



Exploratory Data Analysis of Earthquake Moment Tensor Catalog in Japan Using Non-linear Graph-Based Dimensionality Reduction

HISAHIKO KUBO,¹  TAKESHI KIMURA,¹  and KATSUHIKO SHIOMI¹ 

Abstract—The catalog of moment tensor solutions, which contains the information on spatial location, origin time and fault mechanism type of earthquakes, has been interpreted by researchers based on a wealth of past knowledge. However, the long-term routine analyses have led to the accumulation of a huge amount of data in the moment tensor catalog, and it is worth considering moving away from the artisanal approach. In this study, using dimensionality reduction of unsupervised machine learning, we performed exploratory data analysis of the moment tensor catalog in Japan to objectively obtain comprehensive images of seismic activity and to acquire knowledge on the spatial and temporal characteristics of the earthquake mechanism. Source parameters of the moment tensor catalog in Japan, spatial location (latitude, longitude, and focal depth) and source-mechanism diagram information were embedded in two-dimensional space via a non-linear graph-based dimensionality-reduction method, Uniform Manifold Approximation and Projection. On the embedding map, earthquakes in eastern and western Japan are distributed separately and are further embedded to reflect their characteristic fault mechanism and focal depth in each region. The similarity degree of the earthquakes can be obtained as the distance on the embedding map. This study demonstrates that the data visualization using dimensionality reduction is useful for intuitively and objectively understanding the regional characteristics of earthquake mechanisms. The embedding map can also be employed to visualize temporal changes in regional seismic activity and to perform a similarity search with a past event.

Keywords: Earthquake moment tensor solutions, exploratory data analysis, dimensionality reduction, earthquakes in Japan, data visualization, similarity search.

1. Introduction

The information on spatial location, origin time, magnitude, and fault mechanism of an earthquake is contained in the catalog of moment tensor solutions,

where the earthquake source is described by symmetric second-order moment tensors of six independent components at a point source (Ammon et al., 2020). Moment tensor inversion is a method of seismic source analysis to obtain the moment tensor solution from seismic waveform data, using the linear relationship between seismic waveforms caused by an earthquake and each component of the moment tensor (e.g., Dreger & Helmberger, 1993; Dziewonski et al., 1981). Currently, moment tensor inversion has been routinely performed (e.g., Dziewonski et al., 1981; Ekstrom et al., 2012; Fukuyama et al., 1998; Kubo et al., 2002). Based on the catalog of moment tensors of earthquakes, numerous attempts have been made the research on the tectonic stress field, the spatial and temporal evolution of seismicity, and others (e.g., Terakawa & Matsu'ura, 2010; Shuler et al., 2013; Newton & Thomas, 2020).

To date, from a map plot of daily seismic activity, seismologists have grasped the qualitative characteristics of ongoing seismic activity corresponding to the tectonic stress field, such as characteristic seismic activity for each region, its temporal variation, and similarity among earthquakes, jointly considering seismological knowledge and history of seismic activity (Fig. 1a). However, the long-term routine analyses of moment tensor inversion based on seismic observation networks developed in the second half of the twentieth century have led to the accumulation of a huge amount of data in moment tensor catalog, and it is worth moving away from the artisanal approach. This study aims to objectively obtain images of seismic activity that have been intuitively captured by seismologists and acquire knowledge on the spatial and temporal characteristics of earthquake mechanisms via the exploratory data analysis of the

¹ National Research Institute for Earth Science and Disaster Resilience, 3-1 Tennodai, Tsukuba, Ibaraki 305-0006, Japan. E-mail: hkubo@bosai.go.jp

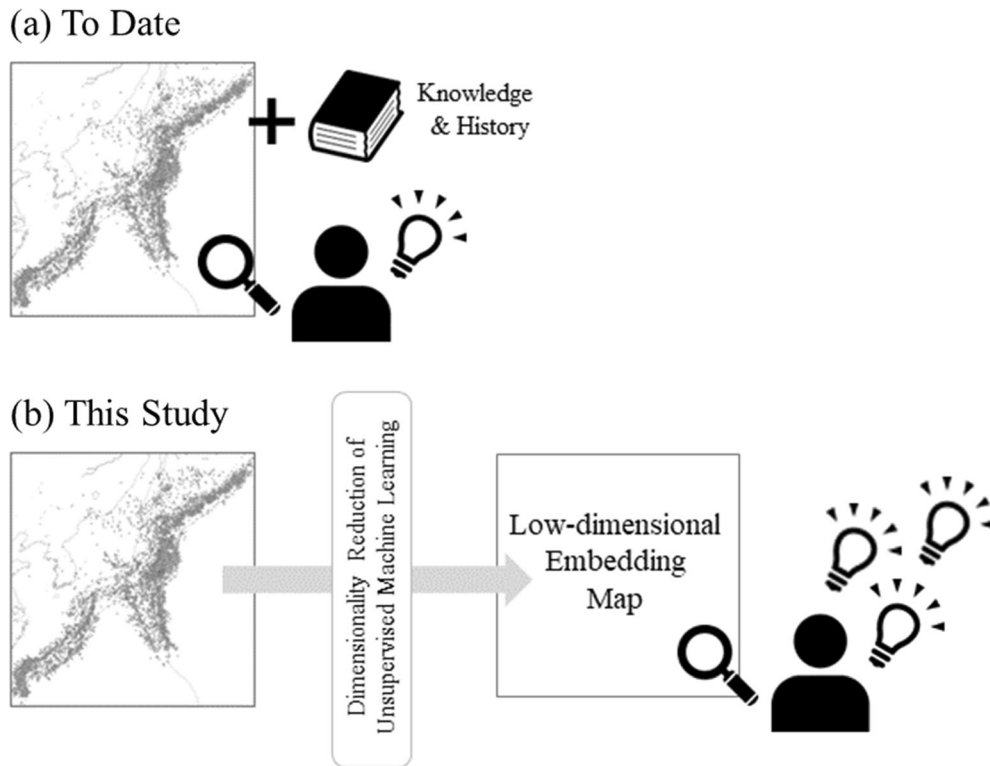


Figure 1

Schematic diagrams to grasp qualitative characteristics of ongoing seismic activity from the catalog of moment tensor solutions (a) to date and (b) in this study

earthquake moment tensor catalog (Fig. 1b). Exploratory data analysis (Tukey, 1977) is an approach of analyzing datasets to summarize the main characteristics using data visualization and statistical graphics, thus providing fruitful insights beyond the existing frameworks. Here, we adopt dimensionality reduction, which is one of the unsupervised machine learning techniques and transforms data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains the meaningful characteristics of the original data (Maaten et al., 2009). It is useful for reducing the amount of data information and for understanding the fundamental structure of data. In this study, we apply this technique to a catalog of moment tensor solutions and investigate the characteristics of its embedding results to confirm its validity. Because seismic activity and rupture mechanisms can be strongly influenced by previous earthquakes (Asano et al., 2011; Hardebeck &

Hauksson, 2001; Yoshida et al., 2015) and the detection of temporal variation of seismic activity is important, we also examine whether the temporal changes in the dataset can be obtained using the learned model of dimensionality reduction. Moreover, we propose a new approach to the similarity search of earthquakes based on the embedding map. The target of the exploratory data analysis in this study is the source parameters of earthquakes around Japan because of the rich spatial diversity in their moment tensor solutions caused by the complex tectonic environment (Kubo et al., 2002; Terakawa & Matsu'ura, 2010) and the large temporal variation of them caused by the active seismicity including the 2011 great Tohoku earthquake (Asano et al., 2011; Yoshida et al., 2015).

2. Data and Methods

2.1. Dataset

National Research Institute for Earth Science and Disaster Resilience (NIED) has developed and managed the broadband seismograph network in Japan, F-net (Aoi et al., 2020; Okada et al., 2004), which consists of approximately 70 stations nationwide. Using the F-net waveform data, NIED has been routinely estimating moment tensor solutions of earthquakes in and around Japan since 1997 (Aoi et al., 2020; Fukuyama et al., 1998; Kubo et al., 2002). In this study, we used the moment tensor solutions of 15,812 earthquakes (Fig. 2) with a moment magnitude $M_w \geq 4$ that occurred during 2003–2019. In 2003, the construction of the observation network was nearly completed and F-net became similar to the current observation system. As shown in Fig. 2, the spatial distribution of earthquakes around Japan is heterogeneous and locally concentrated for tectonic reasons. The fault mechanisms of earthquakes occurring at similar horizontal locations vary with depth caused by the difference of the tectonic stress field and the depth-dependent thermal condition. In Fig. 2, the plots of the focal mechanisms overlap each other, making it difficult to understand the features of focal mechanism distribution directly from this map, and it is necessary to spend an inordinate amount of time creating and reviewing many maps divided by region and time.

