypocentral distribution obtained here shows that the application of the phase classifier reduced false detections and mislocations in especially offshore areas (Fig. 7a, b) and resulted in a slight reduction of the number of events with quality label A (by 4%), but a marked reduction of the numbers of events with labels B, C, and D (by 22%, 26%, and 78% respectively; Table 2a).

We counted the number of automatically picked phases at the same station within ± 0.5 s of the P- and S-wave arrival times in the JMA unified catalog and calculated the percentages of falsely detected P- and S-phases as the proportion of S-phases picked within ± 0.5 s of a P-wave arrival and that of P-phases picked within ± 0.5 s of an S-wave arrival, respectively. A comparison of our results with those of Tamaribuchi and Nakagawa (2020) shows that the application of the phase classifier reduced falsely picked P- and S-waves from 1.6 to 0.6% and from 4.2 to

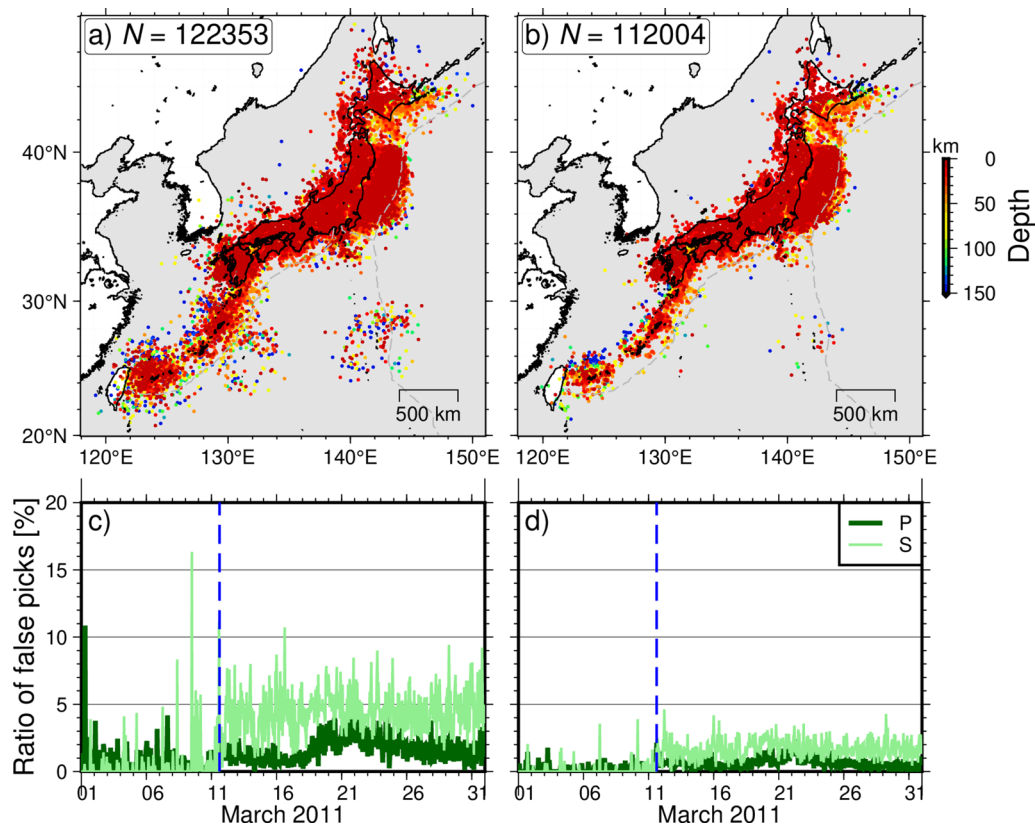


Fig. 7 Comparison of automatically determined hypocenters before and after applying the phase classifier. **a, c** Automatically determined hypocenters from Tamaribuchi and Nakagawa (2020), i.e., without the phase classifier; **b, d** the same data after application of our phase classifier. **a, b** Hypocenter distributions; symbol colors indicate depth. **c, d** Percentage of false P- and S-phase picks within ± 0.5 s of S- and P-wave arrivals, respectively, in the JMA unified catalog. The vertical blue dashed line indicates the origin time of the 2011 Tohoku earthquake

Table 2 Comparison of the results of Tamaribuchi and Nakagawa (2020), abbreviated TN20, before and after the application of the phase classifier in this study

a)								
	All events	TP	FP	Label A	Label B	Label C	Label D	
TN20	122,353	44,575	77,778	104,253	7545	6848	3707	
This study	112,004	44,566	67,438	100,265	5882	5055	802	
b)								
	μ_x (km)	σ_x (km)	μ_y (km)	σ_y (km)	μ_D (km)	σ_D (km)	μ_M	σ_M
TN20	-0.23	3.99	0.09	3.24	-0.42	7.09	0.00	0.18
This study	-0.04	3.74	0.14	3.10	-0.63	6.90	0.00	0.18

(a) TP and FP indicate the numbers of events that do and do not coincide with the JMA unified catalog, respectively. Quality labels A–D are the results of our LightGBM quality control label classifier. (b) Residuals between events in the JMA unified catalog and their corresponding automatically determined hypocenters. μ and σ are the mean and standard deviation, respectively. Subscripts x , y , D , and M indicate longitude, latitude, depth, and magnitude, respectively

1.6%, respectively (Fig. 7c, d). The proportion of false picks in our results remained unchanged after the 2011 Tohoku earthquake.

Table 2 reports the number of events matched with the JMA unified catalog and their residuals and standard deviations. Tamaribuchi and Nakagawa (2020)

detected 122,353 events, whereas we detected 112,004 events, a decrease of nearly 10,000. Of those, however, the number of matches with the JMA unified catalog was almost the same: 44,575–44,566 events, respectively. This result indicates that the number of false detections and mislocations in our results was greatly reduced

by the reduction in false picks, which also resulted in lower residuals between the JMA unified catalog and the matched hypocenters.

Frequency-magnitude distribution

The frequency-magnitude distributions of the JMA unified and our merged catalogs are shown in Figs. 8 and Additional file 1: Fig. S1 to illustrate the change in earthquake detectability in the aftershock area of the 2011 Tohoku earthquake. The number of hypocenters more than tripled from 215,709 in the JMA unified catalog to 701,516 in the merged catalog. We determined M_c based on the “MAXC” approach proposed by Wiemer and Wyss (2000), which considers the highest frequency of events in the non-cumulative frequency-magnitude distribution (MAXC) and defines M_c as MAXC + 0.2 (Woessner and Wiemer 2005).