The F-net moment tensor catalog has information on origin time, latitude, longitude, focal depth, and six moment tensor components. In the F-net analysis, given the epicentral location of an earthquake on the unified hypocenter catalog of the Japan Meteorological Agency, the focal depth is determined by choosing the best solution that has the smallest misfit among several different trial depths. We processed this catalog into a machine learning-friendly form. The origin time is originally written in year, month, day, hour, minutes, and seconds. Because it is desirable to use time information in a continuous manner, the origin time was converted to the elapsed time from 2003–01-01 00:00:00 (UTC). The information on six moment tensor components was decomposed into a double-couple (DC) source and

a compensated linear vector dipole (CLVD) source. Because the isotropic component is assumed to be zero in the moment-tensor analysis of F-net, we neglected it in this process.

From the DC source, we obtained the seismic moment M_0 and M_w as well as the azimuth and plunge angles of the pressure (P) and tension (T) axes. Moreover, we projected the DC sources on the source-mechanism diagram (Kagan, 2005; Kaverina et al., 1996) of the DC earthquake and obtained the x and y components of this diagram that represent the dominant fault-type component (reverse fault, normal fault, or strike-slip fault) (Fig. 3). A positive x value indicates that the reverse fault component is dominant, whereas a negative x value indicates that the normal fault component is dominant. A large absolute value of x indicates that the two nodal planes are close to 45° , whereas a small absolute value of x indicates that one nodal plane is nearly vertical and another nodal plane is nearly horizontal. A large y value indicates that the strike-slip fault component is dominant. Because the x and y values of the source mechanism diagram are calculated from the plunge angles of the P , T , and null (B) axes, information on the strike angle is removed in the representation of the source mechanism diagram, although that on dip and rake angles is retained.

From the CLVD source, we obtained the relationship between the DC and CLVD (\tilde{a}), which is represented by the modified spherical cylindrical projection (j) in the study of Aso et al. (2016). In this formulation, the DC source without the CLVD is located on 0, $-$ CLVD without the DC source is located on -1 , and $+$ CLVD without the DC source is located on $+1$. Compared to the M_0 ratio of DC and CLVD that has been popularly used, an advantage of \tilde{a} representation is that the sign information of the CLVD source is preserved; however, the information on the orientation of the CLVD source is not included in this representation.

Figure 4 shows the distribution of source parameters in the newly developed dataset. The seismic activity in low-angle reverse faulting earthquakes became significant after the 2011 Tohoku earthquake, which was a huge reverse faulting interplate earthquake along the Japan Trench. The relationship

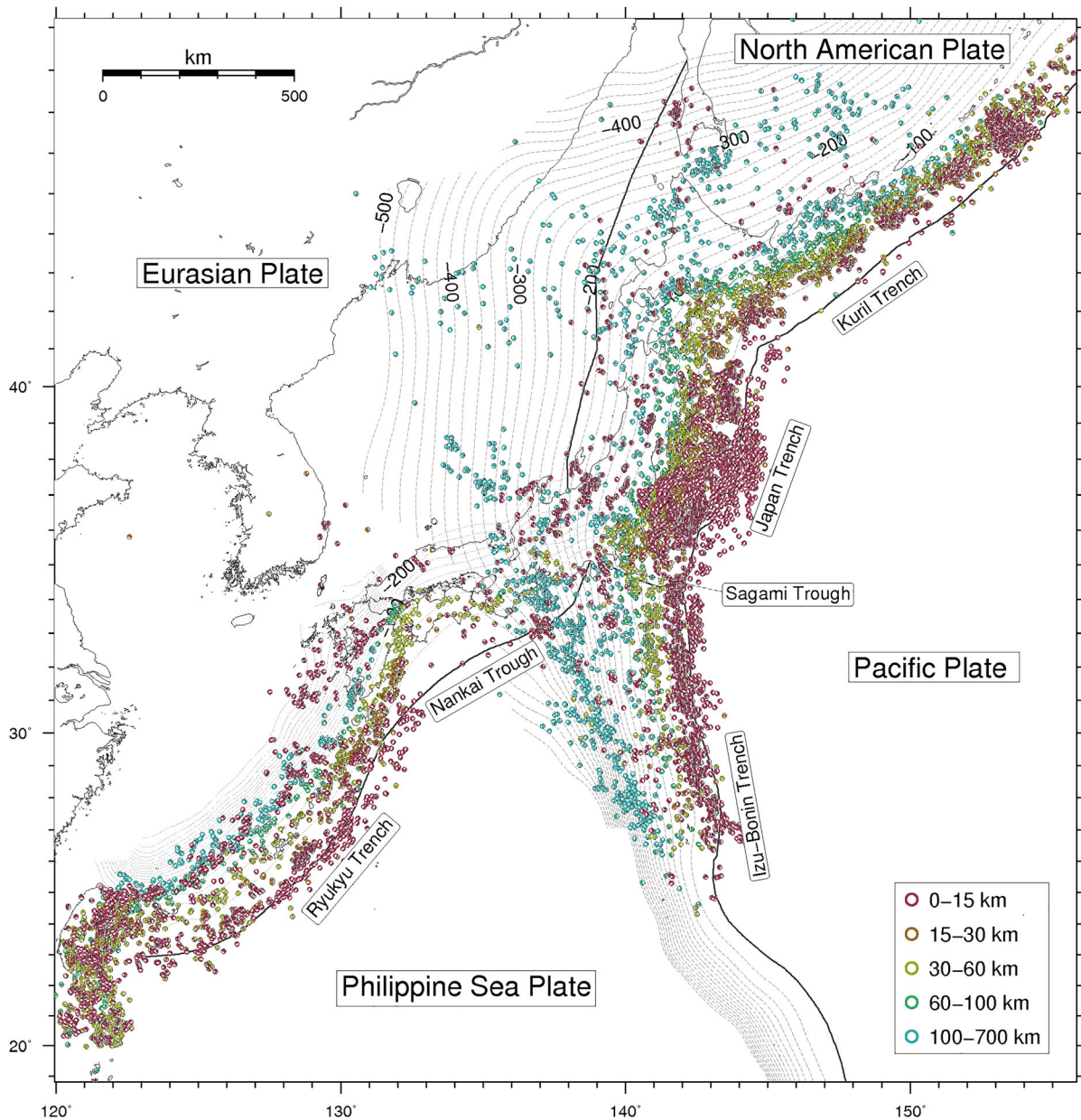


Figure 2

Map of the study area with the focal mechanism of F-net moment tensor catalog in the period of 2003–2019. The color of the focal mechanisms indicates the focal depth. The broken lines and dash-dotted lines indicate the isodepth contours of the upper boundary of the Pacific Plate and the Philippine Sea Plate of Slab2 (Hayes et al., 2018) with a 10 km interval

between the azimuth angles of P and T axes is discontinuous because these parameters are not used in the dimensionality reduction analysis. There appears to be little dependence between \tilde{a} and the

other parameters; therefore, \tilde{a} is not used in the dimensionality reduction analysis.