In our results, M_c in the aftershock area decreased along the coast (areas 1 and 4 in Fig. 8), although the number of large earthquakes was almost unchanged because they were already included in the JMA unified catalog. Farther from the coast (areas 2 and 3 in Fig. 8), our method did not improve detectability because of the sparse station distribution. In contrast, in our results, M_c decreased in many inland areas (Additional file 1: Fig. S1). In particular, the number of hypocenters in the inland areas of the Tohoku region, such as Akita Prefecture (areas 6 and 7 in Additional file 1: Fig. S1), increased by a factor of 4 to 5. Magnitude-time diagrams

(Additional file 1: Fig. S1d) show that microearthquakes during April to July 2011 were not included in the JMA unified catalog, resulting in a large discrepancy in the number of events compared with our merged catalog. Although our estimation of M_c is simplified and that in the JMA unified catalog is clearly underestimated, M_c is significantly lower in our merged catalog. M_c was also decreased at distances up to 800 km from the epicenter of the 2011 Tohoku earthquake (star in Fig. 1), such as in area 23 (Additional file 1: Fig. S1), which seems to reflect the limitations of manual inspection.

Discussion

Swarm migration

Seismicity increased in the northern part of Akita Prefecture (area 6 in Additional file 1: Fig. S1) after 11 March 2011 (Additional file 1: Fig. S1d), even though detectability may have decreased due to station outages during 11–13 March. We applied the double-difference (DD) method (Waldhauser and Ellsworth 2000) to determine the precise locations of events in Akita Prefecture between 1 March and 30 April 2011 in the merged catalog. We used travel-time difference data calculated by waveform cross-correlation. Of 1210 events with labels of K or A, 1118 could be redetermined by the DD method, compared to just 140 in the JMA unified catalog alone. The distribution of the relocated events is shown in Fig. 9, as well as the location of Moriyoshi-zan, a Quaternary volcano. Many of the hypocenters were distributed

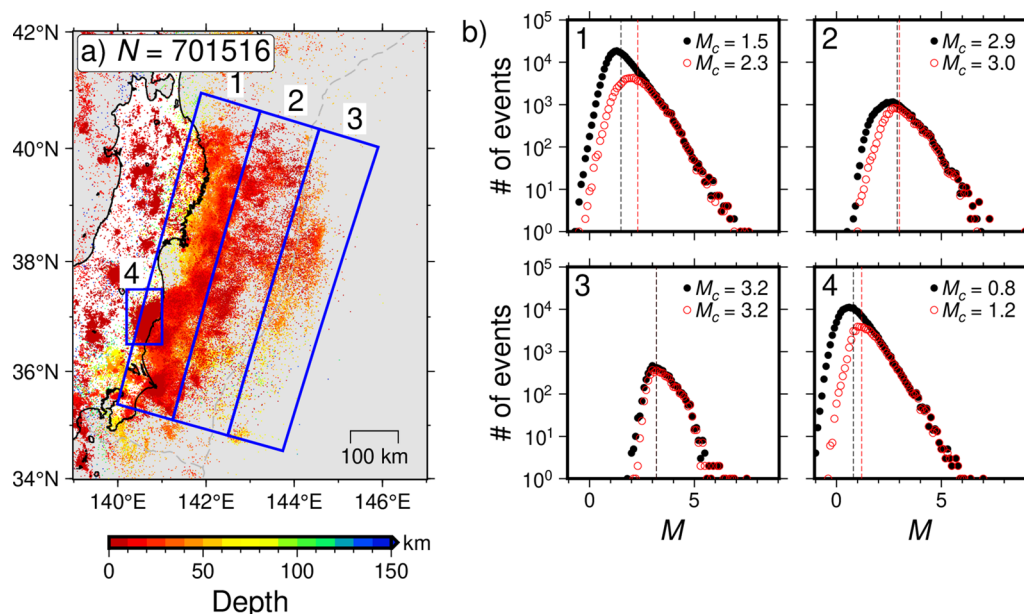


Fig. 8 Changes in earthquake detectability along the Japan Trench. **a** Hypocentral distribution in our merged catalog; symbol colors indicate depth. **b** Frequency-magnitude distribution for each region. Red and black symbols indicate the number of events in the JMA unified catalog and our merged catalog, respectively. M_c is based on MAXC + 0.2 (Woessner and Wiemer 2005). Region 1 does not include the epicenters in region 4

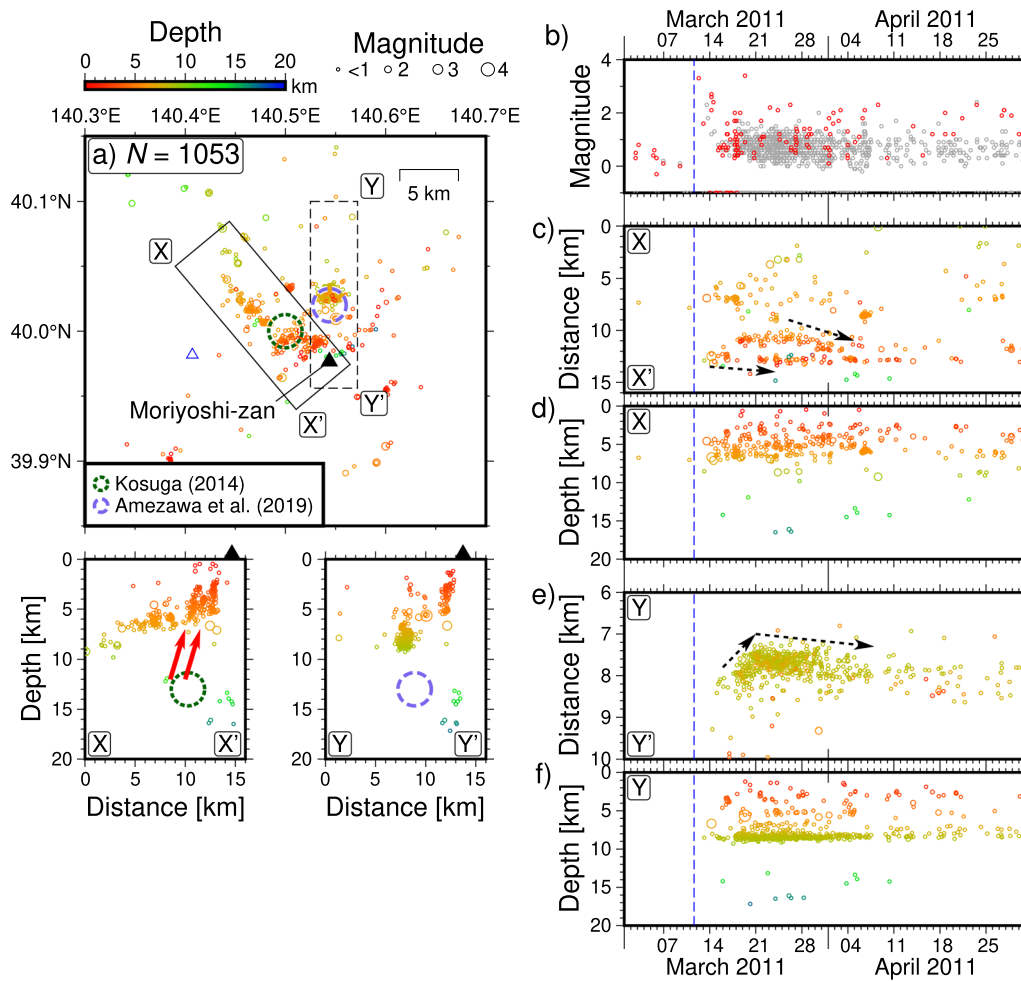


Fig. 9 Precise hypocentral distribution in northern Akita Prefecture (region 6 in Fig. S1); except (b), symbol colors indicate depth. **a** Hypocentral distribution of the merged catalog with events recalculated by the DD method with cross-correlation (1 March–30 April 2011, 0–20 km depth). Blue triangles show observation stations, and the black triangle indicates the Moriyoshi-zan volcano. Bottom panels show cross sections X–X' and Y–Y'. Green and blue dotted circles indicate probable scatterers estimated by Kosuga (2014) and Amezawa et al. (2019), respectively. **b** Magnitude-time diagram of events in (a). Red and gray circles are events in the JMA unified catalog and our merged catalog, respectively. **c, d** Spatiotemporal and depth-time distributions, respectively, projected on line X–X'. **e, f** Spatiotemporal and depth-time distributions, respectively, projected on line Y–Y'. Vertical blue dashed lines indicate the origin time of the 2011 Tohoku earthquake

in clusters aligned on the northwest and north sides of Moriyoishi-zan (cross sections X–X' and Y–Y') and hypocenters aligned on the northwest side were several inclined clusters (red arrows in cross section X–X'). Although the focal depths under Moriyoishi-zan were shallow (0–5 km), they tended to deepen farther from the volcano (~10 km depth, the northwest side in cross section X–X'), suggesting that geothermal activity near Moriyoishi-zan was higher than in the surrounding areas.

Kosuga (2014) applied the DD method to estimate the precise hypocentral distribution of a cluster north of Moriyoishi-zan (cross section Y–Y', Fig. 9) and estimated the hydraulic diffusivity. They also estimated the locations of scatterers from the reflected/scattered phases.

However, their study covered only the period after May 2011. Amezawa et al. (2019) also estimated the location of scatterers north of Moriyoishi-zan (cross section Y–Y', Fig. 9) using a small seismometer array operated during the period from November 2012 to May 2014. Here, we focused on the period from March to April 2011, during which most of the events in the merged catalog were not listed in the JMA unified catalog, and we identified some clusters consistent with the probable scatterers estimated by Kosuga (2014); red arrows in cross section X–X', Fig. 9a). This activity has not been identified before because it subsided by early April 2011. The dashed arrows in Fig. 9c show the complicated migration of the hypocenters. The spatiotemporal distribution projected

on line Y–Y' in Fig. 9e also shows northward migration during the early phase of the activity (16–21 March), followed by a slight southward migration from 21 March to 8 April. The uncertainties on the locations of their estimated scatterers are beyond the scope of our study, but the seismic clusters were distributed shallower than the scatterers they reported. As interpreted by Kosuga (2014), these activities likely reflect fluid migrations from the scatterers, which are probably geofluid reservoirs. Another possibility is that aseismic slip may also be involved in the complex migrations, as suggested in other volcanic areas (e.g., Yukutake et al. 2022). Further investigation using the merged catalog will contribute to our understanding of swarm migrations recently after a giant earthquake.

Foreshock activity

An M 6.7 earthquake (6+ on the JMA intensity scale) occurred in northern Nagano Prefecture at 3:59 on 12 March 2011 (area 14 in Additional file 1: Fig. S1). In this region, the JMA unified catalog did not list any earthquakes until 21:48 on 11 March, but the merged catalog detected earthquakes as early as 14:55 on 11 March. As a result, we detected 86 earthquakes that occurred after the M_w 9.0 Tohoku earthquake at 14:46 on 11 March and preceding the M 6.7 earthquake (Fig. 10), only 10 of which were listed in the JMA unified catalog. During that period, seismicity was observed south of the epicenter of the M 6.7 earthquake, but largely subsided after

the M 6.7 earthquake. The JMA unified catalog did not adequately capture such changes in seismicity. However, there is no evidence of nucleation processes such as the migration or acceleration of foreshocks in this region. Shimojo et al. (2014) detected 139 events during this period by template matching using data from 10 Hi-net stations and suggested that the southern seismicity was likely caused by the migration of geothermal crustal fluids activated by the surface wave of the 2011 Tohoku earthquake. Although we detected fewer hypocenters, the observed trend is consistent with the results of Shimojo et al. (2014). We emphasize that our method does not require a template. We highlight the potential of our automated process to monitor changes in small foreshock activity following a large earthquake. Zaliapin and Ben-Zion (2013) and Tamaribuchi et al. (2018) extracted foreshock activities from many earthquakes in the catalog and contributed to the statistical analysis of foreshocks. Other methods have also been proposed, for example, to statistically extract foreshock conditions from the past catalog and use them to predict mainshocks (Hirose et al. 2021). Because the number of earthquakes in the catalog is important for such analyses, our method of merging earthquake catalogs should prove useful to future contributions on this topic.

Seismicity in and around the Tohoku region

Map and cross-sectional views of the hypocentral distribution in the Tohoku region during the year following