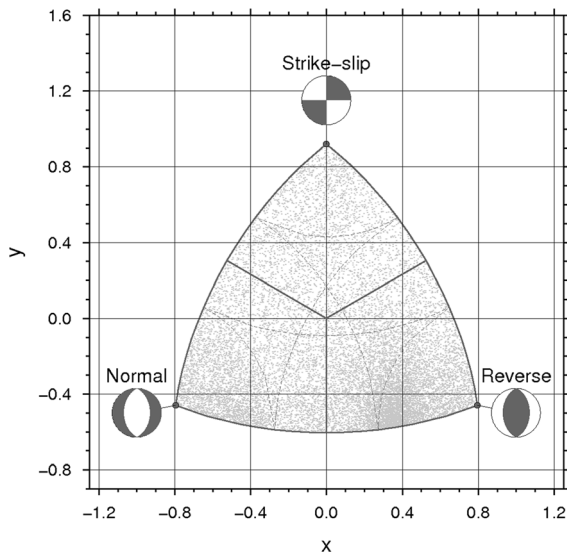


Figure 3

Source mechanism diagram of focal mechanism (Kaverina et al., 1996) with gray dots representing the events used in this study. Gray solid lines are boundaries of strike-slip, normal, and thrust mechanisms. Plunge angles 30° and 60° for all mechanisms are denoted by gray broken lines

2.2. Dimensionality Reduction

The dimensionality reduction method is divided into linear and non-linear approaches. Considering that earthquake spatial distribution is heterogeneous and non-continuous and has a non-linear relationship with the information on source mechanism diagram (Fig. 4), it is desirable to adopt the non-linear approach. Herein, we used a recently developed method of non-linear graph-based dimensionality reduction, namely, Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018), which is a novel manifold learning technique for non-linear dimensionality reduction constructed from a theoretical framework based on Riemannian geometry and algebraic topology. Although UMAP is competitive with t-distributed Stochastic Neighbor Embedding (t-SNE; Maaten & Hinton, 2008) for visualization quality, it arguably preserves more of the global structure with a superior run time performance (McInnes et al., 2018).

From the dataset, five source parameters, namely, latitude, longitude, focal depth, and x and y values of the source mechanism diagram, were embedded in the two-dimensional (2D) space by the

dimensionality reduction of UMAP. Information on the earthquake magnitude was not used in this analysis because the use of magnitude information may lead to the close embedding of an earthquake pair with a similar magnitude and a different focal mechanism, and this situation is inconsistent with the main objective of this study, which is to obtain the qualitative characteristics of seismic activity corresponding to the tectonic stress field. Information on the occurrence time was not included here to exclude the interrelationship of earthquake occurrence such as the relationship among foreshocks, mainshocks, and aftershocks. For data preprocessing of UMAP, we performed axis transformation using principal component analysis (PCA; Jolliffe, 2002) for the latitude and longitude because the spatial distribution of earthquakes around Japan extends in the northeast-southwest direction and there is a strong correlation between latitude and longitude information. Then, we took the common logarithm of the depth and standardized the five explanatory components. The 2D dimensionality-reduced components of the embedding results were standardized so that we could compare the embedding results of the different methods. To optimize the hyperparameters of UMAP, we used the k-nearest neighbor normalized error (k3n-error; Kaneko, 2018), which was developed to compare the visualization performance and optimize the hyperparameters of non-linear visualization methods using only unsupervised data. The size of the local neighborhood used for manifold approximation (n_neighbors) is set to 15, the effective minimum distance between embedded points (min_dist) is set to 0.25, and the metric to compute distances in high dimensional space (metric) is set to Euclidean.

3. Embedding by Dimensionality Reduction

The 2D embedding map is shown in Fig. 5a. The distance in the embedding map reflects the similarity between earthquakes; that is, the shorter the distance on the embedding map, the greater the similarity between earthquakes. The characters A–H in Fig. 5a indicate the embedding location of the nine selected earthquakes shown in Fig. 4b and Table 1. The earthquakes in eastern Japan (Fig. 5a, Events A, B, C,

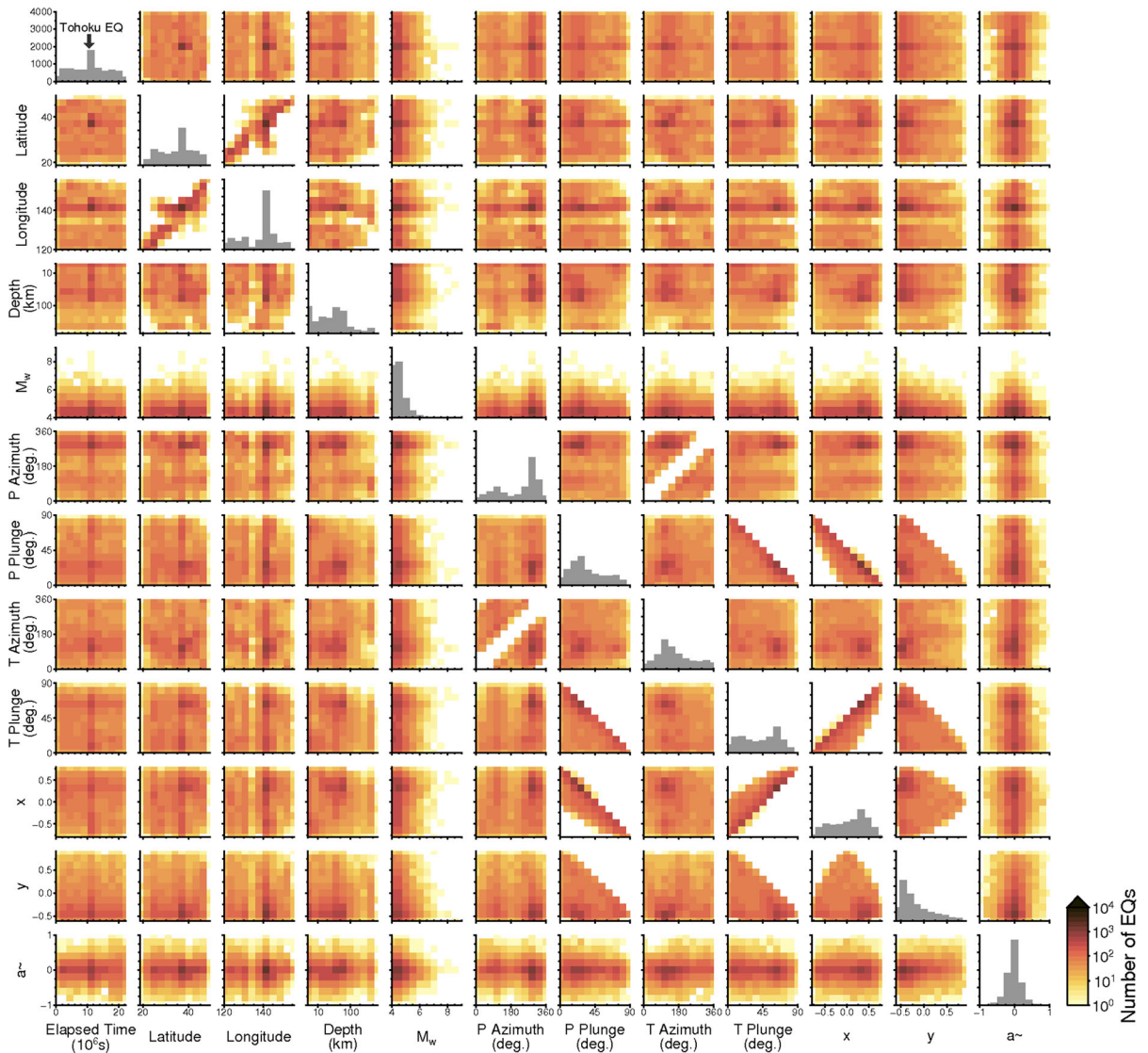


Figure 4

Heat map of the dataset of earthquake source parameters newly developed in this study with each histogram. Here, not only the parameters used in the unsupervised machine learning but also all the parameters obtained during the preparation process are shown

D, and E) and those in western Japan (Fig. 5a, Events G, H, and I) were separately embedded along Component 1. Events A and C, which are interplate reverse faulting earthquakes in eastern Japan, appear close in the embedding map. Event E is a shallow reverse faulting inland earthquake in eastern Japan, and its embedding location is different from those of Events A and C because of the difference in spatial location and dip angle. Events B and D are normal faulting earthquakes in eastern Japan, and their

embedding locations differ along Component 2, which reflects the variety of their spatial locations and dip angles. Event F, a normal faulting deep-focus earthquake occurring in the subducting plate, is located in the group of earthquakes in eastern Japan on the embedding map, which is reasonable because the occurrence of this event was related to the subduction of the Pacific plate. The embedding locations of Events G and I, which are shallow earthquakes in western Japan, are adjacent, although these events