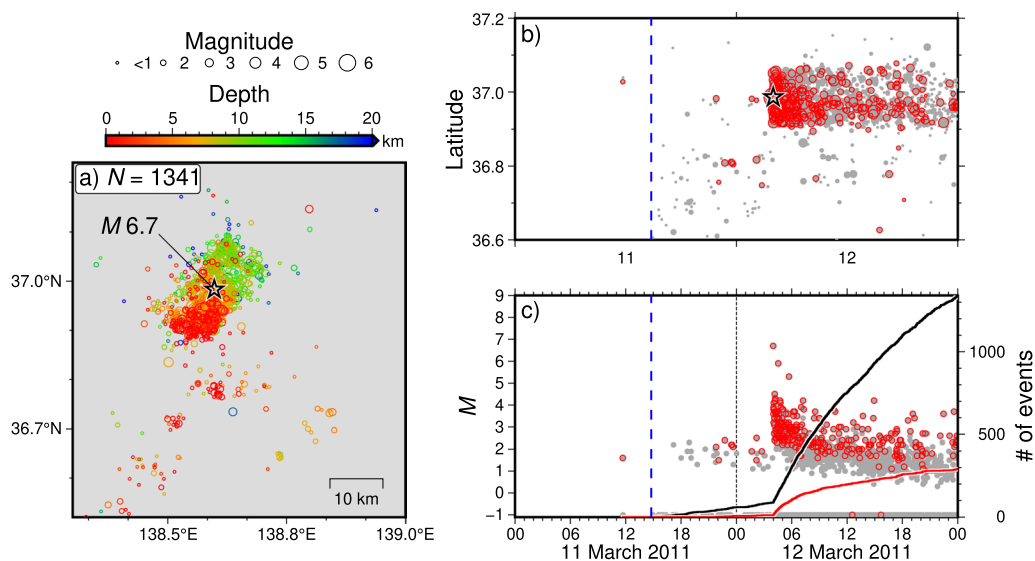


Fig. 10 Hypocentral distribution in northern Nagano Prefecture (region 14 in Additional file 1: Fig. S1). **a** Hypocentral distribution in our merged catalog (11–12 March 2011, 0–50 km depth). Symbol colors indicate depth. **b** Spatiotemporal distribution projected in the N–S direction. **c** Magnitude-time diagram; red and gray circles are events in the JMA unified catalog and our merged catalog, respectively. Curves are plotted against the right vertical axis and show the cumulative numbers of events. Vertical blue dashed lines indicate the origin time of the 2011 Tohoku earthquake

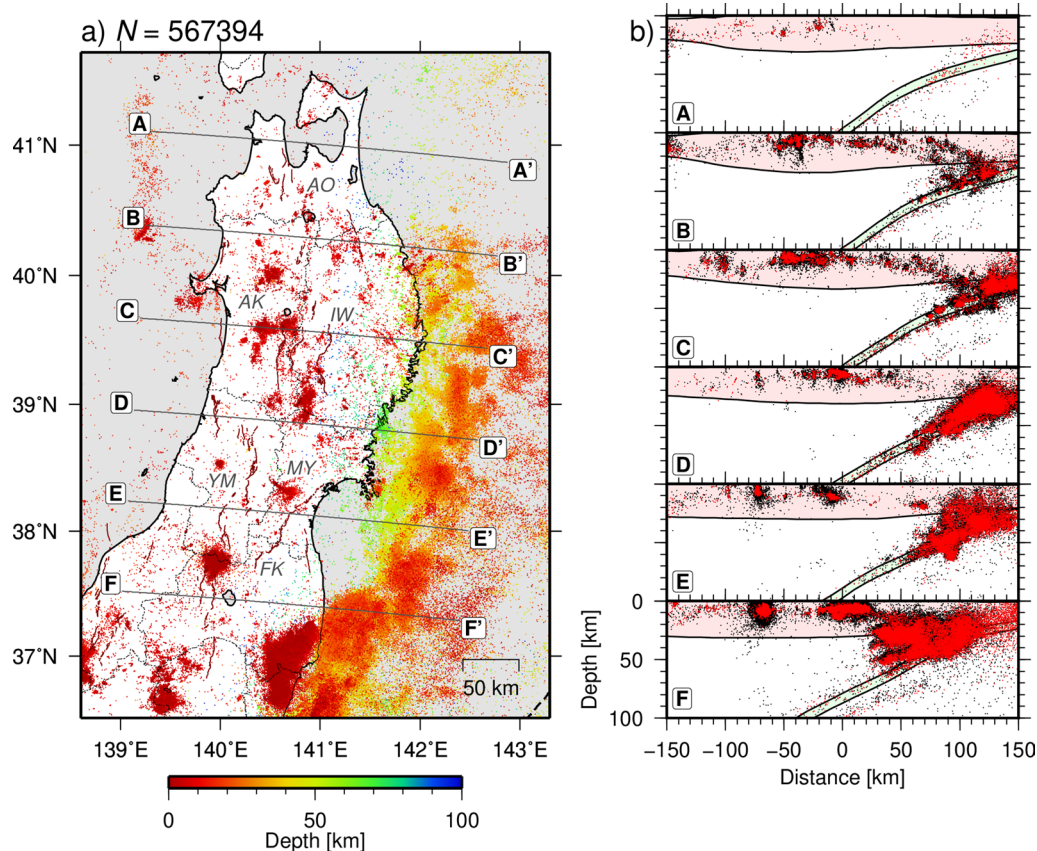


Fig. 11 Hypocentral distribution (1 March 2011 to 29 February 2012, ≤ 100 km depth). **a** Hypocentral distribution of the merged catalog; symbol colors indicate depth. AO, AK, IW, YM, MY, and FK indicate Aomori, Akita, Iwate, Yamagata, Miyagi, and Fukushima Prefectures, respectively. **b** Cross-sectional views of events in the merged catalog (black) and the JMA unified catalog (red). The cross-sectional views show earthquakes occurring within 40 km of each line in (a). The pink shaded area represents the continental crust and the green shaded area represents the oceanic crust of the subducting Pacific Plate from the JIVSM model (Koketsu et al. 2012)

the 2011 Tohoku earthquake (Figs. 11, 12) clearly highlight the associated aftershock activity. The northern limit of the 2011 Tohoku earthquake aftershock area is about 40.3°N , and little seismicity occurred near the plate boundary in cross section A–A' during that time. In contrast, aftershock seismicity was particularly active along the coast from off Iwate to off Fukushima. Although detectability decreases away from the coast, most of these earthquakes were associated with plate subduction at depths of 10–50 km. Figure 12 highlights particularly active areas from off southern Iwate Prefecture to off Miyagi Prefecture (between lines C–C' and E–E') and off Fukushima Prefecture (south of line F–F'). In contrast, relatively little seismicity occurred off northern Iwate Prefecture (between lines B–B' and C–C') and off the Miyagi–Fukushima prefectural border (between lines E–E' and F–F'). Tanioka et al. (1997) explained that the plate boundary off Iwate Prefecture is rough because of a well-developed horst-and-graben structure on the subducting Pacific plate, which can only form small patches

of earthquakes there. Thus, the aftershock seismicity observed here may reflect the characteristics of the subducting plate.

Cross sections B–B' and C–C' in Figs. 11b and 12b show a clear eastward dip from the continental crust to the subducting Pacific plate between 50 and 100 km along the horizontal axis. Such seismicity was also slightly observed in the 11 years preceding the 2011 Tohoku earthquake (Additional file 1: Figs. S2, S3) but made clearer from the merged catalog after the 2011 Tohoku earthquake. Although inland seismicity within Iwate Prefecture was low both before and after the 2011 Tohoku earthquake, making quantitative evaluation difficult, this unique pattern of seismicity may have been due to changes in the stress field caused by the 2011 Tohoku earthquake (Yoshida et al. 2012), in turn changing the stresses on inland faults.