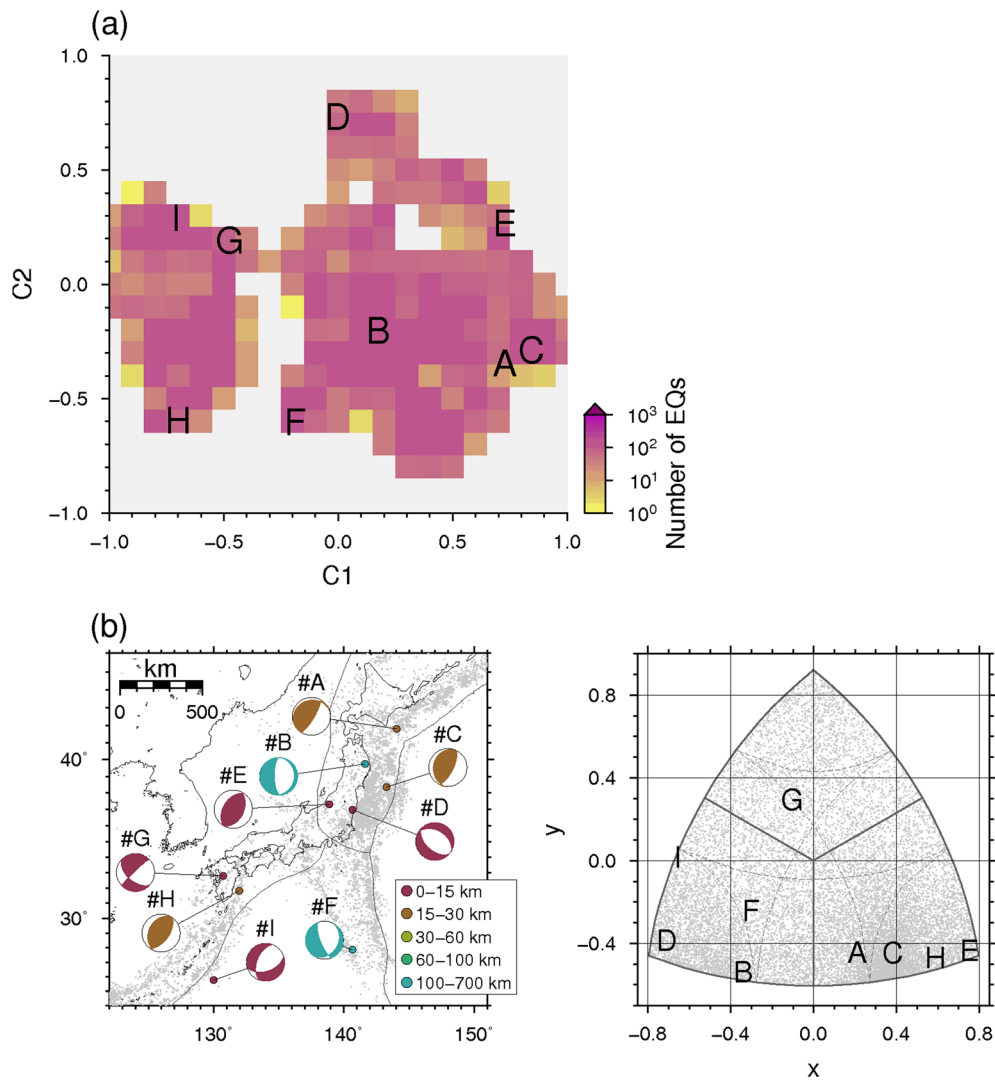


Figure 5

a Heatmap of UMAP embedding. **b** Map plot of the focal mechanisms of the double couple components for nine selected earthquakes shown in **a** on the map and their relationship with x and y values of the source-mechanism diagram. The selected events are listed in Table 1. The color of the focal mechanisms indicates the focal depth. Event A is a reverse faulting interplate earthquake (Honda et al., 2004), Event B is a normal faulting intraslab earthquake (Suzuki et al., 2009), Event C is a reverse faulting interplate earthquake (Ji et al., 2011), Event D is a normal faulting inland earthquake (Tanaka et al., 2014), Event E is a reverse faulting inland earthquake (Honda et al., 2005), Event F is a normal faulting deep-focus earthquake (Kuge, 2017), Event G is a strike faulting inland earthquake (Kubo et al., 2016), Event H is a reverse faulting interplate earthquake, and Event I is a normal faulting outer-rise earthquake. Gray points in Fig. 5b denote the events used in this study

have different fault mechanism types. The focal mechanism of Event G is strike-slip faulting, whereas that of Event I is normal faulting. The reverse faulting interplate earthquake H is embedded far from Events G and I along Component 2, despite the fact

that the spatial location of Event H is close to that of Event G compared with Event I.

The relationships between the 2D dimensionality-reduced components and the five original source parameters are depicted in Fig. 6. Component 1 strongly corresponds to the spatial location of

Table 1
List of earthquakes plotted in Fig. 5

| ID | Origin time (UTC) | Focal depth (km) | M_w | Type* |
|----|-------------------|------------------|-------|-------|
| A | 2003-09-25 19:50 | 23 | 7.9 | R, IP |
| B | 2008-07-23 15:26 | 104 | 6.8 | N, IS |
| C | 2011-03-09 02:45 | 23 | 7.2 | R, IP |
| D | 2011-04-11 08:16 | 5 | 6.6 | N, IL |
| E | 2004-10-23 08:56 | 5 | 6.6 | R, IL |
| F | 2015-05-30 11:23 | 680 | 7.9 | N, DF |
| G | 2016-04-15 16:25 | 11 | 7.1 | S, IL |
| H | 2019-05-09 23:48 | 26 | 6.2 | R, IP |
| I | 2010-05-26 08:53 | 8 | 6.4 | N, OR |

*N normal faulting earthquake, R reverse faulting earthquake, S strike faulting earthquake, IL inland earthquake, IP interplate earthquake, IS intraslab earthquake, OR outer-rise earthquake, DF deep-focus earthquake

earthquakes and is > -0.3 for earthquakes in eastern Japan and < -0.3 for earthquakes in western Japan. A large value of Component 1 corresponds to reverse faulting interplate earthquakes with a low dip angle along the Japan Trench, whereas a small value of Component 1 corresponds to normal faulting and reverse faulting earthquakes in western Japan. Component 2 corresponds to the focal depth and x of the source mechanism diagram. A large value of Component 2 corresponds to shallow normal faulting earthquakes along the Izu-Bonin Trench and the Japan Trench as well as normal faulting inland earthquakes in eastern Japan, whereas a small value of Component 2 corresponds to reverse faulting interplate earthquakes with low dip angle along the Japan Trench, which have a deeper depth and steeper dip angle than events with a large value of Component 1. Reverse faulting interplate earthquakes in western Japan and deep-focus earthquakes also have a small value of Component 2.

Inverse transformation is useful for the interpretation of dimensionality reduction results, which is available in UMAP. Figure 7 shows the inverse transformation of UMAP, in which several representative positions of the embedding map are inversely transformed into the five source parameters. Here, 15 positions were selected so that the whole embedding map is sampled uniformly. Shallow reverse faulting earthquakes in eastern Japan are classified into Groups 1 and 2. Most earthquakes of

Group 1 are located along the Izu-Bonin Trench, whereas many events of Group 2 are located in the inland area of eastern Japan. Event E in Fig. 5 is close to Group 2. Interplate reverse faulting earthquakes in eastern Japan are classified into Groups 3, 4, and 6, which have different horizontal locations and depths. Events A and C are close to Group 3. Group 5 consists of earthquakes along the Kuril Trench of which the reverse faulting component is dominant. Shallow normal faulting earthquakes in eastern Japan are distributed in Groups 7 and 8. Event D is close to Group 7. Shallow strike-slip faulting earthquakes in eastern Japan are distributed in Group 9. Group 10 consists of intermediate-depth earthquakes with a down-dip compressional focal mechanism (Kawakatsu, 1986) in eastern Japan. Deep-focus events in the subducting Pacific plate including Event F are found in Group 11. Groups 12 and 13 include strike-slip faulting and normal faulting shallow earthquakes in western Japan, which correspond to Events G and I. The adjacency between Groups 12 and 13 on the embedding map could be caused by the adjacency of their spatial distribution, although their focal mechanisms are different. Group 14 includes intermediate-depth earthquakes with a down-dip compressional focal mechanism around the islands of Taiwan. Reverse faulting interplate earthquakes along the Ryukyu Trench are found in Group 15, which is close to Event H.

These embedding results demonstrate that earthquakes with similar characteristics can be projected on the embedding map by UMAP. The use of UMAP leads to uniform re-distribution of focal mechanisms as much as possible on the embedding map, although the distribution of focal mechanisms is collapsed in the direct mapping (Fig. 2). On the embedding map, earthquakes in eastern and western Japan were embedded separately and further embedded to reflect their characteristic fault mechanism and focal depth in each region. The degree of similarity of their groups was represented as the distance between the groups on the embedding map. The results indicate that this map can be useful for intuitively and objectively understanding the regional characteristics of earthquake mechanisms and the relationship among earthquake groups.

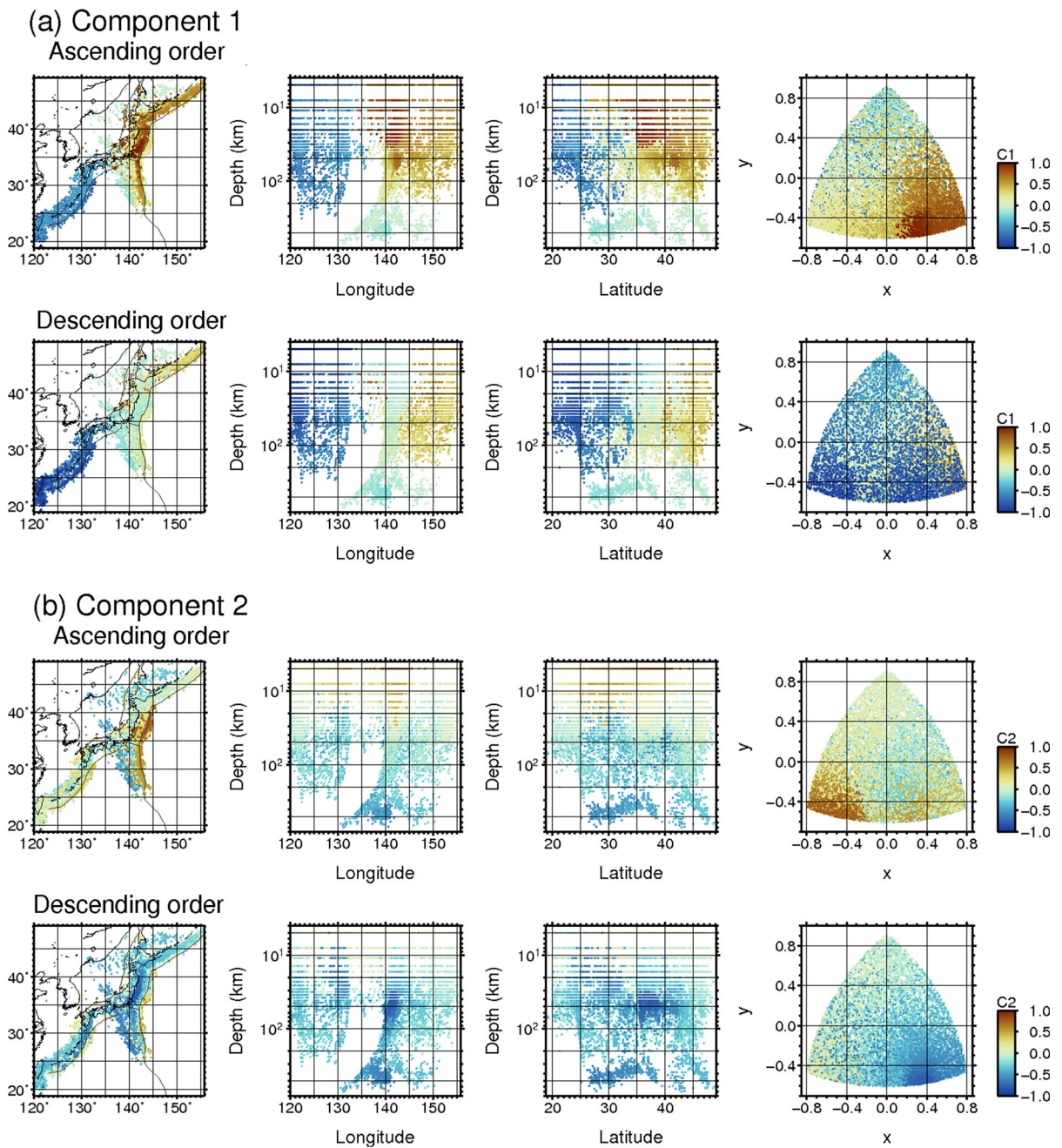


Figure 6

Relationships between the two-dimensional dimensionality-reduced components shown in Fig. 5a and the five original source parameters in the UMAP embedding. “Ascending order” means that the higher the value, the later it is plotted in the sort order, and “Descending order” means the opposite

We also performed a dimensionality reduction of PCA to evaluate the usefulness of the non-linear dimensionality reduction method (Fig. 8). The

distribution of PCA embedding is highly skewed, whereas earthquakes are distributed over the whole region with little bias in UMAP embedding (Fig. 5a).

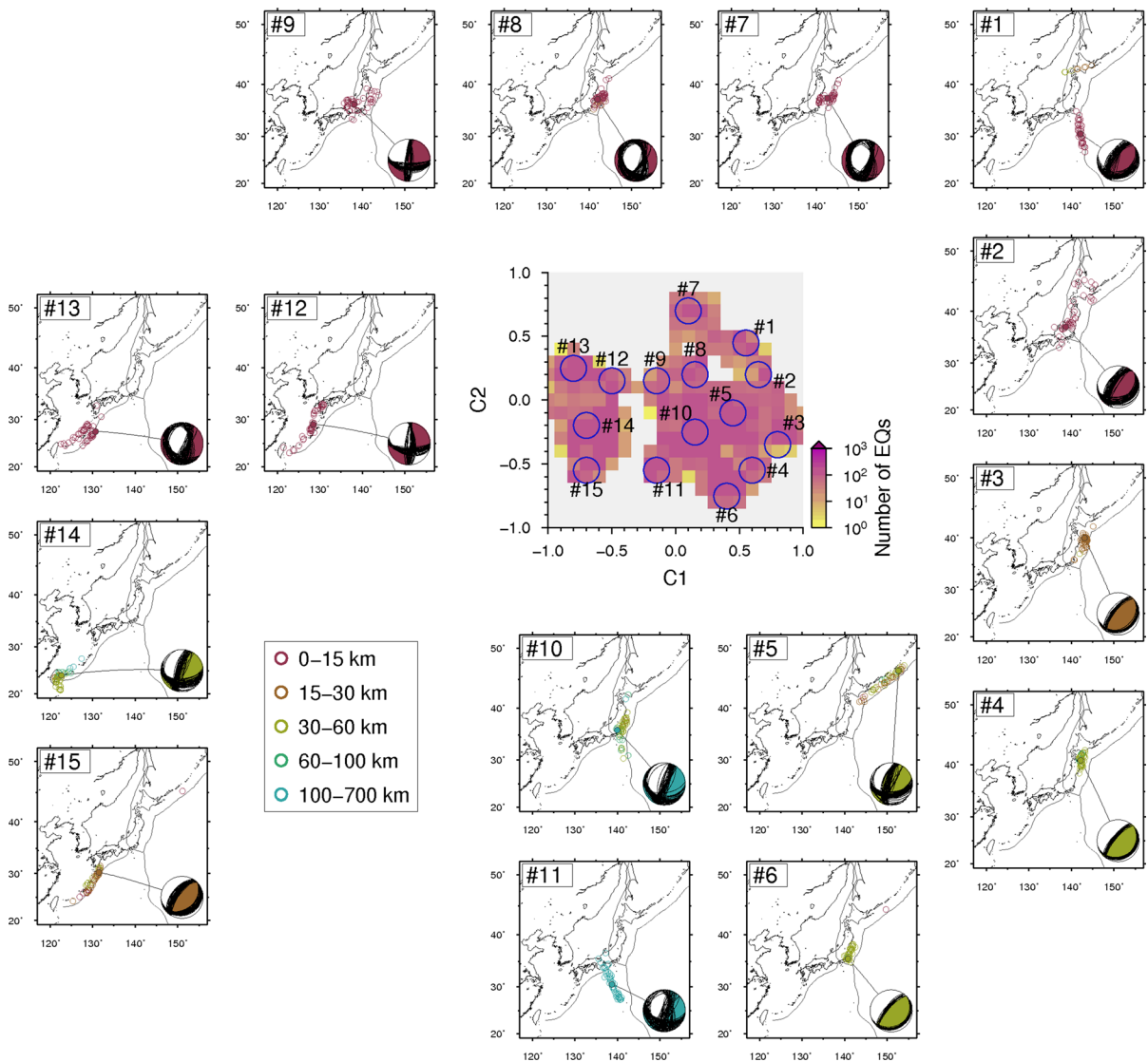


Figure 7

Inverse transformation of UMAP embedding of the five source parameters. The range of the blue circle of each number in the UMAP embedding was inversely transformed into the five original source parameters, showing the mechanism solution and nodal plane. When converting the x and y values of the source mechanism diagram to a double couple solution, the azimuth of the P axis is given at 0° . The color of the circles and focal mechanisms indicates their depth