In cross sections D–D', E–E', and F–F', abundant seismicity is observed in the mantle wedge (Figs. 11, 12), but such activity was also observed before the 2011 Tohoku

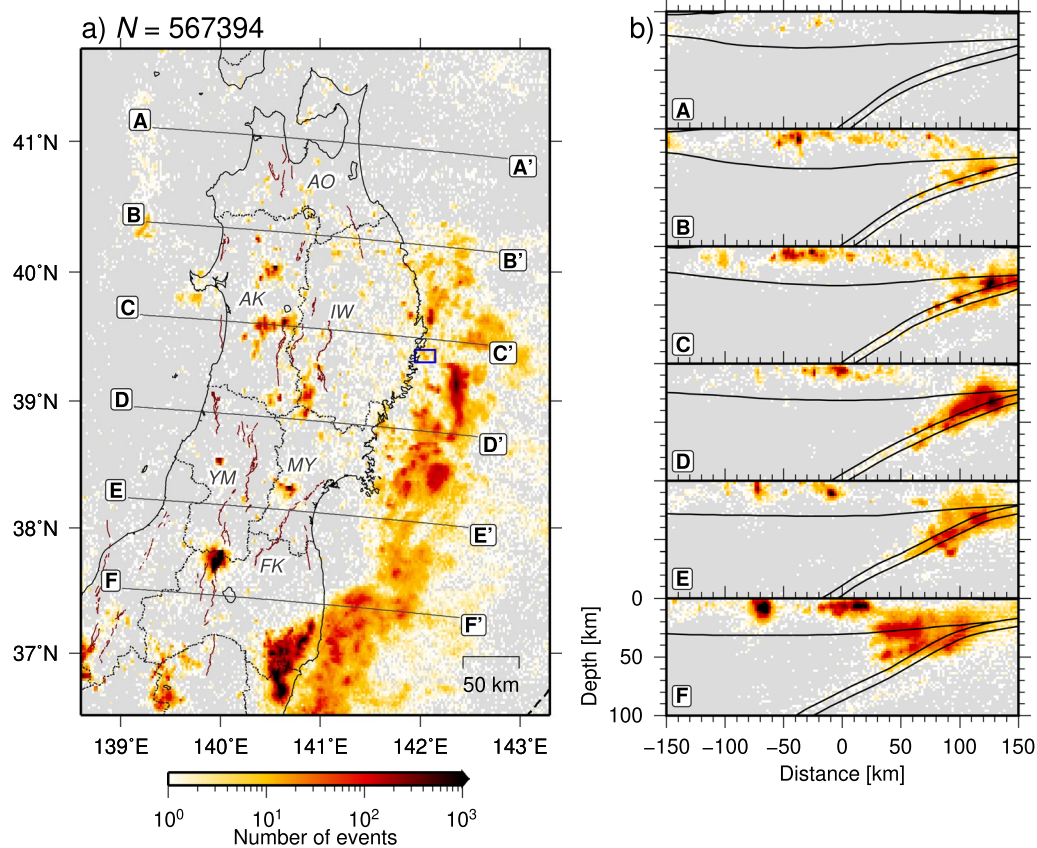


Fig. 12 Heat maps of the events in Fig. 11. **a** Frequency distribution binned in a $0.02^\circ \times 0.02^\circ$ grid. **b** Cross-sectional views of the frequency distributions counted in 2-km bins. Other details are as in Fig. 11. The blue rectangle in (a) marks the region off Kamaishi, Iwate Prefecture, shown in Fig. 13

earthquake (Additional file 1: Figs. S2, S3). Uchida et al. (2010) called these “supraslab” earthquakes and attributed them to seamounts detached from the subducting plate. Because these earthquakes are near the hypocenters of interplate earthquakes, it will be important to evaluate intraplate, interplate, and mantle wedge seismicity separately in future, more detailed hypocenter analyses.

Repeating M 5 earthquakes occur off Kamaishi, Iwate Prefecture, and surrounding seismicity tends to increase in the latter half of the recurrence interval of the M 5 earthquakes (Matsuzawa et al. 2002; Uchida et al. 2007; 2012). Here, we applied the DD method (Waldhauser and Ellsworth 2000) to our merged catalog (labels K and A only) to determine the precise locations of hypocenters in the area $39.3\text{--}39.5^\circ\text{N}$, $141.93\text{--}142.15^\circ\text{E}$, and at 0–70 km depth during 1 March 2011 to 29 February 2012. We used travel-time difference data calculated by waveform cross-correlation. Of 557 events in the merged catalog, 524 could be redetermined by the DD method, compared to only 97 events in the JMA unified catalog alone.

Figure 13b, c show the magnitude-time diagram and cumulative number of events, respectively. Overall, the

number of earthquakes decreased with increasing time following the 2011 earthquake. However, focusing on each cycle as delimited by the repeating M 5 events, seismicity seems to increase before the M 5 earthquakes. When normalized to the total number of earthquakes that occurred within each interval, seismicity was consistently low during the first half of the recurrence interval and increased at an accelerating rate in the latter half of the interval (Fig. 13d). Based on the repeating earthquakes before the 2011 Tohoku earthquake, Uchida et al. (2012) concluded that heterogeneous structures exist in the main asperities of the fault on which repeating M 5 earthquakes occur, and that stress concentrations are caused by aseismic slip in the surrounding region. Okuda et al. (2018) detected microearthquakes off Kamaishi immediately after the 2011 Tohoku earthquake using a template-matching technique and observed similar phenomena over two cycles. We obtained similar results over seven cycles during the 1-year period after the 2011 Tohoku earthquake. This temporal consistency supports the existence of heterogeneous structures, such as in the hierarchical asperity model of Hori and Miyazaki