Moreover, earthquakes in eastern and western Japan are separately distributed in UMAP embedding, which is not the case in PCA embedding. Although the spatial location differs among the three reverse faulting interplate earthquakes (A, C, and H), they are concentrated in one place of the PCA embedding. These results suggest that 2D PCA embedding does not contain the data structure of the original moment

tensor catalog, whereas non-linear dimensionality reduction methods such as UMAP are superior in terms of data visualization of geospatial data including the moment tensor catalog.

In the dimensionality reduction analysis of this section, the azimuth information of P and T axes is not considered for simplicity. Because consideration of the azimuth information is important to determine

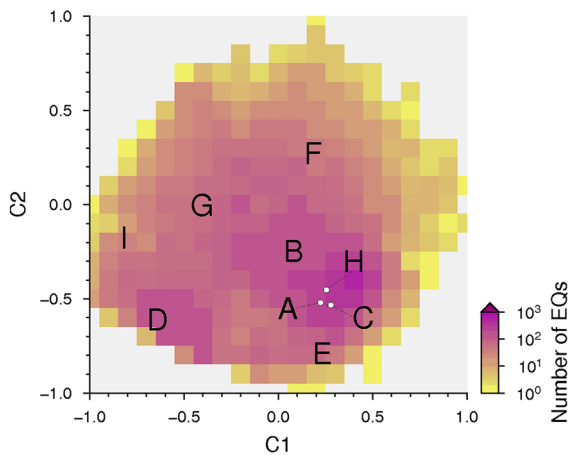


Figure 8

Heatmap of PCA embedding. Characters denote the embedding location of the nine selected earthquakes shown in Fig. 5

the direction of principal stress, the use of the azimuth information in our dimensionality reduction may lead to obtaining new insights into the spatial variations of focal mechanisms through the explanatory analysis. Moreover, there are other source parameters that were not considered in this study such as time of earthquake occurrence, earthquake magnitude, and CLVD source. Although the isotropic component of moment tensor is neglected in the F-net analysis, some studies have pointed out the importance of estimating the isotropic component in the moment tensor analysis (e.g., Menke & Russell, 2020). The finiteness of the earthquake rupture zone should also be considered in the explanatory analysis to account for the spatial extent of the effect of an earthquake, although the source information in moment tensor catalogs is estimated assuming a point source. How to incorporate the other information into the dimensionality reduction is future work.

4. Further Use of Embedding MAP

4.1. Visualization on Temporal Change

A learned model of UMAP can be used to transform new data into the learned space, which is useful for understanding the trend change in new compared to original data. Here, this technique was applied to the visualization of the different-period

data of seismicity to investigate temporal changes in seismic activity. First, we prepared a UMAP model that learned the whole-period dataset (2003–2019), which is identical to the one used in the previous section. Then, we passed four datasets belonging to different periods (2003–2010, 2011–2015, 2016–2019, and 2020–2021) to the learned model and had it transform them in the learned embedding map. The data of the first three datasets were included in the whole dataset, whereas the last dataset was a new one that was newly derived from the original catalog of F-net and consisted of moment tensor solutions of 1447 earthquakes with M_w of ≥ 4 that occurred during 2020–2021. Figure 9 shows the embedding results of each dataset. The residual distribution from the whole-data embedding in each period (bottom panels of Fig. 9) pertains to the relative activity of earthquakes in the target period compared to the whole-period dataset. The relative seismic activity differs among the four periods, and we focus on the nine significant positive regions (P1–P9) in the bottom panels of Fig. 9. In the period from 2003 to 2010, two significant positive regions, P1 and P2, correspond to the activation of seismic activity for earthquakes along the Kuril Trench (Group 5) and shallow reverse faulting crustal earthquakes in eastern Japan (Group 2) including Event E. In the period from 2011 to 2015, three regions are significantly positive (P3, P4, and P5). P3 primarily corresponds to the seismicity for shallow normal faulting earthquakes in eastern Japan (Groups 7 and 8), whereas P4 and P5 correspond to the seismic activity for reverse faulting interplate earthquakes along the Japan Trench (Groups 3 and 6). These trend changes in seismicity after 2011 have also been reported in previous studies; Asano et al. (2011) noted that the occurrence of the 2011 Tohoku earthquake activated the seismic activity of shallow normal faulting earthquakes in eastern Japan and reverse faulting interplate earthquakes along the Japan Trench. In the period from 2016 to 2019, the seismicity in western Japan (P6) was active, especially for shallow strike-slip faulting earthquakes (Group 12), including the 2016 Kumamoto earthquake (Event H) and its sequence. During 2020–2021, three distinct positive regions (P7, P8, and P9) were found. P7 corresponds to shallow strike-slip faulting and normal faulting

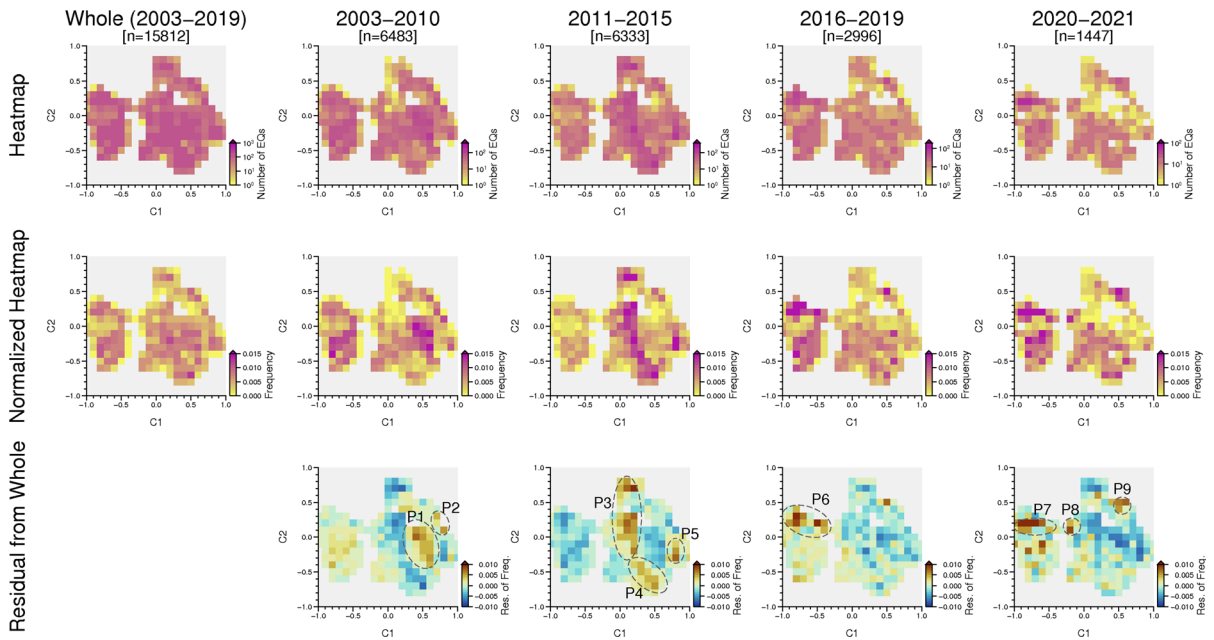


Figure 9

Heatmap of UMAP embedding of five source parameters (top), normalized heatmap (middle), and residual from the normalized heatmap of whole data (bottom) in the periods of 2003–2019 (left), 2003–2010 (second from the left), 2011–2015 (center), 2016–2019 (second from the right), and 2020–2021 (right). The color scale of the heatmaps depends on the total number of events (n) in each period. The normalized heatmaps are made by dividing the heatmaps by n so that the effect of the number of events is removed from the appearance of the heatmaps. In the bottom panels, red color means that the seismic activity of corresponding embedded earthquakes was active compared to that during the period of 2003–2019 (whole-period data), and blue color means the opposite. Nine significant positive regions in the bottom panels (P1–P9) are outlined by dashed circles