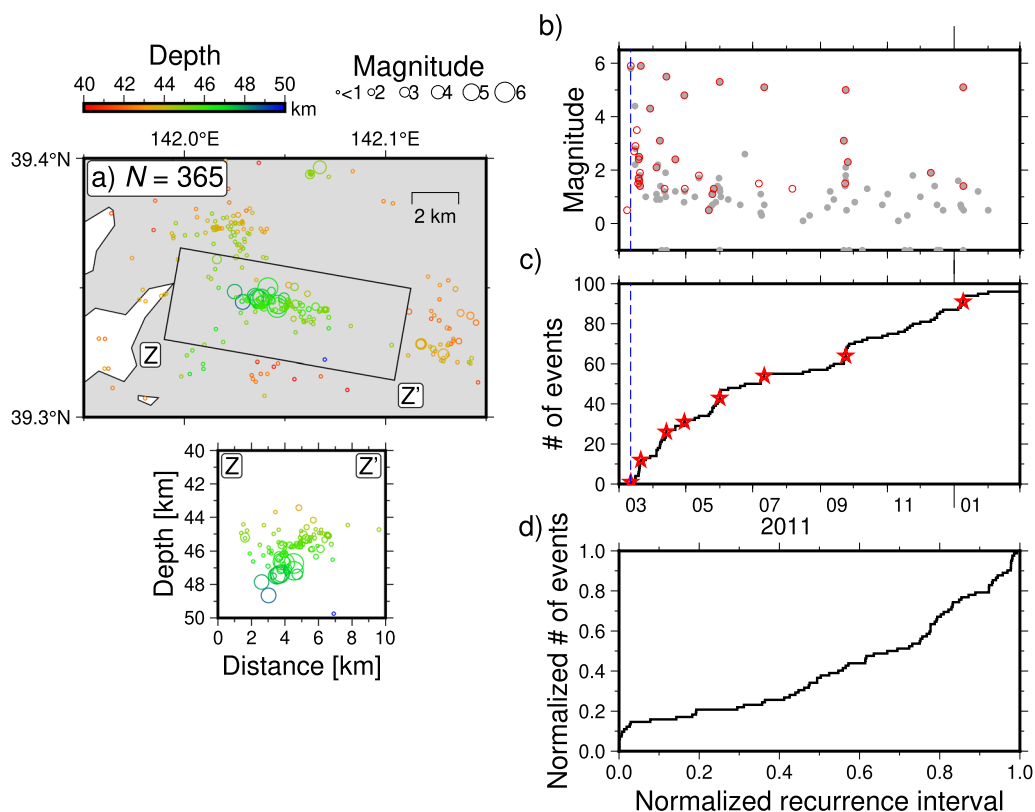


Fig. 13 Precise hypocentral distribution off Kamaishi, Iwate Prefecture. **a** Hypocentral distribution of the merged catalog with events recalculated by the DD method with cross-correlation (1 March 2011–29 February 2012; 40–50 km depth; area indicated in Fig. 12a). The bottom panel is a cross-sectional view of earthquakes occurring within a rectangle in (a). Symbol colors indicate depth. **b** Magnitude-time diagram; red and gray circles are events in the JMA unified catalog and the merged catalog, respectively. **c** Cumulative number of events. Red stars indicate the repeating $M \geq 4.8$ earthquakes (Uchida et al. 2015). Vertical blue dashed lines in (b) and (c) indicate the origin time of the 2011 Tohoku earthquake. **d** Normalized cumulative number of events over seven recurrences. The times immediately after the repeating $M \geq 4.8$ earthquakes and immediately before the next one are represented by $x=0$ and 1 on the horizontal axis, respectively. The $M \geq 4.8$ repeating earthquakes are not included in the count

(2010). Because such repeating earthquakes can be used to estimate stress on their associated fault patches, the microseismic catalog obtained in this study will provide valuable information for such estimations. Further analyses of similar seismicity may provide detailed information on the slip distribution over a wide area.

Summary and future challenges

We incorporated machine learning into automatic hypocenter determination and reduced false detections by approximately 80%. By merging the obtained automatically determined hypocenters with the JMA unified catalog, we created a merged catalog of ~920,000 events, nearly triple the ~320,000 events in the conventional catalog. Our merged catalog contributes to the detection of fine fault structures and small foreshock activities, as well as to high-resolution monitoring of interplate slips based on repeating earthquakes.

We used a CNN alone for noise reduction and P- and S-wave discrimination. However, many deep learning methods have been proposed for phase picking, and incorporating those methods will further improve the identification of small earthquakes. Nonetheless, we note that Park et al. (2023) pointed out that deep learning methods are not easily reproducible. Therefore, it may be necessary to increase model robustness to unsupervised data while continuing to use robust conventional methods.

We focused on eliminating false detections, which is similar to the method used to detect far-field and deep earthquakes that cannot be detected even by conventional methods. Although we made a strong effort to reduce noise, our results include ~1% false detections. It is therefore necessary to improve the phase-association method to achieve sufficient performance for earthquakes with long S–P times. This is a pressing

issue for monitoring large earthquakes that occur near trench axes and subsequent far-field seismicity.

Abbreviations

AR-AIC	Autoregressive-Akaike's information criterion
CNN	Convolutional neural network
DD	Double-difference
JIVSM	Japan integrated velocity structure model
JMA	Japan Meteorological Agency
LightGBM	Light gradient-boosting machine
M	Magnitude
M_c	Magnitude of completeness
M_w	Moment magnitude
PF	Phase combination forward search

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-023-01915-3>.

Additional file 1: Figure S1. Changes in earthquake detectability in and around eastern Japan (≤ 50 km depth). **Figure S2.** Hypocentral distribution of events in the JMA unified catalog during the 11 years preceding the 2011 Tohoku earthquake. **Figure S3.** Heat map of events in the JMA unified catalog during the 11 years preceding the 2011 Tohoku earthquake. **Table S1.** Features used in our LightGBM quality control classifier.

Additional file 2: Dataset S1. Previously undetected earthquakes from our merged catalog.

Acknowledgements

To create the automatically determined hypocenters used in this study, we used data from the following sources: National Research Institute for Earth Science and Disaster Resilience (National Research Institute for Earth Science and Disaster Resilience, 2019a; b), 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), Aomori Prefecture, the Tokyo Metropolitan Government, Shizuoka Prefecture, Yokohama City (Kanagawa Prefecture), the Hot Springs Research Institute of Kanagawa Prefecture, and Japan Meteorological Agency. For analysis, we used the large continuous seismic data analyzing system (Nakagawa et al. 2016) of the Earthquake Research Institute, the University of Tokyo (ERI JURP 2022-F3-12). We used the JMA unified catalog [<https://www.data.jma.go.jp/svd/eqev/data/bulletin/index.html>]. 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 location of Moriyoshi-zan is available from the Volcanoes of Japan database of the Geological Survey of Japan, AIST [https://gbank.gsj.jp/volcano/Quat_Vol/index_e.html]. We used the Python library PyCaret (Ali 2020). We used HypoDD (Waldhauser and Ellsworth 2000) to estimate precise hypocenter locations. We used Generic Mapping Tools (Wessel et al. 2019) to create the figures. This work was supported by JSPS KAKENHI Grant Number JP20K14565. We are also grateful to two anonymous reviewers, and Editor David Shelly for their constructive comments.

Author contributions

KT analyzed and wrote most of the manuscript. SK and KS contributed to phase classification. FH helped prepare the manuscript. 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). The seismic catalog created in this article is included as an Additional file 1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan. ²Seismology and Volcanology Department, Japan Meteorological Agency, 3-6-9 Toranomon, Minato-Ku, Tokyo 105-8431, Japan.

Received: 13 June 2023 Accepted: 28 September 2023

Published online: 10 October 2023

References

- Ali M (2020) PyCaret: an open source, low-code machine learning library in Python. <https://www.pycaret.org>.
- Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68(5):1521–1532. <https://doi.org/10.1785/bssa0680051521>
- Amezawa Y, Kosuga M, Maeda T (2019) Temporal changes in the distinct scattered wave packets associated with earthquake swarm activity beneath the Moriyoshi-zan volcano, northeastern Japan. *Earth Planets Space* 71(1):132. <https://doi.org/10.1186/s40623-019-1115-6>
- Arrowsmith SJ, Trugman DT, MacCarthy J, Bergen KJ, Lumley D, Magnani MB (2022) Big data seismology. *Rev Geophys* 60(2):e2021RG000769. <https://doi.org/10.1029/2021RG000769>
- Baillard C, Crawford WC, Ballu V, Hibert C, Mangeney A (2013) An automatic kurtosis-based P- and S-phase picker designed for local seismic networks. *Bull Seismol Soc Am* 104(1):394–409. <https://doi.org/10.1785/0120120347>
- Beroza GC, Segou M, Mousavi SM (2021) Machine learning and earthquake forecasting—next steps. *Nat Commun* 12(1):4761. <https://doi.org/10.1038/s41467-021-24952-6>
- DeMets C, Gordon RG, Argus DF (2010) Geologically current plate motions. *Geophys J Int* 181(1):1–80. <https://doi.org/10.1111/j.1365-246X.2009.04491.x>
- Feng T, Mohanna S, Meng L (2022) EdgePhase: a deep learning model for multi-station seismic phase picking. *Geochem Geophys Geosyst* 23(11):e2022GC010453. <https://doi.org/10.1029/2022GC010453>
- Gibbons SJ, Ringdal F (2006) The detection of low magnitude seismic events using array-based waveform correlation. *Geophys J Int* 165(1):149–166. <https://doi.org/10.1111/j.1365-246X.2006.02865.x>
- Hamada N, Yoshida A, Hashimoto H (1983) Improvement of the hypocenter determination program of Japan Meteorological Agency (reanalyses of the hypocenter distribution of the 1980 earthquake swarm off the east coast of the Izu peninsula and the Matushiro earthquake swarm). *Quart J Seis* 48:35–55 (in Japanese with English abstract)
- Hirose F, Miyaoka K, Hayashimoto N, Yamazaki T, Nakamura M (2011) Outline of the 2011 off the Pacific coast of Tohoku Earthquake (M_w 9.0)—seismicity: foreshocks, mainshock, aftershocks, and induced activity. *Earth Planets Space* 63(7):513–518. <https://doi.org/10.5047/eps.2011.05.019>
- Hirose F, Tamaribuchi K, Maeda K (2021) Characteristics of foreshocks revealed by an earthquake forecasting method based on precursory swarm activity. *J Geophys Res Solid Earth* 126(9):e2021JB021673. <https://doi.org/10.1029/2021jb021673>