earthquakes in western Japan (Groups 12 and 13), P8 corresponds to shallow strike-slip faulting earthquakes in eastern Japan (Group 9), and P9 corresponds to shallow reverse faulting earthquakes along the Izu-Bonin Trench (Group 1). Thus, the temporal changes in regional seismic activity can be visualized via dimensionality reduction of different-period data with the learned model, which can be useful for evaluating abnormalities in current seismic activity compared to that of the past and the objective detection of changes in the tectonic stress field. Such time-varying trends would be difficult to find from the direct mapping of focal mechanisms because of the overlapping of their plots and the difference in the amount of data from period to period. Recently, new focal mechanism catalogs for hundreds of thousands of small inland earthquakes in Japan have been developed using machine learning techniques (e.g., Uchide, 2020; Uchide et al., 2022). The application of our visualization method to such a catalog is

expected to enable the objective detection of spatiotemporal changes in the tectonic stress field.

4.2. Similarity Search

Based on the UMAP embedding, we proposed a new approach to similarity search of earthquakes. In advance, source information on past earthquakes is embedded by UMAP to obtain an embedding map and pre-trained UMAP model. Here, we used the UMAP embedding as shown in Fig. 5 where earthquakes in the period from 2003 to 2019 were embedded. A given target earthquake was then embedded with the pre-trained UMAP model, and the Euclidean distances between the target event and all past events on the embedding map were calculated. Based on the distance information, similar events were selected. Two types of searches were performed: one with no magnitude limitation (magnitude free) and the other with a limitation of

magnitude range. In the magnitude-limited similarity search, the search range of past earthquakes was limited so that their magnitude was within a magnitude ± 0.5 of that of the target earthquake.

Figure 10 shows an example of the result of similarity search where the target event occurred on 2021-10-07 at 13:41 [UTC] (M_w 5.9), and it was a reverse faulting earthquake with a relatively deep depth (68 km). The similarity search selected the past earthquakes with similar mechanism solutions and spatial locations to the target event, although the fault

strike varies among the target and selected events. Approximately 16 years before the target earthquake, an earthquake of almost the same magnitude (M_w 5.9) at nearly the same location with a similar fault mechanism occurred, which was selected as the event with the top similarity in the magnitude-limited similarity search. The earthquake similarity search as shown here is useful for linking recent earthquakes with past seismic activity to clarify the nature of the recent earthquakes and rapidly estimate damage area for them. Although it is possible to directly compare

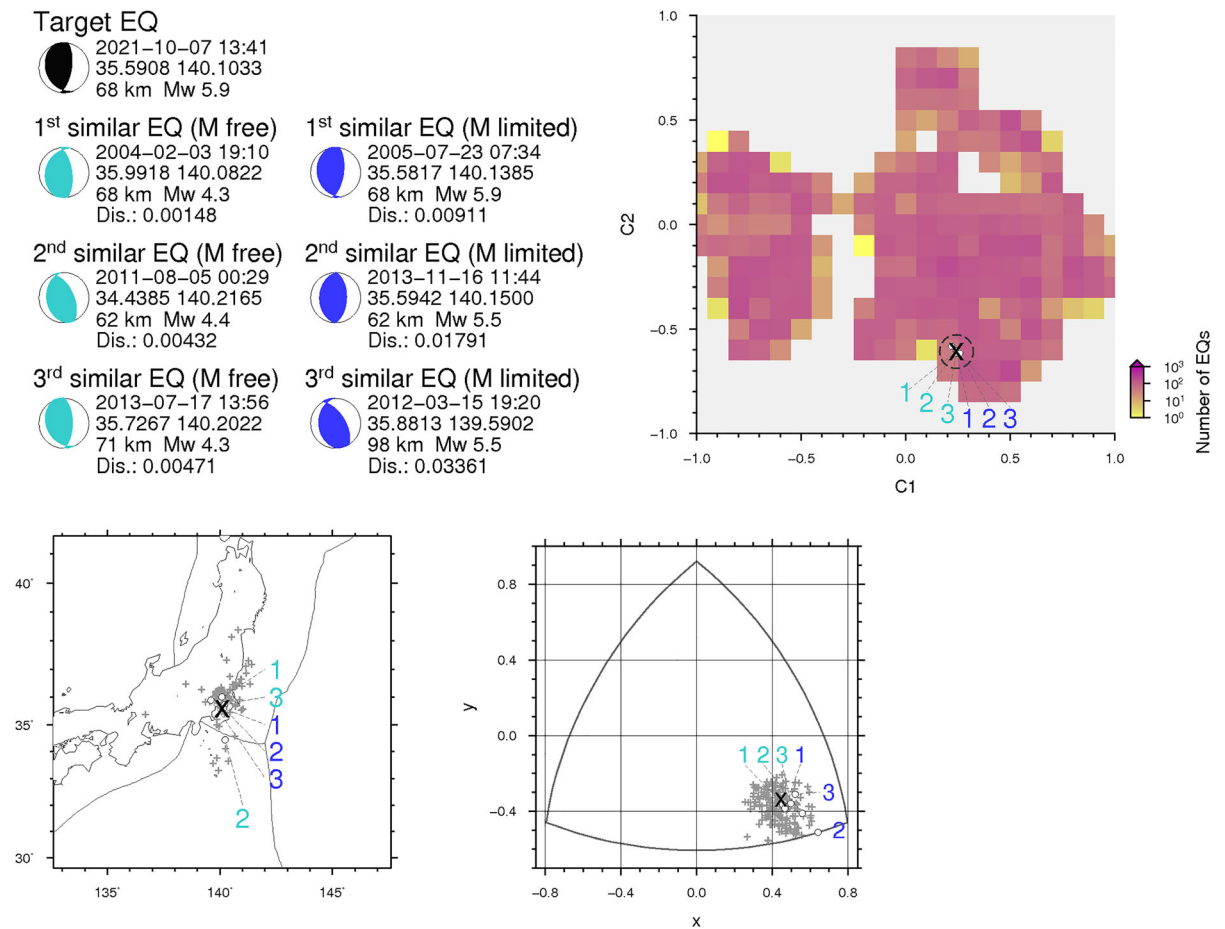


Figure 10

Results of the similarity search for the event on 2021-10-07 at 22:41 [JST] (M_w 5.9). In top left, event information on the target event (black), the top three similar events in the magnitude-free similarity search (cyan), and the top three similar events in the magnitude-limited similarity search (blue) are shown. The top right figure indicates the heatmap of UMAP embedding with the target event (black cross), the top three similar events in the magnitude-free similarity search (cyan characters), and the top three similar events in the magnitude-limited similarity search (blue characters). In the bottom left and right figures, the target event and similar events are plotted on the map and the source-mechanism diagram, respectively. Events in which the distance on the embedding map to the target earthquake is < 0.08 are also plotted by gray crosses in the bottom figures

the source information without using embedded maps, it is challenging to objectively compare various types of seismic information with different geophysical units and estimation methods. A weighting for each source parameter is necessary to comprehensively evaluate the earthquake similarity; however, this weighting is where the uncertainty of the solution arises. Our approach using unsupervised machine learning is superior because the weights can be learned directly from the parameter distribution of the dataset.

Acknowledgements

We thank the anonymous reviewer, the reviewer, Dr. Naofumi Aso, and the editor for their helpful comments. The Generic Mapping Tools (Wessel and Smith, 1998) was used to draw the figures.

Author contributions Conceptualization: HK, TK, KS. Formal analysis: HK. Investigation: HK, TK, KS. Data Curation: HK, TK. Writing—Original Draft: HK. Writing—review and editing: HK, TK, KS. Visualization: HK. Funding acquisition: HK, KS. All authors have read and approved the final manuscript.

Funding

This work was supported in part by MEXT Project for Seismology toward Research Innovation with Data of Earthquake (STAR-E) Grant Number JPJ010217.