- Hori T, Miyazaki S (2010) Hierarchical asperity model for multiscale characteristic earthquakes: a numerical study for the off-Kamaishi earthquake sequence in the NE Japan subduction zone. *Geophys Res Lett* 37(10):L10304. <https://doi.org/10.1029/2010GL042669>
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st international conference on neural information processing systems, pp 3149–3157
- Koketsu K, Miyake H, Suzuki H (2012) Japan integrated velocity structure model version 1. In: Proceedings of the 15th world conference on earthquake engineering, vol 1, pp 1–4
- Kosuga M (2014) Seismic activity near the Moriyoshi-zan volcano in Akita Prefecture, northeastern Japan: implications for geofluid migration and a midcrustal geofluid reservoir. *Earth Planets Space* 66(1):77. <https://doi.org/10.1186/1880-5981-66-77>
- Kudo S, Shimojo K, Tamaribuchi K (2023) Seismic phase detection and picking with 1-D convolutional neural network. *J Seismol Volcanol Relat Eng* 86:4 (in Japanese)
- Matsuzawa T, Igarashi T, Hasegawa A (2002) Characteristic small-earthquake sequence off Sanriku, northeastern Honshu, Japan. *Geophys Res Lett* 29(11):38-1–38-4. <https://doi.org/10.1029/2001GL014632>
- Mousavi SM, Beroza GC (2023) Machine learning in earthquake seismology. *Annu Rev Earth Planet Sci* 51(1):105–129. <https://doi.org/10.1146/annurev-earth-071822-100323>
- Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun* 11(1):3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Nakagawa S, Tsuruoka H, Kato A, Sakai S, Hirata N (2016) A petabyte-scale large continuous seismic data analyzing system. *Bull Earthq Res Inst* 91:1–9. <https://doi.org/10.15083/0000032408>. (in Japanese with English abstract)
- Nakamura S, Takeo M, Okabe Y, Matsuura M (2007) Automatic seismic wave arrival detection and picking with stationary analysis: application of the KM_2O -Langevin equations. *Earth Planets Space* 59(6):567–577. <https://doi.org/10.1186/BF03352719>
- National Research Institute for Earth Science and Disaster Resilience (2019a) NIED F-net. doi:<https://doi.org/10.17598/NIED.0005>
- National Research Institute for Earth Science and Disaster Resilience (2019b) NIED Hi-net. doi:<https://doi.org/10.17598/NIED.0003>
- Okuda T, Uchida N, Asano Y (2018) Detection of small earthquakes using the matched filter method immediately after the Tohoku-oki earthquake. *Zisin2* 71:87–101. <https://doi.org/10.4294/zisin.2017-8>. (in Japanese with English abstract)
- Park Y, Beroza GC, Ellsworth WL (2023) A mitigation strategy for the prediction inconsistency of neural phase pickers. *Seismol Res Lett* 94(3):1603–1612. <https://doi.org/10.1785/0220230003>
- Ross ZE, Meier M-A, Hauksson E (2018a) P wave arrival picking and first-motion polarity determination with deep learning. *J Geophys Res Solid Earth* 123(6):5120–5129. <https://doi.org/10.1029/2017JB015251>
- Ross ZE, Meier MA, Hauksson E, Heaton TH (2018b) Generalized seismic phase detection with deep learning. *Bull Seismol Soc Am* 108(5A):2894–2901. <https://doi.org/10.1785/0120180080>
- Ross ZE, Trugman DT, Hauksson E, Shearer PM (2019a) Searching for hidden earthquakes in Southern California. *Science* 364(6442):767–771. <https://doi.org/10.1126/science.aaw6888>
- Ross ZE, Yue Y, Meier M-A, Hauksson E, Heaton TH (2019b) PhaseLink: a deep learning approach to seismic phase association. *J Geophys Res Solid Earth* 124(1):856–869. <https://doi.org/10.1029/2018JB016674>
- Saragiotis CD, Hadjileontiadis LJ, Panas SM (2002) PAI-S/K: a robust automatic seismic P phase arrival identification scheme. *IEEE Trans Geosci Remote Sens* 40(6):1395–1404. <https://doi.org/10.1109/TGRS.2002.800438>
- Shimojo K, Enescu B, Yagi Y, Takeda T (2014) Fluid-driven seismicity activation in northern Nagano region after the 2011 M9.0 Tohoku-oki earthquake. *Geophys Res Lett* 41(21):7524–7531. <https://doi.org/10.1002/2014GL061763>
- Tamaribuchi K (2018) Evaluation of automatic hypocenter determination in the JMA unified catalog. *Earth Planets Space* 70(1):141. <https://doi.org/10.1186/s40623-018-0915-4>
- Tamaribuchi K, Hirose F, Noda A, Iwasaki Y, Iwakiri K, Ueno H (2021) 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. *Earth Planets Space* 73(1):91. <https://doi.org/10.1186/s40623-021-01411-6>
- Tamaribuchi K, Moriawaki K, Ueno H, Tsukada S (2016) Automatic hypocenter determination for the seismological bulletin of Japan using Bayesian estimation. *Quart J Seis* 79:1–13 (in Japanese with English abstract)
- Tamaribuchi K, Nakagawa S (2020) Evaluation of the seismic activity of automatically processed hypocenters in swarms over wide areas—application to the 2011 Mw 9.0 Tohoku Earthquake. *Quart J Seis* 83:3 (in Japanese with English abstract)
- Tamaribuchi K, Yagi Y, Enescu B, Hirano S (2018) Characteristics of foreshock activity inferred from the JMA earthquake catalog. *Earth Planets Space* 70(1):90. <https://doi.org/10.1186/s40623-018-0866-9>
- Tan YJ, Waldhauser F, Ellsworth WL, Zhang M, Zhu W, Michele M, Chiaraluce L, Beroza GC, Segou M (2021) Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central Italy sequence. *Seism Rec* 1(1):11–19. <https://doi.org/10.1785/0320210001>
- Tanioka Y, Ruff L, Satake K (1997) What controls the lateral variation of large earthquake occurrence along the Japan Trench? *Island Arc* 6(3):261–266. <https://doi.org/10.1111/j.1440-1738.1997.tb00176.x>
- Uchida N, Matsuzawa T, Ellsworth WL, Imanishi K, Okada T, Hasegawa A (2007) Source parameters of a M4.8 and its accompanying repeating earthquakes off Kamaishi, NE Japan: implications for the hierarchical structure of asperities and earthquake cycle. *Geophys Res Lett* 34(20):L20313. <https://doi.org/10.1029/2007GL031263>
- Uchida N, Kirby SH, Okada T, Hino R, Hasegawa A (2010) Supraslab earthquake clusters above the subduction plate boundary offshore Sanriku, northeastern Japan: seismogenesis in a graveyard of detached seamounts? *J Geophys Res Solid Earth* 115:B09308. <https://doi.org/10.1029/2009JB006797>
- Uchida N, Matsuzawa T, Ellsworth WL, Imanishi K, Shimamura K, Hasegawa A (2012) Source parameters of microearthquakes on an interplate asperity off Kamaishi, NE Japan over two earthquake cycles. *Geophys J Int* 189(2):999–1014. <https://doi.org/10.1111/j.1365-246X.2012.05377.x>
- Uchida N, Shimamura K, Matsuzawa T, Okada T (2015) Postseismic response of repeating earthquakes around the 2011 Tohoku-oki earthquake: moment increases due to the fast loading rate. *J Geophys Res Solid Earth* 120(1):259–274. <https://doi.org/10.1002/2013JB010933>
- Ueno H, Hatakeyama S, Aketagawa T, Funasaki J, Hamada N (2002) Improvement of hypocenter determination procedures in the Japan Meteorological Agency. *Quart J Seis* 65:123–134 (in Japanese with English abstract)
- 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(6):1353–1368. <https://doi.org/10.1785/0120000006>
- Wessel P, Luis JF, Uieda L, Scharroo R, Wobbe F, Smith WHF, Tian D (2019) The generic mapping tools version 6. *Geochem Geophys Geosyst* 20(11):5556–5564. <https://doi.org/10.1029/2019gc008515>
- Wiemer S, Wyss M (2000) Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull Seismol Soc Am* 90(4):859–869. <https://doi.org/10.1785/0119990114>
- 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>
- Yano K, Shiina T, Kurata S, Kato A, Komaki F, Sakai S, Hirata N (2021) Graph-partitioning based convolutional neural network for earthquake detection using a seismic array. *J Geophys Res Solid Earth* 126(5):e2020JB020269. <https://doi.org/10.1029/2020JB020269>
- Yokota T, Zhou S, Mizoue M, Nakamura I (1981) An automatic measurement of arrival time of seismic waves and its application to an on-line processing system. *Bull Earthq Res Inst* 56:449–484. <https://doi.org/10.15083/0000033017>. (in Japanese with English abstract)
- Yoon CE, Cochran ES, Vanacore EA, Huerfano V, Báez-Sánchez G, Wilding JD, Smith J (2023) A detailed view of the 2020–2023 Southwestern Puerto Rico seismic sequence with deep learning. *Bull Seismol Soc Am*. <https://doi.org/10.1785/0120220229>
- 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(7):565–569. <https://doi.org/10.5047/eps.2011.05.011>

- Yoshida K, Hasegawa A, Okada T, Iinuma T, Ito Y, Asano Y (2012) Stress before and after the 2011 great Tohoku-oki earthquake and induced earthquakes in inland areas of eastern Japan. *Geophys Res Lett* 39(3):L03302. <https://doi.org/10.1029/2011GL049729>
- Yukutake Y, Yoshida K, Honda R (2022) Interaction between aseismic slip and fluid invasion in earthquake swarms revealed by dense geodetic and seismic observations. *J Geophys Res Solid Earth* 127(4):e2021JB022933. <https://doi.org/10.1029/2021JB022933>
- Zaliapin I, Ben-Zion Y (2013) Earthquake clusters in southern California I: identification and stability. *J Geophys Res Solid Earth* 118(6):2847–2864. <https://doi.org/10.1002/jgrb.50179>
- Zhang M, Ellsworth WL, Beroza GC (2019) Rapid earthquake association and location. *Seismol Res Lett* 90(6):2276–2284. <https://doi.org/10.1785/0220190052>
- Zhu W, Beroza GC (2018) 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>
- Zhu W, McBrearty IW, Mousavi SM, Ellsworth WL, Beroza GC (2022) Earthquake phase association using a Bayesian Gaussian mixture model. *J Geophys Res Solid Earth* 127(5):e2021JB023249. <https://doi.org/10.1029/2021JB023249>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