Data availability

The moment tensor catalog of F-net is available at this website: <http://www.fnet.bosai.go.jp/>. The python code of UMAP is available at the website: <https://umap-learn.readthedocs.io/en/latest/>.

Declarations

Conflict of interest The authors declare no competing interests.

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/>.

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

REFERENCES

- Ammon, C. J., Velasco, A. A., Lay, T., & Wallace, T. C. (2020). *Foundations of Modern Global Seismology* (2nd ed.), Academic Press. <https://doi.org/10.1016/C2017-0-03756-4>
- Aoi, S., et al. (2020). MOWLAS: NIED observation network for earthquake, tsunami and volcano. *Earth, Planets and Space*, 72, 126. <https://doi.org/10.1186/s40623-020-01250-x>
- Asano, Y., et al. (2011). Spatial distribution and focal mechanisms of aftershocks of the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planet and Space*, 63, 29. <https://doi.org/10.5047/eps.2011.06.016>
- Aso, N., Ohta, K., & Ide, S. (2016). Mathematical review on source-type diagrams. *Earth, Planets and Space*, 68, 52. <https://doi.org/10.1186/s40623-016-0421-5>
- Dreger, D. S., & Helmberger, D. V. (1993). Determination of source parameters at regional distances with single station or sparse network data. *Journal of Geophysical Research*, 98, 8107–8125. <https://doi.org/10.1029/93JB00023>
- Dziewonski, A. M., Chou, T. A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, 86, 2825–2852. <https://doi.org/10.1029/JB086iB04p02825>
- Ekstrom, G., Nettles, M., & Dziewonski, A. M. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9. <https://doi.org/10.1016/j.pepi.2012.04.002>
- Fukuyama, E., Ishida, M., Dreger, D. S., & Kawai, H. (1998). Automated seismic moment tensor determination by using on-line broadband seismic waveforms. *Journal of the Seismological Society of Japan (Zisin 2)*, 51(1), 149–156. https://doi.org/10.4294/zisin1948.51.1_149 (in Japanese with English abstract)
- Hardebeck, J. L., & Hauksson, E. (2001). Crustal stress field in southern California and its implications for fault mechanics. *Journal of Geophysical Research*, 106, 21859–21882. <https://doi.org/10.1029/2001JB000292>
- Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., & Smoczyk, G. M. (2018). Slab2, a

- comprehensive subduction zone geometry model. *Science*, 362(6410), 58–61. <https://doi.org/10.1126/science.aat4723>
- Honda, R., Aoi, S., Morikawa, N., Sekiguchi, H., Kunugi, T., & Fujiwara, H. (2004). Ground motion and rupture process of 2003 Tokachi-oki earthquake obtained from strong motion data of K-NET and KiK-net. *Earth, Planets and Space*, 56, 317–322. <https://doi.org/10.1186/BF03353058>
- Honda, R., Aoi, S., Morikawa, N., Sekiguchi, H., Kunugi, T., & Fujiwara, H. (2005). Ground motion and rupture process of the 2004 Mid Niigata Prefecture earthquake obtained from strong motion data of K-NET and KiK-net. *Earth, Planets and Space*, 57, 527–532. <https://doi.org/10.1186/BF03352587>
- Ji, G., Shao, C., & Zhao, D. (2011). Rupture process of the 9 March, 2011 Mw 7.4 Sanriku-Oki, Japan earthquake constrained by jointly inverting teleseismic waveforms, strong motion data and GPS observations. *Geophysical Research Letters*, 38, L00G20. <https://doi.org/10.1029/2011GL049164>
- Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.
- Kagan, Y. Y. (2005). Double-couple earthquake focal mechanism: Random rotation and display. *Geophysical Journal International*, 164, 236–245. <https://doi.org/10.1111/j.1365-246X.2005.02781.x>
- Kaneko, H. (2018). K-nearest neighbor normalized error for visualization and reconstruction? A new measure for data visualization performance. *Chemometrics and Intelligent Laboratory Systems*, 176, 22–33. <https://doi.org/10.1016/j.chemolab.2018.03.001>
- Kaverina, A. N., Lander, A. V., & Prozorov, A. G. (1996). Global creep distribution and its relation to earthquake-source geometry and tectonic origin. *Geophysical Journal International*, 125, 249–265. <https://doi.org/10.1111/j.1365-246X.1996.tb06549.x>
- Kawakatsu, H. (1986). Double seismic zones: Kinematics. *Journal of Geophysical Research*, 91(B5), 4811–4825. <https://doi.org/10.1029/JB091iB05p04811>
- Kubo, A., Fukuyama, E., Kawai, H., & Nonomura, K. (2002). NIED seismic moment tensor catalogue for regional earthquakes around Japan: Quality test and application. *Tectonophysics*, 356(1–3), 23–48. [https://doi.org/10.1016/S0040-1951\(02\)00375-X](https://doi.org/10.1016/S0040-1951(02)00375-X)
- Kubo, H., Suzuki, W., Aoi, S., & Sekiguchi, H. (2016). Source rupture processes of the 2016 Kumamoto, Japan, earthquakes estimated from strong motion waveforms. *Earth, Planets and Space*, 68, 161. <https://doi.org/10.1186/s40623-016-0536-8>
- Kuge, K. (2017). Seismic observations indicating that the 2015 Ogasawara (Bonin) earthquake ruptured beneath the 660 km discontinuity. *Geophysical Research Letters*, 44, 10855–10862. <https://doi.org/10.1002/2017GL074469>
- Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- Maaten, L. V. D., Postma, E., & Herik, J. (2009). Dimensionality reduction: a comparative review. *Journal of Machine Learning Research*, 10, 66–71.
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. <https://doi.org/10.48550/arXiv.1802.03426>. arXiv:1802.03426
- Menke, W., & Russell, J. B. (2020). Non-double-couple components of the moment tensor in a transversely isotropic medium. *Bulletin of the Seismological Society of America*, 110(3), 1125–1133. <https://doi.org/10.1785/0120190319>
- Newton, T. J., & Thomas, A. M. (2020). Stress orientations in the Nankai Trough constrained using seismic and aseismic slip. *Journal of Geophysical Research: Solid Earth*, 125, e2020JB019841. <https://doi.org/10.1029/2020JB019841>
- 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 and Space*, 56(8), xv–xxviii. <https://doi.org/10.1186/BF03353076>
- Shuler, A., Nettles, M., & Ekstrom, G. (2013). Global observation of vertical-CLVD earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth*, 118(1), 138–164. <https://doi.org/10.1029/2012JB009721>
- Suzuki, W., Aoi, S., & Sekiguchi, H. (2009). Rupture process of the 2008 Northern Iwate intraslab earthquake derived from strong-motion records. *Bulletin of the Seismological Society of America*, 99(5), 2825–2835. <https://doi.org/10.1785/0120080331>
- Tanaka, M., Asano, K., Iwata, T., & Kubo, H. (2014). Source rupture process of the 2011 Fukushima-ken Hamadori earthquake: How did the two subparallel faults rupture? *Earth, Planets and Space*, 66, 101. <https://doi.org/10.1186/1880-5981-66-101>
- Terakawa, T., & Matsu'ura, M. (2010). The 3-D tectonic stress fields in and around Japan inverted from centroid moment tensor data of seismic events. *Tectonics*, 29, TC6008. <https://doi.org/10.1029/2009TC002626>
- Tukey, J. W. (1977). *Exploratory data analysis*. Pearson.
- Uchide, T. (2020). Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International*, 223, 1658–1671. <https://doi.org/10.1093/gji/ggaa401>
- Uchide, T., Shiina, T., & Imanishi, K. (2022). Stress map of Japan: Detailed nationwide crustal stress field inferred from focal mechanism solutions of numerous microearthquakes. *Journal of Geophysical Research: Solid Earth*, 127, e2022JB024036. <https://doi.org/10.1029/2022JB024036>
- Wessel, P., & Smith, W. H. F. (1998). New, improved version of generic mapping tools released. *Eos, Transactions American Geophysical Union*, 79, 579.
- Yoshida, K., Hasegawa, A., & Okada, T. (2015). Spatially heterogeneous stress field in the source area of the 2011 Mw 6.6 Fukushima-Hamadori earthquake, NE Japan, probably caused by static stress change. *Geophysical Journal International*, 201(2), 1062–1071. <https://doi.org/10.1093/gji/ggv068>